

Evaluation of Oxford Nanopore Technologies workflows for genomic epidemiology of outbreak-associated bacterial isolates in the clinical setting

Stefan Neuenschwander¹, Loïc Borcard¹, Sonja Gempeler¹, Miguel A. Terrazos Miani¹, Carlo Casanova¹ and Alban Ramette^{1,2,*}

Abstract

Accurate and efficient whole-genome sequencing (WGS) is crucial for clinical diagnostics and surveillance of bacterial infections. Here, we investigate the potential of a new Oxford Nanopore Technologies (ONT) workflow for WGS of clinically relevant bacterial isolates. Specifically, we assess the performance of R10.4.1 flow cells in combination with the V14 version of the transposase-based (RBK) library preparation kit to provide rapid and accurate genomic epidemiological comparisons of bacterial species of clinical importance. We focused on retrospective collections of outbreak-associated *Corynebacterium diphtheriae* (CDIP) and vancomycin-resistant enterococci (VRE) and benchmarked expected performance parameters such as genome assembly quality, genotyping [multi-locus sequence typing (MLST) and core-genome multi-locus sequence typing (cgMLST)], SNP profiling and antimicrobial resistance and virulence prediction, against WGS data obtained routinely by Illumina MiSeq sequencing. Complete concordance with Illumina results was observed for MLST in both species and for cgMLST in CDIP, across all ONT kits and software evaluated. For VRE, however, cgMLST results varied with strain identity, library preparation kit and analysis parameters, likely due to software challenges to correctly call methylated bases. Yet, the use of the latest basecalling models combined with a PCR-based library preparation kit (RPB) reliably reproduced Illumina cgMLST results across all tested VRE strains. By testing two hybrid strategies combining PCR-free and PCR-based library preparation approaches, we also showed that combining PCR-free and PCR-based methods may yield a promising strategy, achieving both high accuracy and assembly completeness. Genomic-based antimicrobial resistance (AMR) prediction was consistent across sequencing methods, and we further highlight advantages and limitations of the PCR-based, PCR-free and mixed assemblies, to inform on the genomic context of AMR genes. This study demonstrates that a Nanopore-only sequencing approach may offer improved accuracy and consistency for classical bacterial typing in outbreak investigations, paving the way to wider use in clinical microbiology laboratories.

Impact Statement

This study demonstrates that an Oxford Nanopore Technologies-only sequencing approach improves the accuracy and consistency of bacterial typing in outbreak investigations, even for difficult cases such as vancomycin-resistant enterococci. This advancement paves the way for wider adoption of this technology in clinical microbiology laboratories, leading to more rapid and effective responses to bacterial infections.

Received 06 May 2025; Accepted 02 January 2026; Published 10 February 2026

Author affiliations: ¹Institute for Infectious Diseases, University of Bern, Bern, Switzerland; ²Multidisciplinary Center for Infectious Diseases, University of Bern, Bern, Switzerland.

***Correspondence:** Alban Ramette, alban.ramette@ifik.unibe.ch

Keywords: bacterial pathogens; core-genome multi-locus sequence typing (cgMLST); clinical setting; genome assembly; genomic epidemiology; outbreak; Oxford Nanopore Technologies; plasmids; R10.4.1.

Abbreviations: AMR, antimicrobial resistance; CDIP, *Corynebacterium diphtheriae*; CT, complex type; HAC, high-accuracy mode; IFIK, institute for infectious diseases; MLST, multi-locus sequence typing; NGS, next-generation sequencing; ONT, Oxford nanopore technologies; SB, strand bias; ST, sequence type; VRE, vancomycin-resistant enterococci; WGS, whole-genome sequencing.

All supporting data, code and protocols have been provided within the article or through supplementary data files. Five supplementary figures and two supplementary tables are available with the online version of this article.

001626 © 2026 The Authors



This is an open-access article distributed under the terms of the Creative Commons Attribution License.

DATA SUMMARY

Illumina MiSeq and Oxford Nanopore Technologies (ONT) sequencing data for the Illumina, SUPD43, SUPDP43 and SUPD50 treatments are available under BioProject accessions PRJNA889706 and PRJNA1230056. NCBI SRA accession numbers for the individual libraries are listed in Table S2. ONT datasets generated using earlier basecaller versions are available upon request. The bioinformatic correction script (Fig. 1) is available at <https://github.com/RametteLab/NanoporeHybridKP>.

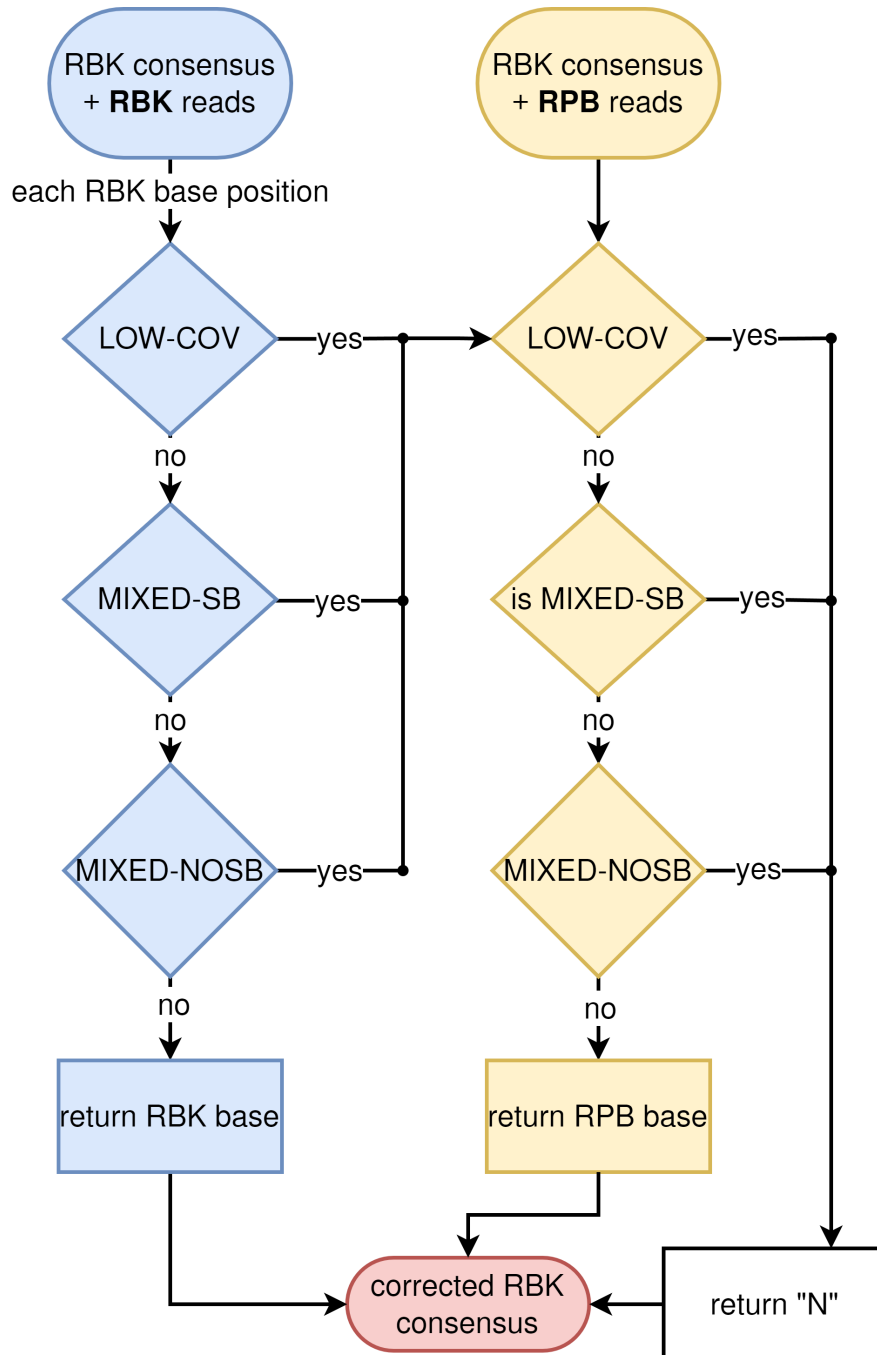


Fig. 1. Correction algorithm of RBK-based assembly using RPB reads. Abbreviations: LOW-COV, coverage below the threshold of 20X; MIXED-SB, mixed base composition where the majority base constitutes more than 67% of all bases on one strand and less than 33% on the other strand; MIXED-NOSB, mixed bases that do not fulfil the criteria for strand bias mentioned above.

INTRODUCTION

Whole-genome sequencing (WGS) has become the gold standard for bacterial strain typing, revolutionizing our ability to elucidate infection-epidemiological connections due to its high discriminatory power [1]. Consequently, WGS is increasingly employed in clinical settings to enhance outbreak detection, implement epidemiological surveillance and improve infection control, leveraging its unparalleled resolution [2–5]. Common approaches for assessing genetic relatedness from WGS data include gene-by-gene comparisons, such as core-genome multi-locus sequence typing (cgMLST) or whole-genome multi-locus sequence typing and SNP analysis [6]. While SNP typing offers high resolution, it is computationally intensive and time-consuming [7, 8]. In contrast, cgMLST-typing provides a standardized nomenclature with lower computational demands [6]. Target-free, k-mer-based methods, like split k-mer analysis, are emerging but are not yet standard for transmission analysis [7, 9]. Despite the widespread use of these methods, a standardized approach for interpreting WGS data in epidemiological investigations remains undefined [7].

Healthcare facilities have traditionally relied on short-read Illumina technology for WGS and genomic pathogen surveillance [10, 11]. However, long-read next-generation sequencing (NGS), particularly provided by Oxford Nanopore Technologies (ONT) [12], is gaining traction. ONT sequencing offers rapid turnaround times, real-time data, long-read capabilities and lower capital costs, potentially reducing outbreak investigation times or enabling genomic surveillance in resource-limited settings [13, 14]. Studies have shown promising results when using ONT sequencing for bacterial sequence typing and epidemiology [15, 16], including largely consistent findings with Illumina for vancomycin-resistant enterococci (VRE) [17], *Mycobacterium tuberculosis* [18], *Staphylococcus aureus* [19] and methicillin-resistant *S. aureus* [20]. Several studies also demonstrate successful Nanopore-based typing in *Salmonella* [21, 22], with one addressing homopolymer error reduction and combined Illumina/Nanopore analysis [22]. Smaller studies also report consistent results between Illumina and Nanopore for highly pathogenic bacteria [23], *Escherichia coli* [10], VRE [24], improved resolution of hospital VRE isolates [25] and *Klebsiella pneumoniae* strain typing [26].

Despite significant improvements in ONT technology, including higher throughput and lower error rates, and increasing use for bacterial genome assembly [27, 28], challenges remain. These include the sensitivity of established typing methods like multi-locus sequence typing (MLST) [29, 30], cgMLST or cgSNP [11] to single base errors potentially introduced by ONT sequencing. Consequently, there is a need to compare data generated with ONT to the more accurate Illumina sequencing for outbreak detection. For clinical diagnostics, where accuracy is key for correct treatment and control measures, these errors can be critical. This re-evaluation is particularly important given the recent improvement in ONT sequencing accuracy due to the development of newer flow cells and basecalling algorithms, making it increasingly reliable for clinical applications [23, 31]. For instance, the R10.4.1 flow cell version has improved modal sequencing accuracy of 99.6% at the read level, which may be crucial for high-resolution genotyping necessary in clinical diagnostics. In addition, ONT transposase-based library preparation kits (e.g. RBK) now offer very fast turnaround time from isolates to sequencing results. Further, the possibility to combine PCR and PCR-free library preparation has not been fully assessed in terms of turnaround time, cost and effects on the genomic sequencing accuracy.

To validate ONT V14 kits for genomic epidemiology in outbreak settings, we performed a retrospective, comparative study using clinical isolates of *Corynebacterium diphtheriae* (CDIP) and VRE, which we obtained from our routine diagnostic laboratory. Beyond their clinical relevance, these isolates were selected because they were managed through a fully integrated workflow spanning isolation, storage, DNA extraction, sequencing and bioinformatic analysis conducted by the same expert team. This approach ensures our evaluations closely replicate real-world clinical diagnostic settings and methodological consistency. Recognizing that Illumina short-read sequencing may represent the accuracy gold standard, but may also be limited in resolving genomic structures, our goal was to determine if ONT long reads provide actionable data for outbreak decision-making. We focused on evaluating ONT performance in key areas: typing (MLST and cgMLST), antimicrobial resistance (AMR) and plasmid analysis in these two bacterial outbreak examples.

METHODS

Sample origin, cultivation and DNA extraction

All CDIP isolates were associated with local outbreaks in Switzerland from July to September 2022 and were described in a previous study [32]. The corresponding Illumina MiSeq data are available as BioProject PRJNA889706. VRE isolates, associated with local hospital outbreaks in a Swiss hospital, were collected between 2017 and 2021, with some of them belonging to ST796 [33]. They were isolated on CHROMagar VRE plates (CHROMagar, Paris, France) during routine analysis by the clinical microbiology laboratory of the Institute for Infectious Diseases (IFIK), University of Bern. Species identification was determined by MALDI-TOF (Bruker Daltonics, Bremen, Germany). Susceptibility testing for all isolates was performed during routine clinical processing at IFIK. All isolates were stored at -80°C and re-grown on CSBA plates before performing genomic extraction, thus minimizing the risk of *in vitro* evolution. Genomic DNA was extracted from

agarose plates using PureLink Genomic DNA kit (Thermo Fisher, Switzerland), or with Maxwell RSC Cultured Cells DNA Kit (Promega, Switzerland).

Whole-genome sequencing

NGS libraries for Illumina sequencing were produced with the Nextera DNA Flex Library Prep Kit (Illumina, Switzerland), according to the manufacturer's recommendations, and sequenced on an Illumina MiSeq sequencer with v2 reagents in 2×150 paired-end mode, at the Next Generation Sequencing Platform of the Institute for Infectious Diseases, Bern. Nanopore sequencing libraries were produced with the rapid barcoding chemistry SQK-RBK114.96 (RBK) and Rapid PCR barcoding chemistry SQK-RPB114.96 (RPB) according to the manufacturer's recommendations. The libraries were loaded onto standard GridION flowcells (FLO-MIN114) and sequenced in batches of 21 samples for 12–72 h on GridION X5 sequencers with real-time basecalling and under high-accuracy mode (HAC). The raw signal files (fast5_pass/pod5_pass directories) were subsequently re-basecalled either on the GridION or with standalone basecallers on a separate Linux workstation. All software versions and basecaller models are provided in Table S1 (available in the online Supplementary Material).

Ethics statement

All CDIP and VRE isolates in the present study have been anonymized, and no patient information is used in the interpretation of the bacterial genomic data. Publication of this analysis does not harm or influence either patients or institutions. Ethical committee approval was, therefore, not requested.

Bioinformatic analyses

Genome assembly and polishing. All genomes were reconstructed with the same software versions and parameters, except for the assembly parameters and for the models used for genome polishing, which were adjusted to the basecalling mode (SUP or HAC), and the basecalling model, respectively (Fig. 2). All reads were filtered to remove sequences shorter than 500 bases and those with an average quality score below 10. In addition, 30 bases were trimmed from both the 5' and 3' ends of each read to remove lower quality bases (*NanoFilt*, version 2.8.0, parameters: -q 10 -l 500 --headcrop 30 --tailcrop 30) [34]. The remaining reads were assembled with *Flye* (version 2.9.2) [35], and polished with *Medaka* (version 1.11.3) [36]. Detailed information, including the sequencing and basecalling software, basecalling models, assembly parameters and Medaka models, can be found in the supplementary material (Table S1). To mask bases with ambiguous base compositions in the corresponding reads, selected assemblies were processed with the software MPOA [37].

Combining PCR-based and PCR-free Nanopore reads. We tested two different approaches to combine RBK and RPB data: (a) Read mixtures. Random subsets of the reads produced with RBK and RPB were combined at theoretical coverage depths of 70× and 30×, respectively, to obtain mixed datasets with total coverage depths of 100×. These read sets were processed as described above for the non-mixed datasets. (b) Assembly correction. Reads sets produced with RBK and RPB were mapped separately against an assembly produced from the RBK reads with the software *minimap2* (version 2.24) [38]. The resulting SAM files were sorted (*samtools*, version 1.15.1) [39] and further analysed with the software *Pysamstats* (version 1.1.2) [40] to obtain per-strand base compositions for each assembly position. Base composition statistics were then processed using a custom Python script as follows (Fig. 1): (1) We assigned a confidence level (high or low) to each base of the original RBK assembly, based on the underlying base compositions of the mapped reads. Assembly positions were considered as 'low confidence' if one or more of the following criteria applied: Low read coverage (<20× coverage depth), mixed base composition (<80% of the reads matching the most frequent base) or strand bias (different dominant bases between the forward and reverse strand). Otherwise, the positions were considered as 'high confidence'. (2a) The RPB majority consensus base was obtained for each base position of the original assembly. (2b) We assigned a confidence level (high or low) to the RPB majority consensus bases, using the same criteria described in step one above. (3) The assembly was corrected for low confidence positions in the original assembly, which were replaced by the majority consensus of the RPB reads if the confidence was high for the latter and, by N, if the confidence in the RPB read set was low too. Each 'high confidence position' of the original assembly was kept.

Typing and comparison of the different treatments. All assemblies were imported into *SeqSphere+* (v.8.4.0; Ridom GmbH, Germany) and processed with the cgMLST schemes *Enterococcus faecium* cgMLST v1.1 [41] and a custom task template encompassing 1,319 core genes for CDIP based on the Institute Pasteur scheme, as described in [32]. Comparison tables (containing allele IDs for each target gene and assembly) were created for the Nanopore and Illumina assemblies of each analysed species and exported for further processing with custom Python and R scripts.

AMR and plasmid analyses. Protein sequences were predicted with the software *Prokka* (version 1.14.6) [42] and used as input for *AMRFinderPlus* (version 3.11.2, database: 2023-08-08.2, parameters -p -g -n --annotation_format prokka --organism) [43]. For plasmid analysis, WGS reads from treatments 'SUPD43', 'SUPDP43' and 'SUPD and P43' were reassembled and analysed with the 'long read' module with *Flye* as assembler and default parameters (*SeqSphere+*, version 10.0.0). Comparison

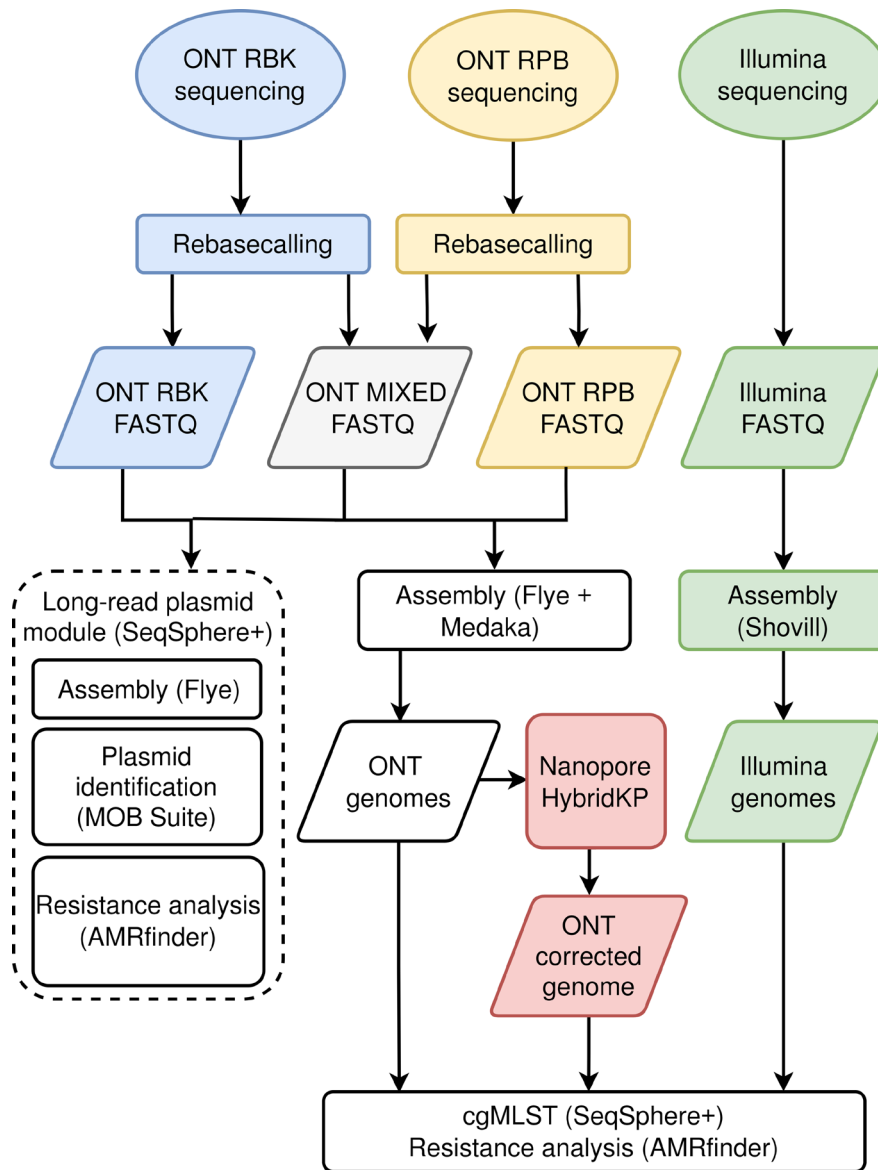


Fig. 2. Overview of the bioinformatic analyses performed in the study.

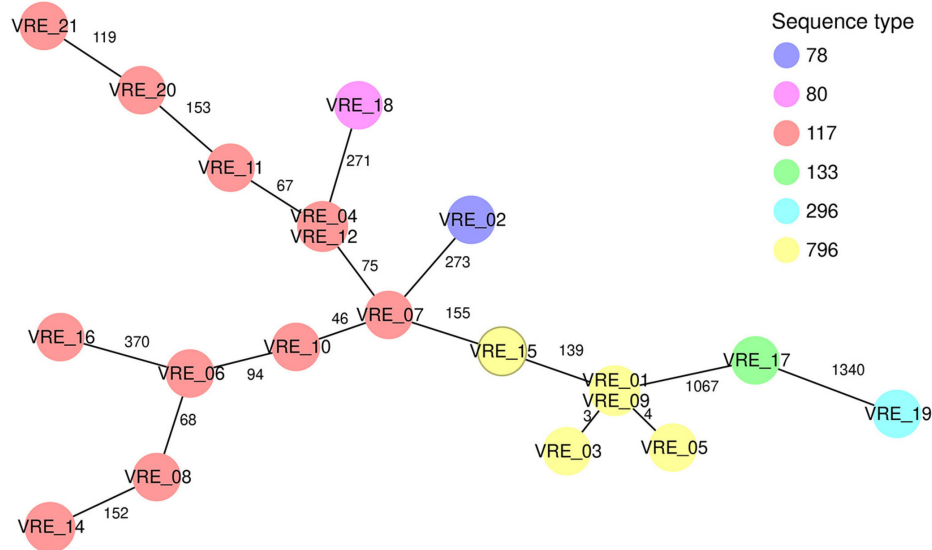
tables including the task template fields ‘Chromosome and plasmid Overview’, as well as a local plasmid database (created via ‘Create Task Template for Plasmid Mash Database’), were generated to facilitate data extraction from SeqSphere+. Assembly statistics (number of replicons, circularity) and plasmid-borne resistance were extracted from the resulting output tables and further analysed using custom Python scripts.

RESULTS

VRE genomic epidemiology using Nanopore vs. Illumina WGS

We choose a representative and diverse collection of clinical VRE isolates, consisting of 6 different MLST sequence types, 3 clonal clusters and 12 different cgMLST complex types based on Illumina assemblies (Fig. 3a). In terms of target gene recovery, we observed near-identical performance between Illumina and the top-performing Nanopore WGS methods: Percentages of successfully detected cgMLST target genes in the assemblies were highest for Illumina (98.80%), closely followed by the PCR-based Nanopore approach SUPDP43 (98.77%) and the most recent PCR-free approach SUPD50 (98.71%). The lowest percentage was found for the HAC43.masked (97.46%), which was still well above the quality threshold of 90% (Fig. 3b).

(a) cgMLST MST Illumina



(b) Percent good cgMLST targets

Treatment	Kit	PCR	Mode	Basecaller	Model version	VRE_16	VRE_14	VRE_12	VRE_21	VRE_02	VRE_04	VRE_05	VRE_18	VRE_07	VRE_08	VRE_11	VRE_10	VRE_17	VRE_19	VRE_20	VRE_03	VRE_09	VRE_01	VRE_15	VRE_06	min	max
HAC4k	RBK	no	HAC	G 6.3.9	default	98.7	99.4	99.2	99.0	99.1	99.2	98.7	97.9	99.2	99.0	99.3	99.3	97.7	94.7	96.3	-	-	98.3	98.2	98.3	94.7	99.4
SUP4k	RBK	no	SUP	G 6.3.9	default	98.7	99.4	99.2	99.0	99.2	99.2	98.7	98.0	99.1	99.2	99.3	99.2	97.7	94.9	97.5	-	-	98.5	98.7	98.6	94.9	99.4
HAC5k	RBK	no	HAC	G 6.5.7	v4.1	98.6	99.4	99.2	99.0	99.2	99.2	98.5	98.0	99.2	99.1	99.2	99.2	97.7	94.9	99.0	98.2	98.4	98.1	98.7	98.7	94.9	99.4
SUPD5k	RBK	no	SUP	D 0.3.3	v4.2	98.7	99.3	99.2	98.9	99.2	99.2	98.6	98.0	99.1	99.2	99.2	99.2	97.5	94.9	99.0	98.2	98.4	98.2	98.5	98.9	94.9	99.3
HAC43	RBK	no	HAC	D 0.5.0	v4.3	98.6	99.2	99.2	99.0	99.0	99.2	98.7	98.0	99.2	99.2	99.3	99.2	97.4	94.7	99.0	98.3	98.5	98.5	98.8	98.9	94.7	99.3
SUPD43	RBK	no	SUP	D 0.5.0	v4.3	98.7	99.3	99.2	98.9	99.2	99.2	98.8	97.9	99.0	99.2	99.1	99.3	97.9	94.9	99.0	98.5	98.7	98.8	99.0	99.2	94.9	99.3
HAC50	RBK	no	HAC	D 0.7.0	v5.0	98.6	99.2	99.2	99.0	99.2	99.2	98.7	97.9	99.2	99.0	99.2	99.2	97.7	94.8	98.9	98.6	98.5	98.4	98.6	98.8	94.8	99.2
SUPD50	RBK	no	SUP	D 0.7.0	v5.0	98.6	99.4	99.2	99.0	99.2	99.2	98.8	97.9	99.2	99.2	99.2	99.3	97.9	94.9	99.0	98.8	98.5	98.8	99.0	99.3	94.9	99.4
HACP43	RPB	yes	HAC	D 0.5.0	v4.3	98.6	99.4	99.1	98.9	99.1	99.2	98.7	97.8	99.2	99.2	99.2	99.2	97.8	94.8	99.0	98.8	98.8	98.8	99.0	99.2	94.8	99.4
SUPDP43	RPB	yes	SUP	D 0.5.0	v4.3	98.7	99.4	99.2	99.0	99.1	99.2	98.8	97.9	99.2	99.2	99.2	99.3	97.9	94.9	99.0	98.9	98.9	98.9	99.0	99.4	94.9	99.4
SUPD&P43	SUPD43 (70%) mixed with SUPDP43 (30%)					98.7	99.4	99.2	98.8	99.2	99.2	98.8	97.8	99.2	99.2	99.3	99.3	98.0	94.9	99.0	98.9	98.8	98.7	99.0	99.4	94.9	99.4
HAC43CC	HAC43 corrected with HACP43 reads					98.6	99.2	99.2	99.0	99.0	99.2	98.7	97.9	99.1	99.2	99.3	99.2	97.5	94.7	99.0	98.4	98.6	98.5	98.8	99.0	94.7	99.3
SUPD43CC	SUPD43 corrected with SUPD43 reads					98.7	99.3	99.2	98.9	99.2	99.2	98.8	97.8	99.0	99.2	99.1	99.3	97.9	94.9	99.0	98.5	98.7	98.8	99.0	99.2	94.9	99.3
Illumina	DNA	yes	-	-	-	98.7	99.4	99.2	99.0	99.2	99.2	98.8	98.3	99.2	99.2	99.3	99.3	98.0	94.9	99.0	98.9	98.9	98.9	99.0	99.4	94.9	99.4

(c) Numbers of mismatching cgMLST targets, and mismatching complex types between Illumina and Nanopore

Treatment	Kit	PCR	Mode	Basecaller	Model version	VRE_16	VRE_14	VRE_12	VRE_21	VRE_02	VRE_04	VRE_05	VRE_18	VRE_07	VRE_08	VRE_11	VRE_10	VRE_17	VRE_19	VRE_20	VRE_03	VRE_09	VRE_01	VRE_15	VRE_06	min	max
HAC4k	RBK	no	HAC	G 6.3.9	default	0	0	0	0	1	0	0	1	4	4	5	4	6	6	19	-	-	27	20	32	0.0	32.0
SUP4k	RBK	no	SUP	G 6.3.9	default	0	0	0	0	0	0	2	1	2	2	1	4	6	11	16	-	-	44	39	39	0.0	44.0
HAC5k	RBK	no	HAC	G 6.5.7	v4.1	0	0	0	0	0	1	1	1	6	5	9	7	4	6	0	11	16	8	15	14	0.0	16.0
SUPD5k	RBK	no	SUP	D 0.3.3	v4.2	0	0	0	0	0	1	0	1	1	2	3	5	16	10	4	30	29	24	28	38	0.0	38.0
HAC43	RBK	no	HAC	D 0.5.0	v4.3	0	0	0	0	0	0	0	1	1	0	0	0	11	9	3	22	30	19	19	27	0.0	30.0
SUPD43	RBK	no	SUP	D 0.5.0	v4.3	0	0	0	0	0	0	0	1	0	0	0	0	2	1	0	0	6	4	6	3	0.0	6.0
HAC50	RBK	no	HAC	D 0.7.0	v5.0	0	0	0	0	0	0	0	1	0	0	0	0	5	6	1	20	16	15	18	20	0.0	20.0
SUPD50	RBK	no	SUP	D 0.7.0	v5.0	0	0	0	0	0	0	1	2	0	1	0	0	0	0	0	0	4	1	0	1	0.0	4.0
HACP43	RPB	yes	HAC	D 0.5.0	v4.3	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0.0	1.0
SUPDP43	RPB	yes	SUP	D 0.5.0	v4.3	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0.0	2.0
SUPD&P43	SUPD43 (70%) mixed with SUPDP43 (30%)					0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0.0	2.0
HAC43CC	HAC43 corrected with HACP43 reads					0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0.0	1.0
SUPD43CC	SUPD43 corrected with SUPD43 reads					0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0.0	1.0

 mismatching complex type (ct) between Illumina and Nanopore

Fig. 3. Comparison of cgMLST results for clinical VRE isolates sequenced with Illumina or Nanopore. The cgMLST scheme contained 1,423 target genes. (a) Minimum spanning tree of the 20 Illumina assemblies. (b) Percentage of the 1,423 cgMLST target genes that were detected (and passed the QC) in the different assemblies. (c) Number of mismatching alleles between the ONT and corresponding Illumina assemblies, and mismatching complex types between the two technologies (indicated by red boxes). Letters G and D in the column Basecaller stand for Guppy and Dorado, respectively. A detailed explanation about treatments is provided in Table S1.

A near-perfect agreement was found between Illumina and the top-performing Nanopore methods for the individual cgMLST allele assignments. The numbers of mismatching alleles between the Illumina and Nanopore assemblies varied strongly between isolates, Nanopore software versions, basecalling models and the library preparation kits: For VRE21, for example, no mismatching alleles were detected, irrespective of the approach used to create the Nanopore assemblies, whereas 0 to 44 mismatching alleles were detected in the assemblies produced from isolate VRE01 (Fig. 3c). Noticeably, PCR-based library preparation outperformed PCR-free alternatives irrespective of the basecalling models used, allowing for a perfect replication of the allelic profiles obtained with Illumina in all but one case, with a maximum number of two mismatches (VRE18). Mixtures of PCR-based and PCR-free reads, as well as PCR-free assemblies that were corrected with PCR-based reads, performed on par with PCR-free reads in these aspects.

The PCR-free approach was also more sensitive to the analysis parameters: SUP models clearly outperformed HAC models, and the most recent iterations of the software (and models) outperformed previous versions. The best combination of parameters resulted in perfectly matching allele assignments in 14 out of 20 cases with a maximum of four mismatches (Fig. 3c). Subsequent masking with the software *MPOA* resulted in the exclusion of all but one of the previously mentioned mismatching alleles (*SeqSphere+* excludes genes with N's from the analysis), while the percentage of successfully detected cgMLST targets dropped by 0.8%, remaining well above the threshold of 90% (SUPD34, 98.68%; SUP43.masked, 97.90%, Fig. S1A, B).

MLST classification into sequence type (ST) was successful across all tested combinations that did not involve masking with *MPOA* (Fig. S1D). The latter resulted in two unassigned STs in the treatment HAC43.masked (Fig. S1D), whereas treatment SUP43.masked did not result in a missing ST assignment. cgMLST classification into complex type (CT) was more sensitive to the variability introduced by the different treatments: The PCR-based treatments SUPDP43 and HACP43 matched the Illumina CT in all cases and in 19 out of 20 cases, respectively. The single CT 1552 (instead of CT 2887) that differed between the PCR-based treatments was a result of one target gene (EFAU004_01770, hypothetical protein), which was exclusively found in SUPDP43 in otherwise identical allelic profiles (Fig. S1D). The mixtures of PCR-based and PCR-free reads matched the Illumina CT in all cases and the PCR-free assemblies that were corrected with PCR-based reads in 19 out of 20 cases, respectively. The single mismatch was caused by the absence of the same target gene, as described above for the PCR-based approaches. Among the PCR-free treatments, SUPD50 performed the best with all 20 correct CTs assigned, followed by SUPD43 with 17/20 correct assignments (Fig. 3c). Subsequent *MPOA* masking (SUP43.masked) corrected two of the remaining mis-assigned CTs, but another CT was assigned for VRE10, while no CT was assigned for it based on the Illumina assembly (Fig. S1).

Noticeably, a large proportion of the mismatching bases between Illumina and Nanopore assemblies were found in assembly positions with conflicting information between the forward and reverse reads (Fig. 4a). This phenomenon is referred to as strand bias (SB) in subsequent sections. We analysed the frequency of SB as a function of its genomic context and MLST types (Fig. 4b). In our dataset, SB was more often observed on plasmids than on chromosomes. No distinct pattern related to ST was identified; instead, SB levels appeared to be strain-specific rather than ST-specific (Fig. 4b).

VRE-predicted AMR and plasmids

To evaluate whether ONT assemblies are suitable for AMR analysis, we applied AMRFinderPlus to the Illumina and ONT assemblies: On the class level, predictions for the Illumina and ONT-based assemblies were matched in all cases (Fig. 5a). On the antibiotic subclass level, discrepancies were found for one isolate (VRE_02), for which gentamicin and tobramycin resistances were exclusively predicted for the ONT assemblies (Fig. 5b). On the gene level, the bifunctional aminoglycoside-modifying enzyme *AAC(6')-Ie-APH(2'')-Ia*, which confers resistance to all commercially available aminoglycosides except streptomycin [44], was exclusively detected in the ONT assemblies in five cases (Fig. 5c). However, in all but one isolate (VRE_02), the C-terminal domain of *AAC(6')-Ie-APH(2'')-Ia* was detected in the Illumina and ONT assemblies, explaining the matching resistance predictions (Fig. 6a, b).

Different Nanopore approaches were compared with respect to their performance for plasmid analysis (PCR-free, PCR-based and mixtures of both). PCR-free and mixed Nanopore methods showed superior chromosome assembly performance than PCR-based Nanopore assemblies, as PCR-based Nanopore assemblies showed a higher degree of chromosome fragmentation than PCR-free or mixed assemblies. Plasmid assembly performance did not differ significantly between the tested methods (Fig. 6).

Discrepancies in the predicted genomic context of resistance genes were identified in 4 out of the 19 isolates when comparing the tested Nanopore approaches: Tetracycline resistance in VRE02, along with gentamicin and tobramycin resistance in VRE16, was determined to be plasmid-mediated in PCR-free and mixed assemblies, but not in PCR-based assemblies. Conversely, amikacin, gentamicin, kanamycin and tobramycin resistance in VRE14 were identified as plasmid-mediated only in the PCR-based assemblies. Vancomycin resistance was predicted to be plasmid-mediated in PCR-based and mixed assemblies of VRE20, but not in the PCR-free variant (Fig. 7).

(a)

	VRE_01	VRE_02	VRE_03	VRE_04	VRE_07	VRE_08	VRE_09	VRE_10	VRE_11	VRE_12	VRE_14	VRE_15	VRE_16	VRE_17	VRE_18	VRE_19	VRE_20	VRE_21
AMINOGLYCOSIDE	111	111	111	111	111	111	111	111	111	111	111	111	111	111	111	111	111	111
BETA-LACTAM	111	111	111	111	111	111	111	111	111	111	111	111	111	111	111	111	111	111
GLYCOPEPTIDE	111	111	111	111	111	111	111	111	111	111	111	111	111	111	111	111	111	111
LINCOSAMIDE/MACROLIDE/STREPTOGRAMIN	111	111	111	111	111	111	111	111	111	111	111	111	111	111	111	111	111	111
LIPOPEPTIDE											111		111					
MACROLIDE				111	111			111	111	111							111	111
MACROLIDE/STREPTOGRAMIN	111	111	111	111	111	111	111	111	111	111	111	111	111	111	111	111	111	111
PHENICOL		111			111			111										
PLEUROMUTILIN														111	111	111		
QUINOLONE	111	111	111	111	111	111	111	111	111	111	111	111	111			111	111	111
STREPTOTHRICIN				111		111			111	111			111			111		111
TETRACYCLINE	111	111	111		111			111					111	111	111			
TRIMETHOPRIM	111	111	111	111	111			111	111	111	111	111			111		111	111

(b)

	VRE_01	VRE_02	VRE_03	VRE_04	VRE_07	VRE_08	VRE_09	VRE_10	VRE_11	VRE_12	VRE_14	VRE_15	VRE_16	VRE_17	VRE_18	VRE_19	VRE_20	VRE_21
AMIKACIN/GENTAMICIN/KANAMYCIN/TOBRAMYCIN	111	110	111		111	111	111	111			111	111	111				111	111
AMIKACIN/KANAMYCIN				111					111	111					111		111	111
AMINOGLYCOSIDE	111	111	111	111	111	111	111	111	111	111	111	111	111	111	111	111	111	111
AZITHROM./ERYTHROM./STREPTOGRAMIN B/TYLOSIN	111	111	111	111	111	111	111	111	111	111	111	111	111	111	111	111	111	111
BETA-LACTAM	111	111	111	111	111	111	111	111	111	111	111	111	111		111		111	111
CHLORAMPHENICOL		111			111			111										
CLINDAMYCIN/ERYTHROMYCIN/STREPTOGRAMIN B	111	111	111	111	111	111	111	111	111	111	111	111	111		111		111	111
DAPTOMYCIN											111		111					
ERYTHROMYCIN				111	111			111	111	111							111	111
PLEUROMUTILIN								111	111	111				111	111	111		
QUINOLONE	111	111	111	111	111	111	111	111	111	111	111	111	111		111		111	111
STREPTOMYCIN		111		111	111	111		111	111	111		111	111		111		111	111
STREPTOTHRICIN				111		111			111	111					111		111	111
TETRACYCLINE	111	111	111		111		111					111	111	111				
TRIMETHOPRIM	111	111	111	111	111		111	111	111	111	111	111			111		111	111
VANCOMYCIN	111	111	111	111	111	111	111	111		111	111	111	111	111	111		111	111

(c)

	VRE_01	VRE_02	VRE_03	VRE_04	VRE_07	VRE_08	VRE_09	VRE_10	VRE_11	VRE_12	VRE_14	VRE_15	VRE_16	VRE_17	VRE_18	VRE_19	VRE_20	VRE_21
aac(6)-I	111	111	111	111	111	111	111	111	111	111	111	111	111	111	111	111	111	111
aac(6)-Ie/aph(2'')-Ia	110	110	110			111	110	110			111	111	111				111	111
ant(6)-Ia		111		111	111	111		111	111	111		111	111		111		111	111
aph(2'')-I		110									111	111						
aph(2'')-Ia	111		111			111	111	111					110					
aph(3')-IIIa				111		111			111	111		111	111		111		111	111
catA		111			111			111										
cls_R218Q											111							
dfrF		111			111			111			111	111			111			
dfrG	111		111	111	111		111	111	111	111		111					111	111
eatA_T450I														111	111	111		
erm(B)	111	111	111	111	111	111	111	111	111	111	111	111	111		111		111	111
gyrA_S83I	111		111			111	111				111				111		111	
gyrA_S83Y		111		111	111			111	111	111		111	111					
liaR_W73C											111		111					
liaS_T120A											111		111					
mef(H)				111	111			111	111	111							111	111
mrs(C)	111	111	111	111	111	111	111	111	111	111	111	111	111	111	111	111	111	111
parC_S80I	111		111	111	111	111	111	111	111	111	111	111	111		111		111	111
parC_S80R		111																
pbb5_M485A	111	111	111	111	111	111	111	111	111	111	111	111	111		111		111	111
sat4				111		111			111	111					111		111	111
tet(L)														111				
tet(M)	111	111	111		111		111					111		111				
vanA		111			111			111						111	111	111		
vanB	111		111	111		111	111				111	111	111				111	111
vanH-A		111			111			111						111	111	111		
vanH-B	111		111	111		111	111				111	111	111				111	111
vanR-A		111			111			111					111	111				
vanR-B	111		111	111		111	111				111	111	111				111	111
vanS-A		111			111			111					111	111				
vanS-B	111		111	111		111	111				111	111	111				111	111
vanW-B	111		111	111		111	111				111	111	111				111	111
vanX-A		111			111			111					111	111	111			
vanX-B	111		111	111		111	111				111	111	111				111	111
vanY-A		111			111			111					111	111	111			
vanY-B	111		111	111		111	111				111	111	111				111	111
vanZ-A		111			111			111					111	111	111			

Fig. 5. Comparison of the predicted resistance genes based on ONT and Illumina assemblies. Tuples indicate presence (1) or absence (0) in the treatments SUBDP_43 (RPB), SUPD_43 (RBK) and ILLUM (Illumina) with (a) antibiotic classes, (b) antibiotic subclasses, (c) resistance genes.

CDIP_18). Specifically, the aminoglycoside phosphotransferase gene *aph(3')-Ia*, conferring kanamycin resistance, was absent in the PCR-based ONT assemblies of both isolates (Fig. S3B, C).

Time and cost estimation

The ONT approaches used in this study are competitive with Illumina both in terms of hands-on time, overall turnaround time and costs per sample at the tested throughput levels (Table 1). For example, preparing 20 samples takes only 0.7 h (RBK) or 1 h (RPB) hands-on time, significantly less than the 1.8 h required for Illumina MiSeq. The overall turnaround time is also notably faster for smaller batches: For four samples, ONT (RBK or RPB) finishes sequencing in 5 h, while Illumina requires 24 h.

(a)

	Chromosome Circular	Chromosome Noncircular	Plasmid Circular	Plasmid Noncircular	AMR Plasmid Noncircular	AMR Plasmid Circular
VRE_08-N	1	1	5	7	1	1
VRE_08-P	0	12	6	9	1	1
VRE_08-X	1	0	6	11	1	1
VRE_09-N	1	0	5	8	1	1
VRE_09-P	0	9	5	9	1	1
VRE_09-X	1	1	5	7	1	1
VRE_10-N	0	3	1	5	1	1
VRE_10-P	0	18	0	9	2	0
VRE_10-X	0	4	1	3	0	1
VRE_11-N	1	1	5	2	0	1
VRE_11-P	0	8	4	4	0	1
VRE_11-X	1	0	3	4	0	1
VRE_12-N	0	5	5	2	0	1
VRE_12-P	0	11	6	3	0	1
VRE_12-X	1	1	4	6	0	1
VRE_14-N	0	3	4	2	0	1
VRE_14-P	0	15	5	5	1	1
VRE_14-X	0	3	4	2	0	1
VRE_15-N	1	1	3	4	2	0
VRE_15-P	0	9	3	4	1	1
VRE_15-X	1	0	3	8	2	0

	Chromosome Circular	Chromosome Noncircular	Plasmid Circular	Plasmid Noncircular	AMR Plasmid Noncircular	AMR Plasmid Circular
VRE_08-N	1	1	5	7	1	1
VRE_08-P	0	12	6	9	1	1
VRE_08-X	1	0	6	11	1	1
VRE_09-N	1	0	5	8	1	1
VRE_09-P	0	9	5	9	1	1
VRE_09-X	1	1	5	7	1	1
VRE_10-N	0	3	1	5	1	1
VRE_10-P	0	18	0	9	2	0
VRE_10-X	0	4	1	3	0	1
VRE_11-N	1	1	5	2	0	1
VRE_11-P	0	8	4	4	0	1
VRE_11-X	1	0	3	4	0	1
VRE_12-N	0	5	5	2	0	1
VRE_12-P	0	11	6	3	0	1
VRE_12-X	1	1	4	6	0	1
VRE_14-N	0	3	4	2	0	1
VRE_14-P	0	15	5	5	1	1
VRE_14-X	0	3	4	2	0	1
VRE_15-N	1	1	3	4	2	0
VRE_15-P	0	9	3	4	1	1
VRE_15-X	1	0	3	8	2	0

	Chromosome Circular	Chromosome Noncircular	Plasmid Circular	Plasmid Noncircular	AMR Plasmid Noncircular	AMR Plasmid Circular
VRE_16-N	1	0	6	0	0	1
VRE_16-P	0	5	7	0	0	1
VRE_16-X	1	2	5	0	0	1
VRE_17-N	1	0	1	0	0	1
VRE_17-P	0	8	1	0	0	1
VRE_17-X	1	1	1	0	0	1
VRE_18-N	1	1	6	3	0	2
VRE_18-P	0	14	4	11	1	1
VRE_18-X	1	0	5	7	1	1
VRE_19-N	1	0	1	0	0	0
VRE_19-P	0	4	1	0	0	0
VRE_19-X	1	0	1	0	0	0
VRE_20-N	0	2	5	1	0	2
VRE_20-P	0	13	3	11	2	0
VRE_20-X	0	2	5	3	1	2
VRE_21-N	1	0	7	1	0	2
VRE_21-P	0	11	7	2	0	2
VRE_21-X	1	0	6	2	0	2

(b)

Number of replicons

[min/max/average]

	RBK	RPB	MIXED
Chromosome circular	0/1/0.68	0/0/0	0/1/0.74
Chromosome noncircular	0/7/1.74	5/19/11.74	0/7/1.63
Plasmid circular	1/7/4.11	0/7/3.84	1/6/3.95
Plasmid noncircular	0/10/3.89	0/14/6.63	0/11/5
AMR_Plasmid_Circular	0/2/1.05	0/2/0.79	0/2/1
AMR_Plasmid_Noncircular	0/2/0.53	0/2/0.84	0/2/0.58

Fig. 6. Assembly performance of PCR-based, PCR-free and mixed Nanopore reads. (a) Numbers of contigs by categories based on replicon type, circularity and presence of AMR genes. (b) Summary of the data provided in (a). Grey shadings indicate significant differences to the other treatments within one category (Mann-Whitney *U* test; corrected *P*-value for multiple testing, *P* < 0.05).

	Amikacin	Chloramphenicol	Clindamycin	Copper	Erythromycin	Gentamicin	Kanamycin	Streptogramin B	Streptomycin	Streptothricin	Tetracycline	Tobramycin	Vancomycin
VRE_01-N	1	0	1	0	1	1	1	1	0	0	0	1	0
VRE_01-P	1	0	1	0	1	1	1	1	0	0	0	1	0
VRE_01-X	1	0	1	0	1	1	1	1	0	0	0	1	0
VRE_02-N	1	1	1	0	1	1	1	1	1	0	1	1	1
VRE_02-P	1	1	1	0	1	1	1	1	1	0	0	1	1
VRE_02-X	1	1	1	0	1	1	1	1	1	0	1	1	1
VRE_03-N	1	0	1	0	1	1	1	1	0	0	0	1	0
VRE_03-P	1	0	1	0	1	1	1	1	0	0	0	1	0
VRE_03-X	1	0	1	0	1	1	1	1	0	0	0	1	0
VRE_04-N	1	0	1	0	1	0	1	1	1	1	0	0	0
VRE_04-P	1	0	1	0	1	0	1	1	1	1	0	0	0
VRE_04-X	1	0	1	0	1	0	1	1	1	1	0	0	0
VRE_05-N	1	0	1	0	1	1	1	1	0	0	0	1	0
VRE_05-P	1	0	1	0	1	1	1	1	0	0	0	1	0
VRE_05-X	1	0	1	0	1	1	1	1	0	0	0	1	0
VRE_06-N	1	0	1	0	1	0	1	1	1	1	0	0	0
VRE_06-P	1	0	1	0	1	0	1	1	1	1	0	0	0
VRE_06-X	1	0	1	0	1	0	1	1	1	1	0	0	0
VRE_07-N	0	1	1	0	1	0	0	1	1	0	0	0	1
VRE_07-P	0	1	1	0	1	0	0	1	1	0	0	0	1
VRE_07-X	0	1	1	0	1	0	0	1	1	0	0	0	1
VRE_08-N	2	0	1	0	1	1	2	1	1	1	0	1	0
VRE_08-P	2	0	1	0	1	1	2	1	1	1	0	1	0
VRE_08-X	2	0	1	0	1	1	2	1	1	1	0	1	0
VRE_09-N	1	0	1	0	1	1	1	1	0	0	0	1	0
VRE_09-P	1	0	1	0	1	1	1	1	0	0	0	1	0
VRE_09-X	1	0	1	0	1	1	1	1	0	0	0	1	0
VRE_10-N	0	2	1	0	1	0	0	1	1	0	0	0	2
VRE_10-P	0	1	1	0	1	0	0	1	1	0	0	0	1
VRE_10-X	0	1	1	0	1	0	0	1	1	0	0	0	1
VRE_11-N	1	0	1	0	1	0	1	1	1	1	0	0	0
VRE_11-P	1	0	1	0	1	0	1	1	1	1	0	0	0
VRE_11-X	1	0	1	0	1	0	1	1	1	1	0	0	0
VRE_12-N	1	0	1	0	1	0	1	1	1	1	0	0	0
VRE_12-P	1	0	1	0	1	0	1	1	1	1	0	0	0
VRE_12-X	1	0	1	0	1	0	1	1	1	1	0	0	0
VRE_14-N	0	0	1	0	1	0	0	1	0	0	0	0	0
VRE_14-P	1	0	1	0	1	1	1	1	0	0	0	1	0
VRE_14-X	0	0	1	0	1	0	0	1	0	0	0	0	0
VRE_15-N	2	0	1	0	1	1	2	1	1	1	0	1	0
VRE_15-P	2	0	1	0	1	1	2	1	1	1	0	1	0
VRE_15-X	2	0	1	0	1	1	2	1	1	1	0	1	0
VRE_16-N	1	0	1	0	1	1	1	1	1	0	1	1	1
VRE_16-P	1	0	1	0	1	0	1	1	1	0	1	0	1
VRE_16-X	1	0	1	0	1	1	1	1	1	0	1	1	1
VRE_17-N	0	0	0	1	0	0	0	0	0	0	0	0	1
VRE_17-P	0	0	0	1	0	0	0	0	0	0	0	0	1
VRE_17-X	0	0	0	1	0	0	0	0	0	0	0	0	1
VRE_18-N	1	0	1	0	1	0	1	1	1	1	0	0	1
VRE_18-P	1	0	1	0	1	0	1	1	1	1	0	0	1
VRE_18-X	1	0	1	0	1	0	1	1	1	1	0	0	1
VRE_20-N	2	0	2	0	2	1	2	2	2	2	0	1	0
VRE_20-P	2	0	2	0	2	1	2	2	2	2	0	1	1
VRE_20-X	2	0	2	0	2	1	2	2	2	2	0	1	1
VRE_21-N	2	0	1	0	1	1	2	1	1	1	0	1	1
VRE_21-P	2	0	1	0	1	1	2	1	1	1	0	1	1
VRE_21-X	2	0	1	0	1	1	2	1	1	1	0	1	1

Fig. 7. Plasmid-borne resistance. Numbers indicate the number of resistance genes detected for each specific antibiotic subclass.

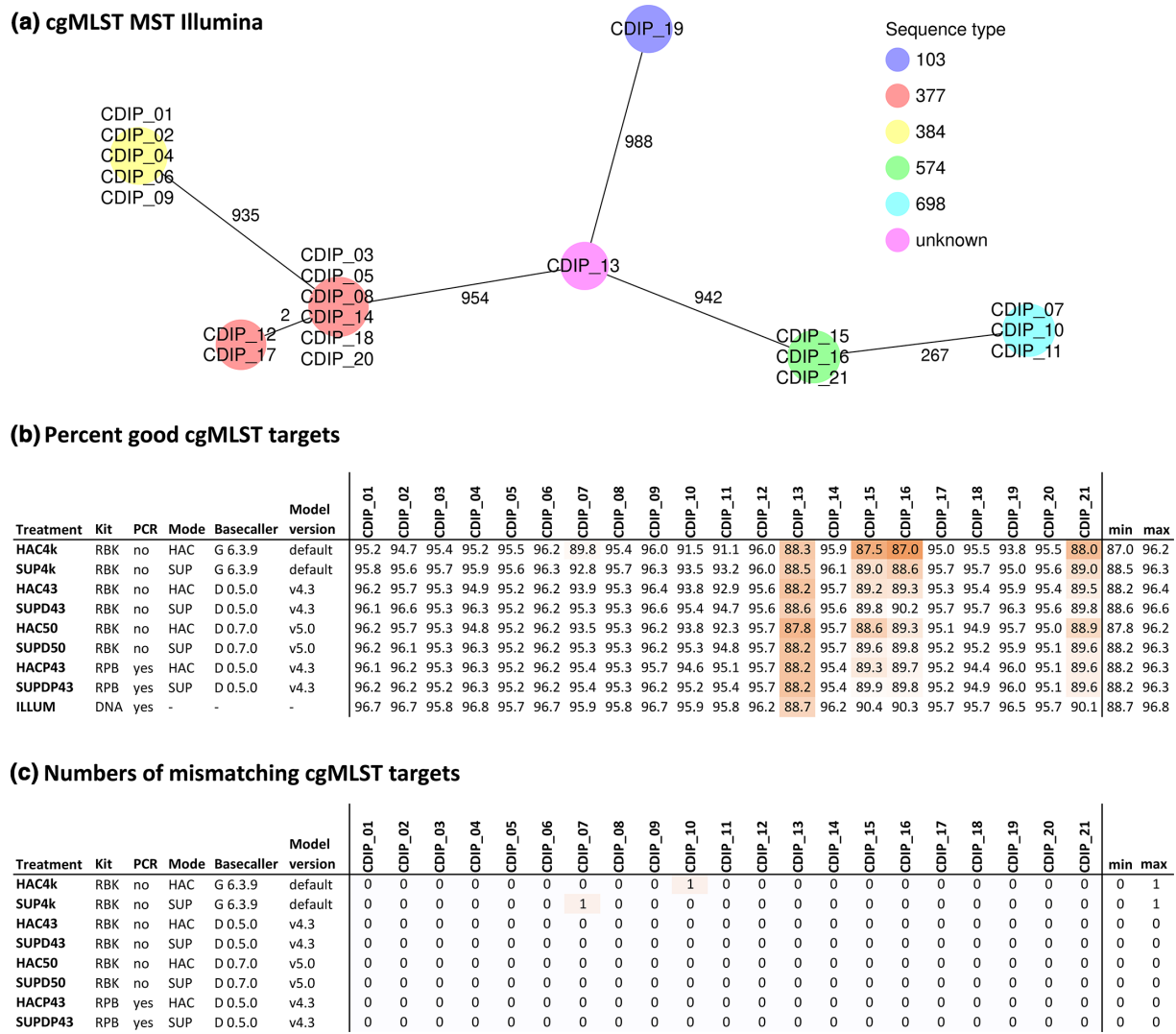


Fig. 8. Comparison of cgMLST results from 21 clinical *C. diphtheriae* isolates sequenced in parallel with Illumina and ONT. The cgMLST scheme contained 1,312 target genes. Treatment: Sequencing and data processing protocol used (Table S1). (a) Minimum spanning tree of the 21 Illumina assemblies. (b) Percentage of the 1,312 cgMLST target genes that were detected (and passed the QC) in the different assemblies. (c) Numbers of mismatching alleles between the ONT and the corresponding Illumina assemblies.

There is a slight cost increase when combining ONT RBK with RPB workflows, as the latter involves the use of long-range PCR in addition to the transposase library preparation as in the RBK protocol. Most critically, the cost reduction is substantial: For batches of 20 samples, the cost is estimated at \$310 (RBK) or \$400 (RPB), compared to \$2,340 for the MiSeq run, leading to a cost difference per sample of \$117 for MiSeq vs. \$36 for the combination of RBK and RPB on the same sample, so a 3.25-fold cost reduction. This significant cost reduction can be obtained given the possibility to wash ONT flow cells with a DNase treatment and to stop the sequencing run earlier when enough reads have been obtained, as compared to the Illumina-based approach, which entails fixed costs and run durations (Table 1).

DISCUSSION

We assessed the recovery of cgMLST target genes, allele assignments, MLST and cgMLST classification, prediction of AMR genes and plasmid content using various Nanopore library preparation methods (PCR-based, PCR-free and mixtures) and analysis pipelines. We evaluated a diverse, clinically relevant collection of outbreak isolates, originating from local outbreaks, which were classified as different MLST sequence types and clonal clusters based on Illumina WGS assemblies. Our main findings indicate that MLST classification was generally successful across Nanopore approaches, while cgMLST CT classification showed more sensitivity to the different Nanopore approaches. We also explored the suitability

Table 1. Turnaround time and cost estimates

	Illumina MiSeq	ONT RBK*	ONT RPB†
Library preparation (hands-on/total)			
<i>n</i> =4 samples	1.5/2.7 h	0.5/1.0 h	0.7/3.0 h
<i>n</i> =20 samples	1.8/2.7 h	0.7/1.2 h	1.0/3.3 h
Sequencing time			
<i>n</i> =4 samples	24 h	5 h	5 h
<i>n</i> =20 samples	24 h	24 h	24 h
Cost estimates per batch (per sample)‡			
<i>n</i> =4 samples	\$468 (\$117)	\$200 (\$50)	\$300 (\$75)
<i>n</i> =20 samples	\$2,340 (\$117)	\$310 (\$16)	\$400 (\$20)

*ONT transposase-based library preparation (RBK).

†ONT transposase-based followed by PCR-based library preparation (RPB).

‡Costs are estimated at \$117 per sample for MiSeq based on 20 isolates per run, excluding personnel costs. For ONT runs, we hypothesized conservatively that 24 h and 5 h runs correspond to half and a fourth of the maximal lifespan of a flow cell, whose total cost is estimated at \$500. The costs are indicative and may vary depending on the country and specific contract with the involved companies.

of Nanopore assemblies for AMR analysis, revealing generally concordant results with Illumina at the class level for AMR but some discrepancies at the subclass and gene levels. The nanopore workflows that performed best in cgMLST were further compared with respect to their ability to predict the genomic context of resistance genes using the SeqSphere+ Long read Data Plasmid Transmission Analysis Module. This analysis yielded largely consistent results, with only a few discrepancies. We selected this approach because it provides a straightforward framework for plasmid characterization and demonstrated higher consistency than a plasmid analysis in SeqSphere+ based on the preassembled contigs (Figs S4 and S5). Finally, the study provided turnaround time and cost estimation, suggesting that the tested Nanopore approaches are competitive with those offered by short-read NGS approaches. With more than 550 resulting assemblies analysed with MLST and cgMLST to determine clonal complexes and accuracy of the epidemiological conclusions, we highlighted how type of WGS data produced combined with the software running on Nanopore sequencers, as well as parameters selected by the operator, can have a major influence on the quality of the resulting sequencing reads and the epidemiological conclusions derived from the WGS data. Our results have practical consequences for genomic studies that aggregate sequencing data originating from different laboratories using different technologies or software versions.

While Nanopore demonstrated near-identical performance to Illumina for cgMLST target gene recovery and MLST typing, some discrepancies were observed in cgMLST allele assignments and AMR gene detection depending on the specific Nanopore approach used. Notably, PCR-based Nanopore library preparation generally outperformed PCR-free methods for allelic profiling in VRE. We concluded that optimized Nanopore sequencing and analysis pipelines offer a competitive alternative to Illumina for genomic surveillance of these pathogens in terms of accuracy, time and cost. Complete concordance with Illumina results was observed for MLST in both tested species and for cgMLST in CDIP across all ONT kits and software evaluated. In contrast, the accuracy of cgMLST results for VRE varied based on the strain, library preparation kit and analysis parameters, likely due to challenges in resolving base methylations. The latest software ($\geq$ Dorado 0.5.0) and basecalling model versions ($\geq$ v4.3), combined with the PCR-based library preparation kit (RPB), reliably reproduced Illumina cgMLST results across all tested VRE strains.

Previous studies assessing ONT performance for accurate genome reconstruction often did not use the latest generation of flow cells, sequencing kits and software. For instance, Foster-Nyarko *et al.* [45] found R9.4.1/V10 chemistry reliable for MLST and AMR in *K. pneumoniae*, but cgMLST/SNP cluster detection remained challenging. Greig *et al.* [10] noted 95% of nt discrepancies between Illumina and ONT data for two isolates of Shiga toxin-producing *E. coli* (STEC) O157:H7, which could be resolved by masking methylated and prophage regions in ONT data. More recent R10.4/V12 studies showed mixed results: [13] found V12 chemistry accuracy promising but throughput insufficient for large-scale studies. [23] found high-resolution cgMLST/cgSNP genotyping feasible for *F. tularensis* and *B. anthracis*, but not *B. suis*. [28] concluded R10.4 cells yield ‘near-finished’ bacterial genomes without short-read polishing, though they did not test high-resolution typing. Conversely in [14], V12 chemistry accuracy was found to be sufficient for *Bordetella pertussis* clinical epidemiology via cgMLST. Despite V12’s accuracy improvements, especially for homopolymer errors, previous generations were limited by reduced data throughput. The latest R10.4.1/V14 chemistry was evaluated by [37] in a *K. pneumoniae* outbreak using cgMLST,

concluding that ONT sequencing caused considerable base errors leading to incorrect exclusions in outbreak tracing. They suggested mitigating these methylation-related errors through PCR-based library prep or masking ambiguous base positions.

Our study thus expands the knowledge on ONT performance for WGS-based bacterial typing by applying a similar comparative framework to two sets of well-characterized outbreak strains from the highly relevant species CDIP and VRE, both with native and PCR-based library preparation kits. In contrast to previous studies, we put a strong focus on the influence of different sequencing and basecalling software on the accuracy of downstream analyses such as cgMLST and AMR gene contextual analyses. Thus, we tested multiple software versions released in 2023 and 2024, in combination with different parameters such as different versions of the GridION software, varying sampling rate of the signal acquisition (4 kHz and 5 kHz), basecaller software (Guppy and Dorado), basecaller versions and basecalling models. All sequencing runs were conducted with real-time basecalling in HAC. The acquired signals were later re-basecalled in HAC and SUP mode, with standalone versions of the basecallers Guppy and Dorado, and different basecalling models. Our study thus reflects and integrates in its evaluation the rapid evolution of the technology over the past 2 years, including both wet laboratory and software changes that initially reduced the accuracy of the tested applications. We show the significant influence such an update may have on ONT datasets when processing native DNA libraries. Further strengths of our study are that we also provided estimates of both turnaround time and associated costs to offer a comprehensive perspective on implementation considerations.

Our study also identified several weaknesses and areas where ONT sequencing showed variability or discrepancies compared to Illumina. The number of mismatching alleles between Illumina and Nanopore assemblies varied strongly depending on the isolate, ONT software versions, basecalling models and library preparation kits. PCR-free library preparation was found to be more sensitive to analysis parameters than PCR-based methods. Masking of ambiguous base positions with MPOA could lead to unassigned STs and a reduction in successfully detected cgMLST target genes. Discrepancies were observed in AMR predictions at the antibiotic subclass and gene levels for certain isolates. Furthermore, the predicted genomic context of resistance genes (chromosomal vs. plasmid-mediated) showed discrepancies between different Nanopore approaches in some cases. However, we cannot exclude that optimizations in the plasmid reconstruction procedure could improve the consistency between different nanopore approaches [46], but this was outside the scope of the study.

Despite its strengths, the PCR-based ONT kit exhibited certain drawbacks, such as a smaller read length, ultimately leading to more fragmented chromosome assemblies. To mitigate this, we tested two hybrid strategies combining PCR-free and PCR-based Nanopore approaches and demonstrated promising results. These combined methods indeed retained the accuracy of the PCR-based approach while achieving assembly performance comparable to PCR-free methods. AMR prediction at the genome level was consistent across all tested sequencing methods; however, resolving the genomic context of AMR genes remains a complex task [47]. Although ONT long-read sequencing may offer significant improvements in determining genomic context compared to short-read technologies, we did observe inconsistencies in the predicted genomic context of selected AMR genes when comparing datasets produced from PCR-based, PCR-free and mixed Nanopore reads. As we chose a retrospective collection of outbreak-related isolates for in-depth comparative analyses, future work would need to provide a prospective evaluation to assess the performance of the new technology in actual clinical or epidemiological settings, thus offering a more accurate reflection of real-world applicability. Also, we limited ourselves to local outbreaks, for which the source of material and reference short-read data were controlled. We may also ask how international collection and inter-laboratory comparison would perform in such a comparative framework.

The results of this study suggest that Nanopore WGS is a promising technology for genomic epidemiology of outbreak bacterial species, such as VRE and CDIP, demonstrating comparable performance to Illumina in key areas such as cgMLST target gene recovery and allele assignment. The strengths in chromosome assembly using PCR-free methods and the competitive time and cost further support this potential. The observed variability related to software and parameter specification and the discrepancies in AMR and plasmid analysis highlight the need for continued optimization of Nanopore sequencing protocols, basecalling algorithms and analysis pipelines to improve accuracy and reliability in genomic epidemiological applications. Future research could focus on refining PCR-free methods to reduce their sensitivity to analysis parameters, developing more robust methods for resolving ambiguous base positions and improving the accuracy of AMR and plasmid prediction from Nanopore assemblies. As the technology continues to evolve, Nanopore sequencing has the potential to serve as a valuable alternative or complementary approach to short-read technologies for comprehensive genomic characterization in epidemiological investigations.

Funding information

AR received funding from the Institute for Infectious Diseases and the Multidisciplinary Center for Infectious Diseases, Bern, Switzerland, to support this study.

Acknowledgements

The authors thank the diagnostic department of the Institute for Infectious Diseases, University of Bern, for collecting and storing the clinical isolates.

Author contributions

Conceptualization: S.N. and A.R. Data curation: S.N., S.G. and M.A.T.M. Formal analysis: S.N. and A.R. Investigation: S.N., L.B., S.G., M.A.T.M. and A.R. Methodology: S.N., M.A.T.M. and A.R. Funding acquisition: A.R. Project administration: A.R. Resources: C.C. and A.R. Software development: S.N., L.B. and A.R. Supervision: A.R. Validation: S.N., S.G., M.A.T.M., C.C. and A.R. Visualization: S.N. and A.R. Writing – original draft: S.N., M.A.T.M. and A.R. Writing – review and editing: S.N., L.B., S.G., M.A.T.M., C.C. and A.R. All authors have read and agreed to the published version of the manuscript.

Conflicts of interest

The authors declare that there are no conflicts of interest.

References

- Van Goethem N, Descamps T, Devleeschauwer B, Roosens NHC, Boon NAM, et al. Status and potential of bacterial genomics for public health practice: a scoping review. *Implement Sci* 2019;14:79.
- Didelot X, Bowden R, Wilson DJ, Peto TEA, Crook DW. Transforming clinical microbiology with bacterial genome sequencing. *Nat Rev Genet* 2012;13:601–612.
- Price V, Ngwira LG, Lewis JM, Baker KS, Peacock SJ, et al. A systematic review of economic evaluations of whole-genome sequencing for the surveillance of bacterial pathogens. *Microb Genom* 2023;9:000947.
- Schadron T, van den Beld M, Mughini-Gras L, Franz E. Use of whole genome sequencing for surveillance and control of foodborne diseases: status quo and quo vadis. *Front Microbiol* 2024;15.
- Simar SR, Hanson BM, Arias CA. Techniques in bacterial strain typing: past, present, and future. *Curr Opin Infect Dis* 2021;34:339–345.
- Deurenberg RH, Bathoorn E, Chlebowicz MA, Couto N, Ferdous M, et al. Application of next generation sequencing in clinical microbiology and infection prevention. *J Biotechnol* 2017;243:16–24.
- 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.
- Nadon C, Van Walle I, Gerner-Smidt P, Campos J, Chinen I, et al. PulseNet International: vision for the implementation of whole genome sequencing (WGS) for global food-borne disease surveillance. *Euro Surveill* 2017;22:30544.
- Harris SR. SKA: Split Kmer Analysis Toolkit for bacterial genomic epidemiology. bioRxiv 2018. DOI: 10.1101/453142.
- Greig DR, Jenkins C, Gharbia S, Dallman TJ. Comparison of single-nucleotide variants identified by Illumina and Oxford Nanopore technologies in the context of a potential outbreak of Shiga toxin-producing *Escherichia coli*. *Gigascience* 2019;8:giz104.
- Schürch AC, Arredondo-Alonso S, Willems RJL, Goering RV. Whole genome sequencing options for bacterial strain typing and epidemiologic analysis based on single nucleotide polymorphism versus gene-by-gene-based approaches. *Clin Microbiol Infect* 2018;24:350–354.
- Deamer D, Akeson M, Branton D. Three decades of nanopore sequencing. *Nat Biotechnol* 2016;34:518–524.
- Sanderson ND, Kapel N, Rodger G, Webster H, Lipworth S, et al. Comparison of R9.4.1/Kit10 and R10/Kit12 Oxford Nanopore flow-cells and chemistries in bacterial genome reconstruction. *Microb Genom* 2023;9:mgen000910.
- Wagner GE, Dabernig-Heinz J, Lipp M, Cabal A, Simantzik J, 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.
- Ahrenfeldt J, Skaarup C, Hasman H, Pedersen AG, Aarestrup FM, et al. Bacterial whole genome-based phylogeny: construction of a new benchmarking dataset and assessment of some existing methods. *BMC Genomics* 2017;18:19.
- Magi A, Semeraro R, Mingrino A, Giusti B, D'Aurizio R. Nanopore sequencing data analysis: state of the art, applications and challenges. *Brief Bioinform* 2018;19:1256–1272.
- Oh S, Nam SK, Chang HE, Park KU. Comparative analysis of short- and long-read sequencing of vancomycin-resistant *Enterococci* for application to molecular epidemiology. *Front Cell Infect Microbiol* 2022;12:857801.
- Hall MB, Rabodoarivelo MS, Koch A, Dippenaar A, George S, et al. Evaluation of Nanopore sequencing for *Mycobacterium tuberculosis* drug susceptibility testing and outbreak investigation: a genomic analysis. *Lancet Microbe* 2023;4:e84–e92.
- Liao Y-C, Wu H-C, Liou C-H, Lauderdale T-LY, Huang I-W, et al. Rapid and routine molecular typing using multiplex polymerase chain reaction and MinION sequencer. *Front Microbiol* 2022;13:875347.
- Ferreira FA, Helmersen K, Visnovska T, Jørgensen SB, Aamot HV. Rapid nanopore-based DNA sequencing protocol of antibiotic-resistant bacteria for use in surveillance and outbreak investigation. *Microb Genom* 2021;7:000557.
- Wu X, Luo H, Ge C, Xu F, Deng X, et al. Evaluation of multiplex nanopore sequencing for *Salmonella* serotype prediction and antimicrobial resistance gene and virulence gene detection. *Front Microbiol* 2022;13:1073057.
- Xian Z, Li S, Mann DA, Huang Y, Xu F, et al. Subtyping evaluation of *Salmonella Enteritidis* using single nucleotide polymorphism and core genome multilocus sequence typing with Nanopore reads. *Appl Environ Microbiol* 2022;88:e0078522.
- Linde J, Brangsch H, Hölzer M, Thomas C, Elschner MC, et al. Comparison of Illumina and Oxford Nanopore Technology for genome analysis of *Francisella tularensis*, *Bacillus anthracis*, and *Brucella suis*. *BMC Genomics* 2023;24:258.
- Tarumoto N, Sakai J, Sujino K, Yamaguchi T, Ohta M, et al. Use of the Oxford Nanopore MinION sequencer for MLST genotyping of vancomycin-resistant enterococci. *J Hosp Infect* 2017;96:296–298.
- Both A, Kruse F, Mirwald N, Franke G, Christner M, et al. Population dynamics in colonizing vancomycin-resistant *Enterococcus faecium* isolated from immunosuppressed patients. *J Glob Antimicrob Resist* 2022;28:267–273.
- Cao MD, Ganesamoorthy D, Elliott AG, Zhang H, Cooper MA, et al. Streaming algorithms for identification of pathogens and antibiotic resistance potential from real-time MinION(TM) sequencing. *Gigascience* 2016;5:32.
- Dilthey AT, Meyer SA, Kaasch AJ. Ultrplexing: increasing the efficiency of long-read sequencing for hybrid assembly with k-mer-based multiplexing. *Genome Biol* 2020;21:68.
- Sereika M, Kirkegaard RH, Karst SM, Michaelsen TY, Sørensen EA, 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.
- Maiden MC, Bygraves JA, Feil E, Morelli G, Russell JE, et al. Multi-locus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc Natl Acad Sci USA* 1998;95:3140–3145.
- Maiden MCJ, Jansen van Rensburg MJ, Bray JE, Earle SG, Ford SA, et al. MLST revisited: the gene-by-gene approach to bacterial genomics. *Nat Rev Microbiol* 2013;11:728–736.
- Delahaye C, Nicolas J. Sequencing DNA with nanopores: troubles and biases. *PLoS One* 2021;16:e0257521.
- Kofler J, Ramette A, Iseli P, Stauber L, Fichtner J, et al. Ongoing toxin-positive diphtheria outbreaks in a federal asylum centre in Switzerland, analysis July to September 2022. *Euro Surveill* 2022;27:2200811.

33. Piezzi V, Wassilew N, Atkinson A, D'Incau S, Kaspar T, *et al.* Nosocomial outbreak of vancomycin-resistant *Enterococcus faecium* (VRE) ST796, Switzerland, 2017 to 2020. *Euro Surveill* 2022;27:2200285.
34. De Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C. NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics* 2018;34:2666–2669.
35. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol* 2019;37:540–546.
36. Chris W. Medaka: sequence correction provided by ONT research. 2024. <https://github.com/nanoporetech/medaka>
37. Lohde M, Wagner GE, Dabernig-Heinz J, Viehweger A, Braun SD, *et al.* Accurate bacterial outbreak tracing with Oxford Nanopore sequencing and reduction of methylation-induced errors. *Genome Res* 2024;34:2039–2047.
38. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 2018;34:3094–3100.
39. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, *et al.* Twelve years of SAMtools and BCFtools. *GigaScience* 2021;10:giab008.
40. Miles A. Pysamstats. Github. 2024. <https://github.com/alimanfoo/pysamstats>
41. de Been M, Pinholt M, Top J, Bletz S, Mellmann A, *et al.* Core genome multilocus sequence typing scheme for high-resolution typing of *Enterococcus faecium*. *J Clin Microbiol* 2015;53:3788–3797.
42. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 2014;30:2068–2069.
43. Feldgarden M, Brover V, Gonzalez-Escalona N, Frye JG, Haendiges J, *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.
44. Arias CA, Murray BE. The rise of the *Enterococcus*: beyond vancomycin resistance. *Nat Rev Microbiol* 2012;10:266–278.
45. Foster-Nyarko E, Cottingham H, Wick RR, Judd LM, Lam MMC, *et al.* Nanopore-only assemblies for genomic surveillance of the global priority drug-resistant pathogen, *Klebsiella pneumoniae*. *Microb Genom* 2023;9:mgen000936.
46. Bouras G, Sheppard AE, Mallawaarachchi V, Vreugde S. Plassembler: an automated bacterial plasmid assembly tool. *Bioinformatics* 2023;39:btad409.
47. Djordjevic SP, Jarocki VM, Seemann T, Cummins ML, Watt AE, *et al.* Genomic surveillance for antimicrobial resistance – a One Health perspective. *Nat Rev Genet* 2024;25:142–157.

The Microbiology Society is a membership charity and not-for-profit publisher.

Your submissions to our titles support the community – ensuring that we continue to provide events, grants and professional development for microbiologists at all career stages.

Find out more and submit your article at microbiologyresearch.org