JAC Antimicrob Resist
https://doi.org/10.1093/jacamr/dlaf069

JAC-Antimicrobial Resistance

Tracking Antimicrobial Resistant Organisms Timely: a workflow validation study for successive core-genome SNP-based nosocomial transmission analysis

Kotaro Aoki[1]*, Kohji Komori [2], Tetsuo Yamaguchi[1], Sohei Harada[1], Mayumi Tsukada[3], Hinako Murakami[3,4] and Kazuhiro Tateda[1,2,3,4]

[1]Department of Microbiology and Infectious Diseases, Toho University School of Medicine, Tokyo, Japan; [2]Division of Collaborative Regional Infection Control, Department of Community Well-being, Toho University School of Medicine, Tokyo, Japan; [3]Department of Infection Prevention and Control, Toho University Omori Medical Center, Tokyo, Japan; [4]Department of Clinical Laboratory, Toho University Omori Medical Center, Tokyo, Japan

*Corresponding author. E-mail: kotaro.aoki@med.toho-u.ac.jp

Received 4 February 2025; accepted 18 April 2025

Background and Objectives: Effective infection prevention and control (IPC) interventions in hospitals require timely information to determine the potential transmission of antimicrobial-resistant (AMR) organisms. We proposed and developed a successive core-genome SNP (cgSNP)-based phylogenetic analysis workflow, 'Tracking Antimicrobial Resistant Organisms Timely' (TAROT), using the Oxford Nanopore Technologies (ONT) sequencer for MRSA, and compared the results with those obtained using the Illumina sequencer.

Methods: We have developed a TAROT workflow for successive phylogenetic analysis using ONT data. We sequenced 34 MRSA strains isolated from Toho University Omori Medical Center using MinION (ONT) and MiSeq (Illumina). Each strain's ONT data were inputted into TAROT (TAROT–ONT), and successive cgSNP-based phylogenetic analyses were conducted. Illumina data were processed with a batched cgSNP-based phylogenetic analysis. Assembly-based analysis identified AMR genes, AMR mutations and virulence genes.

Results: MinION generated an average sequence depth of 262× for the ST8 reference genome within 3 h. TAROT–ONT successively generated 11 phylogenetic trees for 14 ST8 strains, 7 trees for 10 ST1 strains and 2 trees for 5 ST5 strains. Highly suspected transmission pairs (pairwise cgSNP< 5) were detected in trees #6 through #11 for ST8, trees #3, #5 and #7 for ST1, and tree #2 for ST5. Differences in pairwise cgSNP value between TAROT–ONT and Illumina ranged from zero to two within pairs with fewer than 20 cgSNPs using Illumina. TAROT–ONT bioinformatic analysis for each strain required 5–42 min. The identification of AMR genes, mutations and virulence genes showed high concordance between ONT and Illumina.

Conclusions: TAROT–ONT can facilitate effective IPC intervention for MRSA nosocomial transmissions by providing timely feedback through successive phylogenetic analyses based on cgSNPs.

Introduction

The transmission of antimicrobial-resistant (AMR) organisms in hospitals is a critical issue that undermines patient safety and results in substantial economic losses.[1–3] The rising number of AMR organisms identified in clinical microbiology laboratories may signal their spread within hospitals. Among these, MRSA is closely monitored in most hospitals, and a short-term increase in the MRSA isolation rate may indicate active transmission.[4] MRSA clones can frequently spread within hospitals, even when patients carrying MRSA are quarantined, and contact precautions are implemented.[5] Effective prevention of nosocomial transmission requires detailed information to verify whether clonal transmission of MRSA strains is occurring, but this cannot be determined through routine microbiological testing.[6–8]

WGS enables SNP analysis within the core-genome, which is considered the gold-standard method for quantitatively determining genetic relationships among strains. This approach has replaced traditional typing methods for MRSA, such as pulsed-field gel electrophoresis, multilocus variable-number tandem repeat analysis, and *spa*-typing.[7,9–15] Core-genome SNP (cgSNP) analysis was first applied to the investigation of MRSA outbreaks.[16] The method quantitatively evaluates the genetic relationships between strains based on the number of detected cgSNPs. The thresholds of cgSNPs that support short-term nosocomial transmission depend on the evolutionary rate and within-host variation of the species under consideration. Coll *et al.*[17] proposed a conservative cgSNP threshold for MRSA, suggesting that transmission within 6 months can be ruled out if more than 15 cgSNPs are identified.

Using genome epidemiology of AMR organisms with WGS and cgSNP analysis in healthcare settings provides crucial insights to enhance and target infection prevention and control (IPC) interventions. The routine implementation of weekly WGS for MRSA not only improves the accuracy of outbreak detection but also enables the de-escalation of IPC activities by identifying pseudo-outbreaks.[18] However, frequent WGS and timely reporting to the IPC team would have an even greater impact on IPC decision-making. To date, cgSNP analysis has been conducted using short-read data generated by high-throughput sequencers (HTS), specifically the Illumina system, to reduce costs through batch sequencing.[19–21] However, batch sequencing is time-consuming as it requires the collection of a large number of samples for a single analysis, making it challenging to provide cgSNP analysis results promptly and timely enough to influence IPC intervention.[22,23] The latest MinION/Flongle technology [R10.4.1; Oxford Nanopore Technologies (ONT), Oxford, UK] offers a cost-effective and scalable solution for daily, single-strain WGS and recent improvements have addressed its previously lower basecalling accuracy compared with Illumina.[24–27] Nevertheless, most automated bioinformatic tools and workflows for cgSNP analysis and genome epidemiology are optimized for short-read HTS data.[28,29]

To bridge the gap between ONT sequencing and automated bioinformatics workflows for efficient IPC interventions, we propose 'Tracking Antimicrobial Resistant Organisms Timely' (TAROT) (Figure 1) along with a novel bioinformatics workflow that was newly developed in this study (Figures 2 and S2 (available as Supplementary data at *JAC-AMR* Online)). TAROT facilitates automated, successive cgSNP-based phylogenetic analyses by using ONT data on the server side, eliminating the need for high-performance machines on the client side, such as in clinical microbiology laboratories. TAROT includes databases that accumulate core-genome data, categorize it by strain lineage and conduct cgSNP-based phylogenetic analysis based on the input of each strain lineage. In this study, we validated the performance of cgSNP analysis using ONT data within TAROT (TAROT–ONT) by comparing it with cgSNP analysis based on Illumina data, applied to MRSA strain, which serve as a model AMR pathogen.

Materials and methods

Strains

A total of 165 MRSA strains, isolated from both patients (clinical isolates) and ward environments (environmental isolates) at Toho University Omori Medical Center—a 916-bed academic medical centre in Tokyo, Japan—were collected for infection control investigations between May 2011 and February 2022 (Table S1). Prior to this study, draft WGS using MiSeq (Illumina Inc., CA, USA) and cgSNP-based phylogenetic analysis had been conducted to inform IPC interventions (Figure S1). The 34 MRSA strains used in this study—30 clinical isolates and 4 environmental isolates—were purposively selected as a representative set based on their suitability for validating TAROT, with reference to the previously conducted cgSNP-based phylogenetic analysis using MiSeq data. Some strains were selected from the three frequently collected STs (ST8, ST1 and ST5) to assess whether TAROT could identify the number of cgSNPs needed to distinguish strains within the same ST for suspected transmission. Other strains were selected from different STs to verify the functionality of the automated database classification systems across different STs. In total, the strains included 14 belonged to ST8, 10 to ST1, 5 to ST5 and one each to ST97, ST1516, ST2725, ST2764 and ST4143 (Tables 1 and S2).

Ethics

This study did not require ethics approval, as it was conducted as a methodological validation of a successive cgSNP analysis workflow using MRSA strains that had previously been analysed for IPC purposes at Toho University Omori Medical Center. No clinical information that could identify individuals is included in this manuscript.

Sequencing by MiSeq and analysis

DNA was extracted from *S. aureus* using achromopeptidase and phenol/chloroform/isoamyl alcohol (25:24:1) (FUJIFILM Wako Pure Chemical Corporation, Osaka, Japan) and purified using the spin column method with the FastGene Gel/PCR Extraction Kit (NIPPON Genetics Co., Ltd, Tokyo, Japan). DNA library preparation was performed using the Illumina DNA Prep (M) Tagmentation Library Preparation kit and IDT for Illumina DNA/RNA UD Indexes (Illumina). Sequencing was planned with 300 bp paired-end reads using the 600-cycle MiSeq Reagent Kit v3 (Illumina). Reads were trimmed and filtered using Trimmomatic version 0.39,[30] applying a quality score of 20 and a minimum read length of 50 bp, and *de novo* assembly to generate contigs was performed using SPAdes version 3.15.1.[31]

Sequencing by MinION R10.4.1 and analysis

High molecular weight DNA was extracted from *S. aureus* strains by bead-beating with Ez-Beas (Promega K.K., Tokyo, Japan) and processed with the magLEAD system and the magDEA SV and PS protocol (Precision System Science Co., Ltd, Chiba, Japan). For ONT, DNA library preparation was performed using the Rapid Barcoding Kit 24 V14 SQK-RBK114.24 (ONT). Sequencing was conducted on MinION with Flow Cell (R10.4.1) FLO-MIN114 (ONT) for up to 72 h. Basecalling and demultiplexing were performed using Dorado version 0.6.0 with the dna_r10.4.1_e8.2_400bps_sup@v4.1.0 model. Reads were trimmed and filtered with NanoFilt version 2.8.0, applying a quality score of 10 and a minimum read length of 1000 bp.[32] To limit downstream data volume, reads were down-sampled with Seqkit version 2.6.1, ensuring that the total nucleotide count did not exceed 100 times the genome size of *S. aureus*.[33] Self-error correction was carried out using DeChat version 1.0.1, and *de novo* assembly was performed with Flye version 2.9.4.[34,35]

Species identification, basic typing and gene identification

Species identification was primarily conducted using KmerFinder, and subsequently confirmed by calculating the Average Nucleotide Identity (ANI) against the type strain genome using FastANI.[36,37] A threshold of 95.0% ANI was applied as the cut-off for species identification.[38] Identification

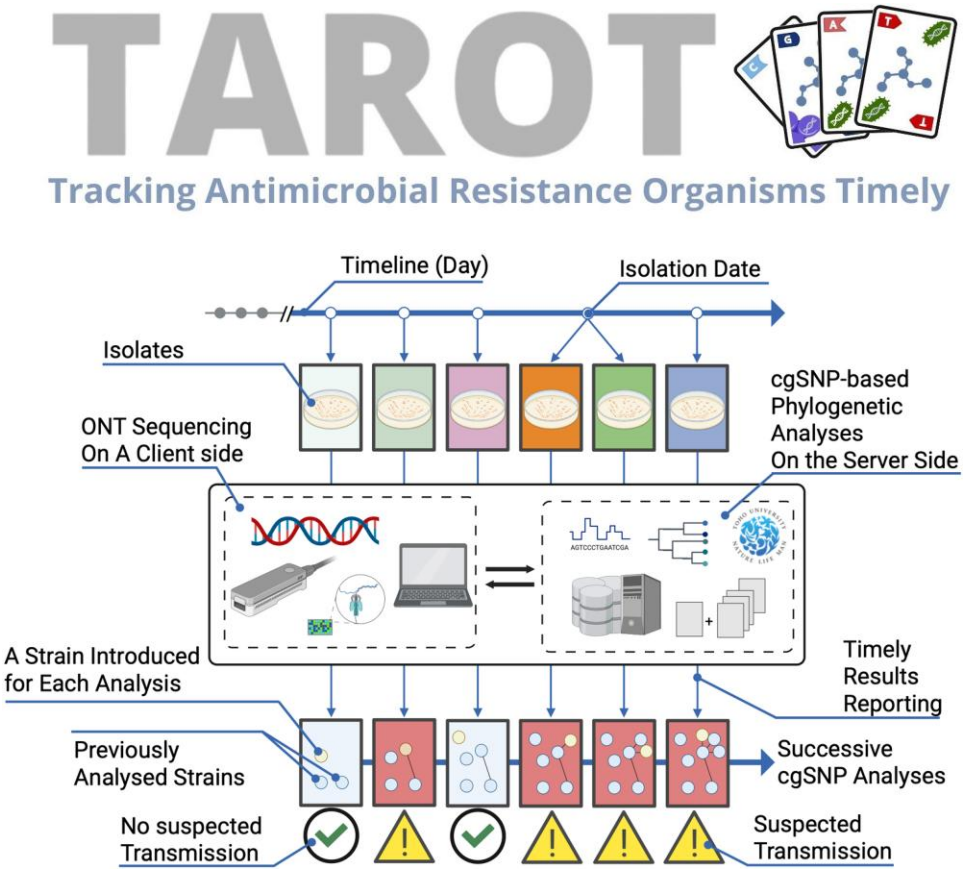

Figure 1. Concept of TAROT. TAROT provides on-demand, successive analyses of cgSNP. On the client side, each isolate is sequenced using an ONT sequencer. The resulting sequence data are then transferred to and analysed on the server side, and the cgSNP analysis results are reported for each isolate. Timely feedback can help enhance IPC interventions in hospitals. This figure was created with BioRender.com.

of acquired AMR genes and nucleotide point mutations in the quinolone resistance-determining regions (QRDRs) of DNA gyrase subunit A (GyrA) and DNA topoisomerase IV subunit A (GrlA), as well as the rifampicin-resistance determining regions (RRDRs) of RNA polymerases subunit B (RpoB), was conducted with ResFinder version 4.5.0.[39] SCC*mec* typing was performed using SCCmecFinder version 1.0.0.[40] MLST was conducted following the PubMLST scheme (https://pubmlst.org/organisms/staphylococcus-aureus). Virulence gene identification was performed with VirulenceFinder version 2.0.5.[41] These analyses were conducted on genomes generated from both Illumina and ONT sequencing. Default parameters were used for all tools in this section.

Successive cgSNP-based phylogenetic analysis in TAROT

TAROT was validated by randomly inputting MinION data one at a time into the workflow and comparing it to Illumina data inputs. The results obtained using ONT were compared with those obtained using Illumina. The TAROT bioinformatic workflow is illustrated in Figures 2 and S2. Successive cgSNP phylogenetic analysis, rather than analysing a single tree per ST, was chosen to enable timely feedback of results to IPC teams in hospitals (Figures 1 and S2). Representative STs of MRSA reference genomes were selected based on a review article and prepared within the TAROT database (Table S3).[42] For mapping analysis, a reference genome was selected based on either MLST analysis results using stringMLST[43] or the best mapping outcome across multiple references using BBsplit (version was last modified 11 June 2018), a tool included in BBTools (https://github.com/kbaseapps/BBTools). Simulations of short reads for 300 bp of 1 000 000 paired-end reads were generated using Wgsim version 1.20.[44] Mapping to the reference genome and generating BAM files was performed with BWA version 0.7.17 using the sw option, and SAMtools version 1.20.[44,45] The BAM files were grouped and organized based on MLST or BBsplit results. The core-genome was extracted from the grouped BAM files using SAMtools with the mpileup option and VarScan version 2.3.9 with the mpileup2cns option.[46] Recombination regions in the core-genome were identified and excluded using ClonalFrameML version 1.12, followed by text editing in the Bash console.[47] Pairwise cgSNP calculations were performed using SNP-dists version 0.8.2 (https://github.com/tseemann/snp-dists). Phylogenetic trees were generated using RAxML version 8.2.12 with the raxmlHPC-PTHREADS option, model GAMMAMI, and 1000 bootstrap replications.[48] The entire time for inputting FASTQ files into TAROT and outputting a phylogenetic tree was measured using the time command on Ubuntu 20.04.

Data availability

The WGS data have been deposited in GenBank under the BioProject accession number PRJNA1090044, and SRA and nucleotide accession numbers are shown in Table S2.

Code availability

The TAROT code is available on GitHub (https://github.com/KotaroAoki/TAROT_MRSA) and can be run in a Docker container (https://hub.docker.com/repository/docker/kotaroaoki/tarot-mrsa-env/general).

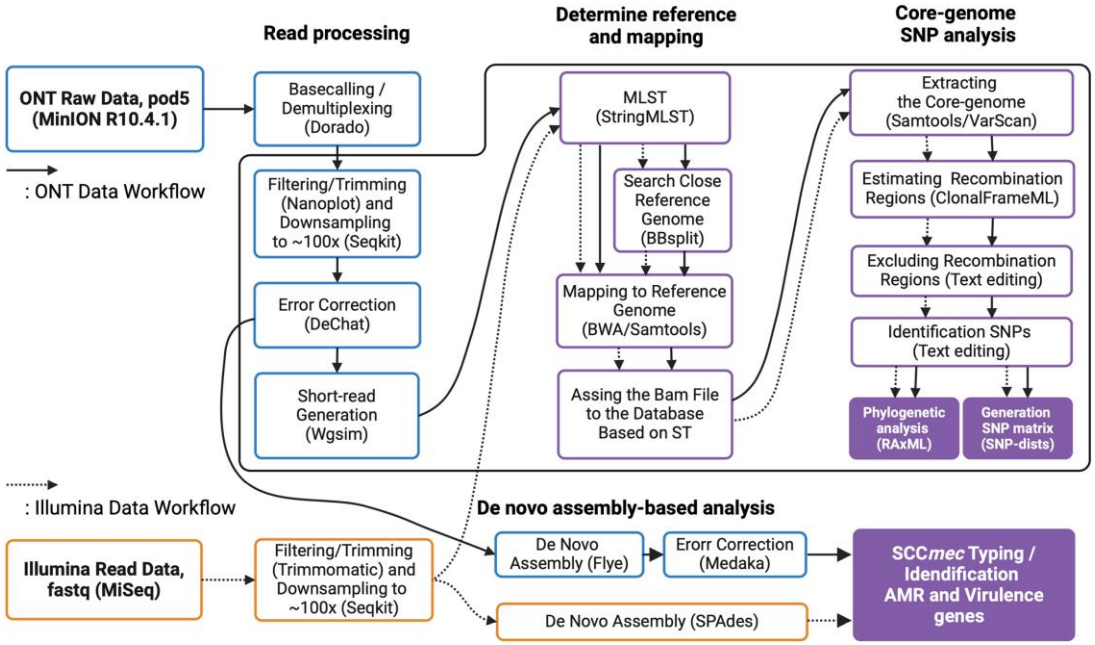

Figure 2. Bioinformatics workflow of TAROT. Solid arrows represent the ONT data workflow, while dashed-arrows indicate the Illumina data workflow. The time taken for the workflow in the solid box processes was measured using the 'time' command in Linux. This figure was created with BioRender.com.

Results

Sequencing yield and assembly results of MinION and MiSeq

The average sequencing depth per MinION run for the ST8 reference genome, obtained from sequencing 34 strains across four runs, was 73× at 1 h, 262× at 3 h, 451× at 6 h, 1901× at 24 h and 3080× at 48 h (Table S4). The average sequencing depths achieved by MinION and MiSeq for each assembled genome were 65× and 98×, respectively (Tables 1 and S2). Assembly using ONT reads provided closed genome replicons in 30 of the 34 strains. Among these, two closed plasmids were detected in six strains and one closed plasmid in 18 strains (Table S2). The results of MLST and SCC*mec* typing showed strict concordance between MinION and MiSeq (Table S2).

Successive phylogenetic analysis using TAROT–ONT

A successive phylogenetic analysis using TAROT–ONT generated 11 trees for 14 strains of ST8, 7 trees for 10 strains of ST1 and 2 trees for 5 strains of ST5 (Figure 3a to c). TAROT–ONT assigned MRSA strains to the same ST as those identified by Illumina. Here, we set the threshold for highly suspected transmission as pairwise cgSNP < 5 and for suspected transmission as pairwise cgSNP < 15. Four pairs of highly suspected transmission events were detected in ST8 (Figure 3a and Table S5), four in ST1 (Figure 3b and Table S5) and two in ST5 (Figure 3c and Table S5). In ST8, three suspected transmission events were identified with TUM22721 added in tree #11 related to one of the highly suspected transmission pairs. Timely detection of highly suspected or suspected transmissions occurred in trees #6 through #11 for ST8, trees #3, #5 and #7 for ST1 and tree #2 for ST5.

Apart from strains belonging to ST8, ST1 and ST5, TUM22730 belonged to ST97 and was assigned to a reference genome of the same ST in the database (Table S6). The remaining four strains were not assigned to any reference genome in the database by MLST and were classified using BBsplit as a reference genome classifier. TUM20915 belonged to ST2725, TUM22749 to ST2764 and TUM22741 to ST4143, which were all classified as ST1_bbsplit. ST1_bbsplit differed from ST1 and was obviously unrelated to nosocomial transmissions; therefore, these strains were not included in the cgSNP-based phylogenic analysis of ST1 strains. TUM22723, belonging to ST1516, was assigned to ST228_bbsplit.

Comparison of cgSNP-based phylogenetic analysis accuracy between TAROT–ONT and Illumina

The proportion of core-genome size over the reference genome, calculated using the maximum number of strains in cgSNP-based phylogenetic analysis for TAROT–ONT and Illumina, were 93.1% and 94.3% for ST8, 97.2% and 98.6% for ST1, and 95.3% and 97.6% for ST5, respectively (Table S6). The differences in cgSNPs identified between TAROT–ONT and Illumina ranged from zero to 17 for ST8, zero to 14 for ST1 and zero to 5 for ST5 (Figure 4). In pairs of cgSNPs with fewer than 100 identified in Illumina, the differences were significantly lower, ranging from zero to two for ST8, zero to eight for ST1 and zero to two for ST5. Additionally, in pairs of cgSNPs with fewer than 20 in Illumina, the differences ranged from zero to two for ST8 and were zero for ST1 and ST5.

Table 1. Strains used for validation of TAROT for MRSA

MLST (Clonal complex)	MLST (Sequence type)	Strain ID (TUM)	SCCmec typing (MiSeq)	Average sequencing depth (MiSeq) (*x*)	Average sequencing depth (MinION) (*x*)
CC8 (*n*=15)	ST8 (*n*=14)	20816	IV	211.7	28.7
		20817	IV	252.7	57.7
		20818	IV	211.3	39.1
		20823	IV	209.1	19.7
		20888	IV	182.4	45.6
		20914	IV	310.7	67.3
		20929	IV	117.7	52.6
		20963	IV	109.5	182.8
		22173	IV	77.2	102.8
		22182	IV	77.3	48.3
		22698	I	82.8	34.4
		22699	I	74.3	28.6
		22705	IV	85.8	68.5
		22721	IV	79.0	88.9
	ST1516	22723	IV	51.5	24.2
CC1 (*n*=13)	ST1 (*n*=10)	20819	IV	64.1	79.0
		20820	IV	52.5	62.6
		20825	IV	84.5	117.0
		20913	IV	77.1	88.2
		20926	IV	57.6	120.9
		20967	IV	66.9	121.7
		22178	IV	44.4	155.4
		22188	IV	51.4	94.2
		22744	IV	48.6	114.4
		22750	IV	51.0	177.2
	ST2725 (*n*=1)	20915	IV	51.9	33.5
	ST4143 (*n*=1)	22741	IV	68.0	41.2
	ST2764 (*n*=1)	22749	IV	52.0	21.6
CC5 (*n*=5)	ST5 (*n*=5)	22457	II	64.9	32.8
		22458	II	52.9	11.1
		22701	II	84.9	18.6
		22702	II	84.4	15.1
		22707	II	74.3	14.7
CC97 (*n*=1)	ST97 (*n*=1)	22730	IV	74.7	16.7

Time taken for each bioinformatic analysis of each strain in TAROT–ONT

The time required for the bioinformatic analysis of each strain, from inputting FASTQ data from ONT to obtaining a phylogenetic tree via the TAROT workflow, ranged from 5 to 42 min (Table S6). Analysis times were significantly shorter when three or fewer strains were assigned to the same reference genome, as phylogenetic analysis was not conducted in those cases. The time required increased with the number of strains assigned to the same reference genome.

Comparison of detection for genes of interest between ONT and Illumina

Using ResFinder with contigs obtained from ONT or Illumina, 10 types of acquired AMR genes were identified in 34 strains (Figure S3). The only false-negative result for ONT, using Illumina as the reference, was *blaZ* in TUM22741. In TUM22741, *blaZ* was detected as part of a gene sequence comprising 681 bp out of the total gene length of 864 bp. The average ONT sequence depth for the contigs of TUM22741 was 19×, which was slightly low (Table S1). Conversely, the false-positive results detected by ONT included *blaZ* in TUM22701, TUM22702, TUM22457 and TUM22458, as well as *ermA* in TUM22701 and TUM22702, all of which were detected only as partial gene length by Illumina (Figure S3). ONT detected multiple copies of resistance genes such as *aac(6′)-aph(2′′)*, *ant(9)-Ia*, *blaZ*, *cat* and *ermA* in several strains (Figure S3). There was no discrepancy between ONT and Illumina in identifying QRDR mutations in GyrA and GrlA, RRDR mutations in RpoB, and virulence genes (Figures S3 and S4).

Comparison of genome structure for single or multi-copy AMR gene-detected replicons

A part of multi-copy gene detection probably results in assembly errors. Two types of *blaZ*-carrying circular plasmids were

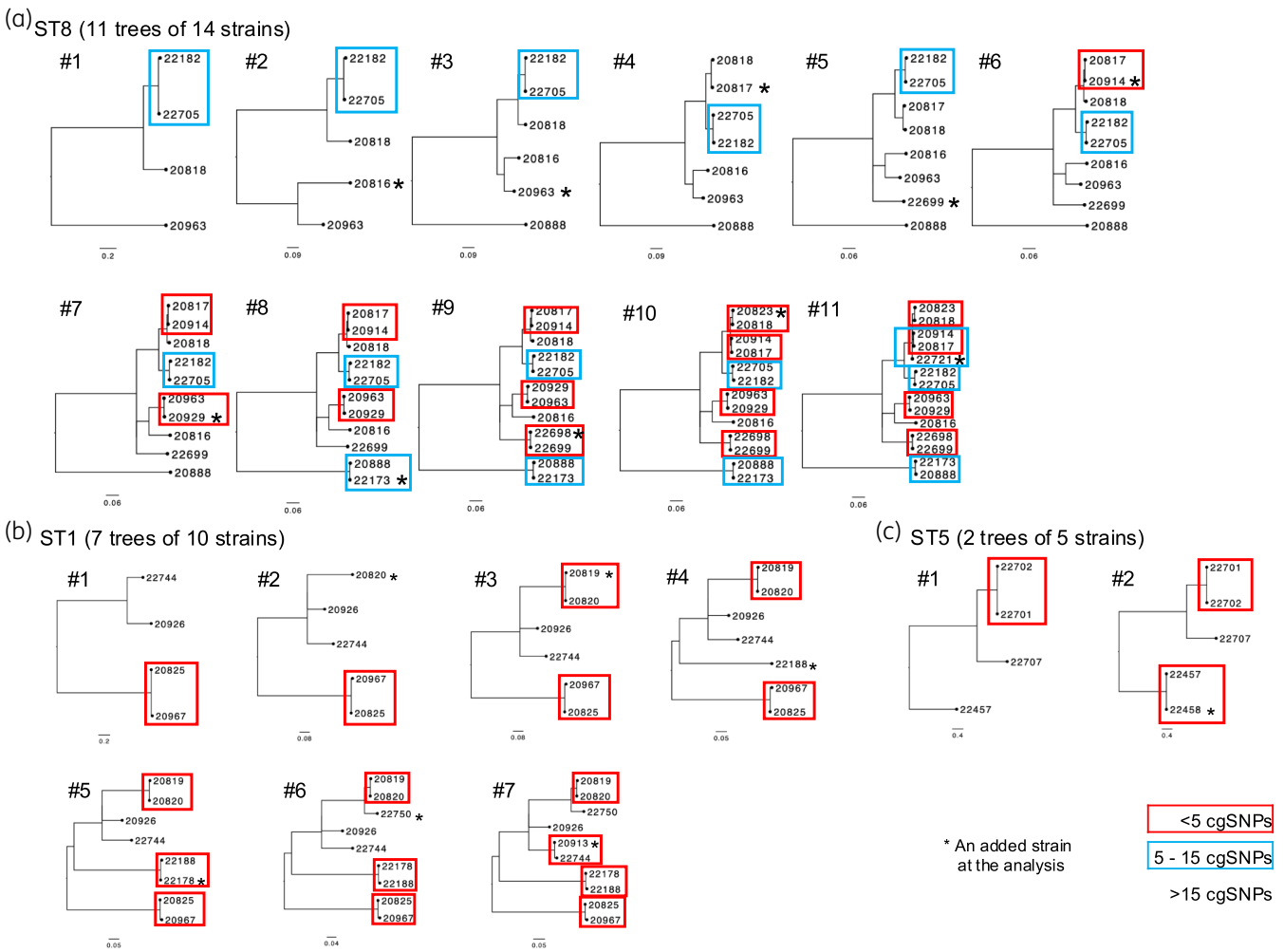

Figure 3. Phylogenetic trees obtained by successive cgSNP-based phylogenetic analysis by TAROT–ONT. (a) Phylogenetic trees for ST8; (b) Phylogenetic trees for ST1; (c) Phylogenetic trees for ST1. The box highlights transmission pairs: highly suspected transmissions involve fewer than five cgSNPs, while suspected transmissions range from five to fifteen cgSNPs. Each asterisk denotes a strain added to the analysis.

identified in ST1 strains (Figure S3). One type of plasmid, carrying a single *blaZ* (~21 326 bp), was identified in five strains (TUM20926, TUM20820, TUM22178, TUM22744 and TUM22188) (Figure S5). The other type of plasmid, carrying two copies of *blaZ* (~42 652 bp), was identified in five additional strains (TUM20913, TUM20819, TUM22750, TUM20967 and TUM20825) (Figure S5). The plasmid carrying two copies of the *blaZ* consisted of two single-copy *blaZ*-carrying plasmids connected in tandem. The smaller plasmids carrying a single *blaZ* appeared to be a structures unit (Figure S5). Mapping analyses to confirm whether assembly errors of ONT data generated the tandem unit sequences revealed no reads longer than the unit structure length (~21 326 bp) in plasmid-harbouring strains with two copies of *blaZ*. In 10 ST8 strains, *blaZ* was located on the chromosome as a single copy (Figure S3).

A plasmid carrying two copies of *aac(6′)-aph(2′′)* in strain TUM20914 (pMTY20914-1) was composed of a tandemly connected plasmid carrying a single copy of *aac(6′)-aph(2′′)* in TUM20817 (pMTY20817-1) (Figures S3 and S6). Mapping analysis of ONT reads from TUM20914 to pMTY20914-1 found only shorter reads than the length of pMTY20817-1. Another plasmid type, carrying multiple copies of *cat* in TUM20817 (pMTY20817-2) and TUM20914 (pMTY20914-2), consisted of a single unit of ~4.2 kb, with three and four copies connected in tandem, respectively (Figures S3 and S7). Like *blaZ*-carrying plasmids, mapping analysis of ONT reads from TUM20817 and TUM20914 to pMTY20817-2 and pMTY20914-2 found only shorter reads than the length of the single unit.

ermA and *ant(9)-Ia* were located adjacent to each other on the chromosomes of 27 strains. This gene pair was a single copy in one strain, two copies in 21 strains and three copies in five strains (Figure S3). These gene pairs were located within an ~6.3 kb unit containing integrase genes. In comparative chromosome structure analysis of representative strains TUM22741 (single-copy), TUM22182 (two-copy) and TUM22721 (three-copy), the insertion positions of the first and second units were identical among the strains (Figure S8).

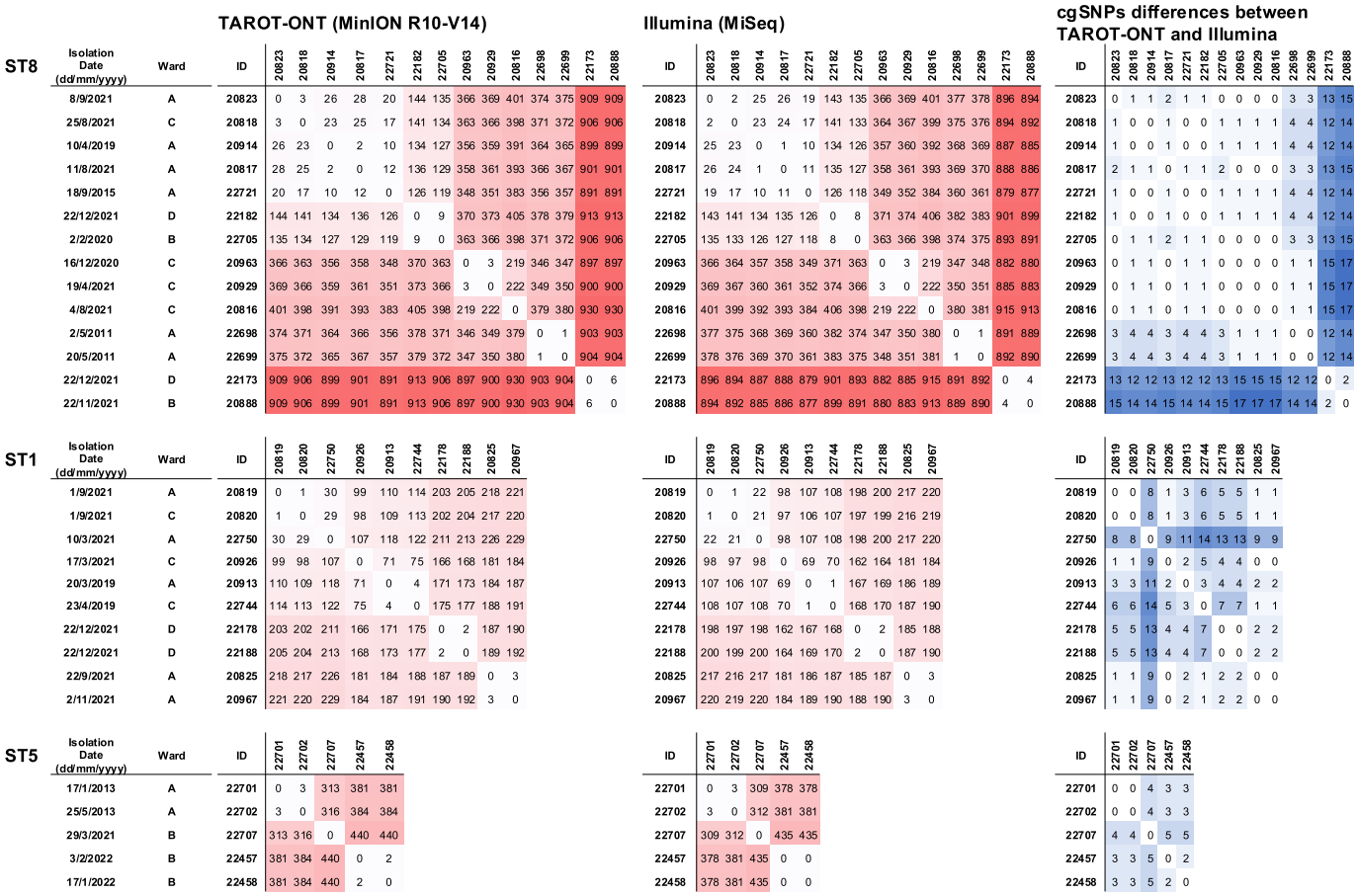

Figure 4. Comparison of cgSNP identification differences between TAROT–ONT and Illumina. The left and middle columns display matrices showing pairwise cgSNP values, while the right column illustrates the differences in cgSNP values between TAROT–ONT and Illumina. Darker colours in the heatmap indicate higher cgSNP values.

Discussions

TAROT has been developed to provide timely feedback on nosocomial transmission of AMR organisms for IPC intervention, using ONT technology instead of batch WGS bacteria processing via Illumina in central laboratories. Recent updates in ONT technology (R10.4.1) have improved its performance to a level comparable to that of Illumina, thereby making the TAROT concept feasible.[25,26] ONT sequencers offer strong scalability and an optimal balance between consumable costs and throughput. For instance, MinION provides adequate throughput to analyse ~12 MRSA strains in a single run. If only one strain is analysed, the Flongle can be used instead to save consumable costs. Regarding data analysis, clinical microbiology laboratories in hospitals generally employ few bioinformaticians, so bioinformatics support is essential. Automating the data analysis workflow helps reduce the workload on clinical laboratory staff, enables timely feedback to IPC teams, and contributes to standardization and quality control of the entire process.

The advantage of the TAROT concept over batch WGS and cgSNP-based phylogenetic analysis is its ability to analyse each strain with ONT and cgSNP methods at a quality level comparable to Illumina, providing feedback within the same day.[25] In a previous study, batch WGS of MRSA and cgSNP-based analysis identified an outbreak within 33 h starting from strain processing.[49] However, batch processing in IPC intervention requires waiting to accumulate sufficient samples for a batch, delaying response time. Validated in this study, TAROT provides a faster turnaround time (TAT), enabling timely feedback that can efficiently inform IPC intervention at the bedside. The entire TAROT process, from DNA extraction of MRSA to the phylogenetic analysis report, can be completed within 6 h (Figure 5). The breakdown is as follows: 1 h for DNA extraction, 1 h for ONT library preparation, 3 h for ONT sequencing, and 1 h for bioinformatic analysis. The average time to achieve a 50× sequence depth per MRSA genome was 44 min (Table S4), potentially allowing the TAROT's TAT to be <6 h. If Flongle, with lower throughput, is used instead of MinION, ~3 h should be allocated for sequencing.

A prospective genomic surveillance study of MRSA in Neonatal Intensive Care Unit patients using batch sample processing with Illumina HTS demonstrated its utility by identifying local transmission events.[50] Genome analyses using ONT are valuable for outbreak investigations not only for MRSA, as used in this study, but also for Gram-negative bacteria such as Shiga toxin-producing

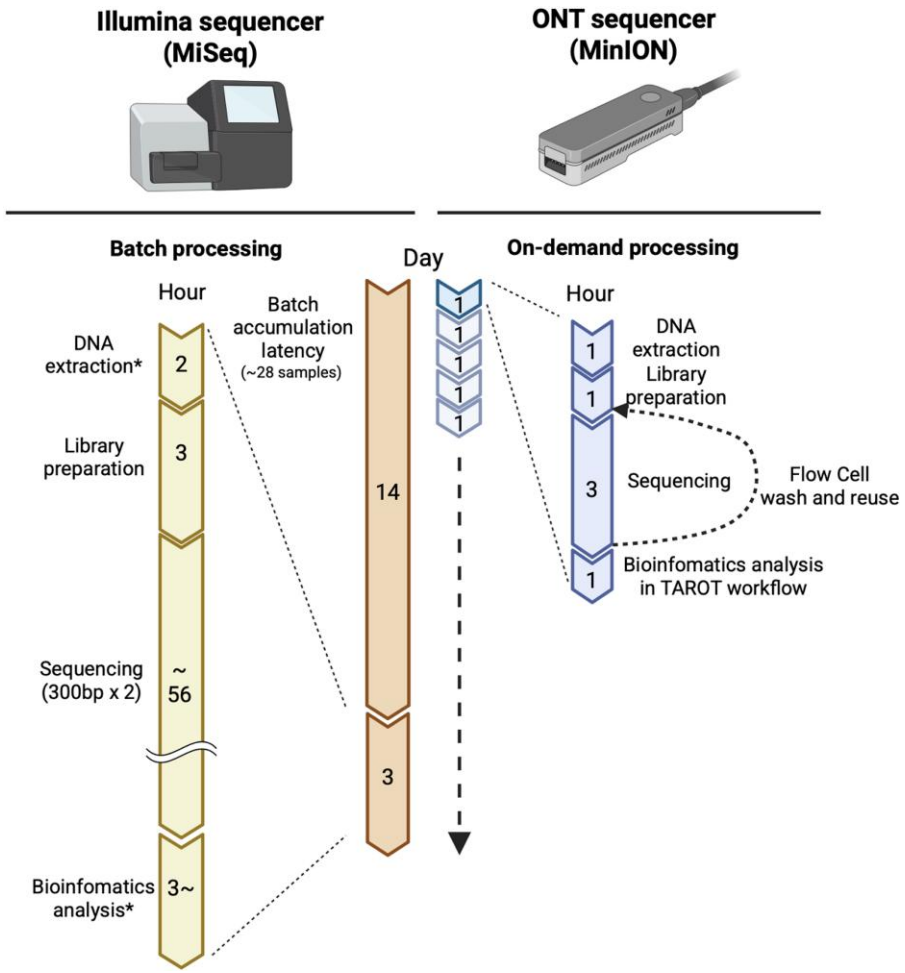

Figure 5. Comparing processing time using Illumina and ONT sequencer. The emergence of two strains per day was assumed in this illustration. The ONT sequencer can save considerable time and provide timely feedback compared with the Illumina sequencer, particularly in the context of IPC in hospitals. *Batch processing requires more hands-on time and CPU time than on-demand processing. This figure was created with BioRender.com.

Escherichia coli, described as a tool for 'closing the gap' from the short-read sequencing era.[25] The updated ONT R10.4.1, which was also used in this study and offers higher nucleotide sequencing accuracy compared with R9.4.1, has proved valuable for bacterial genomic epidemiology.[26,51,52] Our ONT results showed that AMR and virulence gene identification, as well as MLST and SCC*mec* typing results, were highly concordant with Illumina results. However, contigs identified as plasmids carrying AMR genes in certain strains suggested possible assembly errors with the Flye assembler. Comparative plasmid structure analyses using ONT-only *de novo* assemblies may still require visual inspection of read mapping to the generated contigs. Notably, the assembly process branched from the TAROT workflow to minimize the time required for high-priority phylogenetic analysis (Figure 2).

An appropriate threshold of cgSNPs to rule out nosocomial transmission of clonal isolates could support effective IPC interventions. Coll *et al.*[17] proposed a conservative cgSNP threshold of 15 to exclude transmission within 6 months, considering within-host strain variation and evolutionary rate, based on WGS data from MRSA strains isolated from 455 patients using Illumina. According to this threshold, 95% of epidemiologically linked cases could be identified among 775 cases.[17] TAROT–ONT showed high compatibility with Illumina when pairwise cgSNP values were <20, suggesting that this threshold could be applied to TAROT–ONT interpretations. Successive MRSA analyses using TAROT–ONT in high-risk wards for nosocomial transmission can timely identify MRSA strains that require specific transmission prevention interventions. TAROT could help reduce staff workload by enabling focused IPC interventions.

This study has four potential limitations. First, the number of strains and the selection method used to validate TAROT were limited. TAROT could be further enhanced to provide a more robust workflow by analysing more strains, especially when applied for MRSA IPC interventions in hospitals. Second, TAROT's phylogenetic analysis results were not evaluated based on epidemiological links in a hospital setting. TAROT could prioritize cases by identifying MRSA strains with a higher transmission likelihood based on molecular epidemiology. Therefore, future studies should consider evaluating this functionality. Third, we simulated successive phylogenetic analyses using in silico methods that accounted for ONT sequencing consumable kits. Users can choose MinION or Flongle depending on the number of strains analysed

daily. Fourth, while we have published a TAROT script and a Docker container image to build the necessary environment, we have not yet developed web services for TAROT. Such services would be essential for deploying TAROT in hospitals, and we plan to develop a network architecture for authorized members to have secure access.

In conclusion, TAROT–ONT provides accuracy compatible with cgSNP-based phylogenetic analysis using Illumina for MRSA, facilitating effective IPC interventions in hospitals by offering timely feedback.

Acknowledgements

We thank Dr Yoshikazu Ishii for advising us about conducting this study.

Funding

This work was supported by Japan Society for the Promotion of Science KAKENHI grant number Grant-in-Aid for Early-Career Scientists JP21K16327 and Promotion of Joint International Research (Fostering Joint International Research (A)) JP22KK0280.

Transparency declarations

S.H. reports speaker payments from MSD, Shionogi, Pfizer and consulting fees from Shionogi, Pfizer and Denka outside the submitted work. All other authors have no conflicts of interest to declare.

Supplementary data

Figures S1 to S8 and Tables S1 to S6 are available as Supplementary data at *JAC-AMR* Online.

References

1 Zhen X, Lundborg CS, Zhang M *et al*. Clinical and economic impact of methicillin-resistant *Staphylococcus aureus*: a multicentre study in China. *Sci Rep* 2020; **10**: 3900. https://doi.org/10.1038/s41598-020-60825-6

2 Zhen X, Stålsby Lundborg C, Sun X *et al*. Economic burden of antibiotic resistance in China: a national level estimate for inpatients. *Antimicrob Resist Infect Control* 2021; **10**: 5. https://doi.org/10.1186/s13756-020-00872-w

3 Zhen X, Stålsby Lundborg C, Sun X *et al*. Clinical and economic impact of third-generation cephalosporin-resistant infection or colonization caused by *Escherichia coli* and *Klebsiella pneumoniae*: a multicenter study in China. *Int J Environ Res Public Health* 2020; **17**: 9285. https://doi.org/10.3390/ijerph17249285

4 Rocha LEC, Singh V, Esch M *et al*. Dynamic contact networks of patients and MRSA spread in hospitals. *Sci Rep* 2020; **10**: 9336. https://doi.org/10.1038/s41598-020-66270-9

5 Pei S, Liljeros F, Shaman J. Identifying asymptomatic spreaders of antimicrobial-resistant pathogens in hospital settings. *Proc Natl Acad Sci U S A* 2021; **118**: e2111190118. https://doi.org/10.1073/pnas.2111190118

6 Ogihara S, Yamaguchi T, Sato T *et al*. Assessing the discriminability of PCR-based open reading frame typing versus single-nucleotide polymorphism analysis via draft whole-genome sequencing of methicillin-resistant *Staphylococcus aureus* in nosocomial transmission analysis. *J Infect Chemother* 2024; **30**: 951–4. https://doi.org/10.1016/j.jiac.2024.02.021

7 Neoh H-M, Tan X-E, Sapri HF *et al*. Pulsed-field gel electrophoresis (PFGE): a review of the 'gold standard' for bacteria typing and current alternatives. *Infect Genet Evol* 2019; **74**: 103935. https://doi.org/10.1016/j.meegid.2019.103935

8 Kao CM, Fritz SA. Infection prevention-how can we prevent transmission of community-onset methicillin-resistant *Staphylococcus aureus*? *Clin Microbiol Infect* 2024; **31**: 166–72. https://doi.org/10.1016/j.cmi.2024.01.004

9 Uelze L, Grützke J, Borowiak M *et al*. Typing methods based on whole genome sequencing data. *One Health Outlook* 2020; **2**: 3. https://doi.org/10.1186/s42522-020-0010-1

10 Popovich KJ, Snitkin ES. Whole genome sequencing-implications for infection prevention and outbreak investigations. *Curr Infect Dis Rep* 2017; **19**: 15. https://doi.org/10.1007/s11908-017-0570-0

11 Popovich KJ, Aureden K, Ham DC *et al*. SHEA/IDSA/APIC practice recommendation: strategies to prevent methicillin-resistant *Staphylococcus aureus* transmission and infection in acute-care hospitals: 2022 update. *Infect Control Hosp Epidemiol* 2023; **44**: 1039–67. https://doi.org/10.1017/ice.2023.102

12 Popovich KJ, Green SJ, Okamoto K *et al*. MRSA transmission in intensive care units: genomic analysis of patients, their environments, and healthcare workers. *Clin Infect Dis* 2021; **72**: 1879–87. https://doi.org/10.1093/cid/ciaa731

13 Cheng VCC, Wong S-C, Cao H *et al*. Whole-genome sequencing data-based modeling for the investigation of an outbreak of community-associated methicillin-resistant *Staphylococcus aureus* in a neonatal intensive care unit in Hong Kong. *Eur J Clin Microbiol Infect Dis* 2019; **38**: 563–73. https://doi.org/10.1007/s10096-018-03458-y

14 Bartels MD, Petersen A, Worning P *et al*. Comparing whole-genome sequencing with Sanger sequencing for spa typing of methicillin-resistant *Staphylococcus aureus*. *J Clin Microbiol* 2014; **52**: 4305–8. https://doi.org/10.1128/JCM.01979-14

15 Stevens EL, Carleton HA, Beal J *et al*. Use of whole genome sequencing by the Federal Interagency Collaboration for Genomics for Food and Feed Safety in the United States. *J Food Prot* 2022; **85**: 755–72. https://doi.org/10.4315/JFP-21-437

16 Köser CU, Holden MTG, Ellington MJ *et al*. Rapid whole-genome sequencing for investigation of a neonatal MRSA outbreak. *N Engl J Med* 2012; **366**: 2267–75. https://doi.org/10.1056/NEJMoa1109910

17 Coll F, Raven KE, Knight GM *et al*. Definition of a genetic relatedness cutoff to exclude recent transmission of meticillin-resistant *Staphylococcus aureus*: a genomic epidemiology analysis. *Lancet Microbe* 2020; **1**: e328–35. https://doi.org/10.1016/S2666-5247(20)30149-X

18 Blane B, Raven KE, Brown NM *et al*. Evaluating the impact of genomic epidemiology of methicillin-resistant *Staphylococcus aureus* (MRSA) on hospital infection prevention and control decisions. *Microb Genom* 2024; **10**: 001235. https://doi.org/10.1099/mgen.0.001235

19 Kubota KA, Wolfgang WJ, Baker DJ *et al*. PulseNet and the changing paradigm of laboratory-based surveillance for foodborne diseases. *Public Health Rep* 2019; **134**: 22S–8S. https://doi.org/10.1177/0033354919881650

20 Michelacci V, Asséré A, Cacciò S *et al*. European Union Reference Laboratories support the National Food, Feed and Veterinary Reference Laboratories with rolling out whole genome sequencing in Europe. *Microb Genom* 2023; **9**: mgen001074. https://doi.org/10.1099/mgen.0.001074

21 Blane B, Raven KE, Leek D *et al*. Rapid sequencing of MRSA direct from clinical plates in a routine microbiology laboratory. *J Antimicrob Chemother* 2019; **74**: 2153–6. https://doi.org/10.1093/jac/dkz170

22 Madera S, McNeil N, Serpa PH *et al*. Prolonged silent carriage, genomic virulence potential and transmission between staff and patients characterize a neonatal intensive care unit (NICU) outbreak of methicillin-resistant *Staphylococcus aureus* (MRSA). *Infect Control Hosp Epidemiol* 2023; **44**: 40–6. https://doi.org/10.1017/ice.2022.48

23 Reuter S, Ellington MJ, Cartwright EJP *et al*. Rapid bacterial whole-genome sequencing to enhance diagnostic and public health microbiology. *JAMA Intern Med* 2013; **173**: 1397–404. https://doi.org/10.1001/jamainternmed.2013.7734

24 Wang Y, Zhao Y, Bollas A *et al*. Nanopore sequencing technology, bioinformatics and applications. *Nat Biotechnol* 2021; **39**: 1348–65. https://doi.org/10.1038/s41587-021-01108-x

25 Bogaerts B, Van den Bossche A, Verhaegen B *et al*. Closing the gap: Oxford Nanopore Technologies R10 sequencing allows comparable results to Illumina sequencing for SNP-based outbreak investigation of bacterial pathogens. *J Clin Microbiol* 2024; **62**: e0157623. https://doi.org/10.1128/jcm.01576-23

26 Mostafa HH. An evolution of Nanopore next-generation sequencing technology: implications for medical microbiology and public health. *J Clin Microbiol* 2024; **62**: e0024624. https://doi.org/10.1128/jcm.00246-24

27 Ni Y, Liu X, Simeneh ZM *et al*. Benchmarking of Nanopore R10.4 and R9.4.1 flow cells in single-cell whole-genome amplification and whole-genome shotgun sequencing. *Comput Struct Biotechnol J* 2023; **21**: 2352–64. https://doi.org/10.1016/j.csbj.2023.03.038

28 Raven KE, Bragin E, Blane B *et al*. Large-scale evaluation of a rapid fully automated analysis platform to detect and refute outbreaks based on MRSA genome comparisons. *mSphere* 2022; **7**: e0028322. https://doi.org/10.1128/msphere.00283-22

29 Brown NM, Blane B, Raven KE *et al*. Pilot evaluation of a fully automated bioinformatics system for analysis of methicillin-resistant *Staphylococcus aureus* genomes and detection of outbreaks. *J Clin Microbiol* 2019; **57**: e00858-19. https://doi.org/10.1128/JCM.00858-19

30 Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 2014; **30**: 2114–20. https://doi.org/10.1093/bioinformatics/btu170

31 Bankevich A, Nurk S, Antipov D *et al*. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 2012; **19**: 455–77. https://doi.org/10.1089/cmb.2012.0021

32 De Coster W, D'Hert S, Schultz DT *et al*. NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics* 2018; **34**: 2666–9. https://doi.org/10.1093/bioinformatics/bty149

33 Shen W, Le S, Li Y *et al*. SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PLoS One* 2016; **11**: e0163962. https://doi.org/10.1371/journal.pone.0163962

34 Li Y, Chen E, Xu J *et al*. Repeat and haplotype aware error correction in nanopore sequencing reads with DeChat. *Commun Biol* 2024; **7**: 1678. https://doi.org/10.1038/s42003-024-07376-y

35 Kolmogorov M, Yuan J, Lin Y *et al*. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol* 2019; **37**: 540–6. https://doi.org/10.1038/s41587-019-0072-8

36 Larsen MV, Cosentino S, Lukjancenko O *et al*. Benchmarking of methods for genomic taxonomy. *J Clin Microbiol* 2014; **52**: 1529–39. https://doi.org/10.1128/JCM.02981-13

37 Jain C, Rodriguez-R LM, Phillippy AM *et al*. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat Commun* 2018; **9**: 5114. https://doi.org/10.1038/s41467-018-07641-9

38 Richter M, Rosselló-Móra R. Shifting the genomic gold standard for the prokaryotic species definition. *Proc Natl Acad Sci U S A* 2009; **106**: 19126–31. https://doi.org/10.1073/pnas.0906412106

39 Bortolaia V, Kaas RS, Ruppe E *et al*. ResFinder 4.0 for predictions of phenotypes from genotypes. *J Antimicrob Chemother* 2020; **75**: 3491–500. https://doi.org/10.1093/jac/dkaa345

40 Kaya H, Hasman H, Larsen J *et al*. SCCmecFinder, a web-based tool for typing of Staphylococcal cassette chromosome mec in *Staphylococcus aureus* using whole-genome sequence data. *mSphere* 2018; **3**: e00612-17. https://doi.org/10.1128/mSphere.00612-17

41 Malberg Tetzschner AM, Johnson JR, Johnston BD *et al*. In silico genotyping of *Escherichia coli* isolates for extraintestinal virulence genes by use of whole-genome sequencing data. *J Clin Microbiol* 2020; **58**: e01269-20. https://doi.org/10.1128/JCM.01269-20

42 Lakhundi S, Zhang K. Methicillin-resistant *Staphylococcus aureus*: molecular characterization, evolution, and epidemiology. *Clin Microbiol Rev* 2018; **31**: e00020-18. https://doi.org/10.1128/CMR.00020-18

43 Gupta A, Jordan IK, Rishishwar L. stringMLST: a fast k-mer based tool for multilocus sequence typing. *Bioinformatics* 2017; **33**: 119–21. https://doi.org/10.1093/bioinformatics/btw586

44 Li H, Handsaker B, Wysoker A *et al*. The sequence alignment/map format and SAMtools. *Bioinformatics* 2009; **25**: 2078–9. https://doi.org/10.1093/bioinformatics/btp352

45 Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009; **25**: 1754–60. https://doi.org/10.1093/bioinformatics/btp324

46 Koboldt DC, Zhang Q, Larson DE *et al*. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res* 2012; **22**: 568–76. https://doi.org/10.1101/gr.129684.111

47 Didelot X, Wilson DJ. ClonalFrameML: efficient inference of recombination in whole bacterial genomes. *PLoS Comput Biol* 2015; **11**: e1004041. https://doi.org/10.1371/journal.pcbi.1004041

48 Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 2014; **30**: 1312–3. https://doi.org/10.1093/bioinformatics/btu033

49 Ferreira FA, Helmersen K, Visnovska T *et al*. Rapid nanopore-based DNA sequencing protocol of antibiotic-resistant bacteria for use in surveillance and outbreak investigation. *Microb Genom* 2021; **7**: 000557. https://doi.org/10.1099/mgen.0.000557

50 Worley JN, Crothers JW, Wolfgang WJ *et al*. Prospective genomic surveillance reveals cryptic MRSA outbreaks with local to international origins among NICU patients. *J Clin Microbiol* 2023; **61**: e0001423. https://doi.org/10.1128/jcm.00014-23

51 Sanderson ND, Hopkins KMV, Colpus M *et al*. Evaluation of the accuracy of bacterial genome reconstruction with Oxford Nanopore R10.4.1 long-read-only sequencing. *Microb Genom* 2024; **10**: 001246. https://doi.org/10.1099/mgen.0.001246

52 Ye L, Liu X, Ni Y *et al*. Comprehensive genomic and plasmid characterization of multidrug-resistant bacterial strains by R10.4.1 nanopore sequencing. *Microbiol Res* 2024; **283**: 127666. https://doi.org/10.1016/j.micres.2024.127666