

RESEARCH

Open Access



Genomic surveillance of multidrug-resistant organisms based on long-read sequencing

Fabian Landman^{1†}, Casper Jamin^{1†}, Angela de Haan¹, Sandra Witteveen¹, Jeroen Bos¹, Han G. J. van der Heide¹, Leo M. Schouls¹, Antoni P. A. Hendrickx^{1*}  and Dutch CPE/MRSA surveillance study group

Abstract

Background Multidrug-resistant organisms (MDRO) pose a significant threat to public health worldwide. The ability to identify antimicrobial resistance determinants, to assess changes in molecular types, and to detect transmission are essential for surveillance and infection prevention of MDRO. Molecular characterization based on long-read sequencing has emerged as a promising alternative to short-read sequencing. The aim of this study was to characterize MDRO for surveillance and transmission studies based on long-read sequencing only.

Methods Genomic DNA of 356 MDRO was automatically extracted using the Maxwell-RSC48. The MDRO included 106 *Klebsiella pneumoniae* isolates, 85 *Escherichia coli*, 15 *Enterobacter cloacae* complex, 10 *Citrobacter freundii*, 34 *Pseudomonas aeruginosa*, 16 *Acinetobacter baumannii*, and 69 methicillin-resistant *Staphylococcus aureus* (MRSA), of which 24 were from an outbreak. MDRO were sequenced using both short-read (Illumina NextSeq 550) and long-read (Nanopore Rapid Barcoding Kit-24-V14, R10.4.1) whole-genome sequencing (WGS). Basecalling was performed for two distinct models using Dorado-0.3.2 duplex mode. Long-read data was assembled using Flye, Canu, Miniasm, Unicycler, Necat, Raven, and Redbean assemblers. Long-read WGS data with > 40× coverage was used for multi-locus sequence typing (MLST), whole-genome MLST (wgMLST), whole-genome single-nucleotide polymorphisms (wgSNP), in silico multiple locus variable-number of tandem repeat analysis (iMLVA) for MRSA, and identification of resistance genes (ABRicate).

Results Comparison of wgMLST profiles based on long-read and short-read WGS data revealed > 95% of wgMLST profiles within the species-specific cluster cut-off, except for *P. aeruginosa*. The wgMLST profiles obtained by long-read and short-read WGS differed only one to nine wgMLST alleles or SNPs for *K. pneumoniae*, *E. coli*, *E. cloacae* complex, *C. freundii*, *A. baumannii* complex, and MRSA. For *P. aeruginosa*, differences were up to 27 wgMLST alleles between long-read and short-read wgMLST and 0–10 SNPs. MLST sequence types and iMLVA types were concordant between long-read and short-read WGS data and conventional MLVA typing. Antimicrobial resistance genes were detected in long-read sequencing data with high sensitivity/specificity (92–100%/99–100%). Long-read sequencing enabled analysis of an MRSA outbreak.

Conclusions We demonstrate that molecular characterization of automatically extracted DNA followed by long-read sequencing is as accurate compared to short-read sequencing and suitable for typing and outbreak analysis as part of genomic surveillance of MDRO. However, the analysis of *P. aeruginosa* requires further improvement which may be

[†]Fabian Landman and Casper Jamin contributed equally to this work.

*Correspondence:

Antoni P. A. Hendrickx

antoni.hendrickx@rivm.nl

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

obtained by other basecalling algorithms. The low implementation costs and rapid library preparation for long-read sequencing of MDRO extends its applicability to resource-constrained settings and low-income countries worldwide.

Keywords Genomic surveillance, CPE, CPPA, CRAB, MRSA, Long-read sequencing, Nanopore, MLST, WgMLST, WgSNP, MLVA, iMLVA

Background

Multidrug-resistant organisms (MDRO) such as methicillin-resistant *Staphylococcus aureus* (MRSA), carbapenemase-producing *Enterobacterales* (CPE), carbapenemase-producing *Pseudomonas aeruginosa* (CPPA), and carbapenem-resistant *Acinetobacter baumannii* (CRAB) pose a growing public health concern worldwide. The ability to type MDRO accurately and rapidly and to identify resistance determinants is essential for effective surveillance and infection control. National reference laboratories and medical microbiology laboratories historically typed MDRO using various classical typing methods such as phage typing, pulsed-field gel electrophoresis (PFGE) [1, 2], amplified fragment length polymorphism (AFLP) [3], staphylococcal protein A (Spa) typing [4, 5], and multiple-locus variable number of tandem repeat analysis (MLVA) as well as DNA sequencing-based typing methods such as multi-locus sequence typing (MLST) [6–9]. Except for phage typing, these methods rely on the generation of DNA fragment patterns, either of specific DNA fragment sizes (Spa-typing and MLVA) or more random, by DNA restriction (PFGE) or restriction and amplification (AFLP) [10]. These classical typing methods had low typing resolution, were laborious and costly, and were not all suitable for high throughput typing of MDRO on a daily basis.

Another typing method still commonly used is MLST. MLST is based on the comparison of the sequence of typically 7 housekeeping genes. For each unique locus, a number is assigned. The collection of these seven numbers is used to assign a unique MLST sequence type (ST) [11]. The past decade, these classical methods are being phased out by whole-genome sequencing (WGS)-based methods such as core genome (cg) or whole genome (wg) MLST and cg and wg single-nucleotide polymorphisms (SNP), as these methods significantly increased typing resolution [11]. Both cgMLST and wgMLST are an extension on the conventional MLST method but includes typically more loci (2500–7000). For accurate transmission and outbreak detection of MDRO, national reference laboratories increased typing resolution by improved typing methods through the years. Nowadays, besides accurate typing of MDRO, WGS data also enable identification of resistance and virulence genes [11].

Current WGS-based typing methods (cgMLST, cgSNP, wgMLST, wgSNP) typically use short-read sequencing

technologies with read lengths of 150 bases, such as Illumina next-generation sequencing, and have been widely used for public health genomic surveillance, molecular typing, and outbreak detection of MDRO [12, 13]. Short-read WGS costs are high and force concessions for national and local surveillance laboratories to sequence only a limited selection of MDRO [14]. Additionally, short-read sequencing technologies have limitations in detecting structural variations, such as large insertions, deletions, inversions, repetitive sequence elements, and antimicrobial resistance (AMR) plasmids, all common in bacterial genomes [13]. Reconstruction of complete bacterial genomes by de novo assembly is unattainable through short-read sequencing as these structural variations and genomic elements are larger than individual reads. Thus, chromosomes and AMR encoding plasmids fail to assemble into complete assemblies. Hybrid assemblies of short-read and long-read WGS data enabled reconstruction of MDRO genomes but is laborious and more expensive than short-read sequencing alone. As MDRO are rapidly spreading and prevalences increasing worldwide, alternatives are required to enable genomic surveillance of MDRO in national reference laboratories in the AMR era. Long-read sequencing technologies, such as Nanopore long-read sequencing, has recently emerged as a promising alternative to short-read sequencing or hybrid assemblies [15–18], since the accuracy of Nanopore long-read sequencing significantly improved over time [19]. The major objective of this study was to determine whether long-read sequencing can replace short-read sequencing to enable classical and new typing methods such as MLVA, MLST, wgMLST, and wgSNP in the genomic surveillance and outbreak analysis of MDRO for a national reference laboratory.

Methods

Multidrug-resistant organisms (MDRO) from the Dutch national surveillance 2023

Medical microbiology laboratories (MML) in the Netherlands are requested to submit *Enterobacterales*, *P. aeruginosa*, and *A. baumannii* suspected for carbapenemase production or carbapenem resistance, and MRSA isolates cultured for patient care (from symptomatic infections or asymptomatic carriage), to the National Institute for Public Health and the Environment (RIVM) [14, 20]. Between 1 May and 30 September 2023, 3138 suspected

MDRO were received by the RIVM of which the majority 2235 were MRSA isolates. Illumina short-read sequencing and Nanopore long-read sequencing were performed on 356 MDRO in this study (vide infra). In this time period, all carbapenemase-producing Enterobacterales (CPE), carbapenemase-producing *P. aeruginosa* (CPPA), and carbapenem-resistant *A. baumannii-calcoaceticus* complex (CRAB) and a selection of diverse, based on MLVA profiles, methicillin-resistant *Staphylococcus aureus* (MRSA) isolates were retrospectively selected for this study. The MDRO included 106 *Klebsiella pneumoniae* isolates, 85 *Escherichia coli*, 15 *Enterobacter cloacae* complex, 10 *Citrobacter freundii*, 34 *P. aeruginosa*, 16 *A. baumannii*, 69 MRSA, and other species (Table 1, Additional file 1: Table S1).

Automated Oxford Nanopore Technologies long-read sequencing

The genomic DNA of MDRO was isolated using The Maxwell® (Promega), an automated nucleic acid extraction and purification platform. The Maxwell RSC Cultured Cells DNA kit (AS1620) kit was used to extract genomic DNA from up to 48 bacterial isolates per DNA extraction run. Manufacturer’s instructions were followed, except using nuclease-free water instead of TE buffer to create the cell suspension and without RNase treatment. The Oxford Nanopore protocol for rapid sequencing DNA V14 – barcoding SQK-RBK114.24 was used (Oxford Nanopore Technologies [ONT]). In brief, barcoded transposome complexes were used to

tagment the DNA and simultaneous attachment of a pair of barcodes. Routine isolations had shown DNA concentrations within 50–300 ng/μL; thus, for maintaining high-throughput capacity, DNA concentrations were not measured prior to library preparation. Twenty-four samples were pooled, and after clean-up, the sequencing adapters (RAP reagent) were added. Sequencing buffer and library beads were added, and the final library of 24 bacterial genomes was loaded onto a MinION flow cell (FLO-MIN114, R10.4.1) [21]. From Maxwell® DNA isolation until loading, the flow cell routinely takes 3 h following a 48 h sequence run. The sequence run was started with live basecalling and demultiplexing enabled using the MinKNOW software through a GridION device with 5 kHz data acquisition enabled for all samples. However, the actual basecalling for analysis in this study was performed on the resulting barcoded pod5 directories using Dorado 0.3.2 duplex mode with the optional flags –guard-gpus, -b 448 and writing data to Fastq format. We assessed the performance of two basecalling models, dna_r10.4.1_e8.2_400bps_sup@v4.2.0 and res_dna_r10.4.1_e8.2_400bps_sup@2023-09-22_bacterial-methylation. These models are referred to as Dorado [22] and Rerio [23], respectively. Data was assembled using an in-house Snakemake workflow [24]. All bioinformatic tools were used with default parameters unless specified otherwise. First, Chopper v0.6.0 is used to extract all Q12 reads > 1000 bp [25]. Additionally, 80 bp is cropped from both sides to remove possible adapters. Subsequently, FiltLong 0.2.1 was used to keep the 90% best scoring

Table 1 The number of Illumina short-read and Nanopore long-read sequenced MDRO obtained in the Dutch national surveillance from May to September 2023. The number of bacterial isolates reaching 20x, 30x, and 40 × coverage are included

MDRO	All coverages	> 20 × coverage	> 30 × coverage	> 40 × coverage
<i>Acinetobacter baumannii</i>	16	11	7	4
<i>Acinetobacter nosocomialis</i>	1	1	1	1
<i>Acinetobacter ursingii</i>	1			
<i>Citrobacter farmeri</i>	1	1	1	1
<i>Citrobacter freundii</i>	10	9	9	9
<i>Enterobacter cloacae</i> complex	15	14	11	11
<i>Escherichia coli</i>	85	77	69	54
<i>Klebsiella oxytoca</i>	6	5	3	2
<i>Klebsiella pneumoniae</i>	106	84	70	48
<i>Proteus mirabilis</i>	4	4	4	2
<i>Providencia stuartii</i>	3	2	2	2
<i>Pseudomonas aeruginosa</i>	34	23	12	7
<i>Pseudomonas nitroreducens</i>	1			
<i>Raoultella ornithinolytica</i>	4	2	1	
<i>Staphylococcus aureus</i>	45	45	44	39
<i>Staphylococcus aureus</i> (outbreak)	24	23	19	16
Total	356	301	253	196

reads based on length and quality [26]. To assess the performance of different de novo assemblers, long-read data was assembled using the following: (1) Canu v2.2 [27], (2) Flye v2.9.2 [28], (3) Minimap v2 2.26 followed by Miniasm v0.3 [29], Minipolish v0.1.3 collectively referred to as Miniasm throughout this manuscript, (4) Necat v0.0.1 [30], (5) Raven v1.8.1 [31], (6) Redbean wtdbg2.5 [32], (7) and Unicycler v0.5.0 [33] long-read option, referred to as Longcyclor throughout the manuscript. Furthermore, genome assembly polishing with Medaka v1.8.0 [34] was used to test if this improved de novo assemblies, compared to unpolished assemblies. Medaka polishing was only done for Dorado basecalled reads, as there was no Medaka polishing model for Rerio basecalled data available during this study. Quast v5.2.0 [35] was used to determine various assembly QC statistics such as L90 score.

Automated Illumina short-read sequencing

For in-house Illumina short-read sequencing of MDRO, the same DNA isolation method as described above was followed. DNA libraries were prepared using Nextera DNA Flex Library Prep kit (Illumina, San Diego, USA), followed by paired-end sequencing (2×150 bases) on the Illumina NextSeq550 platform (Illumina, USA), according to the manufacturer's instructions. From the start of DNA extraction until loading, the sequencer takes 8 h, and additionally, a sequence run of 26 h was performed. Read quality analysis and de novo assembly were performed with the Juno-assembly v2.0.2 pipeline [36]. Briefly, read quality assessment and filtering were done using FastQC and FastIP [37, 38]. Genomes were assembled using SPAdes [39] and curated with QUAST [35], CheckM [40] and Bbtools [41]. For Illumina short-read sequencing, bacterial isolates failed the quality control if (1) there was less than 30×coverage, (2) more than 4% contamination, (3) more than 120 contigs (MRSA), (4) more than 200 contigs (CPE), or (5) 340 contigs (CPPA).

Molecular typing from long-read and short-read sequencing data and data analyses

Sixty-nine MRSA isolates, including 24 outbreak isolates, were characterized by multiple-locus variable number of tandem repeat analysis (MLVA) in vitro at BaseClear (Leiden, the Netherlands) and simultaneous PCR detection of the *mecA/mecC* and *lukF-PV/lukS-PV* genes, the latter indicative of PVL production, as previously described [42, 43]. In addition, in silico (i)MLVA on MRSA isolates was done in the RIVM on long-read WGS data only [44]. The identification of resistance genes and replicons of MDRO was performed using ResFinder (version 4.1.11) and PlasmidFinder (version 2.1) databases from the Center for Genomic Epidemiology, using ABRicate, with minimal identity 80% and minimal length 80% (v1.0.1 [45]).

The resulting FASTA files were subjected to their corresponding species-specific multi-locus sequence typing (MLST) and whole-genome MLST (wgMLST) schemes using the Ridom software in SeqSphere v9.0.1 if available [46]. The following in-house wgMLST schemes were used for analyses: the *A. baumannii* wgMLST scheme was based on 3473 genes (2390 core genome genes and 1083 accessory genome genes, cut-off 15 wgMLST alleles), using *A. baumannii* strain ACICU (Genbank accession number NC_010611.1, September 2021) as a reference genome. The *K. pneumoniae* wgMLST scheme comprised 4978 genes (3471 core genome and 1507 accessory genome targets, cut-off 20 wgMLST alleles) using *K. pneumoniae* MGH78578 (NC_009648.1) as a reference genome [20]. The *E. coli* wgMLST scheme comprised 4503 genes (3199 core genome and 1304 accessory genome targets, cut-off 25 wgMLST alleles) using *E. coli* 536 (CP000247.1) as a reference genome [20]. The *P. stuartii* scheme comprised 3744 genes (3079 core genome and 665 accessory genome targets, cut-off 15 wgMLST alleles) with *P. stuartii* CP014024.2 as reference genome [47]. The *P. aeruginosa* wgMLST scheme included 6442 genes (6117 core genome and 325 accessory genome targets, cut-off 15 wgMLST alleles) using *P. aeruginosa* PAO1 (NC_002516.2) as a reference genome. The *P. mirabilis* wgMLST scheme included 3517 genes (2675 core genome and 842 accessory genome targets, cut-off 15 wgMLST alleles) using *P. mirabilis* HI4320 (NZ_CP042907.1) as a reference genome. The *C. freundii* wgMLST scheme included 4495 genes (2964 core genome and 1531 accessory genome targets, cut-off 20 wgMLST alleles) using *C. freundii* strain HM38 (CP024672.1) as a reference. For MRSA, the COL-based wgMLST scheme comprised 2567 genes (1861 core genome and 706 accessory genome targets, cut-off 15 wgMLST alleles) was used [14, 48]. For *E. cloacae* complex, a pgMLST scheme comprised of 9829 genes from references CP001918, CP017186, and CP017184 was used, cut-off 20 wgMLST alleles [49]. The *A. baumannii* and *S. aureus* schemes were also available via RIDOM (cgmlst.org). Additionally, for *E. coli* and *K. pneumoniae*, the canonical pubMLST cgMLST schemes (cgmlst.org) were also ran on isolates with 40×coverage or higher and only for Rerio basecalled data, to further determine the usability ONT sequencing and cgMLST for outbreak investigation and genomic surveillance. Both MLST and cg/wgMLST profiles were imported into BioNumerics version 8.1.1 (Applied Maths, Sint-Martens-Latem, Belgium) and used in cluster analyses. Missing data were ignored in the analyses.

The number of whole-genome (wg)SNPs between Illumina short-read and Nanopore long-read sequencing data was determined by using Mintyper (v1.1.2).

Reference genomes per species were selected using ReferenceSeeker. Only Rerio basecalled datasets with at least 40× coverage were used for this analysis.

The Illumina short-read sequencing results were considered gold standard method for MLST sequence type determination, wgMLST allele calling, AMR gene and replicon identification, and long-read assembly methods throughout this study. The AMR gene and replicon detection in Illumina short-read sequencing data was used to calculate sensitivity and specificity for the long-read assembly methods. All further data analyses and visualization was done using Python (v3.11.5), Pandas (v2.0.3) and Plotly (v5.18.0).

Results

Assessing sequencing coverage and basecalling model for wgMLST on long-read sequenced MDRO

From May to September 2023, the National Institute for Public Health and the Environment (RIVM) received

3138 MDRO. In this study, 356 genetically highly diverse MDRO were included: 269 CPE, of which the majority were *K. pneumoniae* ($n=106$) and *E. coli* ($n=85$) and nine other species ($n=78$), CPPA ($n=34$), CRAB ($n=17$), and MRSA ($n=69$) (Table 1, Additional file 1: Table S1). We also included 24 PVL-negative MRSA isolates belonging to a recurring fusidic acid-resistant impetigo-associated MRSA outbreak in the Netherlands [50]. We isolated genomic DNA using the Maxwell and sequenced 356 MDRO using both Nanopore and Illumina WGS platforms. Sequencing coverage for long-read sequencing was assessed from 356 MDRO to determine a suitable cut-off for sequencing depth. The relative number of wgMLST alleles (number of Nanopore alleles divided by the number of Illumina alleles) were 0.45, 0.93, and 0.99 (medians) for coverages up to 0–10x, 10–20x, and 20–30x, respectively (Fig. 1A). The number of wgMLST alleles identified remained at 1.00 with higher sequencing depth, indicating that the same number of alleles

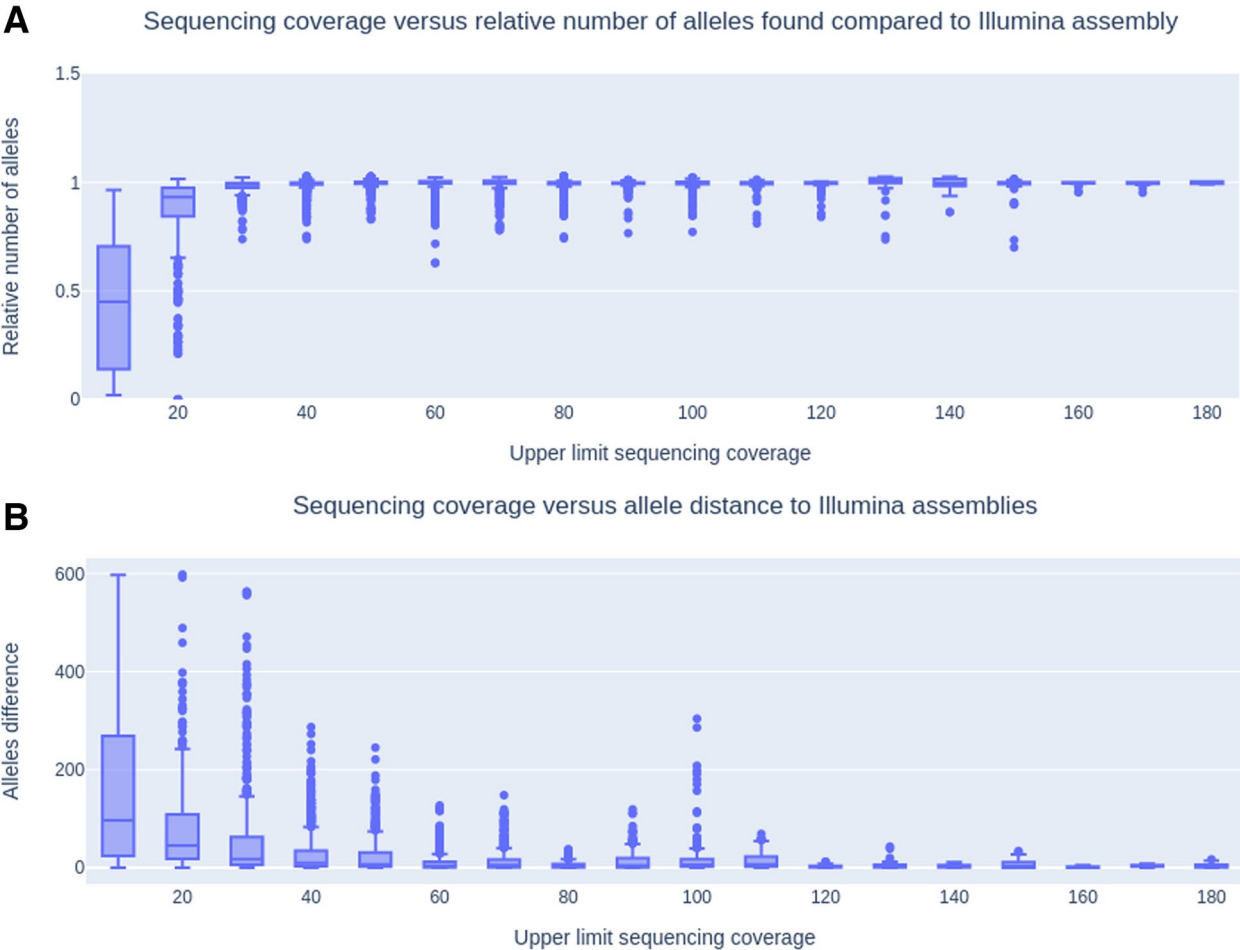


Fig. 1 **A** Nanopore long-read sequencing coverage versus relative number of wgMLST alleles identified in each assembly versus the Illumina genome assembly. **B** Nanopore sequencing coverage versus wgMLST allele distance to Illumina assemblies

were identified in the Nanopore assemblies compared to the Illumina assemblies. The median number of different wgMLST alleles compared to the Illumina assemblies was 97, 45, 17, 10, 7, 4, and 5 for coverages up to 0–10x, 10–20x, 20–30x, 30–40x, 40–50x, 50–60x, 60–70x, respectively (Fig. 1B). The median distance with Illumina assemblies was between 4 to 7 wgMLST alleles, for coverages from 40x and higher. One particular *P. aeruginosa* isolate with 100x coverage had up to 300 alleles different for all the assemblers used in this study.

To determine genome continuity, the L90 score (number of contigs needed to cover 90% of the genome) was assessed for all long-read assemblies. For assemblies with a coverage higher than 40x, the L90 score was one contig except for Canu, which had on average two contigs, across all species (Additional file 2: Fig. S1). In comparison, the average L90 score for Illumina assemblies was 29 contigs. The 196 long-read sequenced MDRO with a coverage higher than 40x were used for further analyses.

To determine which long-read sequencing data base-calling model performed best and if Medaka polishing (for Dorado only) improved assembly when comparing long-read assemblies to the Illumina short-read sequencing gold standard, we determined the number of wgMLST alleles difference for 196 MDRO using these methods (Fig. 2). Across all species, Rerio performed best and had the lowest number of faulty wgMLST allele calls. The median number of wgMLST alleles difference for Dorado duplex, Dorado duplex with Medaka, and Rerio duplex was 2, 2, 2 (*A. baumannii*), 5.5, 3, 1 (*C. freundii*), 2, 1, 1 (*E. cloacae* complex), 6, 3, 3 (*E. coli*), 23, 19.5, 9 (*K. pneumoniae*), 120, 79, 26.5 (*P. aeruginosa*), and 4, 2, 2 (*S. aureus*), respectively. Rerio was able to improve the long-read sequencing data in such a way that the number of faulty wgMLST allele calls dropped from 87 to 148 (depending on the assembler) to merely 0–5 faulty alleles

for one particular *C. freundii* isolate. Therefore, Rerio was used for further analyses.

Comparison of long-read assemblers for MLST, wgMLST, and wgSNP analysis on long-read sequenced MDRO

To assess the best assembler for de novo assembly of long-read only data, Longcycler (Unicycler with long-reads only), Miniasm, Raven, Necat, Canu, Flye, and Redbean were compared based on the median (50 percentile) wgMLST allele difference and the 95 percentile wgMLST alleles difference to Illumina short-read sequencing. With all species combined, the difference between Illumina short-read and Nanopore long-read only sequencing data was 1; 18 (Canu), 1; 21.55 (Flye), 2.5; 15.55 (Longcycler), 2; 17 (Miniasm), 7; 35.55 (Necat), 3; 34.65 (Raven), and 6; 38.65 (Redbean) for the median and 95 percentile, respectively (Fig. 3A, Table 2). Although the median was slightly lower for Miniasm compared to Longcycler, the 95% percentile performed better for Longcycler and was therefore used for subsequent analyses. For each MDRO, the median and 95% percentile allele difference for Longcycler was 1; 2.0 (*A. baumannii*), 1; 4.0 (*C. freundii*), 1; 3.5 (*E. cloacae*), 2; 17 (*E. coli*), 4; 12.6 (*K. pneumoniae*), 2; 6.0 (*S. aureus*), and 25.5; 43 (*P. aeruginosa*), respectively (Table 2). Additionally, for *E. coli* and *K. pneumoniae* the publicly available cgMLST schemes from cgMLST.org were ran for interstudy comparability. For *E. coli*, the median and 95% percentile number of alleles different to Illumina assemblies were 0; 8.4 (Canu), 0.5; 13.05 (Flye), 1; 11.35 (Longcycler), 1; 9.0 (Miniasm), 4; 13.7 (Necat), 2; 16.7 (Raven), and 3; 10.1 (Redbean) (Fig. 3B). For *K. pneumoniae*, the median and 95% percentile number of alleles different to Illumina assemblies were 3; 21.2 (Canu), 3; 21.2 (Flye), 3; 13 (Longcycler), 3; 15.3

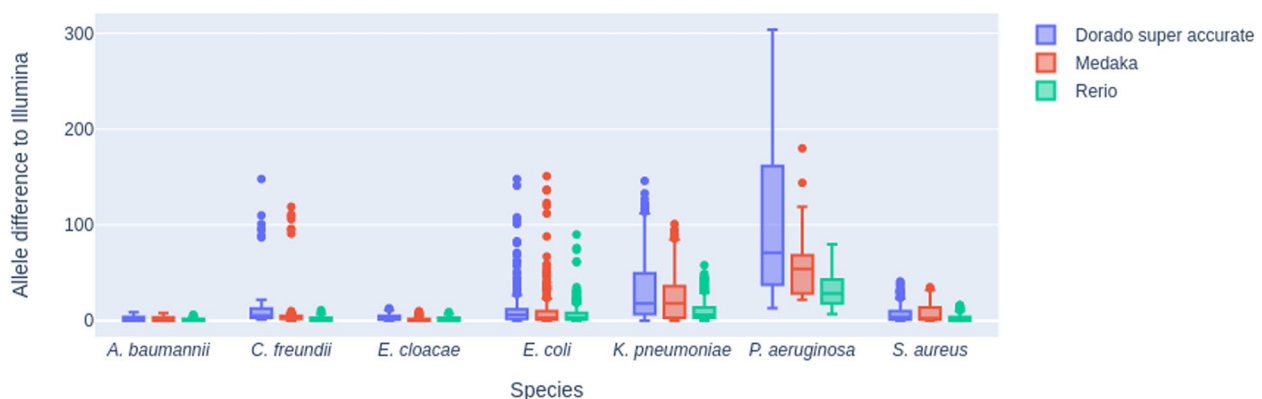


Fig. 2 Nanopore long-read sequencing wgMLST alleles difference compared to Illumina assemblies, for Dorado super accurate, Rerio, and Medaka polishing method

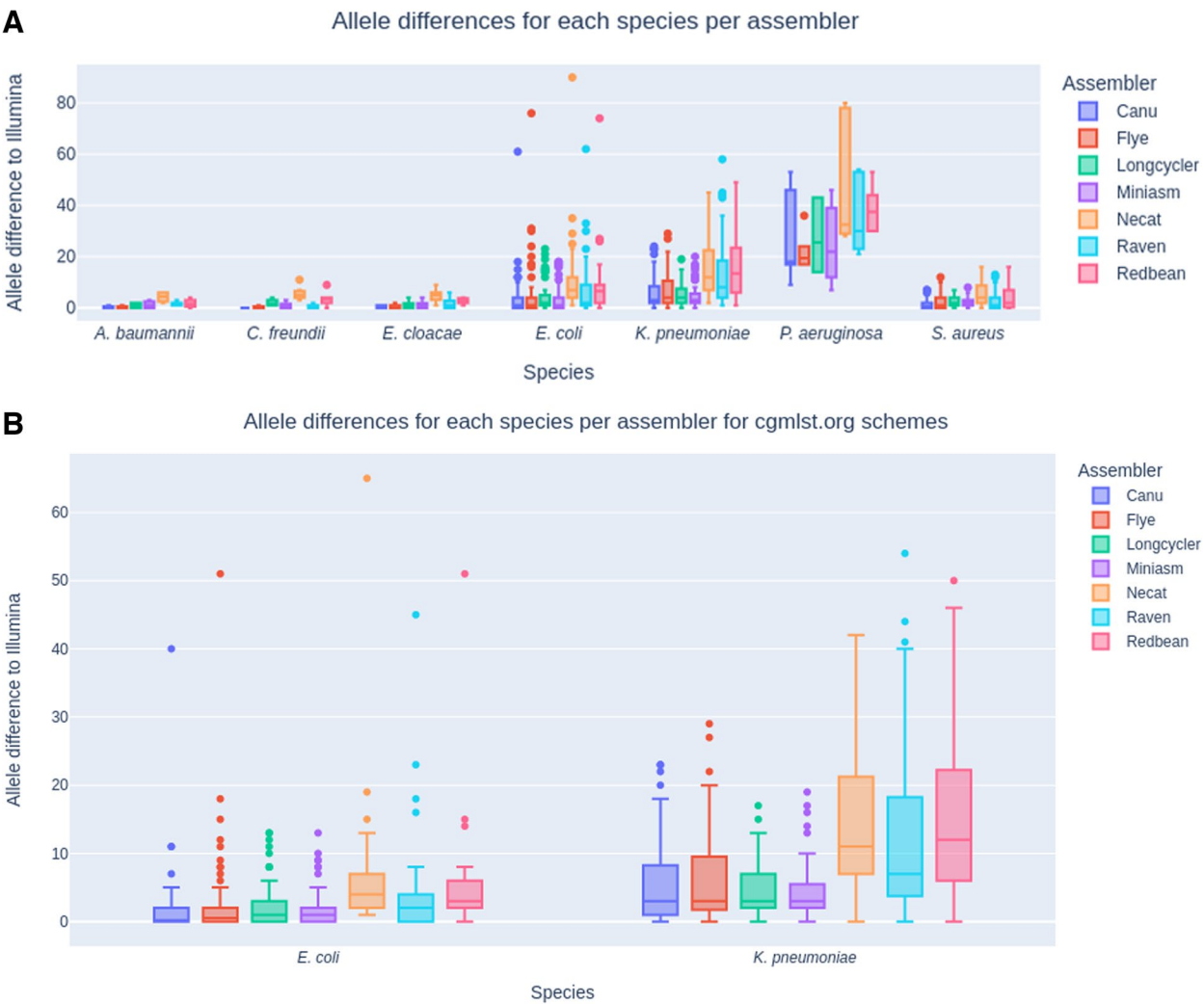


Fig. 3 **A** Nanopore long-read sequencing wgMLST alleles difference relative to Illumina short-read sequences per bacterial species and long-read sequence data assembler. **B** Nanopore long-read sequencing cgMLST alleles difference relative to Illumina short-read sequences for *E. coli* and *K. pneumoniae* cgMLST schemes from cgmlst.org and long-read sequence data assembler

Table 2 The median allele (left number) and 95 percentile wgMLST allele difference (right number) between Illumina short-read and Nanopore long-read sequenced species per de novo assembler

	Canu	Flye	Longcyclor	Miniasm	Necat	Raven	Redbean
<i>A. baumannii</i>	0.0–0.8	0.0–0.8	1.0–2.0	1.0–2.8	4.5–6.0	1.0–2.7	1.5–3.7
<i>C. freundii</i>	0.0–0.0	0.0–1.0	1.0–4.0	0.5–2.6	5.0–9.6	1.0–1.6	4.0–7.2
<i>E. cloacae</i>	0.0–1.0	0.0–1.5	1.0–3.5	1.0–3.5	5.0–8.0	1.0–5.0	3.0–4.0
<i>E. coli</i>	1.0–13.0	1.0–26.1	2.0–17.0	1.0–17.0	7.0–26.4	2.0–25.4	6.5–20.1
<i>K. pneumoniae</i>	3.0–22.3	4.0–21.6	4.0–12.6	3.0–15.6	12.0–40.6	8.0–43.6	13.5–43.3
<i>P. aeruginosa</i>	18.0–51.2	19.5–33.0	25.5–43.0	22.0–44.2	32.5–79.5	30.0–53.8	37.5–50.8
<i>S. aureus</i>	0.0–6.1	1.0–10.0	2.0–6.0	2.0–6.0	4.0–14.2	1.0–10.2	2.0–14.0

(Miniasm), 11; 40.2 (Necat), 7; 40.7 (Raven), and 12; 42.4 (Redbean).

wgSNP analysis was performed using Mintyper to compare the performance of Nanopore long-read sequencing to Illumina short-read sequencing, as the performance of wgMLST for *P. aeruginosa* was poor. For each MDRO, the median and maximum number of SNPs was 8.5 and 10 (*A. baumannii*), 0 and 3 (*C. freundii*), 4 and 6 (*E. cloacae*), 0 and 2 (*E. coli*), 0 and 1 (*K. pneumoniae*), 0 and 3 (*S. aureus*), and 1 and 5 SNPs for *P. aeruginosa*, respectively (Additional file 2: Fig. S2).

Long-read sequencing-derived MLST sequence types (ST) were compared with short-read sequencing derived MLST STs among the MDRO tested, resulting in concordance of sequence types 68%, 95%, 97%, 98%, 93%, 93%, and 89% for Canu, Flye, Longcyclor, Miniasm, Necat, Raven, and Redbean respectively. Observed differences in MLST ST between long-read and short-read assemblies were caused by absence of one or more alleles in the long-read assemblies (Additional file 3: Table S2). For three out of four *A. baumannii*, the *gdhB* gene was not identified in long-read assemblies. BLAST was performed, and two copies of a locus were located on the same contig at different positions for the long-read assemblies. For Illumina short-read sequencing, the same alleles were found but never on the same contig; therefore, a single best scoring allele was called for Illumina MLST. Interruption by IS elements in this element has been previously noted and further explains its absence [51, 52]. For Miniasm only, *gapA* was marked as new and consequently not called in *K. pneumoniae* but otherwise was in concordance with all Illumina MLST STs and thus performed best. All other long-read assemblers also marked *gapA* as new allele. Longcyclor reported the

gene *mdh* as failed for one *E. coli* isolate. Canu performed worst which can be explained due to de novo assembly artifacts resulting in duplicate regions (Additional file 3: Table S2).

Detection of AMR genes and plasmid replicons in long-read sequencing data

Next, we compared the detection of AMR genes and plasmid replicons using both long-read and short-read sequencing platforms. The sensitivity and specificity of AMR gene identification was assessed. The specificity of AMR gene calling for Redbean ranged from 68.2 to 100 (median 87.2) and performed worst of all assemblers tested (Table 3A). Canu had the highest specificity in AMR gene calling (99.7%), followed by Flye (99.6%) and Longcyclor (98.6%). Although Canu seemed to perform best, multiple copies of the same AMR gene were detected (Additional file 2: Fig. S3), also suggesting de novo assembly artifacts. The median number of AMR genes for all species together was 16 for Canu assemblies versus 11 to 13 for all other assemblers (Additional file 2: Fig. S3). The effect of this phenomenon was seen in the median genome size over the entire dataset (5.6-Mbp Canu versus 5.2-Mbp for all other long-read assemblers, Additional file 2: Fig. S4). Furthermore, the number of replicons was at least double for Canu assemblies versus any other assembler (Additional file 2: Fig. S5). The unweighted average sensitivity for AMR gene calling was 98.1% independent of which de novo assembler was used and seemed to be a species-specific effect (Table 3). Only for Flye and Canu assemblers all carbapenemase genes were correctly detected within the MDRO tested (Additional file 4: Table S3). A full overview of discrepant AMR gene calling between Illumina and long-read assemblies

Table 3 (A) Percentage sensitivity (lower panel) and specificity (upper panel) of AMR gene identification for each species and assembler tested in this study. (B) Percentage sensitivity (lower panel) and specificity (upper panel) of plasmid replicon detection for each species and assembler were tested in this study. Only Rerio as basecalling and long-read sequenced isolates with a coverage higher than 40× coverage were used

A Specificity and sensitivity of AMR gene calling							
	Canu	Flye	Longcyclor	Miniasm	Necat	Raven	Redbean
<i>C. farmeri</i>	100	100	100	100	100	100	100
<i>C. freundii</i>	99.2	99.2	97.7	93.2	85	94	87.2
<i>E. cloacae</i>	100	100	100	96.9	91.9	98.8	96.3
<i>E. coli</i>	99.8	99.8	97.1	99.6	98	91.2	81.8
<i>K. oxytoca</i>	100	100	100	100	100	100	75
<i>K. pneumoniae</i>	100	99.6	98.9	98.6	99.3	94.6	84.4
<i>P. aeruginosa</i>	100	100	100	100	100	100	97.8
<i>P. mirabilis</i>	100	100	98.1	98.1	100	98.1	71.2
<i>P. stuartii</i>	100	100	97.1	100	97.1	97.1	91.4
<i>S. aureus</i>	97.3	96.7	96	95.3	95.7	95.7	68.2
<i>C. farmeri</i>	100	100	100	100	100	100	100
<i>C. freundii</i>	100	100	100	100	100	100	100
<i>E. cloacae</i>	95.3	95.3	95.3	95.1	94.9	95.2	95.1
<i>E. coli</i>	99	99.2	99.2	99.2	99.2	99.1	99.3
<i>K. oxytoca</i>	100	100	100	100	100	100	100
<i>K. pneumoniae</i>	98	98	98	97.9	98	97.9	98.7
<i>P. aeruginosa</i>	97.9	97.9	97.9	97.9	97.9	97.9	100
<i>P. mirabilis</i>	92.9	92.9	92.7	92.7	92.9	92.7	90.2
<i>P. stuartii</i>	100	100	100	100	100	100	100
<i>S. aureus</i>	98.6	99	99.3	99	99.3	99.3	99

B Specificity and sensitivity of plasmid replicon calling							
	Canu	Flye	Longcyclor	Miniasm	Necat	Raven	Redbean
<i>C. farmeri</i>	100	100	100	100	100	100	100
<i>C. freundii</i>	100	100	62.2	91.9	62.2	56.8	56.8
<i>E. cloacae</i>	97.9	100	81.2	89.6	83.3	87.5	72.9
<i>E. coli</i>	99.1	92.2	77.2	90	80.8	75.8	57.1
<i>K. oxytoca</i>	100	100	66.7	100	66.7	100	100
<i>K. pneumoniae</i>	100	96.7	82.7	93.5	86	79	63.6
<i>P. aeruginosa</i>	100	100	100	100	100	100	100
<i>P. mirabilis</i>	100	100	66.7	66.7	100	66.7	33.3
<i>P. stuartii</i>	100	100	33.3	66.7	33.3	33.3	33.3
<i>S. aureus</i>	100	95.9	83.5	87.6	89.7	88.7	42.3
<i>C. farmeri</i>	100	100	100	100	100	100	100
<i>C. freundii</i>	100	100	100	100	100	100	100
<i>E. cloacae</i>	97.9	100	100	100	100	100	100
<i>E. coli</i>	99.1	99.1	99.2	99.4	99.4	99.4	99.2
<i>K. oxytoca</i>	100	100	100	100	100	100	100
<i>K. pneumoniae</i>	100	99	100	100	100	100	100
<i>P. aeruginosa</i>	100	100	100	100	100	100	100
<i>P. mirabilis</i>	100	100	100	100	100	100	100
<i>P. stuartii</i>	100	100	100	100	100	100	100
<i>S. aureus</i>	92.4	100	96.4	98.8	100	97.7	97.6

can be found in Additional file 4: Table S3. The specificity of plasmid replicon detection was best for Canu assembled long-read data (Table 3B) but, as mentioned previously, was hampered by the multi-copy assembly artifact issue (Additional file 2: Fig. S3). Next to Canu, Flye performed excellent with an unweighted average specificity of 98.5% over all species (Table 3B). Identifying the correct plasmid replicon was excellent and did not seem to differ much among assemblers as the sensitivity was between 98.9% (Canu) and 99.9% (Flye).

Long-read sequencing enables analysis of an MRSA outbreak

There was an impetigo-associated MRSA outbreak with MLVA-type MT4627 in the Mid-East of the Netherlands in 2019 (Fig. 4A + B) [50]. The number of MRSA MT4627 isolates in the South-West region increased from 14 (in 2021) and 28 (in 2022) to 40 MRSA MT4627 isolates in 2023 and were obtained from 40 persons (Fig. 4A). Characterization of a subset of the 2023 MRSA outbreak isolates using both long-read ($n=16$) and short-read ($n=16$) sequencing-based wgMLST analysis revealed genetic clustering (≥ 2 isolates differing ≤ 15 wgMLST alleles) with 2019 outbreak and isolates from 2020, 2021, and 2022 in the minimum spanning tree (Fig. 4B). Long-read sequenced MRSA outbreak isolates were in close proximity of their short-read sequenced counterparts (Fig. 4B). MLST, MLVA, and in silico MLVA (iMLVA) analyses revealed that MLST sequence type and iMLVA type could be retrieved from long-read sequencing data with $>40\times$ coverage and were concordant with fragment size-based in vitro MLVA types. The outbreak MRSA isolates were from MLST ST121, clonal complex (CC) CC121, MLVA type MT4627, and iMLVA type MT4627. A subset of the MRSA isolates from 2023 have diversified over time. Long-read sequencing also yielded identical AMR genes *aac(6')-aph(2'')*, *mecA*, *fusC*, *blaZ*, and *dfrG*, plasmid replicons, and virulence genes, when compared to short-read sequencing (Fig. 4C, Additional file 5: Table S4). The plasmid-borne *ermC* gene was lacking three long-read sequenced isolates. Long-read sequencing outperformed detection of sortase B substrate genes encoding microbial surface components recognizing adhesive matrix molecules (MSCRAMM) such as *clfB*, *sdrC*, *sdrD*, and *sdrE* (Fig. 4C).

Discussion

We demonstrate that automated DNA extraction followed by Nanopore R10.4.1 long-read sequencing is a reliable method for the molecular typing (iMLVA, MLST, wgMLST, and wgSNP), AMR gene, plasmid replicon, virulence gene identification, and outbreak detection of MDRO and can therefore be applied for genomic

surveillance. A sufficient sequencing depth is required, and in this study, a minimum of $40\times$ sequencing depth was used. Here, we have multiplexed 24 isolates per flow cell. This has been our workflow for hybrid-based de novo assembly and yielded enough sequencing data for completed bacterial genomes. However, for long-read only assemblies, we have demonstrated here that this does not yield enough data for each isolate to reach the required $40\times$ coverage. Therefore, we do not recommend multiplexing 24 isolates per flow cell, but keeping it to a maximum of 16–18 isolates, depending on the genome size of the bacterial species. Previous studies using Illumina short-read sequencing used $30\times$ coverage as a minimum [53, 54]. Older Nanopore flow cell generations had higher error rates of raw Nanopore reads compared to Illumina raw reads and may explain that still a higher coverage is required for R10.4.1 with current models. Long-read based wgMLST seemed to be more efficient for species with a small chromosome, e.g., *S. aureus* than for those with larger chromosomes like *P. aeruginosa*. Unique signatures in *P. aeruginosa* data and or a lack of representation in the training set for the models is one explanation for poor performance. Another possibility is that higher sequencing depth is required for bacteria with large chromosomes to obtain sufficient good quality reads. To overcome this problem, one could reload the DNA library prep during a long-read sequencing run. Or alternatively, a Promethion sequencing run can boost coverage, as the sequencing output is higher, to overcome this problem. Two different basecalling models were tested using Dorado duplex mode with and without Medaka polishing and Rerio; this model has an increased performance for highly methylated DNA motifs. Overall, Rerio performed best for wgMLST allele calling. Research models such as Rerio are continuously updated and if proven successful, implemented into the normal workflow as has been the case for this research model where the same training set was used and now included in dna_r10.4.1_e8.2_400bps_sup@v4.3.0. Although these updates provide a constant improvement of the ability to perform molecular characterization of bacterial isolates, it imposes a problem for laboratories using Nanopore long-read sequencing under quality assurance systems. The implementation of guidelines such as ISO15189 for national reference and medical microbiology laboratories or ISO23418 for food-borne pathogen sequencing requires the validation of every new basecalling model before implementation and is slowing down the introduction of improved methods.

For de novo assembly, Canu was unsuitable for AMR gene detection as it resulted in the erroneous detection of multiple copies of the same gene. This is likely due to the inability of this algorithm to properly overlap contig ends. Although Trycycler [55] is among the best

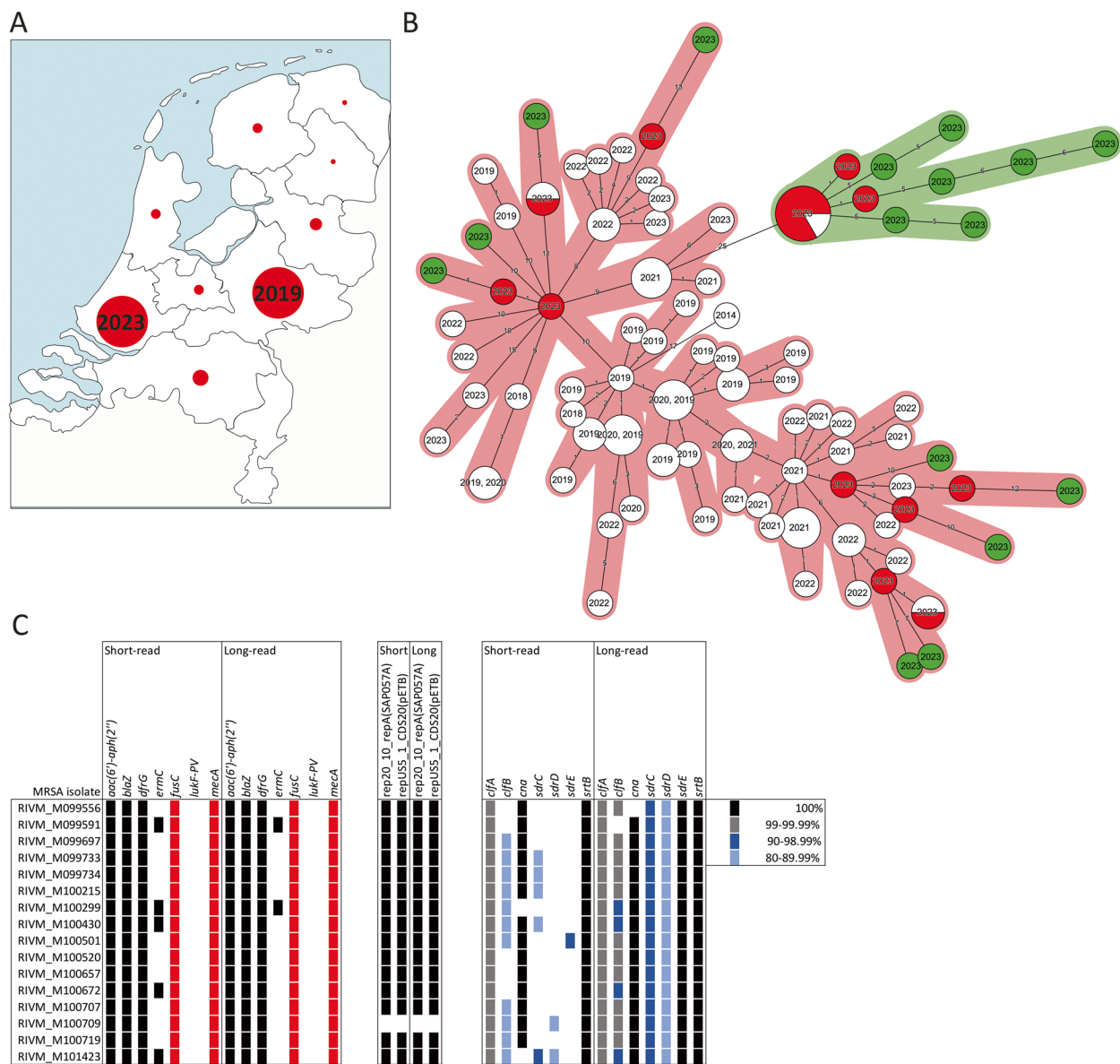


Fig. 4 Long-read sequencing-based analyses of an impetigo-associated MRSA outbreak in the Netherlands 2023. **A** Geographic localization of persons with an MT4627 MRSA in the Netherlands. The initial outbreak in 2019 has been described previously [50]. After this outbreak, this type mainly occurred in the province of Zuid-Holland, The Netherlands, towards the end of 2023. **B** Minimum spanning tree of MT4627 MRSA analyzed by wgMLST. In total 16 isolates from 2023 were long-read sequenced to 40× coverage and included in the analysis (green, Table 1). Of these 16 isolates, the short-read counterpart was included in the figure (red). A wgMLST cluster cut-off of 15 was used. Halo's indicates isolates varying ≤ 15 wgMLST alleles. **C** Comparison of short-read and long-read sequenced resistomes, plasmid replicons, and genes encoding microbial surface components recognizing adhesive matrix molecules (MSCRAMM) of the outbreak-associated cluster isolates

long-read only de novo assembler in the community, this tool was not used as it requires multiple manual curation steps in each assembly, thereby hampering the ability for automated high-throughput bacterial assemblies. The use of Miniasm or Longcyler (Unicycler supplied with long-reads only) resulted in the lowest number of different wgMLST alleles compared to their Illumina short-read

sequenced counterparts. Besides *Pseudomonas*, all other species tested had only a few discrepant wgMLST allele calls (0 to 12 alleles). In addition, when employing wgSNP analyses using Mintyper, excellent results were obtained, as for all species including *P. aeruginosa*, which yielded on average 1 and a maximum of 10 SNPs between Illumina short-read and Nanopore long-read sequencing

data. This was on the same level of variation in a recent multi-center study where Enterobacterales, Enterococci, and Staphylococci isolates were sent to participating laboratories for short-read sequencing and subsequent molecular typing analysis [53, 54]. Mintyper is a bioinformatic tool that was made to address the shortcomings of noisy Nanopore long-read sequencing data, and this likely explains its excellent performance for *P. aeruginosa* compared to wgMLST methods. Therefore, we demonstrate the interoperability of long-read sequencing with existing WGS databases generated by Illumina short-read sequencing for surveillance purposes. For identification of AMR genes and plasmid replicons, Flye performed better than Longcyclus and Miniasm, even though these two assemblers were best for wgMLST allele calling. Overall, genotyping was excellent and on par with other studies investigating the inter-laboratory reproducibility of Illumina short-read sequencing-based genotyping, where they found >99% performance [53, 56]. Notably, long-read sequencing-based de novo assembly methods can better discriminate multi-copy AMR genes in mobile genetic elements such as plasmids, as these regions are impossible to assemble with short-read sequencing. Therefore, the evaluation of the specificity for AMR genes and replicon identification may not be the best metric to evaluate the performance of long-read sequencing methods when using Illumina short-read sequencing data as gold standard. Additionally, short-read assembly is unable to truly identify multi-copy AMR genes. However, no other reference is available. Finally, it should be noted that genotype to phenotype is still difficult to infer, and large discrepancies have been observed among methods used in a multi-center study investigating this challenge [57]. Furthermore, Nanopore long-read sequencing was superior to Illumina short-read sequencing of genes encoding MSCRAMMs. MRSCRAMMs are known to harbor multiple repetitive domains and are implicated in binding to collagen, fibrinogen, and cytokeratin components of the extracellular matrix [58].

Conclusion

For laboratories wanting to implement Nanopore long-read sequencing of MDRO, we recommend using a minimum 40× coverage, Rerio basecalling, and Miniasm or Longcyclus as de novo assembler for molecular typing and outbreak detection for genomic surveillance. For the best performing AMR gene and plasmid replicon detection, we recommend using Flye instead of Canu, Miniasm, or Longcyclus, as Canu generated assembly artifacts and Miniasm and Longcyclus did not perform as good as Flye on sensitivity and specificity. Future studies are needed to optimize the performance

of Nanopore long-read sequencing for *P. aeruginosa*. The use of long-read sequencing can provide additional valuable insights into virulence determinants, resistance plasmids, and resistance gene copy number of MDRO. This may help to inform and guide effective control measures for MDRO which were previously not possible using short-read sequencing. Importantly, the relatively low purchase and implementation costs of long-read sequencing and rapid library preparation not only enables genomic surveillance and outbreak analysis but extends its applicability to resource-constrained settings and low-income countries worldwide.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13073-024-01412-6>.

Additional file 1: Table S1. All isolates used in this study with additional meta data.

Additional file 2: Fig S1. Genome continuity indicated by the L90 score per assembler and species. Fig S2. Number of SNPs between Illumina reads and Rerio basecalled Nanopore reads. Fig S3. The number of AMR genes called per assembler and species. Fig S4. The genome size for each species and assembler. Fig S5. The number of plasmid replicons called for each assembler and species.

Additional file 3: Table S2. Observed differences in MLST ST between long-read and short-read assemblies by absence of one or more alleles in the long-read assemblies.

Additional file 4: Table S3. Genes called per assembler that were either called by long read assembly but not Illumina or called by Illumina but not by long read assembly.

Additional file 5: Table S4. AMR genes, plasmid replicons, virulence genes called by short read and long read for the MRSA outbreak isolates.

Acknowledgements

We thank all the members of the Dutch CPE and MRSA surveillance study Groups and the Dutch medical microbiology laboratories for submitting MDRO isolates to the RIVM for the national CPE/MRSA surveillance program. We thank Dr. Romy D. Zwartink, Dr. Rob Mariman, and Dr. Daan W. Notermans for critical reading of the manuscript.

Members of the Dutch CPE/MRSA Surveillance Study Group:

- A.L.E. van Arkel, ADZ medisch centrum, Department of Medical Microbiology, Goes
- M.A. Leversteijn-van Hall, Alrijne Hospital, Department of Medical Microbiology, Leiden
- W. van den Bijllaardt, Amphibia Hospital, Microvida Laboratory for Microbiology, Breda
- R. van Mansfeld, Amsterdam UMC—location AMC, Department of Medical Microbiology and Infection Prevention, Amsterdam
- K. van Dijk, Amsterdam UMC—location Vumc, Department of Medical Microbiology and Infection Control, Amsterdam
- B. Zwart, Atalmedial, Department of Medical Microbiology, Amsterdam
- B.M.W. Diederens, Bravis Hospital/ZorgSaam Hospital Zeeuws-Vlaanderen, Department of Medical Microbiology, Roosendaal/Terneuzen
- H. Berkhout, Canisius Wilhelmina Hospital, Department of Medical Microbiology and Infectious Diseases, Nijmegen
- D.W. Notermans, Centre for Infectious Disease Control, National Institute for Public Health and the Environment, Bilthoven
- A. Ott, Certe, Department of Medical Microbiology Groningen & Drenthe, Groningen
- K. Waar, Certe, Department of Medical Microbiology Friesland & Noordoostpolder, Leeuwarden

- W. Ang, Comicro, Department of Medical Microbiology, Hoorn
- J. da Silva, Deventer Hospital, Department of Medical Microbiology, Deventer
- A.L.M. Vlek, Diaconessenhuis Utrecht, Department of Medical Microbiology and Immunology, Utrecht
- A.G.M. Buiting, Elisabeth-TweeSteden (ETZ) Hospital, Department of Medical Microbiology and Immunology, Tilburg
- L.G.M. Bode, Erasmus University Medical Center, Department of Medical Microbiology and Infectious Diseases, Rotterdam
- A. Jansz, Eurofins PAMM, Department of Medical Microbiology, Veldhoven
- S. Paltansing, Franciscus Gasthuis & Vlietland, Department of Medical Microbiology and Infection Control, Rotterdam
- A.J. van Griethuysen, Gelderse Vallei Hospital, Department of Medical Microbiology, Ede
- J.R. Lo Ten Foe, Gelre Hospital, Department of Medical Microbiology and Infection Control, Apeldoorn
- M.J.C.A. van Trijp, Groene Hart Ziekenhuis, Department of Medical Microbiology and Infection Prevention, Gouda
- M. Wong, Haga Hospital, Department of Medical Microbiology, 's-Gravenhage
- A.E. Muller, HMC Westeinde Hospital, Department of Medical Microbiology, 's-Gravenhage
- M.P.M. van der Linden, IJsselland hospital, Department of Medical Microbiology, Capelle a/d IJssel
- M. van Rijn, Ikazia Hospital, Department of Medical Microbiology, Rotterdam
- S.B. Debast, Isala Hospital, Laboratory of Medical Microbiology and Infectious Diseases, Zwolle
- E. Kolwijck, Jeroen Bosch Hospital, Department of Medical Microbiology and Infection Control, 's-Hertogenbosch
- N. Al Naiemi, LabMicTA, Regional Laboratory of Microbiology Twente Achterhoek, Hengelo
- T. Schulin, Laurentius Hospital, Department of Medical Microbiology, Roermond
- S. Dinant, Maasstad Hospital, Department of Medical Microbiology, Rotterdam
- S.P. van Mens, Maastricht University Medical Centre, Department of Medical Microbiology, Infectious Diseases & Infection Prevention, Maastricht
- D.C. Melles, Meander Medical Center, Department of Medical Microbiology, Amersfoort
- J.W.T. Cohen Stuart, Noordwest Ziekenhuisgroep, Department of Medical Microbiology, Alkmaar
- P. Gruteke, OLVG Lab BV, Department of Medical Microbiology, Amsterdam
- A. P. van Dam, Amsterdam Health Service, Public Health Laboratory, Amsterdam
- I. Maat, Radboud University Medical Center, Department of Medical Microbiology, Nijmegen
- B. Maraha, Regional Laboratory for Microbiology, Department of Medical Microbiology, Dordrecht
- J.C. Sinnige, Regional Laboratory of Public Health, Department of Medical Microbiology, Haarlem
- E. van der Vorm, Reinier de Graaf Groep, Department of Medical Microbiology, Delft
- M.P.A. van Meer, Rijnstate Hospital, Laboratory for Medical Microbiology and Immunology, Velp
- M. de Graaf, Salto Diagnostic Centre, Department of Medical Microbiology, Utrecht
- E. de Jong, Slingeland Hospital, Department of Medical Microbiology, Doetinchem
- S.J. Vainio, St Antonius Hospital, Department of Medical Microbiology and Immunology, Nieuwegein
- E. Heikens, St Jansdal Hospital, Department of Medical Microbiology, Harderwijk
- M. den Reijer, Star-shl diagnostic centre, Department of Medical Microbiology, Rotterdam
- J.W. Dorigo-Zetsma, TergooiMC, Central Bacteriology and Serology Laboratory, Hilversum
- A. Troelstra, University Medical Center Utrecht, Department of Medical Microbiology, Utrecht
- J. de Vries, VieCuri Medical Center, Department of Medical Microbiology, Venlo
- D.W. van Dam, Zuyderland Medical Centre, Department of Medical

- E.I.G.B. de Brauer, Zuyderland Medical Centre, Department of Medical Microbiology and Infection Control, Heerlen
- R. Steingrover, St. Maarten Laboratory Services, Department of Medical Microbiology, Cay Hill (St. Maarten)
- Analytical Diagnostic Center N.V. Curaçao, Department of Medical Microbiology, Willemstad (Curaçao)

Authors' contributions

Conceptualization and methodology, FL, CJ, and APAH; visualization, FL, CJ, and APAH; data curation, FL, HvdH, SW; formal analysis, FL, CJ, and APAH; laboratory experiments, AdH and JB; manuscript preparation—original draft FL, CJ, LMS, and APAH; review and editing, FL, CJ, and APAH. All authors read and approved the final manuscript.

Funding

This research was funded by the Dutch Ministry of Health, Welfare and Sport (V/150302/22/BR).

Data availability

Raw short-read and long-read sequencing data of 356 Dutch CPE/CPPA/CRAB/MRSA surveillance isolates have been deposited in the SRA database under BioProject numbers PRJNA1076692, PRJNA1076808, and PRJNA903550 (Additional file 1: Table S1). The authors confirm that all supporting data, protocols, and accession numbers have been provided within the article and through supplementary data files.

Declarations

Ethics approval and consent to participate

Ethical approval was not required for the present study, since it is based on genomic surveillance data only. Samples from which the isolates were cultured were all collected as part of routine health care.

Consent for publication

All authors approved the final version of the manuscript prior publication.

Competing interests

The authors do not have any financial or non-financial competing interests that may undermine the objectivity, integrity, and value of this study. Illumina and Oxford Nanopore Technologies were not involved in the design, execution, and analyses nor the interpretation of data or conclusions from this study.

Author details

¹Centre for Infectious Disease Control, National Institute for Public Health and the Environment (RIVM), Bilthoven, the Netherlands.

Received: 19 February 2024 Accepted: 11 November 2024

Published online: 25 November 2024

References

1. Tenover FC, Arbeit RD, Goering RV. How to select and interpret molecular strain typing methods for epidemiological studies of bacterial infections: a review for healthcare epidemiologists. molecular typing working group of the society for healthcare epidemiology of America. *Infect Control Hosp Epidemiol.* 1997;18:426–39.
2. Ichiyama S, Ohta M, Shimokata K, Kato N, Takeuchi J. Genomic DNA fingerprinting by pulsed-field gel electrophoresis as an epidemiological marker for study of nosocomial infections caused by methicillin-resistant *Staphylococcus aureus*. *J Clin Microbiol.* 1991;29:2690–5.
3. Savelkoul PHM, Aarts HJM, de Haas J, Dijkshoorn L, Duim B, Otsen M, et al. Amplified-fragment length polymorphism analysis: the state of an art. *J Clin Microbiol.* 1999;37:3083–91.
4. Frénay HM, Theelen JP, Schouls LM, Vandenbroucke-Grauls CM, Verhoef J, van Leeuwen WJ, et al. Discrimination of epidemic and nonepidemic methicillin-resistant *Staphylococcus aureus* strains on the basis of protein A gene polymorphism. *J Clin Microbiol.* 1994;32:846–7.
5. Harmsen D, Claus H, Witte W, Rothgänger J, Claus H, Turnwald D, et al. Typing of methicillin-resistant *Staphylococcus aureus* in a university

- hospital setting by using novel software for spa repeat determination and database management. *J Clin Microbiol.* 2003;41:5442–8.
6. Enright MC, Day NP, Davies CE, Peacock SJ, Spratt BG. Multilocus sequence typing for characterization of methicillin-resistant and methicillin-susceptible clones of *Staphylococcus aureus*. *J Clin Microbiol.* 2000;38:1008–15.
 7. Maiden MC, Bygraves JA, Feil E, Morelli G, Russell JE, Urwin R, et al. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc Natl Acad Sci U S A.* 1998;95:3140–5.
 8. Maiden MCJ. Multilocus sequence typing of bacteria. *Annu Rev Microbiol.* 2006;60:561–88.
 9. Urwin R, Maiden MCJ. Multi-locus sequence typing: a tool for global epidemiology. *Trends Microbiol.* 2003;11:479–87.
 10. Vos P, Hogers R, Bleeker M, Reijans M, Lee TVD, Hornes M, et al. AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res.* 1995;23:4407–14.
 11. van Alphen LB, von Wintersdorff CJH, Savelkoul PHM. Epidemiological typing using WGS. In: Moran-Gilad J, Yagel Y, editors. *Application and Integration of Omics-powered Diagnostics in Clinical and Public Health Microbiology*. Cham: Springer International Publishing; 2021. p. 69–87. Available from: https://doi.org/10.1007/978-3-030-62155-1_5. Cited 2024 Jan 31.
 12. Aanensen DM, Carlos CC, Donado-Godoy P, Okeke IN, Ravikumar KL, NIHR Global Health Research Unit on Genomic Surveillance of Antimicrobial Resistance. Implementing whole-genome sequencing for ongoing surveillance of antimicrobial resistance: exemplifying insights into *Klebsiella pneumoniae*. *Clin Infect Dis.* 2021;73:S255–7.
 13. Loman NJ, Pallen MJ. Twenty years of bacterial genome sequencing. *Nat Rev Microbiol.* 2015;13:787–94.
 14. Schouls LM, Witteveen S, van Santen-Verheuevel M, de Haan A, Landman F, van der Heide H, et al. Molecular characterization of MRSA collected during national surveillance between 2008 and 2019 in the Netherlands. *Commun Med (Lond).* 2023;3:123.
 15. Stoddart D, Heron AJ, Mikhailova E, Maglia G, Bayley H. Single-nucleotide discrimination in immobilized DNA oligonucleotides with a biological nanopore. *Proc Natl Acad Sci U S A.* 2009;106:7702–7.
 16. Lu H, Giordano F, Ning Z. Oxford nanopore MinION sequencing and genome assembly. *Genomics Proteomics Bioinform.* 2016;14:265–79.
 17. Foster-Nyarko E, Cottingham H, Wick RR, Judd LM, Lam MMC, Wyres KL, et al. Nanopore-only assemblies for genomic surveillance of the global priority drug-resistant pathogen, *Klebsiella pneumoniae*. *Microb Genom.* 2023;9:mgen000936.
 18. Ashton PM, Nair S, Dallman T, Rubino S, Rabsch W, Mwaigwisya S, et al. MinION nanopore sequencing identifies the position and structure of a bacterial antibiotic resistance island. *Nat Biotechnol.* 2015;33:296–300.
 19. Nanopore sequencing accuracy | Oxford Nanopore Technologies. Oxford Nanopore Technologies. Available from: <https://nanoporetech.com/platform/accuracy>. Cited 2024 Nov 8.
 20. Hendrickx APA, Landman F, de Haan A, Witteveen S, van Santen-Verheuevel MG, Schouls LM, et al. bla OXA-48-like genome architecture among carbapenemase-producing *Escherichia coli* and *Klebsiella pneumoniae* in the Netherlands. *Microb Genom.* 2021;7:7.
 21. Rapid sequencing DNA V14 - barcoding (SQK-RBK114.24 or SQK-RBK114.96) (RBK_9176_v114_revN_30Sept2024). Oxford Nanopore Technologies. 2022. Available from: <https://nanoporetech.com/document/rapid-sequencing-gdna-barcoding-sqk-rbk114>. Cited 2024 Nov 8.
 22. Nanoporetech/dorado. Oxford Nanopore Technologies; 2024. Available from: <https://github.com/nanoporetech/dorado>. Cited 2024 Nov 8.
 23. GitHub - nanoporetech/erio: research release basecalling models and configurations. Available from: <https://github.com/nanoporetech/erio>. Cited 2024 Nov 8.
 24. GitHub - RIVM-bioinformatics/Submission-Assembler: snakemake workflow to assemble long read data. Available from: <https://github.com/RIVM-bioinformatics/Submission-Assembler>. Cited 2024 Nov 8.
 25. De Coster W, Rademakers R. NanoPack2: population-scale evaluation of long-read sequencing data. *Bioinformatics.* 2023;39:btad311.
 26. GitHub - rrrwick/Filtlong: quality filtering tool for long reads. Available from: <https://github.com/rrrwick/Filtlong>. Cited 2024 Nov 8.
 27. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* 2017;27:722–36.
 28. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol.* 2019;37:540–6.
 29. Li H. Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences. *Bioinformatics.* 2016;32:2103–10.
 30. Chen Y, Nie F, Xie S-Q, Zheng Y-F, Dai Q, Bray T, et al. Efficient assembly of nanopore reads via highly accurate and intact error correction. *Nat Commun.* 2021;12:60.
 31. Vaser R, Šikić M. Time- and memory-efficient genome assembly with Raven. *Nat Comput Sci.* 2021;1:332–6.
 32. Ruan J, Li H. Fast and accurate long-read assembly with wtdbg2. *Nat Methods.* 2020;17:155–8.
 33. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol.* 2017;13: e1005595.
 34. GitHub - nanoporetech/medaka: sequence correction provided by ONT Research. Available from: <https://github.com/nanoporetech/medaka>. Cited 2024 Nov 8.
 35. QUAST: quality assessment tool for genome assemblies - PubMed. Available from: <https://pubmed.ncbi.nlm.nih.gov/23422339/>. Cited 2024 Jan 31.
 36. RIVM-bioinformatics/juno-assembly. Infectieziekteonderzoek, Diagnostiek en laboratorium Surveillance-Bioinformatics (RIVM, The Netherlands); 2024. Available from: <https://github.com/RIVM-bioinformatics/juno-assembly>. Cited 2024 Nov 8.
 37. Wingett SW, Andrews S. FastQ Screen: a tool for multi-genome mapping and quality control. *F1000Res.* 2018;7:1338.
 38. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics.* 2018;34:i884–90.
 39. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol.* 2012;19:455–77.
 40. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* 2015;25:1043–55.
 41. kbaseapps/BBTools. kbaseapps; 2024. Available from: <https://github.com/kbaseapps/BBTools>. Cited 2024 Nov 8.
 42. Schouls LM, Spalburg EC, van Luit M, Huijsdens XW, Pluister GN, van Santen-Verheuevel MG, et al. Multiple-locus variable number tandem repeat analysis of *Staphylococcus aureus*: comparison with pulsed-field gel electrophoresis and spa-typing. *PLoS ONE.* 2009;4: e5082.
 43. Bosch T, Pluister GN, van Luit M, Landman F, van Santen-Verheuevel M, Schot C, et al. Multiple-locus variable number tandem repeat analysis is superior to spa typing and sufficient to characterize MRSA for surveillance purposes. *Future Microbiol.* 2015;10:1155–62.
 44. RIVM-bioinformatics/in-silico-mlva. Infectieziekteonderzoek, Diagnostiek en laboratorium Surveillance-Bioinformatics (RIVM, The Netherlands); 2024. Available from: <https://github.com/RIVM-bioinformatics/in-silico-mlva>. Cited 2024 Nov 8.
 45. GitHub - tseemann/abricate: :mag_right: mass screening of contigs for antimicrobial and virulence genes. Available from: <https://github.com/tseemann/abricate>. Cited 2024 Nov 8.
 46. Jünemann S, Sedlazeck FJ, Prior K, Albersmeier A, John U, Kalinowski J, et al. Updating benchtop sequencing performance comparison. *Nat Biotechnol.* 2013;31:294–6.
 47. Witteveen S, Hans JB, Izdebski R, Hasman H, Samuelsen Ø, Dortet L, et al. Dissemination of extensively drug-resistant NDM-producing *Providencia stuartii* in Europe linked to patients transferred from Ukraine, March 2022 to March 2023. *Euro Surveill.* 2024;29:2300616.
 48. Leopold SR, Goering RV, Witten A, Harmsen D, Mellmann A. Bacterial whole-genome sequencing revisited: portable, scalable, and standardized analysis for typing and detection of virulence and antibiotic resistance genes. *J Clin Microbiol.* 2014;52:2365–70.
 49. Hendrickx APA, Debast S, Pérez-Vázquez M, Schoffelen AF, Notermans DW, Landman F, et al. A genetic cluster of MDR *Enterobacter cloacae* complex ST78 harbouring a plasmid containing bla VIM-1 and mcr-9 in the Netherlands. *JAC Antimicrob Resist.* 2021;3:dlab046.
 50. Vendrik KEW, Kuijper EJ, Dimmendaal M, Silvis W, Denie-Verhaegh E, de Boer A, et al. An unusual outbreak in the Netherlands: community-onset

impetigo caused by a methicillin-resistant *Staphylococcus aureus* with additional resistance to fusidic acid, June 2018 to January 2020. *Euro Surveill.* 2022;27:2200245.

51. Repizo GD, Espariz M, Seravalle JL, Díaz Miloslavich JI, Steimbrüch BA, Shuman HA, et al. *Acinetobacter baumannii* NCIMB8209: a rare environmental strain displaying extensive insertion sequence-mediated genome remodeling resulting in the loss of exposed cell structures and defensive mechanisms. *mSphere.* 2020;5:e00404–20.
52. Tomaschek F, Higgins PG, Stefanik D, Wisplinghoff H, Seifert H. Head-to-head comparison of two multi-locus sequence typing (MLST) schemes for characterization of *Acinetobacter baumannii* outbreak and sporadic isolates. *PLoS ONE.* 2016;11:e0153014.
53. Jamin C, De Koster S, van Koeveeringe S, De Coninck D, Mensaert K, De Bruyne K, et al. Harmonization of whole-genome sequencing for outbreak surveillance of Enterobacteriaceae and Enterococci. *Microb Genom.* 2021;7: 000567.
54. Mellmann A, Andersen PS, Bletz S, Friedrich AW, Kohl TA, Lilje B, et al. High interlaboratory reproducibility and accuracy of next-generation-sequencing-based bacterial genotyping in a ring trial. *J Clin Microbiol.* 2017;55:908–13.
55. Wick RR, Judd LM, Cerdeira LT, Hawkey J, Méric G, Vezina B, et al. Tricycler: consensus long-read assemblies for bacterial genomes. *Genome Biol.* 2021;22:266.
56. Coolen JPM, Jamin C, Savelkoul PHM, Rossen JWA, Wertheim HFL, Matamoros SP, et al. Centre-specific bacterial pathogen typing affects infection-control decision making. *Microb Genom.* 2021;7: 000612.
57. Doyle RM, O'Sullivan DM, Aller SD, Bruchmann S, Clark T, Coello Pelegrin A, et al. Discordant bioinformatic predictions of antimicrobial resistance from whole-genome sequencing data of bacterial isolates: an inter-laboratory study. *Microb Genom.* 2020;6:e000335.
58. Foster TJ. The MSCRAMM family of cell-wall-anchored surface proteins of gram-positive cocci. *Trends Microbiol.* 2019;27:927–41.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.