Antimicrobial Resistance & Infection Control

RESEARCH

Open Access

Clinical utility and limitations of whole-genome sequencing via nanopore technology and Bayesian modeling for outbreak transmission inference

Yuji Fujikura[1,2*], Nozomi Ichie[2] and Takaaki Hamamoto[3]

Abstract

Background Traditional molecular epidemiological analysis has been limited in its ability to evaluate pathogen homology, making detailed transmission pathway estimation in outbreaks challenging. Bayesian reconstruction methods using whole-genome sequencing and epidemiological data enable estimation of temporal transmission relationships from within-host diversity. This study examined the practical utility and limitations of transmission pathway inference via whole-genome analysis with nanopore sequencers for SARS-CoV-2 and multidrug-resistant Acinetobacter (MDRA) outbreaks.

Methods Two outbreaks at the National Defense Medical College Hospital were analyzed: COVID-19 (9 individuals) and MDRA (22 cases). Transmission pathways were estimated via BadTrIP, an open-source Bayesian package for reconstructing transmission histories from genomic polymorphisms, with whole-genome sequencing performed on an Oxford Nanopore MinION. Input data included sample collection times, infection exposure periods, and nucleotide counts; SARS-CoV-2 genomes were obtained directly from nasopharyngeal swabs, whereas preserved bacterial isolates were used for MDRA.

Results During the SARS-CoV-2 outbreak, one patient and eight staff members were infected within one week, and BadTrIP inferred transmission from the index patient to multiple healthcare workers, followed by onward spread mainly among staff, consistent with descriptive epidemiology. The inferred transmission tree highlighted staff providing intensive oral care to the index patient and identified a staff member who intermittently worked unmasked, aligning with known infection prevention considerations, although mask effectiveness could not be formally quantified. In the MDRA outbreak, the inferred transmission network showed several high-probability links between patients on different wards with limited apparent contact, indicating discrepancies with ward-based spread, and spreader analysis identified only longer hospitalized duration as significantly associated with spreader status.

Conclusions BadTrIP is useful for transmission pathway estimation, although its accuracy depends on pathogen characteristics, outbreak duration, and sample properties. High concordance with clinical and epidemiological

*Correspondence:
Yuji Fujikura
fujikura.yuji@kitasato-u.ac.jp

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

information was achieved for the short-term SARS-CoV-2 outbreak using directly sequenced clinical specimens, whereas the prolonged MDRA outbreak based on preserved isolates showed limited accuracy and epidemiologically difficult-to-explain links. BadTrIP therefore appears most informative for short- to intermediate-term outbreaks with sufficient genetic diversity and direct sampling, while transmission trees from long-term outbreaks using preserved isolates should be carefully interpreted as exploratory.

Keywords Whole genome sequencing, Outbreak investigation, Bayesian inference, Transmission pathway, SARS-CoV-2, Multidrug-resistant Acinetobacter

Introduction

When healthcare-associated outbreaks occur, infection control departments estimate infection sources, transmission pathways, and risk factors from a descriptive epidemiological perspective, evaluate infection control measures, develop new strategies, and work to achieve early outbreak resolution. In this context, molecular epidemiological analysis of microorganisms, particularly bacteria, involves evaluating strain homology via pulsed-field gel electrophoresis [1], rep-PCR [2], POT methods [3], and similar techniques for particular species [4]. However, since the same clone typically spreads in most outbreaks, the impact of homology evaluation via these methods on epidemiological assessment is limited.

Recently, whole-genome sequencing has become more accessible, and attempts to use it in epidemiological investigations of outbreaks have been reported [5–14]. Specifically, these efforts aim to estimate temporal transmission relationships, rather than only strain homology, by inferring evolutionary, mutational, and derivational relationships among pathogens from genomic sequence changes using Bayesian outbreak reconstruction methods [5, 7–9, 11]. Whole-genome analysis thus has the potential to significantly improve the accuracy of infection pathway estimation beyond the limitations of descriptive epidemiology. Methods have been developed to create transmission pathway diagrams by estimating genome evolution through Bayesian reconstruction based on within-host diversity, and these methods have been applied to some outbreaks. Among these approaches, BadTrIP (BAyesian epiDemiological TRansmission Inference from Polymorphisms) [11] is a Bayesian method that integrates epidemiological data (sampling times and exposure periods) with genomic data, including within-host variants, to infer transmission linkages and timing. Instead of relying solely on consensus sequences or pairwise genetic distances, BadTrIP models pathogen populations within each host and incorporates transmission bottlenecks and sequencing error into the transmission inference process.

This study applies this method to SARS-CoV-2 and multidrug-resistant *Acinetobacter* (MDRA), two pathogen types (virus and bacterium), to assess whether transmission pathway diagrams consistent with descriptive epidemiology can be generated in clinical practice and to clarify the limitations of this approach. We used portable, relatively low-cost nanopore sequencers for on-site whole-genome sequencing, constructed transmission pathway diagrams, and evaluated their practical utility and limitations in routine outbreak investigations.

Materials and methods

Study setting and overview

The creation of a transmission pathway prediction diagram for SARS-CoV-2 and MDRA was conducted at the National Defense Medical College Hospital, a tertiary emergency medical facility with 800 beds. The study targeted a COVID-19 outbreak involving nine individuals in the same ward/department in April 2022 and an MDRA outbreak involving 22 cases from 2012 to 2014. Basic attributes, diseases, comorbidities, and medical care information were collected from patients and staff from whom the target microorganisms were detected in these two outbreaks via electronic medical records and interviews. The periods and times when the samples were collected that were considered to contribute to pathogen transmission were also identified. Clinical specimens obtained from each pathogen were processed as described below, and whole-genome analysis was performed via Oxford Nanopore Technologies (ONT) MinION. The study was planned and conducted in accordance with the Declaration of Helsinki and approved by the institutional ethics committee (No. 4456, 5096).

BadTrIP application

An open-source BEAST2 [15] package for Bayesian inference of epidemiological transmission from genomic polymorphisms, BadTrIP [11] was used to estimate pathogen transmission pathways. BadTrIP combines epidemiological data with within-host pathogen evolution and transmission models, requiring three input values: (1) the time when the sample was collected, (2) a time interval for each host describing when it can contribute to the outbreak (infection exposure period), and (3) genetic data from each sample in the form of nucleotide counts for each genome position. When nucleotide counts are unavailable, samples can be added to the analysis by inputting only epidemiological data.

Epidemiological information collection

Electronic medical records were referenced for patients involved in outbreaks, and the time when positive samples were collected was designated the sampling time.

Different criteria were used for the infection exposure period contributing to outbreaks of SARS-CoV-2 versus MDRA. For SARS-CoV-2, we set the introduction time to day − 4 relative to symptom onset (day 0), based on an estimated incubation period of 3.4 days for the Omicron variant [16]. The removal time was defined as the day of absence from work or isolation, and this interval between introduction and removal was used as the infection exposure period.

For MDRA, the introduction time could not be specified based on clinical onset because symptoms were often unclear and many cases were detected through active screening; therefore, we defined the introduction time as the day of hospital admission. Although contact precautions were implemented upon detection, many patients continued to receive intensive care, and infections continued to occur despite these measures, suggesting some degree of infection control failure. Because specific failing patients or situations could not be identified, the removal time was set at hospital discharge, and the infection exposure period for each patient was defined as the interval from admission to discharge. As the MDRA outbreak extended over a long period, BadTrIP was run using weekly rather than daily time steps to reduce computational burden.

SARS-CoV-2 genomic data acquisition

RNA was extracted from PCR residual samples for COVID-19 diagnosis and analyzed via methods similar to those of the National Institute of Infectious Diseases SARS-CoV-2 genome analysis protocol version 1.6 [17]. The NEB ARTIC SARS-CoV-2 Companion Kit (E7660S/L) for Omicron and ONT Native Barcoding Expansion Kits 1–12 (EXP-NBD104) were used for SARS-CoV-2 analysis. Following the manufacturer's instructions, libraries were prepared via an ONT Ligation Sequencing Kit (SQK-LSK109) and an ONT Flow Cell Priming Kit (EXP-FLP002) and then sequenced via ONT MinION Mk1c with flow cells (R.9.4.1, FLO-MIN106D). Basecalling was performed using Guppy (version 4.5, Oxford Nanopore Technologies). Reads with a mean quality score < 10 were excluded from downstream analysis. For SARS-CoV-2, we restricted the analysis to genomic positions with a minimum depth of at least 100 reads.

The fastq data were mapped via SAMtools [18], and strain-specific SNPs and INDELs were identified via VCFtools [19]. For genes, we first identified sample-specific variant sites using VCFtools and then selected eight loci with high coverage (≥ 100×) and clear between-sample variation, while avoiding adjacent positions, to reduce the impact of obvious local linkage. The reference sequence used was Genome assembly ASM985889v3 (NCBI Datasets), and nucleotide counting was performed via IGV ver. 2.14.1 [20].

MDRA genomic data acquisition

The MDRAs were identified from clinical samples via a Microscan Walkaway 96 Plus (Beckman Coulter) with a Neg Combo Panel (Neg Combo 6.1 J). According to Japan's Ministry of Health, Labor and Welfare, MDRA is defined by resistance to three antimicrobial classes: carbapenems, fluoroquinolones, and aminoglycosides [21]. The MDRA isolated during the outbreak was stored in microbanks (IWAKI & Co., Ltd.), and preserved isolates were used in this study. The MDRA isolates were cultured for 24 h on sheep-blood agar plates, and genomic DNA was extracted with MagLEAD 12gC (Precision System Science).

For whole-genome sequencing, DNA libraries were prepared via the ONT Rapid Barcoding Sequencing Kit (SQK-RBK004) following the manufacturer's instructions and then sequenced via ONT MinION Mk1b and Mk1c with flow cells (R.9.4.1, FLO-MIN106D). Basecalling was performed using Guppy (version 4.5, Oxford Nanopore Technologies) and reads with a mean quality score < 10 were excluded from downstream analysis. Because overall coverage was lower for MDRA than for SARS-CoV-2, we selected genomic sites with a SNP quality greater than 100 and an average depth of approximately 20 reads and used only these high-confidence positions for analysis.

For MDRA, sequence types were determined from fastq data obtained via Krocus [22]. Next, as shown in the supplement, mapping was performed via SAMtools [18], and strain-specific SNPs and INDELs were identified via VCFtools [19]. For genes, we identified variant sites using bcftools and selected six loci with high SNP quality and average depth, again avoiding adjacent positions to limit local linkage. The reference sequence usedwas Genome assembly ASM903584v1 (NCBI Datasets), and nucleotide counting was performed via IGV ver. 2.14.1 [20].

Statistical analysis

The chi-square test was used to analyze nonparametric variables, and Fisher's exact test was used when sample numbers were small. Multiple logistic regression was used to estimate risk for multivariate analysis. $p < 0.05$ was considered significant. All the statistical analyses were performed via GraphPad Prism 10.4.2 for Windows (GraphPad Software, La Jolla, California, USA).

For the MDRA outbreak, we operationally defined "spreaders" as cases with more than one outgoing transmission link in the BadTrIP-inferred transmission tree

and compared their clinical characteristics with those of "nonspreaders" using univariate analyses.

Results

SARS-CoV-2 transmission pathway prediction

In April 2022, multiple medical staff working in the same ward developed fever and tested positive for SARS-CoV-2, ultimately leading to detection in 9 individuals, including one patient in the same ward. Positive confirmations were concentrated within one week, indicating that an episode was caused by the same source virus.

Viral nucleic acid extraction was performed from nasopharyngeal swabs, but genome collection was impossible for one of the nine individuals who tested positive only on rapid diagnostic tests. This individual (sample HH8) was entered with a zero genome count on the basis of BadTrIP guidance.

Figure 1 shows a Gantt chart of positive case attributes, onset dates, and isolation dates constituting this outbreak.

Figure 2 shows the transmission prediction diagram created by BadTrIP based on these data. In summary, BadTrIP inferred that the index patient (sample HH2) most likely transmitted SARS-CoV-2 to multiple healthcare workers, with subsequent onward transmission occurring mainly among staff members.

MDRA transmission pathway prediction

MDRA was first detected in August 2012 in a patient who was treated for multiple myeloma, after which MDRA was detected successively in other wards. Detection included both attending physician-ordered culture tests and screening. Despite patient isolation and strict contact precautions, MDRA was detected intermittently hospital-wide, with 23 positive cases identified before resolution. Previous PFGE analysis confirmed the homology among 22 of these isolates [23]. These 22 isolates were the targets of this analysis, with background information for the 22 positive cases shown in Table 1. The detected MDRA was identified as sequence type 2 by Krocus.

Figure 3 shows the transmission prediction diagram created by BadTrIP based on these data. Overall, the inferred MDRA transmission network showed multiple patients acting as spreaders - patients transmitting infection to multiple other patients - however several high-probability links connected cases across different wards with limited apparent epidemiological contact, highlighting inconsistencies between the genomic transmission tree and the expected ward-based spread.

Furthermore, univariate testing results comparing spreader and nonspreader patient backgrounds are shown in Table 2; Fig. 4. The results revealed that only

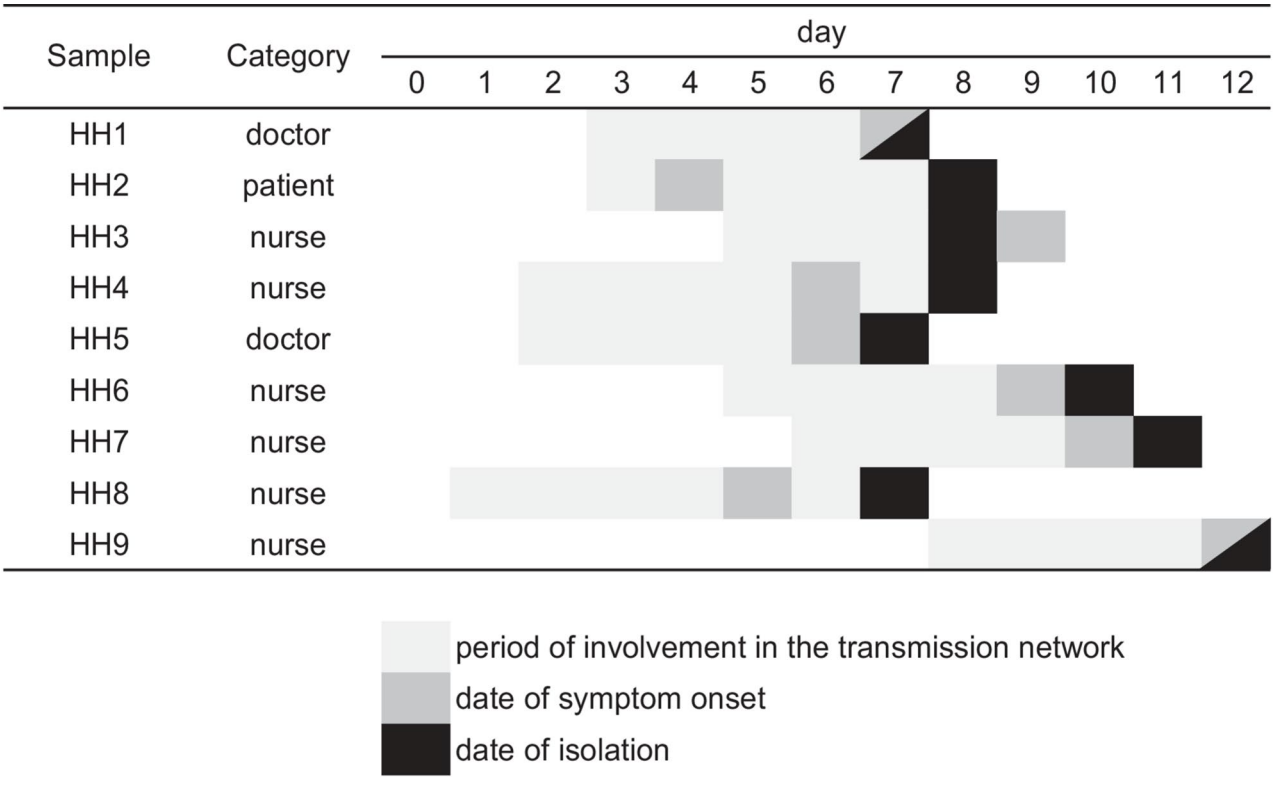

Fig. 1 Gantt chart showing positive case attributes, onset dates, and isolation dates during the SARS-CoV-2 outbreak

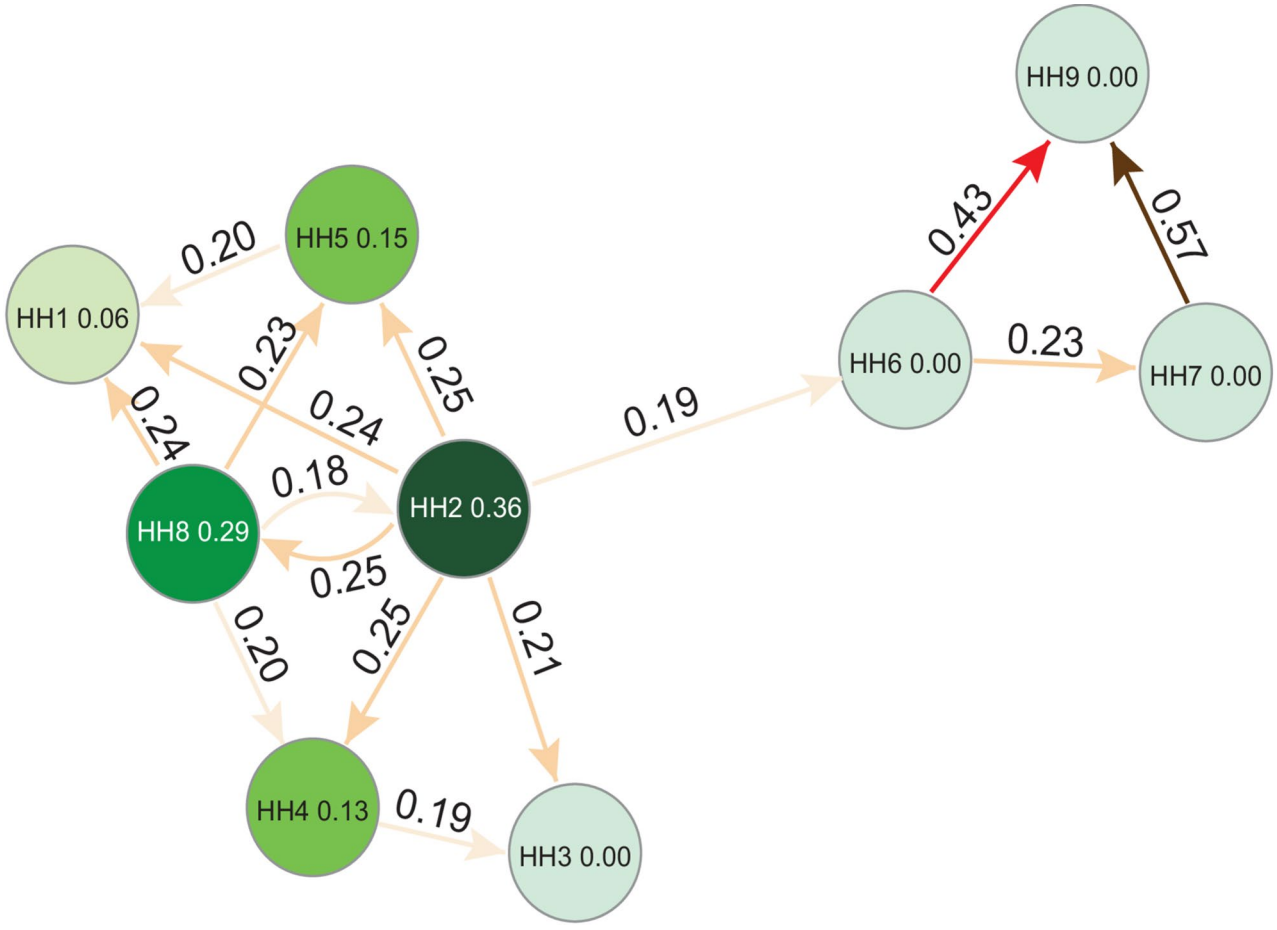

Fig. 2 Reconstruction of the transmission tree for the SARS-CoV-2 outbreak, showing high-probability transmissions inferred by BadTrIP
Transmission events with posterior probabilities higher than 18%, as inferred by BadTrIP. The numbers adjacent to the arrows represent posterior probabilities. Arrow colors correspond to numbers, with low probabilities shown in light colors and high probabilities in dark red. The numbers written in circles with sample numbers represent root probabilities, corresponding to circle colors (light green to dark green)

hospitalized duration was significantly associated with spreader status.

Discussion

In this study, we applied whole-genome sequencing using nanopore technology combined with BadTrIP to reconstruct transmission pathways during two distinct outbreaks: a short-term SARS-CoV-2 outbreak and a prolonged MDRA outbreak. In the SARS-CoV-2 outbreak, analysis of viral genomes directly obtained from clinical specimens enabled reconstruction of transmission pathways consistent with conventional descriptive epidemiology, demonstrating the practical utility of this approach for short-term outbreaks. In contrast, in the MDRA outbreak, reliance on stored isolates and the prolonged duration of the outbreak reduced genomic certainty and inferential precision within BadTrIP, resulting in several inferred transmission pathways that were difficult to reconcile with epidemiological links. These findings indicate that while this combined approach is effective for short-term outbreaks, cautious interpretation is warranted in long-term outbreaks and when only preserved isolates are available.

BadTrIP was deemed suitable for clinical use because it considers sequencing errors and unobserved intermediate hosts [11], enabling complementing descriptive epidemiological information and potentially contribute to outbreak reviews. During the SARS-CoV-2 outbreak, transmission from index-positive cases to multiple healthcare workers was suggested, with further spread among healthcare workers. This positive case required intensive medical care, particularly oral care [24], indicating high transmission risk and supporting recommendations to enhance infection protection for patients requiring intensive oral care. At the time of the outbreak, universal mask use was strongly recommended in all inpatient areas, and interviews indicated that almost all staff reported wearing masks consistently during patient care and when interacting with colleagues. Staff member HH1, who reported working without a mask during part

Table 1 Patient characteristics and medical care

1	Age, Sex	Week number of admission	Week numbers of exposure duration	Week number of sampling	Ward	Disease category	Sample	Mechanical ventilation	Tube feeding	bedridden	Urinary catheter	Hemo-dialysis	Infection†	Carbapenem usage at the time of sampling
	67 M	**4**	**4–17**	**6**	**A**	**Hematology**	**sputum**	**+**		**+**	**+**			**+**
2	77 F	10	10–13	12	B	ENT	blood		+		+		+	+
3	46 M	12	12–18	13	C	Neurosurgery	CSF	+	+	+	+		+	
4	62 M	0	0–14	13	B	General medicine	urine	+	+	+	+		+	+
5	55 M	7	7–15	15	C	Neurosurgery	Urine			+				
6	65 M	13	13–31	24	C	Neurosurgery	sputum		+	+	+	+		
7	76 M	24	24–31	25	C	Neurology	Sputum	+	+	+	+		+	
8	62 M	22	22–26	25	C	Neurosurgery	Sputum	+	+	+	+		+	+
9	66 M	19	19–31	30	C	Neurosurgery	Sputum		+	+	+			
10	91 F	25	25–37	32	C	Neurosurgery	Sputum		+	+	+			
11	69 F	27	27–55	43	C	General medicine	Sputum		+	+	+		+	
12	77 M	40	40–75	43	C→D	Neurosurgery	Sputum	+	+	+	+		+	
13	76 M	29	29–60	45	D	Cardiology	Sputum		+	+		+	+	
14	66 M	43	43–51	51	A	Hematology	Sputum	+		+	+		+	+
15	58 M	52	52–63	56	B	Plastic surgery	wound					+		
16	68 M	61	61–67	63	C	Neurosurgery	Sputum	+	+	+	+		+	
17	48 F	64	64–67	65	C	General medicine	Sputum				+		+	+
18	74 F	60	60–67	66	C	Neurosurgery	Sputum		+	+	+			
19	59 M	60	60–80	67	E	Hepatology	Sputum	+	+	+	+	+	+	+
20	90 M	62	62–71	68	B	Plastic surgery	Sputum		+	+				
21	60 M	63	63–73	70	F	Respiratory	Sputum			+	+		+	+
22	72 M	65	65–76	74	F	Respiratory	Sputum						+	

ENT ear-nose-throat, *CSF* cerebrospinal fluid

†We classified 'infection' as positive (+) when *Acinetobacter* was detected in patients who were receiving antimicrobial therapy for a suspected infection at the time of detection

of the relevant period, therefore represents an exceptional episode that is consistent with the known protective effect of masks against SARS-CoV-2 transmission [25, 26]. However, HH1 was not classified as a spreader, and we did not systematically quantify mask-wearing adherence for all staff; thus, our data do not allow formal estimation of mask effects in this outbreak. Rather, the validity of the inferred transmission pathways is supported primarily by the overall consistency between the BadTrIP transmission tree and the observed temporal and contact patterns on the ward.

In this SARS-CoV-2 outbreak, we directly analyzed viral RNA extracted from clinical nasopharyngeal specimens, which likely reflected the within-host viral population at the time of infection and thus provided genomically reliable input for BadTrIP. Because the outbreak was resolved within a short period, the temporal structure of sampling and exposure windows was well defined, and the inferred transmission tree closely matched known epidemiological contact patterns. This approach reduces reliance on lengthy interviews and exhaustive manual review of patient and contact data, and can rapidly visualize likely transmission pathways, support post-hoc evaluation of infection prevention measures, and identify previously unrecognized epidemiological links. In practical terms, the unique contribution of this study was not to identify oral care or close staff contact as risk factors - these were already apparent from conventional outbreak investigation - but rather to refine and prioritize the likely person-to-person links within the outbreak. By assigning posterior probabilities to each potential transmission route, the method helped to highlight which healthcare workers were most likely centrally positioned in the transmission network. In this way, BadTrIP complements, rather than replaces, traditional descriptive epidemiology.

In contrast, the MDRA outbreak analysis highlighted important limitations of this approach. Because the genomic data were derived from stored isolates that had undergone freezing, thawing, and subculture, the recovered genomes may not have perfectly reflected the bacterial populations present in patients at the time of transmission, thereby reducing genomic certainty. Moreover, the outbreak extended over a long period (more than a year), which is expected to lower the precision of the posterior transmission probabilities in BadTrIP.

Consistent with these methodological challenges, we observed several correlations in which patients on different wards, with little contact, were inferred to have higher posterior probabilities than patients on the same ward with stronger spatiotemporal linkage. Such findings are difficult to explain from a descriptive epidemiology and indicate that, in long term bacterial outbreaks based on preserved isolates, BadTrIP derived transmission trees should be interpreted exploratory with great cautions. Because this was a retrospective analysis, we did not resequence the same MDRA isolates before and after storage and subculture, and we therefore cannot quantify the impact of these procedures on the observed variation. Although only limited additional mutations are generally expected under appropriate storage conditions [27, 28], even rare storage- or passage-related changes could alter inferred relationships between closely related isolates in a prolonged outbreak. Accordingly, the MDRA transmission tree in this study should be regarded as hypothesis generating and interpreted together with detailed epidemiological information.

We also analyzed factors contributing to transmission from specific patients. Multiple factors, including frequent care, tracheal tube/mechanical ventilation, previous antibiotic use and tube feeding, are generally identified as risk factors for the spread of infection [29, 30]. However, in this study there were no significant differences in these factors between spreaders and non-spreaders, and only longerhospitalized duration differed between the groups. Because patient isolation and strict contact precautions were implemented [31] upon detection, transmission-related risks were probably lower than before these measures, and this change in risk over time could not be fully captured by the BadTrIP. From a clinical perspective, patients who remain hospitalized for extended periods inevitably accumulate more opportunities for contact with healthcare workers, shared equipment and the hospital environment, which may not be fully reflected by standard covariates. This outbreak involved only a small number of cases, so the results in Table 2 should be regarded as hypothesis-generating rather than definitive. Nevertheless, for long-stay patients colonized or infected with resistant bacteria, enhanced infection control measures and thorough environmental and device disinfection around these patients may help to reduce transmission risk. However, this spreader classification is therefore exploratory and based solely on the structure of the inferred transmission tree, without incorporating behavioral factors, detailed contact patterns or adherence to infection control measures, which may also be important determinants of transmission risk. Future applications should aim to define spreaders with integration Bayesian reconstruction and infection control information.

BadTrIP characteristically considers within-host variation rather than using pathogen sample phylogenetic trees as proxies for transmission history, reconstructing infection pathways, including infection directionality and timing, through Bayesian approaches without the assumption of single haplotypes (consensus sequences) [10, 11]. For clinicians unfamiliar with Bayesian genomic models, BadTrIP can be viewed as a tool that infers “who

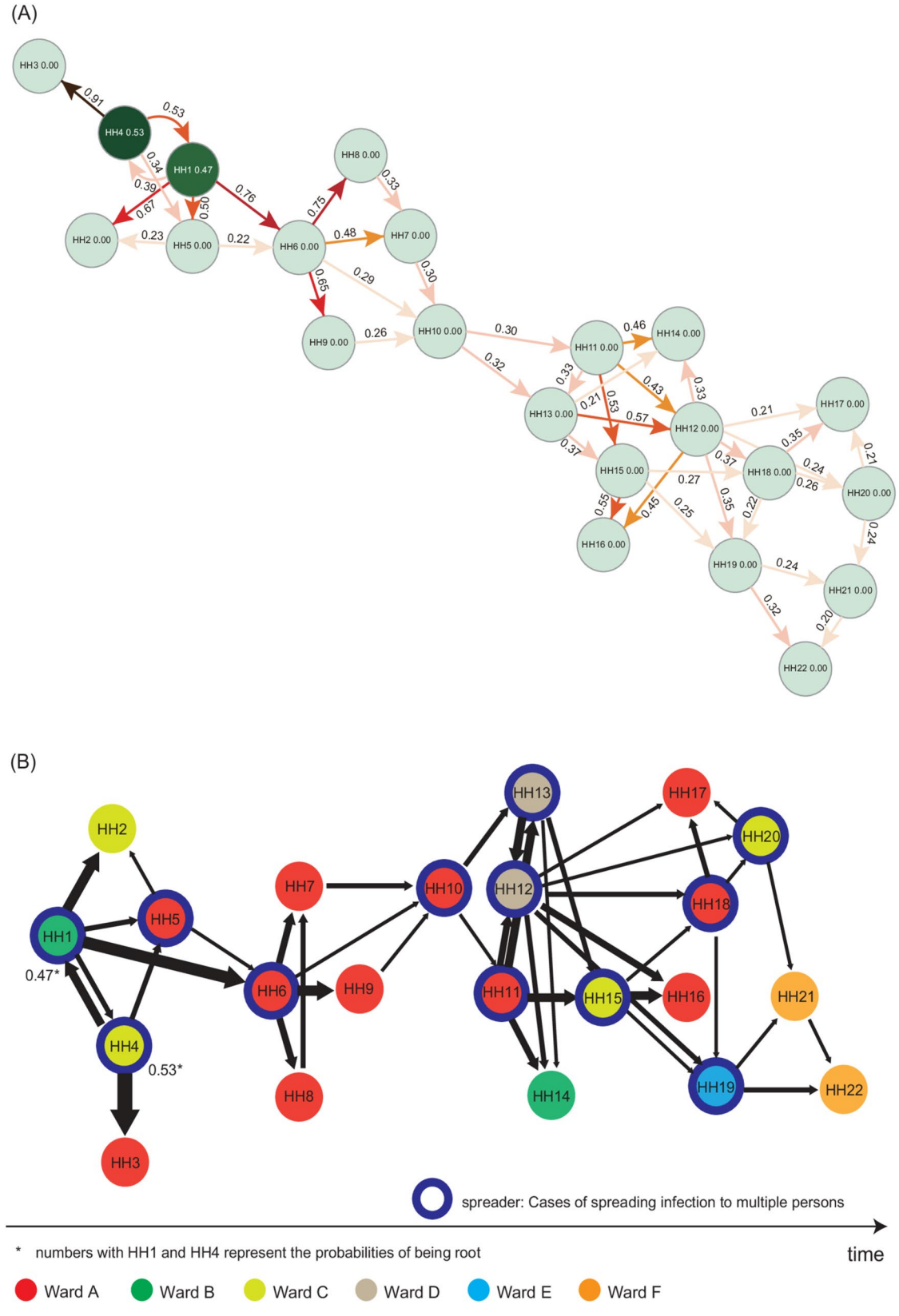

Fig. 3 Reconstruction of the transmission tree for the MDRA outbreak, showing high-probability transmissions inferred by BadTrIP
Transmission events with posterior probabilities higher than 20%, as inferred by BadTrIP. The notation follows that in Fig. 2. Figure 3(A) shows transmission pathways with ward information added to sample numbers. Posterior probability values are omitted and represented by arrow thickness, with root probabilities shown for HH1 and HH4. Cases with transmission arrows to multiple samples are defined as spreaders and are indicated in the figure

Table 2 Risk factors for spreaders

variables	Spreader (*n* = 12)	Non-spreader (*n* = 10)	*p* value	Odds Ratio [OR] (95%CI)/ Mean difference (95%CI)
mechanical ventilation	4	5	0.666	OR 0.50 (95%CI -2.4–1.0)
tube feeding	9	6	0.652	OR 2.0 (95%CI -1.1–2.5)
carbapenem usage	3	5	0.378	OR 0.33 (95%CI -2.9–0.70)
hospitalized duration (weeks)	18	8	0.005	Mean difference 10.0 ± 3.2 (95%CI 3.4–16.6)

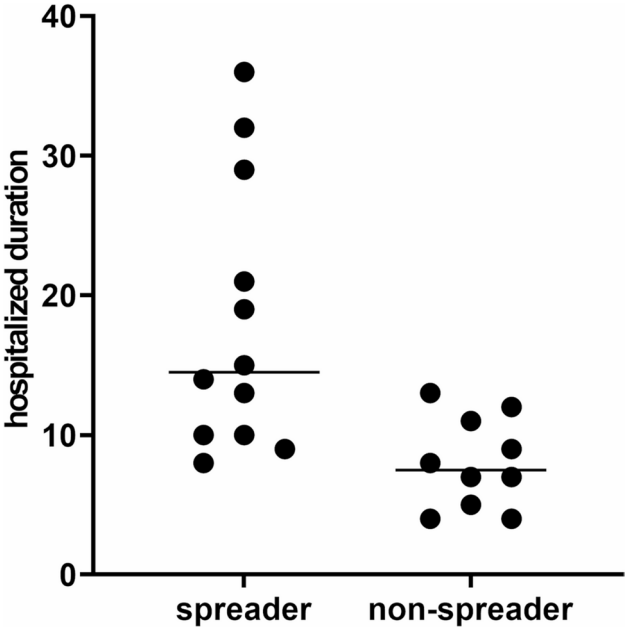

Fig. 4 Distribution of hospitalization duration between spreaders and nonspreaders

infected whom" by combining epidemiological data with genomic data. Instead of relying only on a single consensus sequence per patient, it also uses low-frequency within-host variants, which provide high-resolution to infer plausible transmission directions. For each possible infector–infectee pair, the model computes a probability that the link is compatible with both the epidemiological timelines and the observed genomic patterns; in our figures, darker arrows correspond to links with higher such probabilities. In practical terms, the posterior probability for a given arrow reflects how strongly the combined data supports that particular route compared with alternative pathways, rather than proving that transmission certainly occurred along that link.

While BadTrIP considers sequencing errors [11], the high sequencing error rates common with MinION [32] R9.4.1 platform and Guppy-based basecallings represent limitations for creating transmission diagrams via this method. Additionally, BadTrIP does not allow us to estimate the inoculum size - that is, how many pathogen units succeed in colonizing the next host at transmission (the so-called transmission bottleneck). Furthermore, BadTrIP is designed under the assumption that different genome positions are completely unlinked [11]. And, because only 6–8 high-confidence sites per outbreak were used as input in our dataset, and residual linkage between loci cannot be excluded. Investigating whether extracted genomic data sites are related was difficult, and while adjacent sites were avoided, the impact of genomic position selection on calculation results cannot be ruled out. Therefore, the inferred transmission trees should be interpreted as exploratory rather than definitive. Additionally, this method has other nonnegligible limitations. In our MDRA analysis, increasing the time horizon to cover more than a year and 22 cases substantially increased computation time over days even small genomic site.

In terms of practical use, our SARS-CoV-2 analysis should be regarded as a retrospective application that mainly illustrates the value of BadTrIP for outbreak review in short period, rather than as a real-time decision-making tool. In such settings, the outbreak may have already resolved by the time sequencing and analysis are completed, but a probabilistic transmission tree can still help future lessons and refine infection control strategies. On the other hand, our MDRA example demonstrates that reliance on preserved isolates, and very extended time frames could markedly reduce inference quality. Taken together, these findings suggest that BadTrIP is likely to be most informative in outbreaks of intermediate duration (weeks to a few months) with sufficient within-outbreak genetic diversity, direct sequencing from clinical specimens, and reasonably complete epidemiological data.

In conclusion, transmission pathway estimation in outbreaks using epidemiological information and whole-genome analysis has clinical utility, though with caveats. Specifically, accuracy is high and clinically useful for pathogens such as SARS-CoV-2 that spread over short periods and allow direct genome isolation from clinical specimens. However, analyses of long-term outbreaks such as MDRA or those using preserved isolates may not adequately ensure accuracy. Further development of methods to improve accuracy is needed, including bottleneck adjustment, genomic information selection for calculations, and computational methods. Epidemiological pathway estimation via such genomic analysis is expected to advance further, and establishing appropriate computational protocols will contribute to more accurate transmission inference and effective infection control.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13756-026-01741-8.

Supplementary Material 1.

Supplementary Material 2.

Acknowledgements
We thank all practitioners coordinating with the Department of Infection Control at National Defense Medical College Hospital. We deeply appreciate their contributions and the valuable role they play in preventing outbreaks.

Author contributions
YF analyzed the collected patient data and genomic data, performed data editing, and plotted transmission inference diagrams via open-source program packages. Furthermore, YF is the primary contributor to the writing of this paper. NI collected and analyzed individual epidemiological data related to the outbreak via electronic medical records and interviews. YF and TH extracted genomes from microorganisms and collected genomic data via MinION.

Funding
This work was supported by JSPS KAKENHI Grant Number JP21K17280.

Data availability
The datasets used and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Declarations

Ethics approval and consent to participate
The study was planned and conducted in accordance with the Declaration of Helsinki and approved by the institutional ethics committee (No. 4456, 5096).

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Department of Infectious Diseases, Kitasato University School of Medicine, 1-15-1 Kitazato, Minami-ku, Sagamihara, Kanagawa 252-0374, Japan
[2]Department of Medical Risk Management and Infection Control, National Defense Medical College Hospital, 3-2 Namiki, Tokorozawa, Saitama 359-8513, Japan
[3]Department of Clinical Laboratory, National Defense Medical College Hospital, 3-2 Namiki, Tokorozawa, Saitama 359-8513, Japan

Received: 31 July 2025 / Accepted: 26 March 2026
Published online: 02 April 2026

References

1. Tenover FC, Arbeit RD, Goering RV, Mickelsen PA, Murray BE, Persing DH, et al. Interpreting chromosomal DNA restriction patterns produced by pulsed-field gel electrophoresis: criteria for bacterial strain typing. J Clin Microbiol. 1995;33(9):2233–9.
2. Snelling AM, Gerner-Smidt P, Hawkey PM, Heritage J, Parnell P, Porter C, et al. Validation of use of whole-cell repetitive extragenic palindromic sequence-based PCR (REP-PCR) for typing strains belonging to the Acinetobacter calcoaceticus-Acinetobacter baumannii complex and application of the method to the investigation of a hospital outbreak. J Clin Microbiol. 1996;34(5):1193–202.
3. Suzuki M, Hosoba E, Matsui M, Arakawa Y. New PCR-based open reading frame typing method for easy, rapid, and reliable identification of *Acinetobacter baumannii* international epidemic clones without performing multilocus sequence typing. J Clin Microbiol. 2014;52(8):2925–32.
4. Sabat AJ, Budimir A, Nashev D, Sa-Leao R, van Dijl J, Laurent F, et al. Overview of molecular typing methods for outbreak detection and epidemiological surveillance. Euro Surveill. 2013;18(4):20380.
5. Ypma RJ, van Ballegooijen WM, Wallinga J. Relating phylogenetic trees to transmission trees of infectious disease outbreaks. Genetics. 2013;195(3):1055–62.
6. Worby CJ, Lipsitch M, Hanage WP. Shared genomic variants: identification of transmission routes using pathogen deep-sequence data. Am J Epidemiol. 2017;186(10):1209–16.
7. Didelot X, Gardy J, Colijn C. Bayesian inference of infectious disease transmission from whole-genome sequence data. Mol Biol Evol. 2014;31(7):1869–79.
8. Mollentze N, Nel LH, Townsend S, le Roux K, Hampson K, Haydon DT, et al. A Bayesian approach for inferring the dynamics of partially observed endemic infectious diseases from space-time-genetic data. Proc Biol Sci. 2014;281(1782):20133251.
9. Jombart T, Cori A, Didelot X, Cauchemez S, Fraser C, Ferguson N. Bayesian reconstruction of disease outbreaks by combining epidemiologic and genomic data. PLoS Comput Biol. 2014;10(1):e1003457.
10. Worby CJ, Lipsitch M, Hanage WP. Within-host bacterial diversity hinders accurate reconstruction of transmission networks from genomic distance data. PLoS Comput Biol. 2014;10(3):e1003549.
11. De Maio N, Worby CJ, Wilson DJ, Stoesser N. Bayesian reconstruction of transmission within outbreaks using genomic variants. PLoS Comput Biol. 2018;14(4):e1006117.
12. Hall M, Woolhouse M, Rambaut A. Epidemic Reconstruction in a Phylogenetics Framework: Transmission Trees as Partitions of the Node Set. PLoS Comput Biol. 2015;11(12):e1004613.
13. Gaythorpe K, Morris A, Imai N, Stewart M, Freeman J, Choi M. Chainchecker: An application to visualise and explore transmission chains for Ebola virus disease. PLoS ONE. 2021;16(2):e0247002.
14. Didelot X, Kendall M, Xu Y, White PJ, McCarthy N. Genomic Epidemiology analysis of infectious disease outbreaks using transphylo. Curr Protoc. 2021;1(2):e60.
15. Bouckaert R, Heled J, Kuhnert D, Vaughan T, Wu CH, Xie D, et al. BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Comput Biol. 2014;10(4):e1003537.
16. Wu Y, Kang L, Guo Z, Liu J, Liu M, Liang W. Incubation Period of COVID-19 Caused by Unique SARS-CoV-2 Strains: A Systematic Review and Meta-analysis. JAMA Netw Open. 2022;5(8):e2228008.
17. SARS-CoV-. 2 Genome Analysis Protocol [https://id-info.jihs.go.jp/relevant/manual/010/SARS-CoV2_genome_analysis_manual_Nanopore_NEB_ver_1_6_220127.pdf].
18. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO et al. Twelve years of SAMtools and BCFtools. Gigascience 2021, 10(2).
19. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. Bioinf (Oxford England). 2011;27(15):2156–8.
20. Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nat Biotechnol. 2011;29(1):24–6.
21. [multidrug-resistant Acinetobacter infection] (in Japanese) [https://www.mhlw.go.jp/bunya/kenkou/kekkaku-kansenshou11/01-05-140912-4.html]
22. Page AJ, Keane JA. Rapid multi-locus sequence typing direct from uncorrected long reads using Krocus. PeerJ. 2018;6:e5233.
23. Fujikura Y, Yuki A, Hamamoto T, Ichimura S, Kawana A, Ohkusu K et al. Evaluation and validity of a polymerase chain reaction-based open reading frame typing method to dissect the molecular epidemiology for Acinetobacter baumannii in an epidemiologic study of a hospital outbreak. Am J Infect Control 2016.
24. Santigli E, Lindner M, Kessler HH, Jakse N, Fakheran O. Seroprevalence of SARS-CoV-2 antibodies among oral health care workers with natural seroconversion: a systematic review and meta-analysis. Sci Rep. 2025;15(1):7848.
25. Chu DK, Akl EA, Duda S, Solo K, Yaacoub S, Schunemann HJ, et al. Physical distancing, face masks, and eye protection to prevent person-to-person transmission of SARS-CoV-2 and COVID-19: a systematic review and meta-analysis. Lancet. 2020;395(10242):1973–87.
26. Greenhalgh T, MacIntyre CR, Baker MG, Bhattacharjee S, Chughtai AA, Fisman D, et al. Masks and respirators for prevention of respiratory infections: a state of the science review. Clin Microbiol Rev. 2024;37(2):e0012423.
27. Sprouffske K, Aguilar-Rodriguez J, Wagner A. How archiving by freezing affects the genome-scale diversity of *Escherichia coli* populations. Genome Biol Evol. 2016;8(5):1290–8.

28. Iversen XES, Norman A, Folkvardsen DB, Svensson E, Rasmussen EM, Lillebaek T. Successful direct whole genome sequencing and revivification of freeze-dried nontuberculous mycobacteria after more than half a century of storage. Microbiol Spectr. 2022;10(3):e0031022.
29. Fukuta Y, Muder RR, Agha ME, Clarke LG, Wagener MM, Hensler AM, et al. Risk factors for acquisition of multidrug-resistant *Acinetobacter baumannii* among cancer patients. Am J Infect Control. 2013;41(12):1249–52.
30. Thom KA, Rock C, Jackson SS, Johnson JK, Srinivasan A, Magder LS, et al. Factors Leading to Transmission Risk of Acinetobacter baumannii. Crit Care Med. 2017;45(7):e633–9.
31. Tacconelli E, Cataldo MA, Dancer SJ, De Angelis G, Falcone M, Frank U, et al. ESCMID guidelines for the management of the infection control measures to reduce transmission of multidrug-resistant Gram-negative bacteria in hospitalized patients. Clin Microbiol Infect. 2014;20(Suppl 1):1–55.
32. Zhang T, Li H, Ma S, Cao J, Liao H, Huang Q, et al. The newest Oxford Nanopore R10.4.1 full-length 16S rRNA sequencing enables the accurate resolution of species-level microbial community profiling. Appl Environ Microbiol. 2023;89(10):e0060523.

Publisher's Note