MICROBIAL GENOMICS

RESEARCH ARTICLE

Cottingham *et al.*, *Microbial Genomics* 2026;12:001657

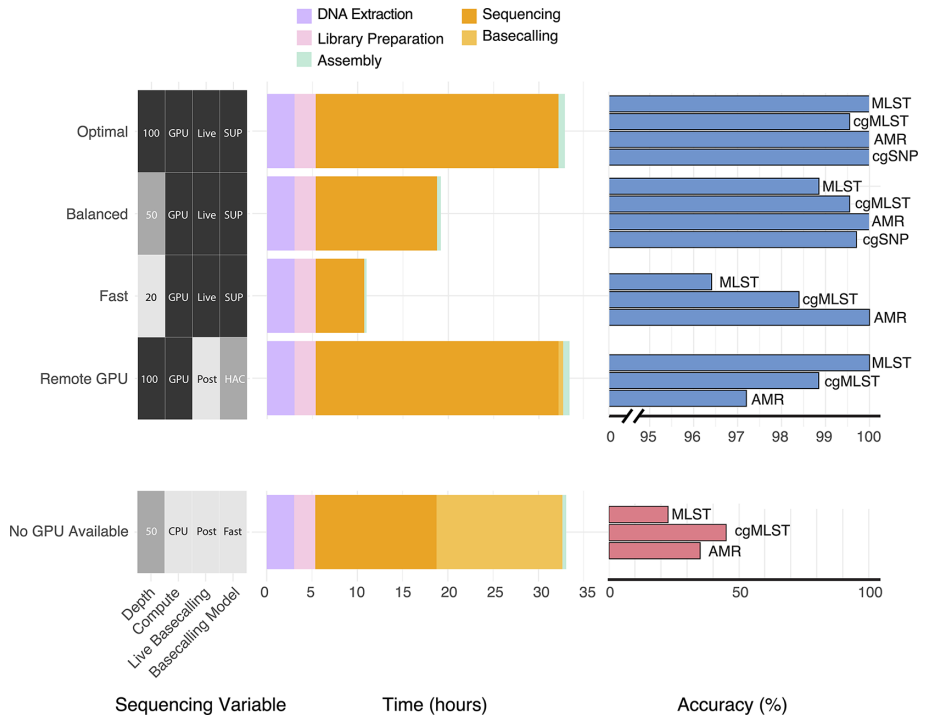
DOI 10.1099/mgen.0.001657

MICROBIOLOGY SOCIETY

OPEN DATA OPEN ACCESS

Nanopore sequencing enables highly accurate genotyping and identification of resistance determinants in key nosocomial pathogens

Hugh Cottingham[1], Louise M. Judd[1], Taylor Harshegyi-Hand[1], Jessica A. Wisniewski[1], Luke V. Blakeway[1], Tiffany Hong[1], Matthew H. Parker[1], Anton Y. Peleg[1,2,3], Kathryn E. Holt[1,4], Jane Hawkey[1,2,†] and Nenad Macesic[1,2,*,†]



Graphical Abstract

Time requirements and expected analysis accuracy across five methods for multiplexed ONT sequencing of 20 bacterial isolates. Accuracy metrics summarize the benchmarking outcomes using R10.4.1 Dorado polished assemblies for MLST, cgMLST and AMR, except for AMR typing at 20× depth, where raw reads are recommended. cgSNP accuracy was summarized from simulated R10.4.1 readsets. Accuracy metrics are coloured by GPU availability (blue, available; red, unavailable). DNA extraction and library preparation time requirements are sourced from Genfind and ONT's protocol approximations [1, 2]. Sequencing time was calculated based on all ONT runs that contributed to the benchmarking dataset. Basecalling and assembly time requirements were benchmarked on an NVIDIA A100 GPU with 60 CPUs and 40 GB of RAM.

Abstract

Whole-genome sequencing of bacterial pathogens can positively impact infectious disease management in clinical contexts, both in individual settings and by assisting infection prevention efforts. However, logistical issues have often prevented its

[Continued on next page]

translation into clinical settings. Oxford Nanopore Technologies (ONT) platforms are flexible and affordable and now offer accuracy comparable to other sequencing platforms, making them uniquely well-suited for clinical bacterial isolate sequencing. We sought to determine the best methods for implementing ONT sequencing into clinical settings by benchmarking multi-locus sequence typing (MLST), core genome multi-locus sequence typing (cgMLST), antimicrobial resistance (AMR) and core genome SNP (cgSNP) typing against the genomic automated gold standard. We sequenced 199 *Enterobacterales* isolates with Illumina and ONT platforms and assessed performance based on sequencing chemistry, basecaller, basecalling model, assembly status, assembly polishing and sequencing depth. Modern ONT data generated perfect MLST and AMR allelic variant calls and correctly classified a median of 99.5% of cgMLST loci. Illumina and kmer-based SNP typing failed to call 9–28 SNPs per 1,000 sites due to poor sensitivity in repetitive regions of the reference genome, while ONT's long reads generated perfect SNP calls across the entire genome using simulated readsets. Using real sequencing data to identify putative transmission pairs, ONT read-based methods were concordant with traditional Illumina approaches in 155–158/158 (98.1–100%) of isolate pairs. We also provide specific recommendations on sequencing depth and basecalling model based on the time and computational resources available to the user. This study demonstrates the viability of modern ONT data for highly accurate characterization of bacterial pathogens to support their future integration into clinical settings.

Impact Statement

Whole-genome sequencing (WGS) of bacterial pathogens has historically been dominated by Illumina platforms due to their high accuracy and throughput. Oxford Nanopore Technologies (ONT) offers several benefits for clinical WGS, including real-time analysis, long reads, portability and lower purchasing costs, but has historically suffered from a high error rate. In this study, we build on previous evidence suggesting that this accuracy has improved and demonstrate that it is now a viable alternative for key characterization tasks across three common Gram-negative pathogens. We also provide recommendations on sequencing depth and basecalling model depending on the turnaround time, computational resources and accuracy requirements of the user. These findings reinforce the potential for widespread implementation of ONT WGS into clinical settings to improve patient outcomes and surveillance efforts.

DATA SUMMARY

All data used in this study are freely available via Figshare (https://doi.org/10.26180/c.7942409.v1). Additionally, Illumina reads and the highest quality Oxford Nanopore Technologies reads have been deposited in the National Centre for Biotechnology Information (NCBI) Sequence Read Archive (SRA) or European Nucleotide Archive (ENA), with accessions listed in Table S1. All custom scripts and instructions to run analyses ('commands.sh') are available at https://github.com/HughCottingham/ONT_benchmarking_paper. The assembly pipeline used in this paper is available at https://github.com/HughCottingham/clinopore-nf. The authors confirm all supporting data, code and protocols have been provided within the article or through supplementary data files.

INTRODUCTION

Whole-genome sequencing (WGS) of human pathogens has become a crucial tool in research and public health settings [3]. Despite the potential advantages of sequencing platforms such as Illumina and Oxford Nanopore Technologies (ONT) for

Received 26 October 2025; Accepted 10 February 2026; Published 27 March 2026
Author affiliations: [1]Department of Infectious Diseases, The Alfred and School of Translational Medicine, Monash University, Melbourne, Victoria 3004, Australia; [2]Centre to Impact AMR, Monash University, Melbourne, Victoria, Australia; [3]Department of Microbiology, Infection Program, Monash Biomedicine Discovery Institute, Monash University, Clayton, VIC, Australia; [4]Department of Infection Biology, London School of Hygiene & Tropical Medicine, London WC1E 7HT, UK.
***Correspondence:** Nenad Macesic, Nenad.Macesic1@monash.edu
Keywords: antimicrobial resistance; benchmarking; genomic surveillance; Gram-negative bacteria; long-read sequencing; Oxford Nanopore.
Abbreviations: AF, allele frequency; AMR, antimicrobial resistance; cgMLST, core genome multi-locus sequence typing; cgSNP, core genome SNP; ECC, *Enterobacter* cloacae complex; GPU, graphics processing unit; HAC, high accuracy; KpSC, *Klebsiella pneumoniae* species complex; MLST, multi-locus sequence typing; ONT, Oxford Nanopore Technologies; SUP, super; VCF, variant call format; WGS, whole-genome sequencing.
All data used in this study are freely available via Figshare (https://doi.org/10.26180/c.7942409.v1). Additionally, Illumina reads and the highest quality ONT reads have been deposited in the SRA or ENA, with accessions listed in Table S1.
†These authors contributed equally to this work.
All supporting data, code and protocols have been provided within the article or through supplementary data files. Twenty supplementary figures and ten supplementary tables are available with the online version of this article.

identifying resistance mechanisms and monitoring suspected outbreaks, translation into clinical settings has been more difficult. Illumina has historically been more widely used in research contexts due to its higher accuracy rate (~99.9%) compared to ONT (~95–97%) [4, 5]. However, it has not been routinely implemented into clinical settings due to a combination of high instrument costs, prohibitive consumables costs for small batch sizes and analysis limitations due to its short (max 301 bp) read lengths.

ONT is affordable (USD $600–800 per flowcell) with consistent consumable costs regardless of batch size, offers real-time analysis and generates long (20 kbp+) reads that assist in various downstream analyses. The release of the 'super' basecalling model and R10.4.1 flowcells has also improved accuracy to levels approaching that of Illumina, achieving ~99% read accuracy and generating assemblies with less than one error per 100,000 bp [5, 6]. ONT platforms allow users to customize various aspects of data generation and analysis, including sequencing duration, basecalling algorithms, assembly methods and characterization software. While this flexibility is part of what makes ONT uniquely viable for clinical implementation, it is not yet clear how different combinations of sequencing depth, basecalling algorithms and assembly methods perform in genotyping and identifying resistance determinants in bacterial isolates. Previous benchmarking has found modern R10.4.1 ONT data to be comparable to Illumina in multi-locus sequence typing (MLST), core genome multi-locus sequence typing (cgMLST), antimicrobial resistance (AMR) and core genome SNP (cgSNP)-based typing of bacterial pathogens [7–16]. However, these studies did not adequately assess the performance of (i) using different basecalling software and models, (ii) using raw reads vs. assemblies as input and (iii) characterizing allelic variants of clinically important AMR genes.

In this study, we addressed this gap by investigating the effects of flowcell pore type, basecalling model, sequencing depth, assembly vs. raw reads and assembly polishing on MLST, cgMLST, AMR gene detection and cgSNP-based analyses. This involved sequencing 199 bacterial *Enterobacterales* isolates and comparing ONT-only results to hybrid assembly genomic automated gold standards. We also provide recommendations for the best combination of sequencing and analysis methods depending on the time and computational resources of the user. This work provides an important roadmap for the use of modern ONT sequencing as a frontline surveillance and diagnostic technology in clinical settings.

METHODS

Bacterial isolate collection and sequencing

We first chose 268 *Klebsiella pneumoniae* species complex (KpSC), *Enterobacter cloacae* complex (ECC) and *Escherichia coli* strains from our institutional biorepository with existing Illumina sequencing data and performed additional ONT sequencing. All DNA extractions were done using GenFind kits according to manufacturer guidelines [17]. ONT sequencing was performed on new DNA extracts since it was typically done several years after the original Illumina sequencing. Illumina sequencing was performed on the HiSeq 2000 and Novaseq 6000 platforms, while ONT sequencing was performed on R9.4.1 (legacy) and R10.4.1 flow cells using the ligation (LSK-109 for R9.4.1) and native (NBD-196 for R10.4.1) barcoding kits.

Basecalling and assembly

Raw ONT fast5 files were processed in several steps, including basecalling, read subsetting, assembly and characterization using commonly used tools (Fig. 1). All fast5 files were basecalled using ONT's legacy basecaller Guppy v6.2.1, and R10.4.1 sequences were also basecalled using the newer software Dorado v0.5.1 [18]. Separate fastq files were generated with all three basecalling models (SUP, HAC and Fast). R9.4.1 basecalling was performed on an NVIDIA T4 graphics processing unit (GPU) with 16 CPUs and 64GB of RAM. R10.4.1 basecalling was performed on an NVIDIA A100 with 60 CPUs and 40 GB of RAM or the NVIDIA V100 GPU present in ONT's gridION device.

To ensure we were using high-quality data, we first filtered out any isolates with sequencing depth <40× (n=48 R9.4.1 isolates, n=6 for R10.4.1). We then ran proch-n50 v1.5.8 [19] to capture read length and NanoStat v1.6.0 [20] to capture median read quality; both with default settings. Read length was high (median 22,469 bp, interquartile range (IQR) 17,219–28,067), and read quality was in line with typical scores relative to sequencing chemistry and basecalling algorithms (Fig. S1, Table S1, available in the online Supplementary Material) [5, 16]. As expected, R10.4.1 reads (median quality score of 17.2–20.1 using Dorado SUP) were of higher quality than R9.4.1 reads (median quality score of 11.2–15.8 using SUP) and quality increased from Fast to HAC to SUP models (Fig. S1).

We then used R9.4.1 Guppy and R10.4.1 Dorado SUP basecalled fastq files to generate hybrid automated gold standard assemblies using a custom-built nextflow pipeline [21] based on the assembly method recommendations in the Trycycler [22] and Polypolish [23] studies. The pipeline filters reads using Filtlong v0.2.1 [24] with the parameters --min_length 1000 and --keep_percent 95 and then assembles using Flye v2.9 [25] with the --nano-hq tag. It then performs long-read polishing using Medaka v1.7 [26] with a user-specified model (-m) depending on the input dataset and short-read polishing using Polypolish v0.5.0 [23] and Polca v3.4.2 [27] with default parameters (if short-read data are available). We used contig number and MLST calls as quality control metrics for filtering out poor-quality automated gold standard assemblies. Any isolates with more than 20 contigs in the automated gold standard assembly (n=0 for R9.4.1, n=2 for R10.4.1) and/or missing/incomplete sequences in any MLST loci for

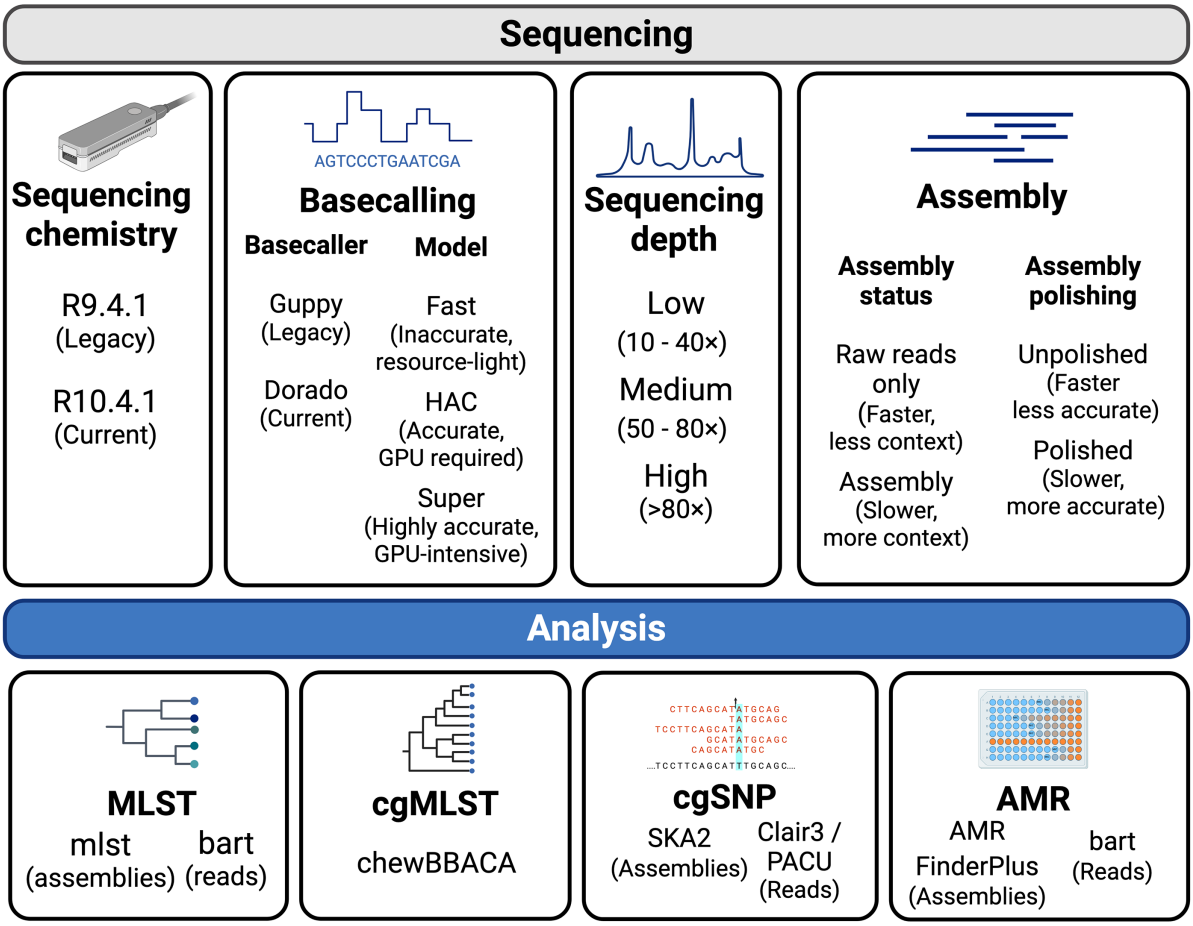


Fig. 1. Overview of analysis methods used to generate an ONT-only dataset to benchmark MLST, cgMLST, cgSNP and AMR typing performance.

automated gold standard assemblies (n=2 for R9.4.1, n=7 for R10.4.1) were excluded. We also excluded three R9.4.1 isolates due to obvious isolate mismatches between ONT and Illumina data, resulting from mislabelling or errors in the sequencing process. The final benchmarking dataset included 199 isolates (Tables 1 and S1).

To generate subset assemblies, we randomized the order of reads in full fastq files and then took increasing proportions of the full readset to reach depth levels from 10× to 100× using the subseq command of Seqtk v1.4 [28]. The order of reads was consistent

Table 1. Summary statistics for isolates included in benchmarking analyses

Species group	Total isolate		Unique ST		Prominent ST (n=no. isolates)	
	R9.4.1 (legacy)	R10.4.1	R9.4.1 (legacy)	R10.4.1	R9.4.1	R10.4.1
KpSC	60	25	39	9	ST37 (n=4) ST323 (n=5)	ST29 (n=10) ST323 (n=9)
E. coli	27	33	16	10	ST10 (n=3) ST38 (n=4) ST617 (n=3)	ST95 (n=14) ST131 (n=12)
ECC	27	27	9	4	ST93 (n=7) ST114 (n=7) ST190 (n=9)	ST93 (n=13) ST190 (n=13)

ECC, *Enterobacter cloacae* complex; KpSC, *Klebsiella pneumoniae* species complex; ST, Sequence type.

between depth levels to ensure that each increasing depth level contained the same reads as the last, plus an additional set to reach the next depth level. These subsetted readsets were then fed into the Nextflow pipeline [21] for assembly. Reads generated from all three basecalling models (Fast, HAC and SUP) were processed in this way. Since Medaka does not have a model designed for Dorado Fast basecalling, we used the HAC model to perform long-read polishing of Fast assemblies.

Typing analyses

For automated gold standard and subset assembly MLST calls, we used mlst v2.23.0 [29] with the relevant -s flag dependent on the organism being analysed. We used bart v0.1.2 [30] with the -r ont flag and appropriate -s species flag for read-based MLST.

For cgMLST, we used chewBBACA v3.3.10 [31] using the *E. coli* and KpSC databases from cgmlst.org [32]. This involved downloading the species database fasta file; running the PrepExternalScheme command with default settings, the AlleleCall command on automated gold standard assemblies while allowing inferred alleles to be added to the database (default settings); and then running AlleleCall on the subset assemblies without adding inferred alleles to the database (--no-inferred) to ensure that each group of subset assemblies was run on the same database. To assign strains to clusters based on cgMLST typing, we used cgMLST-dists v0.4.0 [33] with default settings to generate distance matrices and then assigned strains to clusters based on average distance thresholds using hclust [34] in R. To visualize these clusters, we used igraph [35] and ggplot2 [36] in R.

For AMR calling from hybrid assemblies, we ran AMRFinderPlus v3.12.8 [37] using default settings and the NCBI curated database. For subset assemblies, we used an identity threshold of 70% to avoid false negatives in low-depth assemblies that were expected to be of poor quality. When comparing subset performance to the automated gold standard for determining AMR allelic variants, we only used strains for which an allelic variant was called using AMRFinderPlus on the automated gold standard assembly. We used bart v0.1.2 [30] with the -r ont flag and additional -amr flag for read-based AMR calling.

To compare the MLST, cgMLST and AMR profiles of ONT-only data inputs to genomic automated gold standards of the same isolate, we removed confidence symbols from the tool outputs and summarized the number of correct alleles using string-matching in R. For read-based analyses, bart often reports several possible allelic combinations in order of decreasing confidence. For MLST, we took the top match. For AMR, we took the top *n* hits per gene by alignment depth, where *n* represents the number of allelic variants of that gene present in the automated gold standard assembly. In most cases, *n*=1; however, some automated gold standards contained multiple allelic variants (e.g. $bla_{\text{CTX-M-14}}$ and $bla_{\text{CTX-M-15}}$). For these isolates, we took the top two hits by depth from the bart output. For cgMLST analyses, we removed loci with no allele call in five or more automated gold standard assemblies (*n*=176 for KpSC, *n*=93 for *E. coli*) as well as isolates with ten or more loci with no allele call (*n*=2 for KpSC and *E. coli*).

To calculate AMR gene retention rate in subset assemblies, we first extracted all AMR genes identified by AMRFinderPlus in the automated gold standard assemblies from the AMRFinderPlus database using Seqtk v1.4 [28]. We then aligned all subset assemblies to those genes using Minimap2 v2.26 [38] using the --secondary=no flag and reported whether they generated an alignment or not. All benchmarking results were parsed and plotted in R using the dplyr [39], reshape2 [40] and ggplot2 [36] packages.

Data preparation for variant calling

Determining the ground truth in highly sensitive analyses such as variant calling can be difficult due to imperfections in any assembly method [22]. To test variant callers with an established ground truth, we began by benchmarking performance with simulated reads. This included selecting a random *E. cloacae*, *E. coli* and *K. pneumoniae* hybrid assembly and introducing 1,000 random SNP sites into its chromosome using a custom Python script ('generate_mutants.py' in https://github.com/HughCottingham/ONT_benchmarking_paper). We generated ten mutant assemblies per reference, where each mutant assembly had SNPs in 0–1,000 of these sites.

We then simulated ONT and Illumina raw reads for each of the mutant genomes. For ONT, we simulated reads at 40×, 70× and 100× depth using NanoSim v2.0.0 [41]. This involved training the simulation model using the read_analysis.py genome command with default parameters and the hybrid reference genome with its corresponding R10.4.1 Dorado SUP reads as inputs. Next, we ran NanoSim's simulator.py genome command with the following parameters: -rf mutant_genome.fasta, -c training_file -n genome_size/100 (or 70 or 40 depending on fold coverage) --fastq -med 10000 -sd 0.3. We generated 100× Illumina HiSeq reads using ART v2016.06.05 [42] with the following parameters: -ss HS20 -i mutant_genome.fasta -p -m 300 l 100 s 50 f 100. We assembled all simulated ONT readsets as described earlier using our Nextflow pipeline [21].

To ensure that we had isolates from a range of STs, we sequenced seven additional isolates to complement the existing benchmarking dataset for assessing variant calling with real datasets (Table S1).

Variant calling

Once we had both Illumina and ONT R10.4.1 Dorado SUP reads (simulated and real), we used the chromosome of one automated gold standard genome per ST as a reference and called SNPs using four different variant callers (Table S2). We used Reddog [43]

to call variants from Illumina reads, which has been extensively used across several previous studies [44, 45]. Reddog aligns reads to a reference genome using Bowtie 2 v2.4.1 [46], calls SNPs using Samtools v1.9 [47] and reports variants with quality score>=30. For ONT assemblies, we chose to assess performance using the split-kmer tool SKA2 due to its simplicity for future clinical utility, combined with high reported accuracy with Illumina assemblies [48]. We used the ska build and ska map commands from SKA2 v0.3.7 [48] with default settings as described in their documentation. For ONT reads, we used PACU v0.0.6 [9], which was the only pipeline to our knowledge specifically built for cgSNP typing from R10.4.1 ONT reads. PACU calls SNPs in the reference genome using BCFtools v1.17 and filters for quality score and low depth regions in a similar fashion to Reddog and other established cgSNP tools. Additionally, we also assessed the performance of the Clair3 v1.0.10 [49] variant caller when combined with our own manual filtering, as this was recently found to be the most accurate variant caller for R10.4.1 reads [7]. We initially set allele frequency (AF) to 0.9 and the quality threshold to 30 for PACU and Clair3 to mirror Reddog. However, we found these thresholds to be too stringent due to the differences in the sequencing platform and variant calling process. Instead, we relaxed the quality and AF thresholds to better suit each tool and improve concordance (Table S2). In addition to these adjustments, we ran Clair3 using the r1041_e82_400bps_sup_v430 model found on Rerio [50] and included the parameters --haploid_precise to filter out heterozygous calls and --no_phasing_for_fa to skip phasing, which is not relevant in haploid organisms. For benchmarking variant calling performance, we chose to focus on SNPs due to their clinical utility in calculating pairwise distance matrices rather than other metrics such as indels and large structural variants.

Each of these tools outputs a variant call format (VCF) file showing variant sites in the reference genome. Using a custom Python script ('alleles2mfasta.py'), we generated whole-genome alignment files where each sequence was a copy of the reference genome with SNPs introduced according to those identified in each of the strains. We then fed this whole-genome alignment file into Gubbins v3.4 [51] with default parameters. Finally, we converted the VCF output of Gubbins into final allele matrices using another custom Python script ('vcf2csv.py'). The commands and scripts used for these processes can be found at https://github.com/HughCottingham/ONT_benchmarking_paper.

To compare pairwise distances between methods, we generated fasta alignment files from the allele matrices using a custom Python script ('csv2fasta.py') and fed those files into snp-dists v0.8.2 [52] with default settings. We then plotted these distances in R using dplyr [39] and ggplot2 [36]. We summarized concordance of SNP sites using upsetR [53] with the allele matrices as input and summarized identity concordance using dplyr and ggplot2. To determine whether SNP sites were located in repeat regions of the reference genome, we followed the approach outlined by Hall *et al.* [7], which built on using mummer to align the genome to itself and output alignments with identity 60% or greater. To determine concordance in identifying putative transmission pairs, we used a 25 SNP threshold [54] for Gram-negative bacteria. We did not have any *E. coli* strains within these thresholds, so they were not part of this analysis. To generate phylogenies, we ran RAxML-NG v1.2.1 [55] on the core genome alignment fasta files with the following parameters: --all --model GTR+G+ASC_LEWIS --bs-trees 1000. To calculate the similarity of tree topology between variant callers, we used the normalized Robinson–Foulds distance, which describes homology from 0 (identical) to 1 (no similarity) by calculating the number of splits in topology/total splits across both trees. We used ETE3 v3.1.3 [56] to perform midpoint rooting and distance calculations.

RESULTS

Modern ONT data generated perfect MLST calls across tested species

Basecalling model was the most important factor for MLST accuracy: high accuracy (HAC) and super (SUP) assemblies generated a correct call in a mean 96.1% and 98.8% of isolates, respectively, across all species, basecallers and chemistries, compared to the Fast model achieving this in only 41.7% of isolates (Fig. S2, Table S3). R10.4.1 sequences basecalled with Dorado generated perfect MLST calls in every isolate using both HAC and SUP models. We compared these results to legacy R9.4.1 data, which had four incorrect calls in KpSC with SUP and errors in all three species groups with HAC.

We then assessed the performance of modern ONT data (R10.4.1, Dorado, HAC or SUP models) when unpolished assemblies, polished assemblies and raw reads were used (Fig. 2a, Table S4). We noted that assembly polishing was not necessary, with both unpolished and Medaka-polished assemblies generating 100% MLST calls across all R10.4.1 isolates using both HAC and SUP. Read-based MLST performed almost as well as assemblies, with only one incorrect locus in an *E. coli* isolate (icd locus in AH20I049) across the entire dataset. This appeared to be an error generated by the read-based tool bart [30], which aligns reads to the MLST database using KMA [57] and parses its output. Running KMA by itself against the candidate alleles at this locus generated a higher alignment score for the correct allele, suggesting that the error was made by Bart rather than the ONT basecaller Dorado.

To assess the effects of sequencing depth and basecalling model, we generated MLST calls using subsets of each readset up to 100× depth (see the 'Methods' section). Raw reads generated comparable or better calls compared with assembly methods for ECC and KpSC, generating 100% MLST calls in ECC and KpSC at all depths greater than 10× using both HAC and SUP (Fig. 2b). For *E. coli*, SUP assemblies outperformed raw reads, generating perfect calls from 30× and above. KpSC and *E. coli* assembly errors at submaximal depths were predominantly due to adenine insertion errors in small homopolymer regions of MLST allele-encoding regions. In KpSC, this occurred in both HAC and SUP assemblies, while in *E. coli*, it only occurred in HAC assemblies.

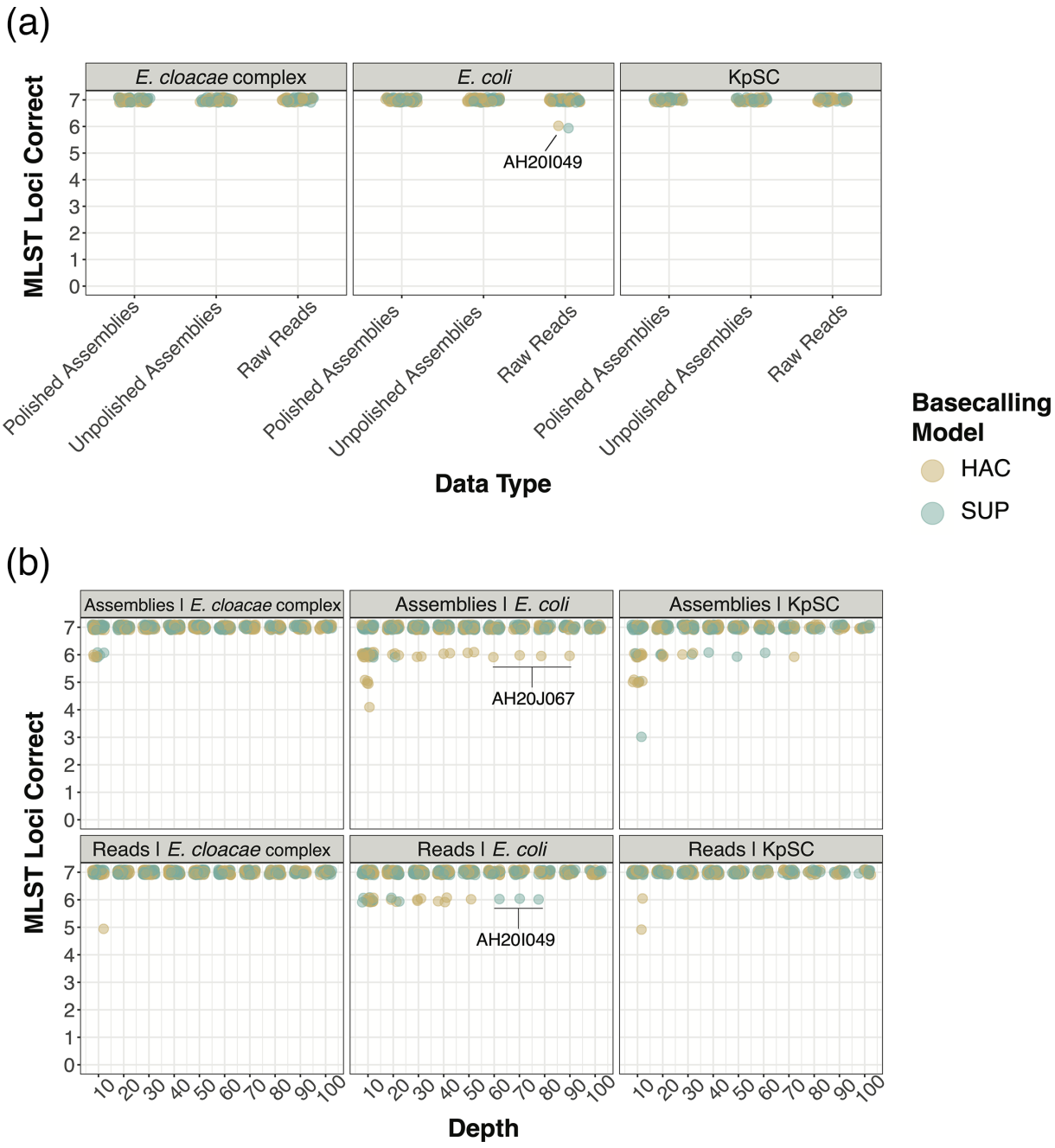


Fig. 2. Accuracy of MLST genotyping using R10.4.1 Dorado (a) raw reads, unpolished assemblies and polished assemblies at maximum available depth (100× or below) for each isolate and (b) raw reads and polished assemblies at depths from 10× to 100×. Labelled *E. coli* isolates (AH20J067 and AH20I049) refer to strains that had incorrect calls at multiple depths. KpSC strains are not labelled as errors that came from a variety of strains.

Modern ONT data reliably classified cgMLST loci in *E. coli* and KpSC isolates

While MLST is a traditional strain typing method, it lacks resolution when dealing with highly diverse STs or with species that have a poorly defined MLST scheme. In these cases, cgMLST can be used to differentiate strains based on allelic profiles across thousands of genetic loci, as opposed to a limited number used in MLST [58]. We generated cgMLST calls in *E. coli* and KpSC subset assemblies and compared them to the automated gold standard (see the 'Methods' section). *E. cloacae* does not have an established cgMLST scheme, so it was not included in this analysis.

Using our optimal approach (R10.4.1 chemistry, maximum available depth for each strain, Dorado SUP basecalling and Medaka polishing), both species had a median of 11 incorrect loci (0.5% of total loci) per isolate (Table S5). For *E. coli*, most errors came

in one strain (ECamp083) that had incorrect allelic calls in 118 loci (Fig. S3A) or 6 loci (b0197, b0241, b1704, b2172, b4354 and b1819) where most strains did not generate any allele call (Fig. S3B). We found exact matches between ONT-only and hybrid assemblies in these six genes following alignment with Minimap2 [38], suggesting that errors likely arose during CDS prediction by the cgMLST tool (chewBBACA [31]). Sequencing depth (47×) for ECamp083 was amongst the lowest of all isolates and likely contributed to the higher error rate. For KpSC, the overall random error rate was higher, although there was no particularly error-prone strain comparable to ECamp083 in *E. coli* (Fig. S3A). Allele calls were not generated for two loci (KP1_RS00725 and KP1_RS25355), similar to the issues detailed for six loci in *E. coli* (Fig. S3B).

When we evaluated the impact of basecalling model, we found that using SUP basecalling provided a higher increase in performance over HAC compared to MLST typing (median 11 incorrect cgMLST loci per isolate for both species using SUP and median 22 and 29 incorrect for *E. coli* and KpSC, respectively, using HAC) (Table S5). We also found that accuracy improved with increasing depth before reaching a plateau above 50×, similar to the results of MLST (Fig. S4).

A common downstream analysis after calling cgMLST alleles is to group strains into genetic clusters based on the number of genes by which they differ. For *E. coli* and KpSC, clusters are often defined as strains that have ten or fewer gene differences across all cgMLST loci [59]. In our dataset, there were no *E. coli* strains that shared a cluster, so we focused on KpSC.

When applying this clustering threshold of ten allele differences to the optimal ONT-only KpSC assemblies (R10.4.1, Dorado, SUP, Medaka-polished, max depth), we found that most clusters were not effectively resolved when compared to the automated gold standard assemblies (Fig. 3). As shown in Fig. S3A, several ONT-only assemblies differed from their automated gold standard assembly by up to 40 loci, indicating that a clustering threshold of ten alleles may be too sensitive. We then evaluated a clustering threshold of 50 alleles. We found that cgMLST clusters 1, 3 and 4 contained the same isolates as the automated gold standard. Meanwhile, cluster 2 still contained both false positive and false negative errors (Fig. 3).

When comparing modern ONT data to legacy datasets in cgMLST typing accuracy, R10.4.1 data (median 98.65% of loci correct) outperformed R9.4.1 (median of 93.9% of loci correct) using HAC and SUP models (Table S6). Fast basecalling again performed poorly across R9.4.1 and R10.4.1 datasets (median 53.3% of loci correct).

Modern ONT data led to 100% accurate typing of allelic variants in acquired AMR genes

To assess AMR allelic variant calling in our benchmarking dataset, we analysed all isolates for which the automated gold standard assembly contained the clinically important $bla_{\text{CTX-M}}$,bla_{IMP},bla_{OXA} or bla_{TEM} acquired beta-lactamase genes. We ran assembly-based and read-based AMR detection tools on ONT-only datasets and compared the allele calls to those of the automated gold standard genome (see the 'Methods' section). R10.4.1 SUP datasets generated 100% allelic matches across all four genes using both raw reads and assemblies (Table 2, Table S7). Raw HAC reads outperformed HAC assemblies, generating 100% allelic matches compared to 95.5% in unpolished assemblies and 96.8% in polished assemblies. Raw reads also excelled at lower depths, generating 100% allelic matches from 30× and above across all four genes (Fig. S5). SUP assemblies reached 100% accuracy at 50× and above, while HAC showed one to two errors in at least one gene at all depths except 100×.

When comparing modern ONT data to legacy datasets, we found that R10.4.1 HAC assemblies showed similar accuracy (96.8%) to R9.4.1 SUP assemblies (95.5%) (Fig. S6, Table S8). The Fast basecalling model performed worse than HAC and SUP for all datasets, generating a correct allele call in an average of just 62.6% of cases across all genes and chemistries compared to 96.3% and 97.7% for HAC and SUP, respectively (Fig. S6, Table S8). R9.4.1 Fast assemblies (84.4% allelic match) outperformed R10.4.1 Fast assemblies (40.8%), although this was likely due to there being no Medaka polishing model for Dorado Fast data types.

While the antimicrobial activity of these beta-lactamase genes is often dependent on allelic variation, for many other AMR genes, resistance phenotype can be predicted from gene presence alone. We evaluated AMR gene detection in R10.4.1 assemblies at depths up to 100× (see the 'Methods' section). We noted that structural errors in ONT-only assemblies led to the omission of AMR genes, particularly at lower depths (Fig. S7). Omission rates across all genes with low-depth assemblies (40× and below) were 9.2%, 6.2% and 7.4% for Fast, HAC and SUP, respectively. Medium-high depth assemblies (50× and above) still had missing AMR genes in some cases, although at a lower rate of 1.2% across all genes and models. These findings also appeared to be gene-specific, with *tetA*, bla_{OXA} and $bla_{\text{CTX-M}}$ showing omission rates of 7.1%, 1.5% and 3.1%, respectively, even at 50× and above. Location of AMR genes also impacted omission rates: plasmid-located genes had a mean 1.5% omission rate across all genes vs. 0.07% for chromosomally located genes at depths of 50× and above.

ONT read-based variant callers outperformed Illumina reads and ONT assemblies using simulated data

One of the limitations of legacy ONT sequencing was concerns surrounding read accuracy, making its use in analyses such as variant calling impractical. Read accuracy of modern ONT data has now risen to ~99% [5], but further data are needed to confirm that this leads to more accurate variant calling. To gain an understanding of how variant callers performed under highly controlled conditions with modern ONT data, we benchmarked four variant calling tools on simulated R10.4.1 Dorado SUP

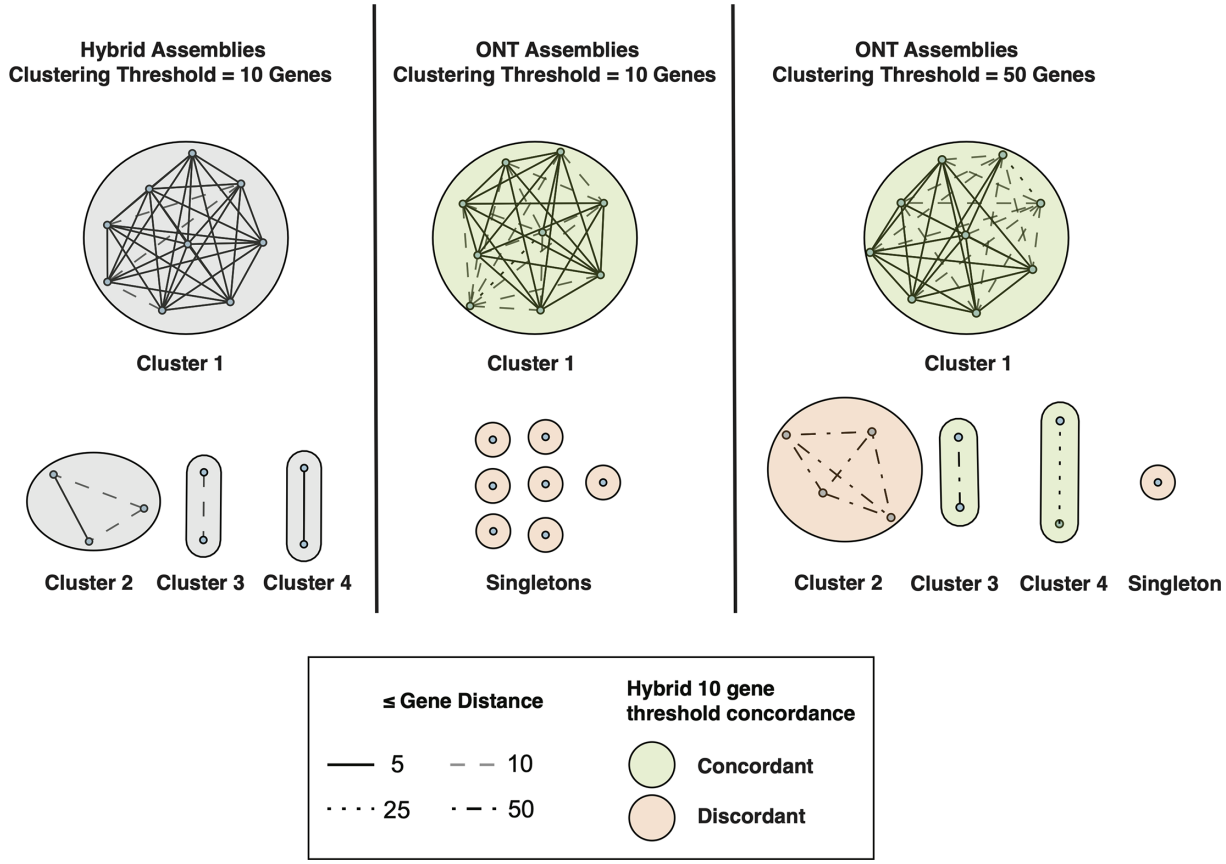


Fig. 3. Hierarchical clustering of KpSC hybrid and ONT-only assemblies based on varying cgMLST distance thresholds. Each node represents an isolate. Clusters with two or more strains are numbered. Gene distances within clusters are depicted as varied line types depending on the distance value. Singletons are isolates that clustered using the hybrid assembly ten gene threshold but did not cluster when using ONT assemblies.

readsets generated from *E. cloacae*, *E. coli* and *K. pneumoniae* genomes with known SNP sites (see the 'Methods' section). ONT read-mapping tools (Clair3, PACU) outperformed ONT assembly (SKA2) and Illumina read-mapping (Reddog) at 40×, 70× and 100×, generating perfect SNP profiles at 100× compared to 9–22 and 14–28 errors per 1,000 SNPs for Reddog and SKA2, respectively (Fig. 4). Reddog and SKA2 had an elevated false negative rate (Fig. 4a) caused by an inability to resolve repeat regions longer than the read or max kmer length for Reddog and SKA2, respectively (see the 'Methods' section for description of defining repetitive regions). Clair3 and PACU utilized ONT's long reads to accurately call SNPs in these regions and at 100× depth made no errors. Nearly all false positive errors (Fig. 4b) came from SKA2. These sites were heterozygous for the alternative and reference allele, which can be identified using mapping approaches but not from standard assembly methods. Incorrect SNP allele calls (Fig. 4c) were mostly generated by Clair3 and PACU at lower sequencing depths. This is indicative of ONT's slightly higher raw read error rate compared to Illumina reads (Reddog) and ONT consensus accuracy (SKA2).

Reliable SNP profiling with ONT read-based variant callers was validated with real sequencing data

We then validated our synthetic findings using real-world R10.4.1 Dorado SUP readsets, as their estimated ~99% accuracy [5] offers the best chance for accurate variant calling. We compared pairwise SNP distributions for different variant calling tools across six *E. cloacae*, *E. coli* and *K. pneumoniae* species/ST combinations (Figs. S8–S10; see the 'Methods' section). While all three species showed high concordance between tools in most cases, *E. cloacae* STs showed near identical SNP distributions across all four tools (Fig. S8). For *E. coli* ST131 (Fig. S9A) and *K. pneumoniae* ST323 (Fig. S10B), mapping-based approaches (Reddog, Clair3, PACU) showed high concordance, but SKA2 suffered from elevated SNP distances.

We then assessed sites in the reference genome that were identified as variants using different variant calling tools, while also assessing whether all strains were assigned the same allele (reference or variant) at these shared sites (Figs S11–S13). We noted that while distance distributions were highly concordant across tools, the specific SNP sites and allele calls that led to those distances were not consistent. *E. cloacae* was the most concordant species across all variant calling tools (Fig. S11). In ST93, 84.7%

Table 2. Accuracy of AMR allelic variants using dorado R10.4.1 raw reads, unpolished assemblies and polished assemblies

Maximal depth (100× and below) was used for each isolate. No allele call refers to instances where AMRFinderPlus detected the gene but could not provide a confident allele call

Model	Gene	Data type	Allele match	Incorrect allele	No allele call	Not in assembly
HAC	$bla_{\text{CTX-M}}$	Raw reads	48 (100%)	0	0	0
		Unpolished assembly	46 (95.8%)	0	2 (4.2%)	0
		Polished assembly	47 (97.9%)	1 (2.1%)	0	0
	bla_{IMP}	Raw reads	28 (100%)	0	0	0
		Unpolished assembly	27 (96.4%)	0	1 (3.6%)	0
		Polished assembly	27 (96.4%)	0	1 (3.6%)	0
	bla_{OXA}	Raw reads	44 (100%)	0	0	0
		Unpolished assembly	41 (93.2%)	0	2 (4.5%)	1 (2.3%)
		Polished assembly	42 (95.4%)	0	1 (2.3%)	1 (2.3%)
	bla_{TEM}	Raw reads	57 (100%)	0	0	0
		Unpolished assembly	55 (96.5%)	0	2 (3.5%)	0
		Polished assembly	56 (98.2%)	0	1 (1.8%)	0
SUP	$bla_{\text{CTX-M}}$	Raw reads	48 (100%)	0	0	0
		Unpolished assembly	48 (100%)	0	0	0
		Polished assembly	48 (100%)	0	0	0
	bla_{IMP}	Raw reads	28 (100%)	0	0	0
		Unpolished assembly	28 (100%)	0	0	0
		Polished assembly	28 (100%)	0	0	0
	bla_{OXA}	Raw reads	44 (100%)	0	0	0
		Unpolished assembly	44 (100%)	0	0	0
		Polished assembly	44 (100%)	0	0	0
	bla_{TEM}	Raw reads	57 (100%)	0	0	0
		Unpolished assembly	57 (100%)	0	0	0
		Polished assembly	57 (100%)	0	0	0

(61/72) of total SNP sites in the reference genome were called in all four tools (Position Concordance in Fig. S11A). In addition, for each site that was shared between two tools, the same allelic profile was called across all strains in nearly every case (Identity Concordance in Fig. S11A). All four variant callers were concordant in 68.8% (1979/2878) and 55.6% (15/27) of total sites for *E. coli* ST95 (Fig. S12A) and *K. pneumoniae* ST29 (Fig. S13A), respectively. *K. pneumoniae* ST323 (Fig. S13B) showed the most discordant results due to three separate factors:

(1) False positive SNP sites with SKA2 not called with any other tool (n=252 sites). This was due to a combination of random assembly errors (153/252 false positive sites) and an inability to identify heterozygous sites (99/252 false positive sites).
(2) Lowered SNP sites using Clair3 that were called using all other methods (n=20). Despite passing quality, depth and allele frequency thresholds, these sites were defined as heterozygous using Clair3 while being passed as homozygous using other variant callers. Clair3 did not provide any additional information regarding these heterozygous calls, but it was likely the result of a combination of read quality, indel misalignments and/or other extraneous factors. Clair3 offers the ability to pass heterozygous variants, but this resulted in far more false positives than false negatives and was therefore not retained in our methods. The ONT data itself was unlikely to be the cause, as PACU showed high concordance with other methods while using the same readsets.
(3) Lowered SNP sites identified using Reddog and SKA2 identified using PACU and Clair3 (n=7 sites). This was due to poor sensitivity in repetitive regions of the genome using kmer and short-read methods described in the simulated read section.

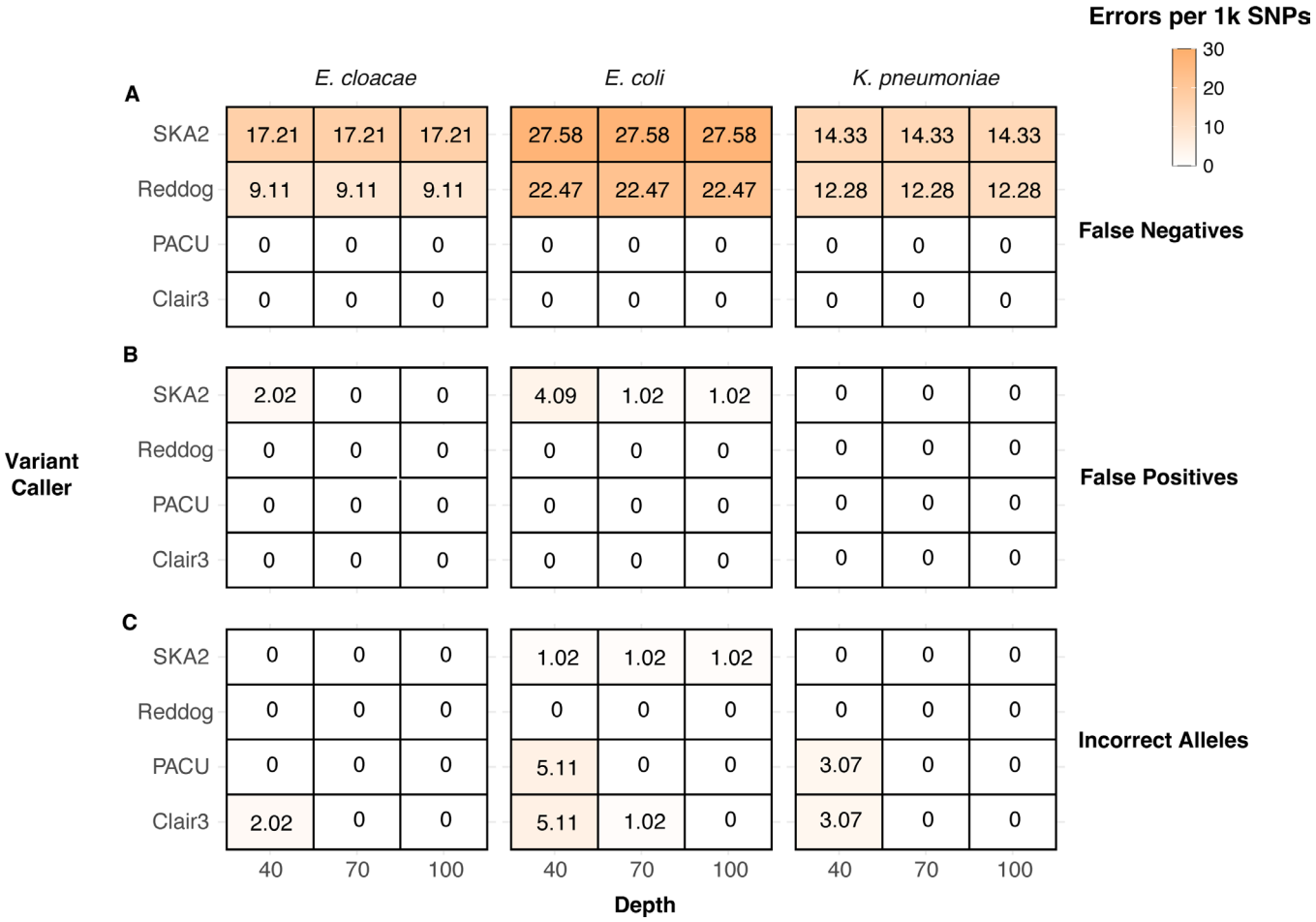


Fig. 4. Errors per 1,000 SNPs of variant calling tools using simulated Illumina and ONT data. False negatives (a) refer to instances where a SNP was missed, while false positives (b) refer to instances where a SNP was called when none was present in the mutant genome. Incorrect alleles (c) refer to instances where a tool correctly identified the presence of a SNP at a given position in the reference genome, but the identity of the SNP was incorrect.

Meanwhile, *E. coli* ST131 (Fig. S12B) showed elevated SNP sites using Clair3 (n=397 sites) and SKA2 (n=405 sites). These sites were called using all four tools but were only filtered out by Gubbins for Reddog and PACU. It is unclear why these sites made it through Gubbins for SKA2 and Clair3. However, pre-Gubbins variant calling generated over 40,000 SNPs using each tool, so it should be noted that the ~400 SNPs that were not consistently filtered were only ~1% of all SNPs.

ONT mapping methods reliably identified putative transmission pairs in hospital pathogens

We then applied these results to identifying putative transmission pairs. Historically, if two Gram-negative isolates have a SNP distance of 25 or less, they are determined as putative transmission pairs [54]. Of 158 pairs that met these thresholds using Reddog, ONT mapping approaches met the threshold in 155 (98.1%) and 158 (100%) of pairs for Clair3 and PACU, respectively (see the 'Methods' section). The discordant Clair3 pairs narrowly missed the cutoff with distances of 26, 28 and 29 (Fig. 5). Meanwhile, due to the elevated SNP sites identified using SKA2, 5/6 *K. pneumoniae* ST323 pairs did not meet the 25 SNP threshold when using SKA2.

We then assessed the impact of ONT sequencing on the construction of SNP-based phylogenies. We compared ONT and Illumina by generating maximum-likelihood phylogenies for each variant caller and compared their topologies using normalized Robinson–Foulds distance and tree visualization (see the 'Methods' section). We noted high concordance in more diverse STs/subtrees but lower concordance in highly related STs or genetic clusters (Table S9, Figs S14–S19). For example, *E. cloacae* ST190 showed lower topological concordance, particularly within clusters of highly related strains (Fig. S15), while the highly divergent *E. coli* ST95 had identical topology using Reddog, Clair3 and PACU (Fig. S16). In highly related STs, small differences in SNP sites or variant profiles between methods are enough to drastically change the topology of the tree.

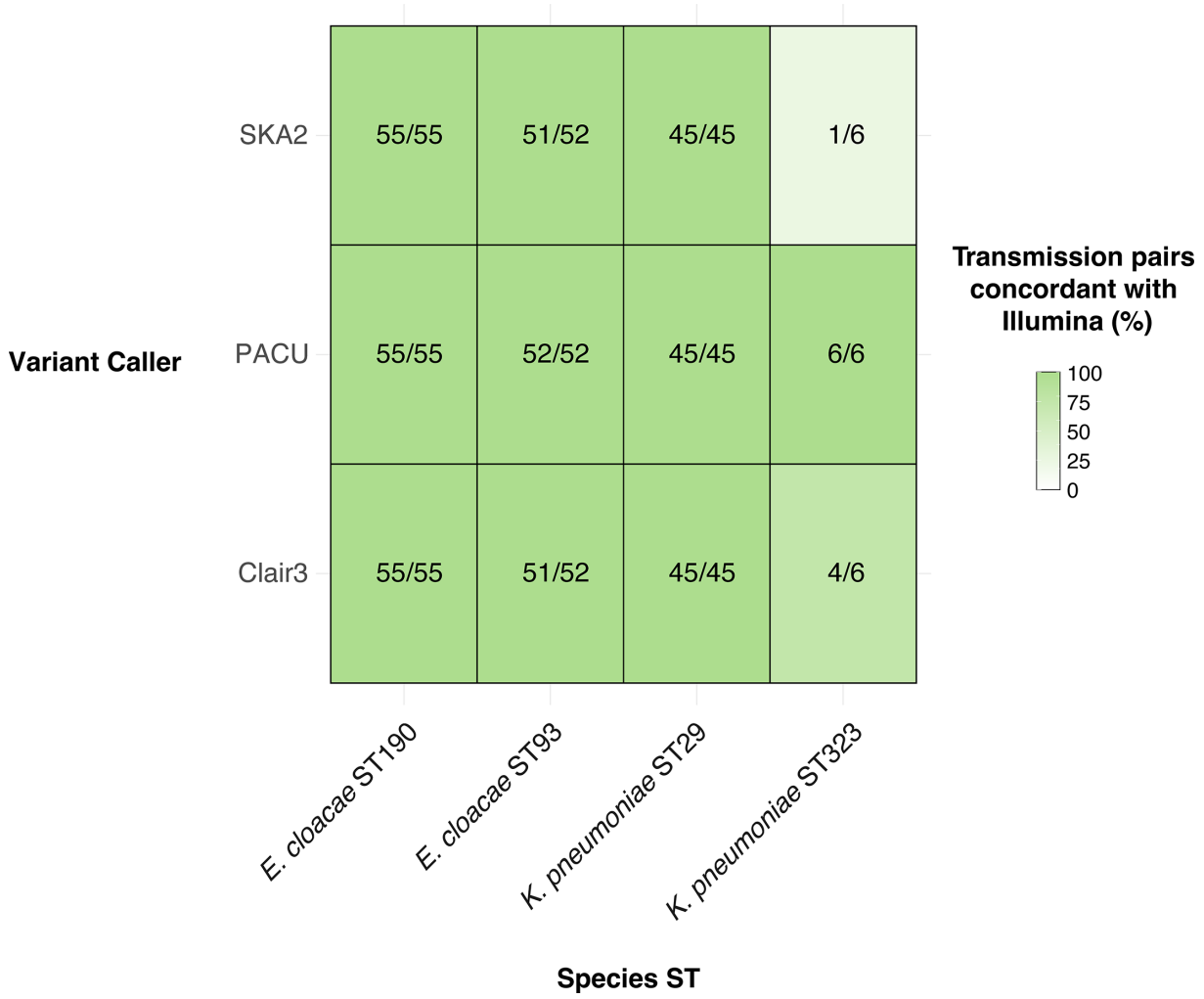


Fig. 5. Concordance between ONT methods and traditional Illumina methods in identifying putative transmission pairs. Included are all isolate pairs that would be assigned as putative transmission pairs using a traditional Illumina approach (distance≤25 SNPs). Pairs are classified as concordant with Illumina if the pairwise distance using ONT methods falls under the defined threshold for each species. No *E. coli* are shown here as no isolates in our dataset were putative transmission pairs.

Recommendations for bacterial isolate sequencing with ONT

Due to the high error rate of legacy ONT reads, assembly has typically been a default step prior to typing analyses such as AMR, MLST and cgMLST [60]. However, our work shows that modern ONT reads perform comparably to assemblies for MLST and AMR typing, with superior AMR typing performance at depths of 40× and below. Additionally, read-based tools performed better than assemblies at SNP calling due to the increased amount of quality information at each SNP site and improved performance in repetitive regions. Based on these findings, we recommend that both read and assembly-based tools are viable for AMR and MLST. At low sequencing depth, we recommend read-based analyses for AMR typing, while assembly is better suited to investigating genome structure. To our knowledge, there is no open-source ONT read-based cgMLST tool available, so we currently recommend assembly for cgMLST. We recommend that read-based tools such as PACU [9] or Clair3 [49] are used for variant calling over assembly-based methods.

For clinical implementation of ONT WGS, time requirements should not play a major role in deciding on assembly and analysis methods, regardless of the urgency of a result. The major factors that should be considered if time is a relevant factor (for example, in the case of a hospital outbreak or severe infection) are sequencing depth and, if live basecalling is not available, the basecalling model. These factors are the main determinants of result turnaround time; the decision of whether to assemble or not likely changes this time by no more than 30 min.

Taking these factors into consideration, we have provided five recommended combinations of sequencing depth and basecalling model. We illustrate different time and resource constraints, paired with expected turnaround times and accuracy. A more detailed breakdown of accuracy per gene and species can be found in Fig. S20.

If a result is non-urgent and accuracy is most important, or variant calling is required, the best results ('Optimal' in) are obtained using at least 100× depth and SUP basecalling. With live basecalling, this option takes ~33 h when including DNA extraction, library preparation, sequencing and analysis. For faster results with high accuracy, 50× depth and SUP basecalling ('Balanced') achieves 98.8–100% accuracy across all analyses with ~17 h required. If live basecalling is available and a rapid outcome is required, 20× depth with SUP basecalling ('Fast') is likely sufficient for AMR, MLST and/or cgMLST. At 20× depth, we recommend raw reads for AMR typing rather than assembly due to better benchmarking results. We are hesitant to recommend variant calling at low depths based on the known higher error rates of ONT raw reads, but we did not test variant calling at these read depths. Some clinical laboratories do not have access to live basecalling, instead using a remote GPU and basecalling after sequencing. In this scenario (and assuming time to result is relevant), we recommend sequencing to 100× depth and using HAC basecalling ('Remote GPU'). This approach provides highly accurate typing results in ~33 h, with HAC basecalling requiring one-fifth of SUP basecalling time based on our benchmarking (Table S10). Finally, if there is no GPU access ('No GPU'), we do not recommend attempting typing analyses due to the poor accuracy of Fast basecalling (the only viable model to use without a GPU).

DISCUSSION

In this study, we demonstrate that modern R10.4.1 ONT data are now a viable choice for the characterization of bacterial pathogens in clinical settings. We noted very high concordance with genomic automated gold standards and comparable or better performance than Illumina. R10.4.1 sequences basecalled with the SUP model enabled perfect MLST and AMR calls across 85 *Enterobacterales* isolates. SUP datasets had a median cgMLST locus distance of 11 genes to the automated gold standard genomes of KpSC and *E. coli* isolates; an improvement over legacy ONT datasets but not yet enough for reliable clustering. ONT read-mapping variant callers outperformed traditional Illumina methods and kmer-based methods using simulated read data due to improved performance in repetitive regions of the genome. ONT mapping approaches showed high concordance with Illumina approaches using real sequence data, generating similar distance distributions, SNP sites and identification of putative transmission pairs. Phylogenies constructed with ONT mapping-based methods resembled those generated with Illumina for diverse STs, but often showed several topological changes in STs with smaller genetic distances due to the increased impacts of any singular discordant SNP call. Our findings have allowed us to provide specific recommendations for clinical ONT sequencing regarding depth, basecalling model and analysis methods depending on the goals and resources available to the user.

To our knowledge, a single prior study [8] has tested MLST accuracy of modern ONT data. This study also noted perfect concordance between Illumina and ONT for all assemblies containing a full MLST profile. Five per cent of ONT assemblies in this study were missing one or more MLST loci, similar to the 3.7% (10/268) isolates removed from our benchmarking dataset due to unresolved MLST profiles in automated gold standards. In this situation, read-based MLST profiling or an alternative assembly algorithm (e.g. Hybracter [61] or Autocycler [62]) should be used. We found the cgMLST accuracy of our ONT-only datasets (median of 11 cgMLST loci incorrect per isolate) to be comparable to other studies that followed a similar approach of Flye assembly followed by Medaka polishing [10, 12–14]. Some studies have sought to further optimize this accuracy through using methods such as duplex ONT sequencing [8], specific cgMLST optimization tools [14], raw read error correction software [63, 64] and bacteria-specific Medaka models [10]. While these optimizations were outside of the scope of this study, prior studies found that cgMLST performance increased to near-perfect (0–2 gene distance to automated gold standard), suggesting that our increased threshold for defining cgMLST clusters (Fig. 3) may not be necessary with future advancements.

The excellent performance of ONT-only assemblies in identifying AMR allelic variants shown here is comparable to recent benchmarking performed in 12 Gram-negative isolates [16]. However, the imperfect retention of AMR genes in Flye assemblies (Fig. S7), particularly in plasmids, requires monitoring in future work. Flye and other long-read assembly algorithms are known to have difficulties resolving small plasmids, which frequently carry AMR genes [62, 65]. While Landman *et al.* [8] did test for AMR gene retention and found very few issues, their genomes had fewer AMR genes (6 genes across all isolates, compared to 17 genes in our dataset). Multi-input and plasmid-specific assembly algorithms such as Autocycler [62], Hybracter [61] and Plassembler [66] (included in Hybracter), or a combination of read and assembly-based approaches, may offer a more robust method for identifying, typing and contextualizing AMR allelic variants. We also focused benchmarking on key beta-lactamase genes that are frequently observed in hospital contexts, but further benchmarking is needed to thoroughly explore accuracy across a wider range of AMR genes.

The poor sensitivity of variant calling in repetitive regions using Reddog and SKA2 shown here was in keeping with recent research by Hall *et al*. [7], who demonstrated that Illumina approaches could not reliably resolve repetitive regions longer than their short-read lengths. A similar limitation is discussed in the SKA2 paper [48], with recognition that repeat regions longer than the max kmer length can lead to false negatives. False positive SNPs observed in this study were dominated by SKA2, due to an inability to identify heterozygous sites or errors introduced during the assembly process. In STs where this occurred, such as *E. coli* ST131 (Fig. S12B) and *K. pneumoniae* ST323 (Fig. S13B), the number of these SKA2-only sites was roughly equal to the consensus error rate (~0.5 errors per 100 kbp) of modern ONT data [67]. These errors point to limitations of assembly rather than SKA2 itself. Assemblers do not provide quality information that would allow for identification of heterozygous sites and filtering of sites with low quality scores, sequencing depth or similar. In addition, SKA2 was designed for Illumina data with a higher consensus accuracy and may not be suited to more error-prone ONT data. For these reasons, read mapping is a more reliable method for variant calling from ONT data.

The dataset used in this study had three potential limitations. Firstly, ONT and Illumina sequencing were performed on different extracts, potentially allowing the bacterial strain to mutate between extractions. This did not appear to have a major impact on any of the benchmarking, as we found ONT accuracy to be comparable to other benchmarking studies in most cases. In instances where ONT datasets performed worse than previously observed or expected, these discrepancies could be explained by a combination of the lack of assembly optimization (cgMLST) and inherent limitations of assembly for variant calling (SKA2) or Clair3-specific inconsistencies not observed in the other ONT read-based variant caller (PACU). The second limitation relates to the basecalling and assembly algorithms used in the study. Flye was the best performing fully automated assembly algorithm at study commencement [68]. However, Hybracter [61] and Autocycler [62] have since been released and provide fully automated approaches for leveraging the strengths of multiple assembly programmes. Autocycler or Hybracter assemblies may lead to lowered rates of presumable false positives with SKA2 and improved cgMLST performance due to a lower error rate in assembly overall. They would likely provide a more accurate automated gold standard for identifying AMR gene determinants, as the automated gold standards used in this study were Flye assemblies polished with long and short reads. While we did not find any evidence of missing resistance determinants in our gold standards using $bla_{\text{CTX-M}}$ as an example with a large sample size and known issues in ONT-only assemblies (Table S11), future benchmarking should utilise Autocycler or Hybracter as well as the new bacteria-specific Dorado and Medaka models to further optimize accuracy. Reads <1 kbp length and in the bottom fifth percentile by quality were removed to improve assembly accuracy and computational efficiency. An alternative approach may be to filter purely on quality score (e.g. selecting reads with quality score >20). We anticipate that this would have a minimal impact on assembly accuracy due to the corresponding loss of depth. However, to our knowledge, the impact of read filtering methods on ONT bacterial genome assembly has not been systematically evaluated and could form part of future benchmarking studies. Finally, this study focused on common Gram-negative pathogens. Future studies should validate benchmarking in Gram-positive organisms and less common bacterial taxa, where ONT benchmarking algorithms may have different performance.

The benchmarking shown in this study highlights the extremely high accuracy of modern R10.4.1 ONT data and its improvement over legacy ONT chemistries and software. While assembly has historically been a default step prior to MLST or AMR typing, we found that read-based tools may be better equipped to avoid omissions of MLST loci or plasmid-associated AMR genes and provide superior variant calling performance due to detection of heterozygous loci and avoiding assembly errors. ONT long reads showed improved performance over Illumina in capturing repetitive regions of the genome and had comparable performance to Illumina outside of repetitive regions, suggesting that they may be the best available method for characterizing outbreaks in hospital settings. When combining the improved accuracy of ONT sequencing with its existing cost and analysis benefits, it presents an opportunity for improved genomic surveillance and diagnostics in clinical settings.

Funding information
This work was supported by the National Health and Medical Research Council of Australia (Emerging Leader 1 Fellowship APP1176324 to N.M. and Practitioner Fellowship APP1117940 to A.Y.P.).

Acknowledgements
This research was supported by Monash eResearch capabilities, including M3. This research was also supported by the use of the Nectar Research Cloud, a collaborative Australian research platform supported by the NCRIS-funded Australian Research Data Commons (ARDC).

Author contributions
Conceptualization: N.M. Data curation: H.C., A.Y.P., K.E.H. and N.M. Formal analysis: H.C. Funding acquisition: N.M. and A.Y.P. Investigation: H.C., L.M.J., T.H.-H., J.A.W., L.V.B., T.H.-H. and M.H.P. Methodology: H.C., N.M. and J.H. Project administration: N.M., J.H., A.Y.P. and K.E.H. Resources: N.M., A.Y.P. and K.E.H. Software: H.C. Supervision: N.M., J.H., A.Y.P. and K.E.H. Validation: H.C. Visualization: H.C. Writing – original draft: H.C. Writing – review and editing: H.C., N.M., J.H., A.Y.P. and K.E.H.

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

References

1. Coulter B. GenFind V3 Genomic DNA Extraction Kit; (n.d.). https://www.beckman.com.au/reagents/genomic/dna-isolation/from-blood
2. Nanopore O. Native Barcoding Kit 24 V14; (n.d.). https://store.nanoporetech.com/native-barcoding-kit-24-v14.html
3. Baker KS, Jauneikaite E, Hopkins KL, Lo SW, Sánchez-Busó L, *et al*. Genomics for public health and international surveillance of antimicrobial resistance. *Lancet Microbe* 2023;4:e1047–e1055.
4. Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. *Nature Reviews Genetics* 2016;17:333–351.
5. Sanderson ND, Kapel N, Rodger G, Webster H, Lipworth S, *et al*. Comparison of R9.4.1/Kit10 and R10/Kit12 oxford nanopore flowcells and chemistries in bacterial genome reconstruction. *Microbial Genomics* 2023;9:mgen000910.
6. Wick RR, Judd LM, Holt KE. Assembling the perfect bacterial genome using oxford nanopore and illumina sequencing. *PLoS Computational Biology* 2023;19:e1010905.
7. Hall MB, Wick RR, Judd LM, Nguyen AN, Steinig EJ, *et al*. Benchmarking Reveals Superiority of Deep Learning Variant Callers on Bacterial Nanopore Sequence Data. *eLife* 2024.
8. Landman F, Jamin C, de Haan A, Witteveen S, Bos J, *et al*. Genomic surveillance of multidrug-resistant organisms based on long-read sequencing. *Genome Medicine* 2024;16:137.
9. Bogaerts B, Van den Bossche A, Verhaegen B, Delbrassinne L, Mattheus W, *et al*. Closing the gap: oxford nanopore technologies R10 sequencing allows comparable results to Illumina sequencing for SNP-based outbreak investigation of bacterial pathogens. *Journal of Clinical Microbiology* 2024;62:e0157623.
10. Weigl S, Dabernig-Heinz J, Granitz F, Lipp M, Ostermann L, *et al*. Improving Nanopore sequencing-based core genome MLST for global infection control: a strategy for GC-rich pathogens like *Burkholderia pseudomallei* . *Journal of Clinical Microbiology* 2025;63:e01569–01524 .
11. 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. *Journal of Clinical Microbiology* 2023;61:e01631–01622 .
12. Dabernig-Heinz J, Lohde M, Hölzer M, Cabal A, Conzemius R, *et al*. A multicenter study on accuracy and reproducibility of nanopore sequencing-based genotyping of bacterial pathogens. *Journal of Clinical Microbiology* 2024;62:e0062824.
13. 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 Research* 2024;34:2039–2047.
14. Prior K, Becker K, Brandt C, Cabal Rosel A, Dabernig-Heinz J, *et al*. Accurate and reproducible whole-genome genotyping for bacterial genomic surveillance with nanopore sequencing data. *Journal of Clinical Microbiology* 2025;63:e0036925.
15. Wu C-T, Shropshire WC, Bhatti MM, Cantu S, Glover IK, *et al*. Rapid whole genome characterization of antimicrobial-resistant pathogens using long-read sequencing to identify potential healthcare transmission. *Infection Control & Hospital Epidemiology* 2025;46:129–135.
16. Lerminiaux N, Fakharuddin K, Mulvey MR, Mataseje L. Do we still need Illumina sequencing data? evaluating Oxford Nanopore Technologies R10.4.1 flow cells and the Rapid v14 library prep kit for Gram negative bacteria whole genome assemblies. *Canadian Journal of Microbiology* 2024;70:178–189.
17. Beckman Coulter. GenFind V3 Reagent Kit; (n.d.). https://www.beckman.com/reagents/genomic/dna-isolation/from-blood/c34881
18. Nanopore O. Dorado. GitHub; (n.d.). https://github.com/nanoporetech/dorado
19. Telatin A, Fariselli P, Birolo G. SeqFu: a suite of utilities for the robust and reproducible manipulation of sequence files. *Bioengineering (Basel)* 2021;8:59.
20. 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.
21. Cottingham H. clinopore-nf. GitHub; 2023. https://github.com/HughCottingham/clinopore-nf
22. Wick RR, Judd LM, Cerdeira LT, Hawkey J, Méric G, *et al*. Trycycler: consensus long-read assemblies for bacterial genomes. *Genome Biology* 2021.
23. Wick RR, Holt KE. Polypolish: short-read polishing of long-read bacterial genome assemblies. *PLoS Computational Biology* 2022;18:e1009802.
24. Wick RR. Filtlong. GitHub; 2017. https://github.com/rrwick/Filtlong
25. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. *Nature Biotechnology* 2019;37:540–546.
26. Medaka. GitHub; 2017. https://github.com/nanoporetech/medaka
27. Zimin AV, Marçais G, Puiu D, Roberts M, Salzberg SL, *et al*. The masurca genome assembler. *Bioinformatics* 2013;29:2669–2677.
28. lh3 seqtk. GitHub; 2016. https://github.com/lh3/seqtk
29. Seemann T. mlst. GitHub; 2015. https://github.com/tseemann/mlst
30. Stanton TD. bart. GitHub; 2021. https://github.com/tomdstanton/bart
31. Silva M, Machado MP, Silva DN, Rossi M, Moran-Gilad J, *et al*. chewBBACA: A complete suite for gene-by-gene schema creation and strain identification. *Microbial Genomics* 2018;4.
32. Ridom cgmlst.org; (n.d.). https://www.cgmlst.org/ncs
33. Seemann T. cgmlst-dists. GitHub; 2020. https://github.com/tseemann/cgmlst-dists
34. Nowakowski S. Hclust1d: hierarchical clustering of univariate (1d) data. 2024.
35. Csárdi G, andNepusz T. The igraph software package for complex network research. 2006.
36. Wickham H. *Ggplot2: Elegant Graphics for Data Analysis Springer-Verlag*. 2016.
37. 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. *Scientific Reports* 2021;11:12728.
38. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 2018;34:3094–3100.
39. Henry L, Wickham HFR, Müller K. dplyr: A Grammar of Data Manipulation. 2023.
40. Wickham H. Reshaping Data with the **reshape** Package. *Journal of Statistical Software* 2007;21.
41. Yang C, Chu J, Warren RL, Birol I. NanoSim: nanopore sequence read simulator based on statistical characterization. *GigaScience* 2017;6:gix010 .
42. Huang W, Li L, Myers JR, Marth GT. ART: a next-generation sequencing read simulator. *Bioinformatics* 2012;28:593–594.
43. Holt KE. reddog-nf. GitHub; (n.d.). https://github.com/katholt/reddog-nf
44. Lam MMC, Wyres KL, Duchêne S, Wick RR, Judd LM, *et al*. Population genomics of hypervirulent *Klebsiella pneumoniae* clonal-group 23 reveals early emergence and rapid global dissemination. *Nature Communications* 2018;9:2703 .
45. Gorrie CL, Mirčeta M, Wick RR, Judd LM, Lam MMC, *et al*. Genomic dissection of *Klebsiella pneumoniae* infections in hospital patients reveals insights into an opportunistic pathogen. *Nature Communications* 2022;13:3017 .
46. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nature Methods* 2012;9:357–359.

47. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, *et al*. Twelve years of SAMtools and BCFtools. *Gigascience* 2021;10:giab008.
48. Derelle R, von Wachsmann J, Mäklin T, Hellewell J, Russell T, *et al*. Seamless, rapid, and accurate analyses of outbreak genomic data using split *k*-mer analysis. *Genome Research* 2024;34:1661–1673.
49. Zheng Z, Li S, Su J, Leung AW-S, Lam T-W, *et al*. Symphonizing pileup and full-alignment for deep learning-based long-read variant calling. *Nature Computer Science* 2022;2:797–803.
50. Nanopore O. Rerio. GitHub. 2024. https://github.com/nanoporetech/rerio
51. Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, *et al*. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Research* 2015;43:e15.
52. Seemann T. Snp-dists. GitHub. 2017. https://github.com/tseemann/snp-dists
53. Conway JR, Lex A, Gehlenborg N. UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics* 2017;33:2938–2940.
54. Gorrie CL, Mirceta M, Wick RR, Edwards DJ, Thomson NR, *et al*. Gastrointestinal carriage is a major reservoir of *Klebsiella pneumoniae* infection in intensive care patients. *Clinical Infectious Diseases* 2017;65:208–215.
55. Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* 2019;35:4453–4455.
56. Huerta-Cepas J, Serra F, Bork P. ETE 3: reconstruction, analysis, and visualization of phylogenomic data. *Molecular Biology and Evolution* 2016;33:1635–1638.
57. Clausen PTLC, Aarestrup FM, Lund O. Rapid and precise alignment of raw reads against redundant databases with KMA. *BMC Bioinformatics* 2018;19:307.
58. Gona F, Comandatore F, Battaglia S, Piazza A, Trovato A, *et al*. Comparison of core-genome MLST, coreSNP and PFGE methods for *Klebsiella pneumoniae* cluster analysis. *Microbial Genomics* 2020;6:e000347.
59. Martak D, Guther J, Verschuuren TD, Valot B, Conzelmann N, *et al*. Populations of extended-spectrum *β*-lactamase-producing *Escherichia coli* and *Klebsiella pneumoniae* are different in human-polluted environment and food items: a multicentre European study. *Clinical Microbiology and Infection* 2022;28:447..
60. 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. *Microbial Genomics* 2023;9.
61. Bouras G, Houtak G, Wick RR, Mallawaarachchi V, Roach MJ, *et al*. Hybracter: enabling scalable, automated, complete and accurate bacterial genome assemblies. *Microbial Genomics* 2024;10:001244.
62. Wick RR, Howden BP, Stinear TP. Autocycler: long-read consensus assembly for bacterial genomes. *Bioinformatics* 2025;41:btaf474.
63. Liu Y, Li Y, Chen E, Xu J, Zhang W, *et al*. Repeat and haplotype aware error correction in nanopore sequencing reads with DeChat. *Communications Biology* 2024;7:1678 .
64. Stanojevic D, Lin D, Nurk S, Florez De Sessions P, Sikic M. Telomere-to-telomere phased genome assembly using HERRO-corrected simplex nanopore reads. *Bioinformatics* 2024. DOI: 10.1101/2024.05.18.594796.
65. Johnson J, Soehnlen M, Blankenship HM. Long read genome assemblers struggle with small plasmids. *Microbial Genomics* 2023;9:mgen001024.
66. Bouras G, Sheppard AE, Mallawaarachchi V, Vreugde S. Plassembler: an automated bacterial plasmid assembly tool. *Bioinformatics* 2023;39:btad409.
67. Sanderson ND, Hopkins KMV, Colpus M, Parker M, Lipworth S, *et al*. Evaluation of the accuracy of bacterial genome reconstruction with Oxford Nanopore R10.4.1 long-read-only sequencing. *Microbial Genomics* 2024;10:001246.
68. Wick RR, Holt KE. Benchmarking of long-read assemblers for prokaryote whole genome sequencing. *F1000Res* 2019;8:2138.