## RESEARCH

## **Open Access**



Rhys T. White<sup>1\*</sup>, Michelle Balm<sup>2,3</sup>, Megan Burton<sup>2</sup>, Samantha Hutton<sup>2</sup>, Jamaal Jeram<sup>2</sup>, Matthew Kelly<sup>3</sup>, Donia Macartney-Coxson<sup>1</sup>, Tanya Sinha<sup>2</sup>, Henrietta Sushames<sup>3</sup>, David J. Winter<sup>1</sup>, and Maxim G. Bloomfield<sup>2,3</sup>

## Abstract

**Background** *Klebsiella variicola* has been implicated in neonatal intensive care unit (NICU) outbreaks previously and can be misidentified as *Klebsiella pneumoniae*. An increased incidence of *K. pneumoniae* bacteremia on the NICU of our institution was notified to the infection prevention and control (IPC) team in May 2024. The four isolates involved displayed wild-type susceptibility, so had not been detected via multidrug-resistant organism surveillance. This triggered investigation with a nanopore-based decentralized whole-genome sequencing (dWGS) system in operation at our laboratory.

**Methods** Since early 2022, the hospital laboratory at Wellington Regional Hospital has been performing dWGS using the Oxford Nanopore MinION device. This allows for prospective genomic surveillance of certain hospital-associated organisms, but also rapid reactive investigation of possible outbreaks. Isolates are sequenced in the hospital laboratory and undergo multilocus sequence typing (MLST). If transmission events are suspected, sequence data are transferred to the reference laboratory, the Institute for Environmental Science and Research (ESR) for high-resolution bioinformatic analysis.

**Results** Within 48 h of notification isolates had been subcultured and sequenced. This showed that three of four isolates were in fact *K. variicola*, and two of these were sequence type (ST)6385. This sequence type had not been seen previously at our institution, so transmission was suspected. Environmental sampling revealed *K. variicola* ST6385 in two sink traps on the unit, and prospective sequencing of all *K. pneumoniae* isolates from NICU samples revealed two further infants with *K. variicola* ST6385. Subsequent phylogenetic analysis at ESR using original sequence data showed tight clustering of these isolates, confirming an outbreak. Sink traps were disinfected, environmental cleaning procedures were updated, and a strict focus on hand hygiene was reinforced on the ward. No further isolates were detected, and the outbreak was closed after two months.

\*Correspondence: Rhys T. White rhys.white@esr.cri.nz

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

© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/. **Conclusions** Access to dWGS at the level of the local hospital laboratory permitted rapid identification of an outbreak of an organism displaying no unusual antimicrobial resistance features at a point where there were only two known cases. This in turn facilitated a rapid IPC response.

Keywords Genomics, Disease outbreaks, Phylogenetic analyses, Infection control, Multilocus sequence typing

## Background

*Klebsiella variicola* is an emerging pathogen with a broad host range, causing infections in humans, animals, plants, while persisting in hospitals and other environments [1-3]. The *K. variicola* genome can harbor numerous virulence factors and antimicrobial resistance (AMR) determinants [4]. Additionally, the *K. variicola* genome can contain genes involved with nitrogen fixation, plant colonization, and mutualistic insect relationships [4]. These features enable widespread environmental persistence of *K. variicola*, potentially serving as a reservoir for AMR across diverse ecological niches [4].

Identifying K. variicola can be problematic for clinical microbiology laboratories because some commercial systems misidentify it as Klebsiella pneumoniae, leading to under-recognition [5, 6]. The clinical relevance of this misidentification is uncertain. However, some evidence suggests K. variicola is associated with more severe bloodstream infection than K. pneumoniae [7], and there are concerns regarding its propensity for spreading AMR [4, 6]. Notably, the prevalence of IncF plasmids, particularly the  $IncFIB_{K}$  subtype, in *K. variicola* raises concerns due to their role in disseminating extended-spectrum beta-lactamases (ESBLs), carbapenemases, and colistin resistance genes, posing significant challenges for treatment and infection control efforts [8]. K. variicola has been implicated in nosocomial infection and hospital outbreaks, including several reports in neonatal intensive care units (NICUs), suggesting its potential for nosocomial transmission like other *Klebsiella* species [9–11].

In early 2022 the hospital laboratory at Wellington Regional Hospital (WRH) established a decentralized whole-genome sequencing (dWGS) system using Oxford Nanopore sequencing to support local hospital infection prevention and control (IPC) [12]. In this context 'decentralized' refers to sequencing and initial bioinformatics analysis performed in the hospital diagnostic laboratory, with sequence data transferred to a central reference laboratory for high-resolution bioinformatic analysis if initial results suggest possible transmission events [13]. The Genomics and Bioinformatics Team and the Antibiotic Reference Laboratories at the Institute of Environmental Science and Research (ESR) serves the function of the reference laboratory for WRH. A decentralized model was adopted because the hospital laboratory lacks the on-site bioinformatics expertise for fine-scale WGS analysis.

In May 2024, WRH NICU observed an increased incidence of invasive *K. pneumoniae* infections, prompting further investigation using dWGS. This revealed that several isolates were actually *K. variicola*, belonging to sequence type (ST)6385, a lineage not previously reported in clinical settings. Here we describe the rapid genomic investigation of this outbreak.

## Methodology

## Setting

WRH provides tertiary services to a population of approximately 500,000 in the lower North Island/Te Ikaa-Māui of New Zealand/Aotearoa. The WRH NICU is a Level 3 facility resourced for 37 infants, providing specialized medical and surgical care for extremely premature infants (as early as 23 weeks gestation) and those requiring ventilation, intravenous feeding, and other intensive care support. The ward has six multi-cot rooms, each with two sinks, and four single isolation rooms.

Awanui Labs Wellington (an International Accreditation New Zealand ISO15189 accredited medical laboratory) provides on-site clinical diagnostic services to WRH and the surrounding region. The microbiology and molecular departments handle around 300,000 samples yearly. Identification of cultured bacterial isolates is performed using the Vitek<sup>®</sup> MS PRIME (bioMérieux) with database v3.3.0. Phenotypic susceptibility testing of Enterobacterales is performed using the Vitek<sup>®</sup> II instrument (bioMérieux) and the AST-N311 card, according to the European Committee on Antimicrobial Susceptibility Testing (EUCAST) guidelines [14].

#### Decentralized whole-genome sequencing program

In early 2022, Awanui Labs Wellington established a prospective dWGS program [12]. Sequencing is performed weekly, allowing for both 'proactive' and rapid 'reactive' sequencing. 'Proactive' refers to the sequencing of certain IPC target organisms (e.g., *Clostridioides difficile*) in realtime as they are identified in the laboratory [15]. 'Reactive' refers to the sequencing of organisms suspected by IPC to be involved in hospital transmission events, which was the case with the outbreak presented here.

Details on DNA extraction are available in the Supplementary Materials. For isolates sequenced before March 2023, library preparation used 50 ng of genomic DNA with the Oxford Nanopore Technologies rapid barcoding kit 96 (SQK-RBK110-96) as per the manufacturer's instructions. The library was loaded onto a R9.4 flow cell (FLO-MIN106) and run on an Oxford Nanopore MinION device for 20–40 h with MinKNOW v22.10.10. After March 2023, including during the outbreak, libraries were created using 50–100 ng of genomic DNA (prepared with the rapid barcoding kit 96 (SQK-RBK114-96)), and sequenced on an R10.4.1 flow cell (FLO-MIN114) with MinKNOW v23.04.5. The initial bioinformatic analysis performed in the hospital laboratory is determination of a multilocus sequence type (MLST) using Krocus v1.0.3 [16] with default settings to query the raw FASTQ files against the *Klebsiella* PasteurMLST sequence definition database [17] hosted on BIGSdb v1.47.0 [18].

#### Genome annotation for kv240612\_barcode19

As part of this investigation, and in keeping with our decentralized approach, sequence data were transferred to the Genomics and Bioinformatics Team at ESR for further analysis. In the absence of an available ST6385 reference genome, we selected kv240612\_barcode19, a clinical isolate from a neonatal blood culture, as our reference genome (index case on the WRH NICU identified in May 2024). The FASTQ files outputted from the basecalling were corrected with single-read error correction using the HERRO algorithm (using the 'dorado correct' function using model-v1, dorado v0.7.2) [19]. HERROcorrected reads underwent de novo assembly using Flye v2.9.2 [20] using a genome size estimate of 5.5 Mb and three polishing iterations. The assembly representing strain kv240612\_barcode19 was annotated using Prokka v1.14.6 [21]. Prophage regions were identified using PHASTER [22] and then annotated using Pharokka v1.6.1 [23]. Mobile genetic elements were identified using IslandViewer 4 [24] and ISsaga v2.0 [25] (ISfinder platform [26]), followed by manual curation using Artemis v18.2.0 [27].

#### **Retrospective genome analyses**

During 2022–2023, before this outbreak, one of the target organisms for proactive dWGS was K. pneumoniae, whereby all isolates from hospital inpatients were systematically sequenced. On further analysis of the sequence data, many of these isolates were identified as K. variicola and were therefore used to genomically contextualize the current outbreak. Original Fast5 sequence files were converted to Pod5 using pod5 v0.3.2 (https:// github.com/nanoporetech/pod5-file-format, accessed on 01 August 2024) and then basecalled using Dorado v0.7.2 (https://github.com/nanoporetech/dorado, accessed on 01 August 2024). The base-calling process was carried out using the 'super accuracy' models, with a batch size of 2008 and a chunk size of 1000, while other parameters were kept at their default settings. Further details of laboratory methods including: nanopore read quality control, genome assembly, MLST, and virulence and antibiotic resistance gene genotyping, are available in Supplementary Materials.

#### Taxonomic classification of the K. variicola genomes

To identify the nearest taxonomic neighbors to the suspected *K. variicola* genomes, the assemblies were input into a Genome Taxonomy Database Toolkit (GTDB-Tk) genome-based taxonomy (GTDB-Tk v2.1.1 with GTDB package R207\_v2 [28]). Only sequences within the genus *Klebsiella* and *Escherichia* (g\_, genus) were extracted from the GTDB-Tk and used to construct a concatenated reference alignment of 117 bacterial marker genes. The taxonomic tree was inferred using maximum-likelihood approximation with FastTree v2.1.7 [29] under the WAG model [30] of protein evolution with gamma-distributed rate heterogeneity [31] (+GAMMA).

## Assembly-based variant detection and *K. Variicola* specieslevel phylogeny

A core-genome alignment was generated from the 50 local *K. variicola* genomes using Parsnp v1.7.4 [32] with the kv240612\_barcode19 chromosome (GenBank: CP165787) serving as the reference to call single-nucleotide variants (SNVs). Resulting SNV alignments were used to reconstruct phylogenies. A maximum-parsimony tree was reconstructed using the heuristic search feature of PAUP v4.0a [33] before adding 1,000 bootstrap replicates. A maximum-likelihood phylogenetic tree was reconstructed using RaxML v8.2.12 [34] (GTR-GAMMA correction) by optimizing 20 distinct, randomized maximum-parsimony trees before adding 1,000 bootstrap replicates. The resulting phylogenetic trees were visualized using FigTree v1.4.4 (http://tree.bio.ed.ac.uk/software/fig tree/, accessed on 01 August 2024).

## High-resolution read-mapping approach to variant calling and SNV density analysis

The quality trimmed and filtered reads for the five other outbreak-associated K. variicola ST6395 isolates were aligned to the reference kv240612\_barcode19 chromosome using minimap2 v2.24 [35, 36] configured with the 'map-ont' preset for long-read data. The resulting Sequence Alignment/Map (SAM) file was converted to Binary Alignment/Map (BAM) format using SAMtools v1.20 [37]. The BAM file was sorted and indexed to facilitate downstream analysis. The median and range of sequence read data depth were calculated using the 'mpileup' tool from SAMtools, with the '-aa' flag and mapping quality of  $\geq 1$ . Variant calling was conducted using Clair3 v1.0.8 [38] with the 'r1041\_e82\_400bps\_sup\_ v500' model, and flags '--include\_all\_ctgs', '--no\_phasing\_for\_fa', and '--haploid\_precise'. Variants were filtered using BCFtools v1.20 [39] to exclude those with a quality score below 30. To visualize SNV density across the kv240612\_barcode19 reference chromosome, we used the ggplot2 v3.4.0 library [40] package in the R package v4.2.2 [41]. A histogram was plotted with binwidth set to 100 base pairs to represent SNV density along the genome. Consensus sequences for each of the five other ST6385 genomes were created by integrating strainspecific SNVs into the kv240612\_barcode19 reference chromosome. A multiple-pseudo-genome alignment (5,500,654 bp) was then input into the Gubbins algorithm v3.2.1 [42] (default settings, "raxml mode" with the General Time Reversible (GTR) GAMMA correction) to identify recombination.

## Results

#### **Outbreak response**

In late May 2024, the IPC team were alerted by the NICU of four cases of *K. pneumoniae* bacteremia over the preceding month. These isolates displayed wild-type phenotypic antimicrobial susceptibility (resistant to amoxicillin only), so had not been identified via multidrug-resistant organism surveillance. This represented an increase above baseline incidence, so the isolates were retrieved and subcultured on the same day. The following day they were added to the weekly sequencing run, which was brought forward. By the next day, less than 48 h after

notification, initial results showed that three of the four isolates were actually *K. variicola*, and two of these were ST6385. ST6385 had not been seen at the hospital previously, so transmission was suspected. Following this, any *K. pneumoniae* isolated from infants in NICU (or recent discharges) were prospectively sequenced. This revealed two further infants colonized with *K. variicola* ST6385 (Fig. 1). One of the infants with *K. variicola* bacteremia died shortly after the infection was detected. The remaining three infants with *K. variicola* ST6385 recovered and were eventually discharged. Following initial mitigations, no further *K. variicola* isolates were identified within the unit and the outbreak was closed after two months.

Environmental sampling was carried out on sinks/ faucets in the three rooms infants with *K. variicola* had spent the most time in, and the two ultrasound machines used on the unit (Fig. 1, and Supplementary Materials for methods). Sinks/faucets were sampled due to proximity to bed spaces of affected infants, while ultrasound machines were sampled because the outbreak spanned several rooms and these machines were the primary shared equipment used between rooms. *K. variicola* ST6385 was isolated from two of six sink-trap water samples, with all other remaining environmental samples *K. variicola* negative. As part of the response, cleaning



**Fig. 1** Line chart of the *Klebsiella variicola* ST6385 outbreak in the neonatal intensive care unit (NICU). The x-axis represents days, with day 0 being the day the hospital infection prevention and control (IPC) team was alerted by the NICU of four cases of *Klebsiella* spp. associated bacteremia over the preceding month (actual dates are omitted for patient privacy). Each horizontal shaded bar represents the days spent by the numbered infant in the room as indicated on the y-axis. Infants with non-ST6385 *K. variicola* (lighter-shaded grey bars), or *Klebsiella pneumoniae* (darker-shaded grey bars) are labelled with letters instead of numbers. Infant A was born on day – 76 and infant B was born on day – 63

schedules for the ultrasound machines were reviewed and updated, additional cleaning of cot spaces was implemented, and affected sink traps were disinfected using a quaternary ammonium compound. Strict adherence to Standard Precautions was emphasized for clinical care for all infants. The outbreak totaled four infants, two associated with bacteremia (Table 1).

## A high-quality reference genome was generated for the index *K. Variicola* ST6385 case

Sequencing of kv240612\_barcode19 initially produced 48,767 single-ended reads with a median read length of 4,223 bp (N50 = 26,672 bp) and a median read quality score of 21.0. HERRO correction refined this to 23,193 single-ended reads (444.5 total megabases) with a median length of 13,593 bp (N50 = 27,418 bp).

De novo assembly revealed a 5,500,654 bp circular chromosome with 57.42% GC content (Supplementary Materials, Figure S1a). Using HERRO-corrected reads, the chromosome was sequenced to a median depth of  $75\times$  (range: 44 to 641×). Multiple antibiotic resistance genes were present, although no clinically relevant phenotypic resistance was observed (Supplementary Materials, Table S1). Virulence factors included *fim* and *mrk* operons (fimbrial adhesins), Klebsiella K and O loci (capsule and O-antigen synthesis), and iron acquisition systems (iroE, iutA, and iron-enterobactin operon). The capsule polysaccharide locus was not typable due to an insertion flanked by ISEcl1 and ISEc33 downstream of *wcaJ* (Supplementary Materials, Figure S2). The lipopolysaccharide locus was typed as O3/O3a serotype. The chromosome also contained several efflux pump systems (aaeRXAB, emrBAR, emrAB-OMF, marORAB-Ea, mdtABCD-baeSR, acrAB), and metal resistance operons (arsBCR, cusABFCRS). Additional genomic features included five prophage regions, a genomic island integrated into *thr*R, and a type VI secretion system.

The complete genome of kv240612\_barcode19 included the plasmid pkv240612\_barcode19A (169,125 bp, FIB plasmid), carrying several key resistance and transport systems (Supplementary Materials, Table S1). Using HERRO-corrected reads, the plasmid was sequenced to a median depth of 145× (range: 129 to 311×). pkv240612\_barcode19A harbored the *sil*E gene encoding a silver-binding protein, and the *cus*ABFCRS operon, both contributing to putative silver resistance. Putative copper resistance was conferred by the *cop*-ABCD operon. The plasmid also carried the *ars*RBCH operon for arsenic resistance. Additionally, two transport systems were identified: the *urt*ABCDE operon, an ABC-type urea permease, and the *fec*IRABCDE operon for iron uptake.

# *K. Variicola* ST6385 is genomically distinct from other local *K. variicola*

We analyzed 58 *K. variicola* genomes from isolates collected in 2022 (n=39), 2023 (n=11), 2024 (n=1), and throughout the outbreak (n=7) (Supplementary Materials, Table S2). Of the 58 genomes, eight were excluded: five due to insufficient sequencing depth (<15×) and three due to genome lengths outside the extended interquartile range (i.e., 5,261,825 bp and 6,092,438 bp) (Supplementary Materials, Table S3). After quality control, 50 genomes were retained for further analysis (Supplementary Materials, Table S4). The analyzed genomes were predominantly from 2022 (n=34) and 2023 (n=8), with one from 2024, and the seven outbreak isolates. The *de novo* assembly quality metrics for these genomes are presented in the Supplementary Materials (Table S5).

The GTDB-tk taxonomy analysis confirmed that these 50 genomes clustered with K. variicola and not K. pneumoniae (Fig. 2). Following this, the core-genome analysis identified 178,995 SNVs across the 50 genomes (Fig. 3). These SNVs were called against the chromosome of isolate kv240612\_barcode19 using assembly-based variant detection (see Methods). A maximum-likelihood phylogenetic tree was constructed and rooted at genome kv220615\_barcode15 (Fig. 3; see branch lengths expressed as SNVs in Supplementary Materials, Figure S3), corresponding to the root of Klebsiella africana (RefSeq: GCF\_900978845), K. pneumoniae (RefSeq: GCF\_000742135), and Klebsiella quasipneumoniae (RefSeq: GCF\_000751755) (Fig. 2). The genome for isolate kv240612\_barcode18 was classified as ST3938, which is genomically distinct from the other K. variicola genomes collected in this outbreak. The remaining six

Table 1 Genomic characteristics of Klebsiella variicola strains from a neonatal outbreak in Wellington, New Zealand

Sample ID	Collection date	Source	Sample type	MLST	Capsule locus	Lipopolysaccharide locus	SRA accession no.
kv240612_barcode18	Apr 2024	Human	Blood	ST3938	KL71	O3/O3a	SRR30084997
kv240612_barcode19	May 2024	Human	Blood	ST6385	Untypeable	O3/O3a	SRR30084996
kv240612_barcode20	May 2024	Human	Blood	ST6385	Untypeable	O3/O3a	SRR30084995
kv240612_barcode24	Jun 2024	Human	Urine	ST6385	Untypeable	O3/O3a	SRR30084994
kv240612_barcode25	Jun 2024	Sink	-	ST6385	Untypeable	O3/O3a	SRR30084993
kv240709_barcode65	Jun 2024	Human	Eye swab	ST6385	Untypeable	O3/O3a	SRR30084992
kv240709_barcode69	Jun 2024	Sink	-	ST6385	Untypeable	O3/O3a	SRR30084991

MLST, multilocus sequence type; SRA, Sequence Read Archive; ST, Sequence Type



Fig. 2 Taxonomic tree of the genus *Klebsiella*. The maximum-likelihood approximation was reconstructed using a concatenated alignment of 117 conserved bacterial markers. Taxonomy is shown according to the GTDB (s\_, species) from NCBI taxonomy. GenBank assembly accession numbers are displayed after the genome names. The blue taxon represents the strain sequenced in this report. The tree is rooted according to the outgroup species *Escherichia coli* 

genomes, representing isolates from four infants and two sink traps, were classified as ST6385 and formed a distinct cluster. The long branch leading to the ST6385 cluster shows these six genomes are genomically distinct from previously sequenced *K. variicola* strains from the NICU, indicating a common origin. The six outbreak isolates were also sent to an affiliated laboratory that uses the Bruker MALDI Biotyper<sup>®</sup> (Bruker Daltonik, v9 of the MBT 8468 database) which correctly identified all isolates as *K. variicola*.

# Minimal SNV differences and recombination events across human- and sink-derived *K. variicola* ST6385 genomes

For the six outbreak isolates, the SNV differences compared to the index isolate (kv240612\_barcode19, infant 1) were 0 (kv240709\_barcode69, sink trap), 1 (kv240612\_ barcode25, sink trap), 5 (kv240709\_barcode65, infant 4), 15 (kv240612\_barcode24, infant 3) and 71 (kv240612\_barcode20, infant 2). The larger number of SNV differences for the latter two isolates were located predominantly in the urea transport operon, with Gubbins analysis suggesting these were due to recombination events (Supplementary Materials, Figure S4), with minimal SNV differences otherwise.

### Discussion

In previous investigations, we have shown the value of proactive dWGS for early identification of hospital outbreaks [15, 43]. Here we demonstrate the value of reactive dWGS for rapidly identifying and precisely defining an outbreak. Within 48 h of the initial notification, rapid MLST analysis performed at the local laboratory level provided a sufficient degree of suspicion to trigger IPC investigations and actions, whilst simultaneously



Fig. 3 Maximum-likelihood phylogeny of *Klebsiella variicola* isolates from Wellington Regional Hospital. The phylogeny was inferred from 178,995 coregenome single-nucleotide variants (SNVs) from 50 genomes. SNVs were derived from a core-genome alignment of 4,283,084 bp and were called against the chromosome of sample kv240612\_barcode19 (GenBank: CP165787). The phylogenetic tree is rooted according to the kv220615\_barcode15 outgroup. Bootstrap values > 80% (1,000 replicates) are shown

excluding two unrelated cases from the outbreak. This revealed further affected infants, and an environmental reservoir of the organism, which could then be addressed. Furthermore, the outbreak was detected when there were only two known cases and the organism displayed no distinctive AMR or other phenotypic features, highlighting the high level of precision rapid dWGS can bring to IPC investigations. Had rapid dWGS not been available it seems likely that this outbreak would have progressed unnoticed to a larger outbreak with further invasive infections. Having access to dWGS in the hospital laboratory gives a high degree of flexibility to respond promptly to local IPC needs. In this case, communication between IPC and the laboratory meant isolates could be retrieved immediately, the weekly sequencing run could be brought forward, and actionable results produced in a short timeframe, with high-resolution off-site genomic analyses following later. Had isolates themselves needed to be shipped to the reference laboratory for sequencing and analysis this speed and agility of response would not have been possible. Many hospital laboratories have sufficient molecular experience to perform sequencing but do not have access to expert bioinformaticians for highresolution WGS analysis, a major advantage of a dWGS model.

Organisms in the genus *Klebsiella*, predominantly *K. pneumoniae*, are important and well-described NICU pathogens [44], with the ability to transmit between infants, cause severe nosocomial infection and harbor key AMR genes [45]. More recently, several reports have

implicated K. variicola as a cause of clinical infection and outbreaks among neonates [9–11]. This investigation supports the importance of K. variicola in the NICU setting and serves as a reminder that it may be misidentified as K. pneumoniae. The organism in this outbreak displayed a propensity for invasive infection and environmental persistence. Several identified virulence factors, such as outer membrane protein A, K locus for capsule synthesis (albeit novel in our study), outer lipopolysaccharide O antigen, and siderophores, are similar to those previously described in hypervirulent K. pneumoniae [46, 47]. However, while these factors are associated with virulence, the strain in this outbreak does not meet the full criteria for hypervirulence, such as the presence of specific genes (e.g., iucA, rmpA) [46]. As such, further phenotypic investigations are required to confirm hypervirulence.

The organism also harbored genes associated with environmental persistence (through biofilm formation) [48], including heavy metal tolerance, efflux pumps, type 1 and 3 fimbriae, and a type VI secretion system. Furthermore, plasmid-borne elements may have enhanced the isolate's metal resistance profile and environmental nutrient acquisition, such as iron uptake systems. These features likely support bacterial survival in hospital environments and may indirectly enhance virulence by facilitating infection establishment. The presence of a conjugative F plasmid *tra* operon (Figure S1b) indicates potential horizontal gene transfer, raising concerns about the dissemination of AMR and virulence-associated genes in the hospital setting. The ability of *K. variicola* to survive in the hospital environment is of IPC concern not only because of its innate virulence, but previous reports demonstrate an ability to acquire and disseminate plasmids containing important AMR genes in the NICU environment [10].

Recent reports have implicated sinks in the transmission of important hospital pathogens and questioned the role of sinks in healthcare [49]. There have been moves to create 'waterless' augmented care units including in neonatal settings [50]. Although it is unclear from our analysis the precise role sinks played in transmission on the unit, their ability to act as reservoirs for the outbreak organism was demonstrated. As part of our response, the affected sinks and traps were chemically disinfected, however recent evidence suggests that this may have limited effectiveness, albeit with a different compound [51]. This has generated ongoing discussions between the IPC and NICU teams regarding the need for sinks on the unit.

A limitation of this analysis was that we did not actively screen infants for the outbreak organism, and as such more infants than those identified may have been affected. This was due both to the lack of validated screening methods for this fully susceptible organism and the preference to utilize a unit-wide focus on strict hand hygiene and environmental cleaning as the core response. The environmental sampling was also performed opportunistically, rather than using specific environmental sampling techniques such as touch plates. As such, the sensitivity of sampling may have been compromised and other environmental sources of the organism missed. Had further invasive infections occurred, systematic screening of infants and the environment may have been necessary.

## Conclusions

In conclusion, this analysis has demonstrated the ability of decentralized whole-genome sequencing to provide rapid outbreak detection to facilitate timely infection prevention and control action. Speed and agility are key factors of an outbreak response, and in this case likely prevented further cases in a highly vulnerable population. This investigation revealed that *Klebsiella variicola*, often misidentified as *Klebsiella pneumoniae*, can cause severe disease and transmission in the NICU setting, with environmental persistence particularly in sink traps.

## Abbreviations

AMR	Antimicrobial resistance
dWGS	Decentralized whole genome sequencing
ESBLs	Extended-spectrum beta-lactamases
ESR	Institute for Environmental Science and Research
EUCAST	European Committee on Antimicrobial Susceptibility Testing
GTDB-Tk	Genome Taxonomy Database Toolkit
HERRO	Haplotype-aware ERRor cOrrection
IPC	Infection prevention and control

 MLST
 Multi locus sequence typing

 NICU
 Neonatal intensive care unit

 SNV
 Single-nucleotide variant

 ST
 Sequence type

 WRH
 Wellington Regional Hospital

#### Supplementary Information

The online version contains supplementary material available at https://doi.or g/10.1186/s13756-025-01529-2.

Supplementary Material 1

#### Acknowledgements

We acknowledge the facilities, and the scientific and technical assistance of staff at Awanui Labs, formerly known as Southern Community Laboratories (Wellington, New Zealand). We thank Juliet Elvy, Megan Smith, and James Ussher from Awanui Labs (Dunedin, New Zealand) for confirming species identification using the Bruker MALDI Biotyper<sup>6</sup> system. We acknowledge and thank the following staff at the Institute of Environmental Science and Research (ESR) for their valuable feedback: Ernest Williams (Wallaceville Science Centre) and William Taylor (Christchurch Science Centre). This research was made possible by the Computational Science Team and the High-Performance Compute (HPC) platform at ESR, with specific thanks to Russell Smithies and Shane Sturrock for HPC.

#### Author contributions

Writing– Original Draft: R.T.W. and M.G.B. Writing– Review & Editing: R.T.W., M.Ba, M.Bu., S.H., J.J., M.K., D.M-C., T.S., H.S., D.J.W., and M.G.B. Conceptualization: R.T.W. and M.G.B. Investigation: R.T.W. and M.G.B. Methodology: R.T.W. Data Curation: R.T.W. and M.G.B. Visualization: R.T.W. Formal Analysis: R.T.W. All authors have read and agreed to the published version of the manuscript.

#### Funding

This study was supported by internal departmental funds at Awanui Laboratories Wellington, the Institute of Environmental Science and Research (ESR), and Genomics Aotearoa through funding from the Ministry of Business Innovation and Employment (MBIE).

#### Data availability

DNA sequence data are available in the National Center for Biotechnology Information (NCBI) BioProject accession numbers: PRJNA1142676 (2022 samples), PRJNA1142678 (2023 samples), and PRJNA1142680 (2024 samples). The raw Oxford Nanopore Technologies (ONT) sequence reads are available under accession numbers: SRR30085170 to SRR30085208 (2022 samples), SRR30087320 to SRR30087330 (2023 samples), SRR30084991 to SRR30084998 (2024 samples). The complete assembly for strain kv240612\_barcode19 has been deposited to GenBank under the accession numbers CP165787 and CP165788.

### Declarations

#### Ethics approval and consent to participate

Ethics waiver was granted under New Zealand Health and Disability Ethics Committee requirements.

#### **Consent for publication**

Not applicable.

#### **Competing interests**

R.T.W received travel expenses from Oxford Nanopore Technologies to attend their annual, London Calling, meeting in May 2024 (as do all speakers). R.T.W received a travel bursary from Oxford Nanopore Technologies and a travel grant from the UK Microbiology Society to present at conferences. The authors declare no other conflicts of financial, general, or institutional competing interests.

#### Author details

<sup>1</sup>Health Group, Institute of Environmental Science and Research, Porirua 5022, New Zealand <sup>2</sup>Department of Microbiology and Molecular Pathology, Awanui Labs Wellington, Wellington 6021, New Zealand <sup>3</sup>Infection Prevention and Control, Te Whatu Ora/Health New Zealand, Capital, Coast & Hutt Valley, Capital, Wellington 6021, New Zealand

### Received: 16 September 2024 / Accepted: 28 January 2025 Published online: 07 February 2025

#### References

- Rosenblueth M, Martinez L, Silva J, Martinez-Romero E. Klebsiella variicola, a novel species with clinical and plant-associated isolates. Syst Appl Microbiol. 2004;27:27–35. https://doi.org/10.1078/0723-2020-00261.
- Podder MP, Rogers L, Daley PK, Keefe GP, Whitney HG, Tahlan K. Klebsiella species associated with bovine mastitis in Newfoundland. PLoS ONE. 2014;9:e106518. https://doi.org/10.1371/journal.pone.0106518.
- Barrios-Camacho H, Aguilar-Vera A, Beltran-Rojel M, Aguilar-Vera E, Duran-Bedolla J, Rodriguez-Medina N, et al. Molecular epidemiology of Klebsiella variicola obtained from different sources. Sci Rep. 2019;9:10610. https://doi.or g/10.1038/s41598-019-46998-9.
- Srinivasan VB, Rajamohan G. Comparative genome analysis and characterization of a MDR Klebsiella variicola. Genomics. 2020;112:3179–90. https://doi.or g/10.1016/j.ygeno.2020.06.00.
- de Campos TA, de Almeida FM, de Almeida APC, Nakamura-Silva R, Oliveira-Silva M, de Sousa IFA, et al. Multidrug-resistant (MDR) Klebsiella variicola strains isolated in a Brazilian hospital belong to new clones. Front Microbiol. 2021;12:604031. https://doi.org/10.3389/fmicb.2021.604031.
- Rodriguez-Medina N, Barrios-Camacho H, Duran-Bedolla J, Garza-Ramos U. Klebsiella variicola: an emerging pathogen in humans. Emerg Microbes Infections. 2019;8:973–88. https://doi.org/10.1080/22221751.2019.1634981.
- Maatallah M, Vading M, Kabir MH, Bakhrouf A, Kalin M, Naucler P, et al. Klebsiella variicola is a frequent cause of bloodstream infection in the Stockholm area, and associated with higher mortality compared to K. pneumoniae. PLoS ONE. 2014;9:e113539. https://doi.org/10.1371/journal.pone.0113539.
- Duran-Bedolla J, Rodríguez-Medina N, Dunn M, Mosqueda-García D, Barrios-Camacho H, Aguilar-Vera A, et al. Plasmids of the incompatibility group FIBK occur in Klebsiella variicola from diverse ecological niches. Int Microbiol. 2023;26:917–27. https://doi.org/10.1007/s10123-023-00346-0.
- Hani S, Lalaoui A, Bennaoui F, Soraa N, Slitine NEI, Maoulainine FMR. Nosocomial Klebsiella variicola infection in neonatal intensive care: a new emerging pathogen. Open J Pediatr. 2023;13:450–8. https://doi.org/10.4236/ojped.2023 .133050.
- Farzana R, Jones LS, Rahman MA, Andrey DO, Sands K, Portal E, et al. Outbreak of hypervirulent multidrug-resistant Klebsiella variicola causing high mortality in neonates in Bangladesh. Clin Infect Dis. 2019;68:1225–7. https://doi.org/ 10.1093/cid/ciy778.
- Piepenbrock E, Higgins PG, Wille J, Xanthopoulou K, Zweigner J, Jahn P, et al. Klebsiella variicola causing nosocomial transmission among neonates - an emerging pathogen? J Med Microbiol. 2020;69:396–401. https://doi.org/10.1 099/jmm.0.001143.
- Bloomfield M, Hutton S, Velasco C, Burton M, Benton M, Storey M, et al. Oxford nanopore next generation sequencing in a front-line clinical microbiology laboratory without on-site bioinformaticians. Pathology. 2024;56:444–7. https://doi.org/10.1016/j.pathol.2023.07.014.
- Beukers AG, Jenkins F, van Hal SJ. Centralised or localised pathogen whole genome sequencing: lessons learnt from implementation in a clinical diagnostic laboratory. Front Cell Infect Microbiol. 2021;11:636290. https://doi.org/ 10.3389/fcimb.2021.636290.
- 14. European Committee on Antimicrobial Susceptibility Testing. European Committee on Antimicrobial Susceptibility Testing breakpoint tables for interpretation of MICs and zone diameters. Version 13.1. Växjö, Sweden: The European Committee on Antimicrobial Susceptibility Testing. 2023. https://w ww.eucast.org/clinical\_breakpoints accessed 21 August 2024.
- Bloomfield M, Hutton S, Burton M, Tarring C, Velasco C, Clissold C et al. Early identification of a ward-based outbreak of Clostridioides difficile using prospective multilocus sequence type-based Oxford Nanopore genomic surveillance. Infect Control Hosp Epidemiol. 2024:1–7. https://doi.org/10.1017 /ice.2024.77
- Page AJ, Keane JA. Rapid multi-locus sequence typing direct from uncorrected long reads using Krocus. PeerJ. 2018;6:e5233. https://doi.org/10.7717/ peerj.5233.

- Diancourt L, Passet V, Verhoef J, Grimont PA, Brisse S. Multilocus sequence typing of Klebsiella pneumoniae nosocomial isolates. J Clin Microbiol. 2005;43:4178–82. https://doi.org/10.1128/jcm.43.8.4178-4182.2005.
- Jolley KA, Bray JE, Maiden MCJ. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. Wellcome Open Res. 2018;3:124. https://doi.org/10.12688/wellcomeopenres.14826.1.
- Stanojevic D, Lin D, De Florez P, Sikic M. Telomere-to-telomere phased genome assembly using error-corrected Simplex nanopore reads. bioRxiv: 594796v1 [Preprint]. 2024. Available from: https://www.biorxiv.org/content/h ttps://doi.org/10.1101/2024.05.18.594796v1
- Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol. 2019;37:540–6. https://doi.org/10.1038/s 41587-019-0072-8.
- 21. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014;30:2068–9. https://doi.org/10.1093/bioinformatics/btu153.
- Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. PHAST: a fast phage search tool. Nucleic Acids Res. 2011;39:W347–52. https://doi.org/10.1093/nar/gkr48
   5.
- Bouras G, Nepal R, Houtak G, Psaltis AJ, Wormald PJ, Vreugde S. Pharokka: a fast scalable bacteriophage annotation tool. Bioinformatics. 2023;39. https:// doi.org/10.1093/bioinformatics/btac776.
- Bertelli C, Laird MR, Williams KP, Simon Fraser University Research Computing Group, Lau BY, Hoad G, et al. IslandViewer 4: expanded prediction of genomic islands for larger-scale datasets. Nucleic Acids Res. 2017;45:W30–5. https://doi .org/10.1093/nar/gkx343.
- Varani AM, Siguier P, Gourbeyre E, Charneau V, Chandler M. ISsaga is an ensemble of web-based methods for high throughput identification and semi-automatic annotation of insertion sequences in prokaryotic genomes. Genome Biol. 2011;12:R30. https://doi.org/10.1186/gb-2011-12-3-r30.
- Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M. ISfinder: the reference centre for bacterial insertion sequences. Nucleic Acids Res. 2006;34:D32–6. https://doi.org/10.1093/nar/gkj014.
- Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, et al. Artemis: sequence visualization and annotation. Bioinformatics. 2000;16:944–5. ht tps://doi.org/10.1093/bioinformatics/16.10.944.
- Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk v2: memory friendly classification with the genome taxonomy database. Bioinformatics. 2022;38:5315–6. https://doi.org/10.1093/bioinformatics/btac672.
- Price MN, Dehal PS, Arkin AP. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. Mol Biol Evol. 2009;26:1641– 50. https://doi.org/10.1093/molbev/msp077.
- Whelan S, Goldman N. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. Mol Biol Evol. 2001;18:691–9. https://doi.org/10.1093/oxfordjournal s.molbev.a003851.
- Yang Z. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. J Mol Evol. 1994;39:306– 14. https://doi.org/10.1007/BF00160154.
- Treangen TJ, Ondov BD, Koren S, Phillippy AM. The Harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. Genome Biol. 2014;15:524. https://doi.org/10.1186/s1305 9-014-0524-x.
- Wilgenbusch JC, Swofford D. Inferring evolutionary trees with PAUP\*. Curr Protocols Bioinf. 2003;6. https://doi.org/10.1002/0471250953.bi0604s00.
- Stamatakis A. Bioinformatics. 2014;30:1312–3. https://doi.org/10.1093/bioinfo rmatics/btu033. RAxML version 8: a tool for phylogenetic analysis and post-a nalysis of large phylogenies.
- Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 2018;34:3094–100. https://doi.org/10.1093/bioinformatics/bty191.
- Li H. New strategies to improve minimap2 alignment accuracy. Bioinformatics. 2021;37:4572–4. https://doi.org/10.1093/bioinformatics/btab705.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25:2078–9. https:/ /doi.org/10.1093/bioinformatics/btp352.
- Zheng Z, Li S, Su J, Leung AW, Lam TW, Luo R. Symphonizing pileup and fullalignment for deep learning-based long-read variant calling. Nat Comput Sci. 2022;2:797–803. https://doi.org/10.1038/s43588-022-00387-x.
- Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. Gigascience. 2021;10. https://doi.org/10.109 3/gigascience/giab008.
- Wickham H. gplot2: elegant graphics for data analysis. New York, NY, USA: Springer; 2016.

- R Core Team. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. 2023. https://www. r-project.org/ accessed 21 August 2024.
- Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, Bentley SD, et al. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. Nucleic Acids Res. 2015;43:e15. https://d oi.org/10.1093/nar/gku1196.
- White RT, Bakker S, Burton M, Castro ML, Couldrey C, Dyet K, et al. Rapid identification and subsequent contextualization of an outbreak of methicillinresistant Staphylococcus aureus in a neonatal intensive care unit using nanopore sequencing. Microb Genomics. 2024;10. https://doi.org/10.1099/m gen.0.001273.
- 44. Chen Y, Brook TC, Soe CZ, O'Neill I, Alcon-Giner C, Leelastwattanagul O, et al. Preterm infants harbour diverse Klebsiella populations, including atypical species that encode and produce an array of antimicrobial resistance- and virulence-associated factors. Microb Genomics. 2020;6. https://doi.org/10.109 9/mgen.0.000377.
- Mukherjee S, Mitra S, Dutta S, Basu S. Neonatal Sepsis: the impact of carbapenem-resistant and hypervirulent Klebsiella pneumoniae. Front Med. 2021;8:634349. https://doi.org/10.3389/fmed.2021.634349.
- Zhu J, Wang T, Chen L, Du H. Virulence factors in hypervirulent Klebsiella pneumoniae. Front Microbiol. 2021;12:642484. https://doi.org/10.3389/fmicb. 2021.642484.
- 47. Rodríguez-Medina N, Rodríguez-Santiago J, Alvarado-Delgado A, Sagal-Prado A, Silva-Sánchez J, De la Cruz MA, et al. Comprehensive study reveals

phenotypic heterogeneity in *Klebsiella pneumoniae* species complex isolates. Sci Rep. 2024;14:5876. https://doi.org/10.1038/s41598-024-55546-z.

- Marian C, Sanjar F, Maxwell C, Sabitu F, Hubbard N, Ratib N, et al. A paradigm for the contextual safety assessment of agricultural microbes: a closer look at Klebsiella variicola. Front Industrial Microbiol. 2024;2:1412302. https://doi.org/ 10.3389/finmi.2024.1412302.
- Parkes LO, Hota SS. Sink-related outbreaks and mitigation strategies in healthcare facilities. Curr Infect Dis Rep. 2018;20:42. https://doi.org/10.1007/s1190 8-018-0648-3.
- Ming LJ, Xia W, Mei YNP, Amin Z. A journey towards water less neonatal intensive care unit (NICU). Int J Infect Dis. 2023;130:S52. https://doi.org/10.10 16/j.ijid.2023.04.129.
- Catho G, Cave C, Grant R, Carry J, Martin Y, Renzi G, et al. Controlling the hospital aquatic reservoir of multidrug-resistant organisms: a cross-sectional study followed by a nested randomized trial of sink decontamination. Clin Microbiol Infect. 2024;30:1049–54. https://doi.org/10.1016/j.cmi.2024.05.008.

### **Publisher's note**

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