Evaluation of Fourier Transform Infrared spectroscopy (IR Biotyper) as a complement to Whole genome sequencing (WGS) to characterise *Enterobacter cloacae*, *Citrobacter freundii* and *Klebsiella pneumoniae* isolates recovered from hospital sinks

1.1 Author names

Aranega-Bou P[1](ORCID: 0000-0001-7415-6698), Cornbill C[1], Rodger G[2,3] (ORCID: 0000-0001-6051-8985), Bird M[2, 4] (ORCID: 0000-0001-9560-2861), Moore G[1] (ORCID: 0000-0001-6969-3528), Roohi A[2], Hopkins KL[4] (ORCID 0000-0001-7279-5322), Hopkins S[4] (0000-0001-5179-5702), Ribeca P[4,5,6] (ORCID: 0000-0001-5599-3933), Stoesser N[2,3#] (ORCID: 0000-0002-4508-7969), Lipworth SI[3#] (0000-0001-5872-9214)

[#] Contributed equally

1.2 Affiliation(s)

[1] Biosafety, Air and Water Microbiology Group, UK Health Security Agency, Porton Down, UK

[2] NIHR Health Protection Unit in Antimicrobial Resistance and Healthcare-associated Infection, University of Oxford, Oxford, UK

[3] Nuffield Department of Medicine, University of Oxford, Oxford, UK

[4] UK Health Security Agency, Colindale, UK

[5] NIHR Health Protection Unit in Genomics and Enabling Data, University of Warwick, Warwick, UK

[6] NIHR Health Protection Unit in Gastrointestinal Infections, University of Liverpool, Liverpool, UK

1.3 Corresponding author and email address

Paz Aranega Bou - Paz.AranegaBou@ukhsa.gov.uk

Alternative: Sam Lipworth - samuel.lipworth@ndm.ox.ac.uk

1.4 Keywords

Fourier Transform Infrared (FT-IR) spectroscopy, bacterial typing, genomics, outbreaks, healthcare environments, hospital sinks

1.5 Repositories:

Raw sequencing data are available in the NCBI repository (BioProject accession: PRJNA925315)

Figshare repository available at https://doi.org/10.6084/m9.figshare.22016009.v1

2. Abstract

Whole genome sequencing (WGS) of healthcare-associated pathogens is recognised as the gold standard for isolate typing and the recognition of transmission networks and outbreaks. However, it remains reasonably expensive to process small numbers of isolates in real-time, and frequently requires specific expertise to enable both sequencing and the analysis of sequencing outputs, limiting its generalisability and turnaround. Spectrometry has revolutionised species identification in clinical laboratory workflows, and has more recently been applied to strain-level identification to facilitate low-cost, routine strain typing in clinical laboratories. However, studies to date of its clinical performance for strain-level typing are conflicting, and limited evaluation has been undertaken on environmental healthcare-associated isolates. We therefore compared its performance with WGS for *Enterobacter cloacae*, *Citrobacter freundii* and *Klebsiella pneumoniae* isolated from sink drains across nine hospitals and investigated whether it could be used as a screening tool prior to WGS. We found its sensitivity and specificity to cluster isolates when compared with WGS were generally poor and highly variable dependent on species and the single nucleotide polymorphism (SNP) distance threshold used to cluster isolates.

3. Impact statement

Enterobacterales are key pathogens of concern in healthcare-associated infections and have been shown to be disseminated via environmental reservoirs in hospitals. Accurately evaluating strain-based transmission networks amongst patients and hospital environments is critical to optimising infection prevention and control interventions. The current gold standard approach to characterise these transmission networks is to use genome sequencing of isolates, but this remains expensive and challenging to do routinely in many clinical laboratories.

Bacterial characterisation based on protein spectra has been widely implemented for species identification in diagnostic microbiology, and devices such as Bruker's matrix-assisted laser desorption/ionization time-of-flight (MALDI-ToF) mass spectrometry system (MALDI Biotyper) have revolutionised workflows, reducing both costs and turnaround. Bruker have introduced another device, known as the IR Biotyper, which aims to use infrared spectroscopy to characterise strain-level variability and facilitate outbreak investigation. Our study shows that when compared to whole genome sequencing, the IR Biotyper cannot be used reliably to type collections of isolates collected from the hospital sink environment, which might limit its suitability to investigate outbreaks associated with this environmental reservoir.

4. Data summary

The authors confirm all supporting data, code and protocols have been provided within the article or through supplementary data files.

Raw isolate sequencing data are deposited in NCBI (BioProject accession: PRJNA925315).

Supplementary data files, phylogenetic and IR Biotyper trees are deposited in Figshare https://doi.org/10.6084/m9.figshare.22016009.v1

5. Introduction

Healthcare-associated infections (HCAIs) increase length of hospital stay, morbidity, mortality and costs; they are an important safety issue in all countries, regardless of income level [1]. Healthcare settings provide conditions that facilitate the transmission of pathogenic microorganisms, such as an immunosuppressed patient population, invasive procedures and multiple potential environmental reservoirs [2, 3]. HCAIs can result from sporadic events or from outbreaks affecting several patients [2].

Outbreak-causing HCAIs can be due to a variety of bacterial, fungal or viral pathogens [3, 4]. Preventing and containing hospital-associated outbreaks requires timely identification of the causative agent and investigation of possible transmission pathways to implement effective infection control measures [2]. However, identifying the source of an outbreak can be difficult [4], particularly when an environmental source is suspected [5]. Hospital sinks have emerged as a key reservoir linked to outbreaks caused by Enterobacterales and other Gram-negative organisms [6].

Several typing techniques can be useful for outbreak investigations in hospitals. Genotypic techniques such as pulsed-field gel electrophoresis (PFGE), repetitive-element palindromic PCR (rep-PCR), multi-locus sequence typing (MLST) [7] and variable-number tandem repeat (VNTR) schemes (Katie Hopkins, personal communication) have been used traditionally. However, these techniques can lack discriminatory power and/or reproducibility and therefore analyses have increasingly been replaced by whole genome sequencing (WGS) [7]. WGS has been widely used to study outbreaks retrospectively [8-11] and, to some extent, prospectively [12, 13], and is considered the gold standard approach to determining isolate relatedness. However, implementation in the routine clinical setting is still limited by availability, turnaround time, relative cost, and the specialist bioinformatics expertise required to process data and interpret outputs. A rapid, low-cost and user-friendly typing method to inform outbreak management in a timely manner would still therefore be of use. Such an approach could also complement WGS-based surveillance in some settings by enabling the targeted selection of isolates for WGS to reduce costs and turnaround times.

The Bruker IR Biotyper uses Fourier Transform Infrared (FT-IR) spectroscopy to generate spectra from bacterial isolates and establish relationships between them according to

spectral similarities [14]. While the technology can quantify the absorption of all molecules present in the sample (carbohydrates, lipids, proteins, and nucleic acids) to generate the spectra, the region reflecting the variations in carbohydrates has been chosen as the default for strain typing on the IR Biotyper, which has been used alongside genotypic methods to type collections of clinical isolates of *Klebsiella pneumoniae* [15-19] and *Enterobacter cloacae* [17, 20, 21] among others. While the results of some of these studies have shown concordance between the techniques, they have often relied on WGS to define the most appropriate cut off value for clustering [15-20] on the IR Biotyper. Other studies have found little correlation with WGS for *E. cloacae* [21] and MLST for *Pseudomonas aeruginosa* [22]. One aspect that has received little attention to date is the use of the technique "out-of-the-box" (i.e. using the manufacturer's recommended settings) on isolates recovered from the hospital environment, thus simulating the "real-world" conditions in which the IR Biotyper might be used in clinical practice without any additional labour-intensive and bespoke optimisation of thresholds.

In this study we assessed the performance of the IR Biotyper using a collection of 162 *E. cloacae*, *Citrobacter freundii*, and *K. pneumoniae* isolates collected in hospital sinks from nine different hospitals in England, UK. We evaluated the correlation between clusters determined using the IR Biotyper with those obtained from nucleotide-level (i.e. single nucleotide polymorphisms [SNP]) threshold analysis from WGS data.

6. Methods

Bacterial isolates

Drain biofilms and waste trap water from ninety hospital sinks in nine different hospitals sampled in England between March and June 2019 were cultured on MacConkey n°3 agar, Brilliance™ *E. coli*/coliform agar, and chromID™ CARBA agar overnight at 37°C. Species identification of morphologically distinct cultured isolates was performed using matrix-assisted laser desorption/ionization time-of-flight (MALDI-ToF) mass spectrometry (Bruker Daltonik MALDI Biotyper) with the direct transfer method. Antibiotic resistance profiling was carried out straight after isolation and identification using the disc diffusion method (gentamicin, amikacin, ciprofloxacin, ceftazidime, meropenem and ertapenem) following current EUCAST guidelines [23]. All Enterobacterales isolates from each sample with a distinct antimicrobial resistance profile were stored as pure sub-cultures in cryobeads at -80°C and those identified by the MALDI Biotyper as members of the *E. cloacae* complex, *C. freundii* complex and *K. pneumoniae/K. variicola* were used in this study. For simplicity we refer to members of the *E. cloacae* complex as *E. cloacae*, members of *C. freundii* complex as *C. freundii* and *K. pneumoniae*/*K. variicola* isolates as *K. pneumoniae* throughout the manuscript.

IR Biotyper: sample preparation, spectrum acquisition and analysis

Sample preparation was carried out according to manufacturer's instructions. In brief, all isolates were grown on tryptic soy agar overnight at 37°C and kept at 4°C as stock plates. Isolates were subsequently sub-cultured from the stock plates and grown at 37 °C for 24 hours. A full 1µl loopful of each isolate, taken from an area of confluent growth, was suspended in 50µl of 70% (v/v) ethanol in a suspension vial with inert metal cylinders (Bruker) and vortexed. After the bacteria were homogenised, 50µl of sterile water was added, suspensions were vortexed again, and 15µl was plated in quintuplicate onto an IR

Biotyper silicon target plate. Bruker infrared test standards (IRTS1 and IRTS2) were also applied to the plate for each run. Each plate was dried at 37 °C for 15 minutes before being analysed.

Spectra were acquired and processed by OPUS v7.5 software (Bruker GmbH) and analysed using the IR Biotyper software (version 3.0) with default analysis settings where the observed spectral range focuses on polysaccharide regions evaluating changes in the carbohydrate composition of bacterial cell walls (1,200-900 cm^{-1}). To determine appropriate cut-off values, 10-11 isolates for each species, expected to be different strains based on antibiograms and hospital location, were independently analysed in quintuplicate by three researchers on three different days. The isolates originated either from this study, pilot work conducted before this study in another hospital, or an unrelated UK Health Security Agency collection (two *K. pneumoniae* strains). Details on isolates used for the determination of IR Biotyper cut-off values for cluster assignation can be found in Table S1. The cut-off values determined (0.198 for *E. cloacae*, 0.191 for *C. freundii* and 0.291 for *K. pneumoniae*) were subsequently applied to the dendrograms to assess clustering.

Dendrograms were built using Euclidean distance and average linkage clustering with the IR Biotyper software (version 3.0). Dendrograms for each species were built for the overall collection of isolates (overall) and for each individual hospital (hospital-level) using Euclidean metric distance and the average linkage method.

Illumina sequencing

Isolates were recovered from beads and DNA was extracted using QIAamp DNA Mini Kit. Libraries were prepared for MiSeq sequencing using the Illumina DNA Prep protocol (Illumina, Document # 1000000025416 v09, June 2020). Manual library normalization was performed to ensure even sample coverage, based on the library's DNA concentration and average size, as measured by the Qubit and 2200 TapeStation. The samples were batched per flow-cell and paired-end sequencing was performed using the MiSeq reagent kit v3, with 2 × 300 bp with one water control on each run. Set A IDT for Illumina DNA/RNA UD Tagmentation Indexes were used.

Bioinformatics/statistics

Genomes were assembled using Shovill (v1.1.0) (https://github.com/tseemann/shovill) with default settings. For further analysis we included only those genomes with assembly size >4Mbp and <6.5Mbp that were also identified as belonging to either *K. pneumoniae*, *C. freundii* or *E. cloacae* using the MLST tool (https://github.com/tseemann/mlst). Raw reads were subsequently mapped to reference genomes (CP000647.1 *K. pneumoniae*, NZ_CP033466.1 *E. cloacae* and CP016762.1 *C. freundii*) using Snippy (v4.6.0) (https://github.com/tseemann/snippy). Whole genome SNP alignments (padded with the reference bases at invariant sites) were created using Snippy-core and recombination masking was performed using Verticall (v0.4.1) (https://github.com/rrwick/Verticall). Maximum likelihood trees were subsequently created using IQ-tree (v2.1.3, using the general time reversible model). We evaluated the ability of the IR Biotyper to correctly cluster isolates at doubling SNP thresholds from 10 to 1280, where in this analysis clusters derived

from WGS data were taken to be the ground truth. Overall correlations between pairwise SNP distances (derived from WGS data) and Euclidean distances (output from the IR Biotyper software as above) were assessed using Spearman correlation coefficients.

True positives were defined as those cases where at a given SNP threshold (*t*) the pairwise SNP distance between two isolates was ≤*t* and the isolates were clustered by the IR Biotyper; similarly true negatives were called where the pairwise SNP distance was >*t* and the isolates were not clustered by the IR Biotyper. False positives were those cases where two isolates were incorrectly clustered by the IR Biotyper but the pairwise SNP distance was >*t* and false negatives those cases where the isolates were incorrectly not clustered by the IR Biotyper but the pairwise SNP distance was ≤*t*.

We then calculated sensitivity and specificity using R version 4.1.0 [24]. Sensitivity was defined by calculating true positives/(true positives + false negatives), and specificity by calculating true negatives/(true negatives + false positives). Calculations were undertaken considering clusters derived using both the overall and hospital-level thresholds.

All data visualisations were produced using ggplot2 [25]. Computation was performed on the Oxford Biomedical Research Computing Cluster.

7. Results

Microbiology and sequence data processing

We sampled a total of 90 sinks across 9 hospitals, where *C. freundii* was cultured from 51 (57%) sinks, *E. cloacae* from 42 (47%) sinks, and *K. pneumoniae* from 28 (31%) sinks. All isolates originating from a different sample with a unique antibiotic resistance profile across the six antibiotics evaluated (i.e. 69 *E. cloacae*, 61 *C. freundii* and 32 *K. pneumoniae)* were typed using the IR Biotyper and subsequently sequenced; 54/69 (78%), 45/61 (74%) and 29/32 (91%) of *E. cloacae*, *C. freundii* and *K. pneumoniae* sequencing datasets passed QC and were analysed.

Isolate relatedness by IR Biotyper

The IR Biotyper identified 13, 7 and 3 clusters for *E. cloacae*, *C. freundii* and *K. pneumoniae* respectively when isolates were evaluated at the hospital-level. When all isolates for each species were evaluated overall and plotted on the same dendrogram, the number of clusters was 14, 12 and 4 respectively (Table S2). Clusters identified within hospitals were not always maintained in the overall dendrograms for *K. pneumoniae* (1 out of 3 clusters) and *E. cloacae* (3 out of 13 clusters).

Isolate relatedness by WGS

Median (IQR) pairwise SNP distances overall were 408,099 (170,587-580,301) for *E. cloacae*, 80,089 (52,774-95,009) for *C. freundii,* and 44,497 (41,898-302,451) for *K.*

pneumoniae highlighting the genetic diversity of these isolates. Genetic relatedness of isolates cultured varied by hospital (Table S3).

Overall correlation between IR Biotyper and SNP distances

For all three species there was a significant but relatively weak correlation between SNP distances (calculated from WGS) and Euclidean distances (from the IR Biotyper matrices; *E. cloacae* r^2 = 0.23 [95% CI: 0.18-0.28], p<0.001; *C. freundii* r^2 = 0.38 [95% CI: 0.32-0.43], p<0.001; *K. pneumoniae* r^2=0.15 [95% CI: 0.05-0.24], p=0.003, Figure S1.

Agreement between clusters derived by IR Biotyper versus whole genome sequencing

Whilst overall for all three species there was some evidence that pairwise distances between isolates clustered by the IR Biotyper were smaller than those not clustered (*K. pneumoniae* – median [IQR] SNP distance for clustered: 45,949 (27,088-305,875) vs non-clustered: 44,371 (41,935-301,982), p=0.67; *E. cloacae* 98,236 [23,494-489,956] vs 408,668 [174,781-583,287], p<0.001; *C. freundii* 595 [284-1,176] vs 88,739 [57,770-95,356], p<0.001), there were also multiple examples of genomically highly similar isolates inappropriately not clustered and genomically divergent isolates inappropriately clustered by the IR Biotyper (Figure 1).

There was substantial heterogeneity in the performance of the IR Biotyper across the three species included in this study. For example, at a ≤80 SNP cut-off, the IR Biotyper correctly identified 0/2 (0%) isolate pairs clustered by WGS for *K. pneumoniae* and 4/6 (67%) *C. freundii*. At this same threshold the IR Biotyper incorrectly clustered 8/63/102 isolates for *K. pneumoniae*/*E. cloacae*/*C. freundii* respectively which were found to be >80 SNPs apart in the WGS analysis. Similar patterns were observed for the whole range of SNP cut-offs, suggesting that it is unlikely that the arbitrary choice of SNP cut-off threshold to determine genetic relatedness effects the result (Table 1).

8. Discussion

The IR Biotyper is designed to provide rapid typing of isolates that can inform surveillance and implementation of infection prevention and control measures during outbreaks. Isolates recovered from the hospital environment could be processed alongside clinical isolates to investigate environmental reservoirs that might be contributing to transmission. This might also allow screening of isolates prior to WGS to save money by filtering out those which are clearly genetically divergent from the outbreak clade of interest. However, our findings suggest that the IR Biotyper has limited sensitivity for typing isolates recovered from hospital sinks.

Several previous studies have used the IR Biotyper to explore retrospectively the similarity among collections of Gram-negative clinical isolates collected during routine surveillance or outbreak investigations, often in intensive care units [15-22, 26] with different degrees of success. In these studies, comparisons have typically been against WGS, PFGE and/or MLST. In contrast to WGS however, MLST is a relatively low-resolution typing approach analysing variation in only small numbers of housekeeping genes, and given that isolates with the same ST can be substantially genetically divergent at the nucleotide-level, using this as a reference standard has limitations. For those studies using WGS as the reference, there was some degree of correlation between the techniques, but FT-IR had more limited

discriminatory power [15, 16, 18-21] including amongst highly genetically related isolates obtained longitudinally from the same patients [20, 21]. Another study found that two isolates of *Klebsiella pneumoniae* with a SNP distance of 27,006 formed part of the same IR Biotyper cluster [16]. Hypotheses for this lack of concordance include the impact of culture conditions and the presence of a mucoid phenotype [16, 21], which can be common amongst Enterobacterales. Wang-Wang et al., [19] have recently argued that use of the IR Biotyper should be compared to conventional epidemiology, which constitutes the most common outbreak investigation tool in many healthcare settings. In their study, using WGS as the gold standard, the IR Biotyper inferred more true genomic relations than conventional epidemiology [19].

The manufacturer recommends determining clustering cut-off thresholds locally in each laboratory, to account for variability associated with the local conditions (media, incubator, operators etc.). For this evaluation, we selected isolates based on phenotypic information (MALDI-ToF identification and antibiotic disk diffusion assays). In contrast, some earlier studies have calibrated IR Biotyper clustering thresholds against genotypic methods such as WGS [19, 20] or MLST [17, 22]. Whether cut-off values determined for one isolate collection are appropriate for subsequent typing of new isolates in the same institution remains to be determined but Vogt et al., [20] observed lower concordance between WGS and IR Biotyper outputs when a larger collection of *E. cloacae* isolates from the same group of patients was considered (239 vs. 24 isolates used in the initial evaluation).

Although we standardized the culture conditions, we did not undertake any bespoke optimization or selection of different features of the spectra, which might improve the performance of the IR Biotyper outputs, seeking rather to evaluate this assay in a pragmatic, “out-of-the-box” way to mimic the approach most users would have. Strengths of this study include an evaluation of the sensitivity and specificity of the IR Biotyper on three species at different thresholds of genetic relatedness (i.e. different SNP thresholds). We have also evaluated performance when considering isolates clustered at the hospital-level, or clustered across all hospitals, with some differences, mimicking what users might wish to do in practice (i.e. evaluating intra- versus inter-hospital outbreaks). However, we were not able to assess the performance of the IR Biotyper in an epidemiologically defined outbreak setting and/or include clinical and environmental isolates associated with an outbreak as part of this study.

A limitation of studies to date, including this one, is the relatively low number of isolates included per species, and the relatively low number of highly genetically related isolates (by WGS). For *K. pneumoniae* the size of previous studies has ranged from 16 to 68 [15-19]) and for *E. cloacae* from 23 to 239 [17, 20, 21]; we found no published peer-reviewed studies applying the IR Biotyper to *C. freundii* isolates. Moreover, the inclusion of Gram-negative isolates recovered from the hospital environment in Fourier transform infrared spectroscopy analysis has not been frequent. The greater diversity of isolates we observed here may adversely affect performance metrics; however, typing methods need to be able to correctly rule-in genetically similar and rule-out genetically divergent cases. Many studies have shown that sinks contribute to nosocomial Enterobacterales transmission networks, and so being able to characterise isolates in these reservoirs accurately is important [27, 28]. However, previous studies have rarely incorporated isolates from the hospital environment. Dieckmann et al., [29] analysed ten bacterial isolates of *Klebsiella oxytoca* recovered from two lots of an intrinsically contaminated liquid hand soap product and showed that all isolates clustered together, concordant with WGS. Wang-Wang et al., [19] included two environmental isolates in their analysis (one from a mattress and one from a sink) which did not cluster with any of

the clinical isolates. This is the first study to include a collection of isolates representing the diversity of Enterobacterales from hospital sinks.

In conclusion, we demonstrate in this study that, when compared to WGS as a gold standard and for the three species investigated, the IR Biotyper is not able to reliably cluster isolates when using WGS-derived genetic relatedness as the reference standard. This might limit its suitability to investigate outbreaks, particularly in cases associated with hospital sink reservoirs.

9. Figures and tables

Figure 1. Pairwise isolate SNP distance distributions by species for isolates clustered by the IR Biotyper at the hospital-level (top panel) and overall (bottom panel). Note SNP distances on the x-axis are on a log scale and these scales are different for each species.

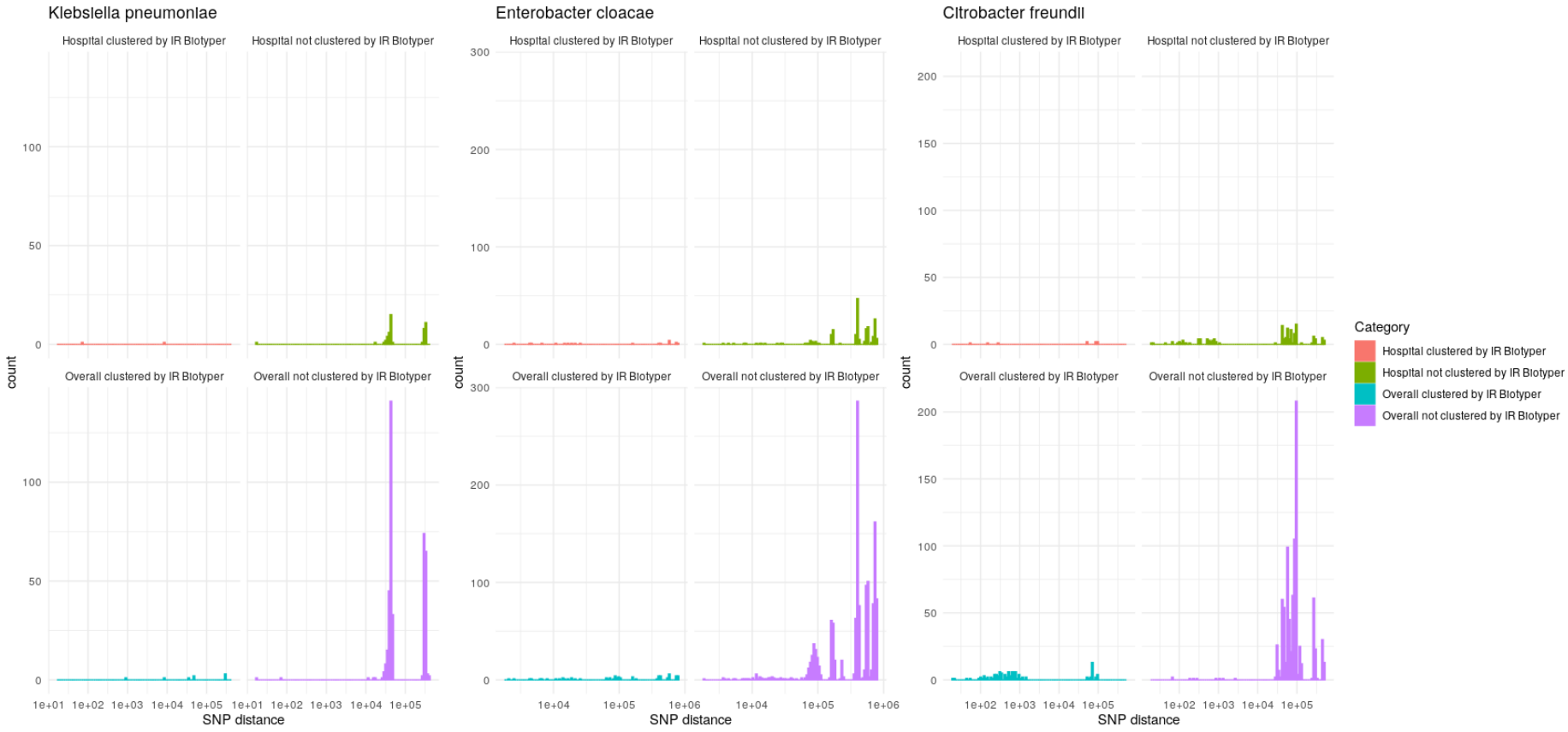

Table 1 - Performance characteristics of clustering using the IR Biotyper compared to WGS. At each SNP threshold we calculated the true positives (pair of isolates clustered by IR Biotyper and WGS), false positives (clustered by IR Biotyper but not WGS), true negatives (not clustered by either method) and false negatives (clustered by WGS but not IR Biotyper). Sensitivity (TP/(TP+FN)) and specificity (TN/(TN+FP)) were subsequently estimated. The “overall” IR Biotyper clusters where used for this analysis.

SNP threshold	Number of clusters defined by WGS	Number of isolate pairs clustered at SNP threshold	Number of isolate pairs not clustered at SNP threshold	True Positives	False Positives	True Negatives	False Negatives	Sensitivity	Specificity
K. pneumoniae n = 29									
10	0	0	406	NA	8	398	NA	NA	0.98
20	1	1	405	0	8	397	1	0	0.98
40	1	1	405	0	8	397	1	0	0.98
80	2	2	404	0	8	396	2	0	0.98
160	2	2	404	0	8	396	2	0	0.98
320	2	2	404	0	8	396	2	0	0.98
640	2	2	404	0	8	396	2	0	0.98
1280	3	3	403	1	7	396	2	0.33	0.98
E. cloacae n = 54									
10	0	0	1431	NA	63	1368	NA	NA	0.96
20	0	0	1431	NA	63	1368	NA	NA	0.96
40	0	0	1431	NA	63	1368	NA	NA	0.96
80	0	0	1431	NA	63	1368	NA	NA	0.96
160	0	0	1431	NA	63	1368	NA	NA	0.96
320	0	0	1431	NA	63	1368	NA	NA	0.96
640	0	0	1431	NA	63	1368	NA	NA	0.96
1280	0	0	1431	NA	63	1368	NA	NA	0.96
C. freundii n = 45									
10	0	0	990	NA	106	884	NA	NA	0.89
20	0	0	990	NA	106	884	NA	NA	0.8
40	1	2	988	2	104	884	0	1	0.89
80	2	6	984	4	102	882	2	0.67	0.9

160	4	16	974	14	92	882	2	0.88	0.91
320	5	38	952	33	73	879	5	0.87	0.92
640	6	63	927	58	48	879	5	0.92	0.95
1280	9	88	902	80	26	876	8	0.91	0.97

361

Author statements

9.1 Author contributions

P A-B, CC, GM, SIL, and NS designed the study; all authors participated in refinements of the study design. P A-B, CC and GM undertook the sampling, microbiology work and the IR Biotyper lab workflows. GR sequenced the isolates. P A-B, GM, PR, NS and SIL designed the analyses, which were performed by P A-B and SIL. P A-B and SIL drafted the first version of the manuscript which was refined by NS; all authors reviewed and improved the final draft.

9.2 Conflicts of interest

The authors acknowledge Bruker for loaning an IR Biotyper instrument for the purposes of this analysis and for technical support during the course of the study; Bruker had no role in the design of the study or the interpretation of the results.

9.3 Funding information

This work was funded by the National Institute for Health Research Health Protection Research Unit (NIHR HPRU) in Healthcare-Associated Infections and Antimicrobial Resistance at the University of Oxford in partnership with UK Health Security Agency (UKHSA) [grant NIHR200915]. The computational aspects of this research were funded from the NIHR Oxford BRC with additional support from a Wellcome Trust Core Award Grant [grant 203141/Z/16/Z]. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, the Department of Health or the UK Health Security Agency. NS is an NIHR Oxford BRC Senior Research Fellow and an Oxford Martin Fellow.

9.4 Ethical approval

No ethical approval was required for this study.

9.5 Acknowledgements

We are grateful to the hospital sites who permitted sampling of the sink drains.

10.References

1. **Haque M, Sartelli M, McKimm J, Abu Bakar M**. Health care-associated infections - an overview. *Infect Drug Resist* 2018;11:2321-2333.
2. **Kanamori H, Weber DJ, Gergen MF, DiBiase LM, Sickbert-Bennett EE et al.** Epidemiologic characteristics of health care–associated outbreaks and lessons learned from multiple outbreak investigations with a focus on the usefulness of routine molecular analysis. *American Journal of Infection Control* 2018;46(8):893-898.
3. **Sood G, Perl TM**. Outbreaks in Health Care Settings. *Infect Dis Clin North Am* 2016;30(3):661-687.
4. **Gastmeier P, Stamm-Balderjahn S, Hansen S, Nitzschke-Tiemann F, Zuschneid I et al.** How outbreaks can contribute to prevention of nosocomial infection: analysis of 1,022 outbreaks. *Infection control and hospital epidemiology* 2005;26(4):357-361.

5. **Suleyman G, Alangaden G, Bardossy AC**. The Role of Environmental Contamination in the Transmission of Nosocomial Pathogens and Healthcare-Associated Infections. *Current infectious disease reports* 2018;20(6):12.
6. **Parkes LO, Hota SS**. Sink-Related Outbreaks and Mitigation Strategies in Healthcare Facilities. *Current infectious disease reports* 2018;20(10):42.
7. **Nutman A, Marchaim D**. How to: molecular investigation of a hospital outbreak. *Clinical Microbiology and Infection* 2019;25(6):688-695.
8. **Decraene V, Phan HTT, George R, Wyllie DH, Akinremi O et al.** A large, refractory nosocomial outbreak of Klebsiella pneumoniae carbapenemase (KPC)-producing Escherichia coli demonstrates carbapenemase gene outbreaks involving sink sites require novel approaches to infection control. *Antimicrobial agents and chemotherapy* 2018.
9. **Gentile B, Grottola A, Orlando G, Fregni Serpini G, Venturelli C et al.** A Retrospective Whole-Genome Sequencing Analysis of Carbapenem and Colistin-Resistant Klebsiella Pneumoniae Nosocomial Strains Isolated during an MDR Surveillance Program. *Antibiotics (Basel)* 2020;9(5):246.
10. **Parcell BJ, Oravcova K, Pinheiro M, Holden MTG, Phillips G et al.** Pseudomonas aeruginosa intensive care unit outbreak: winnowing of transmissions with molecular and genomic typing. *The Journal of hospital infection* 2018;98(3):282-288.
11. **Sundermann AJ, Chen J, Miller JK, Saul MI, Shutt KA et al.** Outbreak of Pseudomonas aeruginosa Infections from a Contaminated Gastroscope Detected by Whole Genome Sequencing Surveillance. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 2021;73(3):e638-e642.
12. **Mellmann A, Bletz S, Böking T, Kipp F, Becker K et al.** Real-Time Genome Sequencing of Resistant Bacteria Provides Precision Infection Control in an Institutional Setting. *J Clin Microbiol* 2016;54(12):2874-2881.
13. **Sherry NL, Gorrie CL, Kwong JC, Higgs C, Stuart RL et al.** Multi-site implementation of whole genome sequencing for hospital infection control: A prospective genomic epidemiological analysis. *The Lancet regional health Western Pacific* 2022;23:100446.
14. **Novais Â, Freitas AR, Rodrigues C, Peixe L**. Fourier transform infrared spectroscopy: unlocking fundamentals and prospects for bacterial strain typing. *European journal of clinical microbiology & infectious diseases : official publication of the European Society of Clinical Microbiology* 2019;38(3):427-448.
15. **Dinkelacker AG, Vogt S, Oberhettinger P, Mauder N, Rau J et al.** Typing and Species Identification of Clinical Klebsiella Isolates by Fourier Transform Infrared Spectroscopy and Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry. *J Clin Microbiol* 2018;56(11).
16. **Hu Y, Zhou H, Lu J, Sun Q, Liu C et al.** Evaluation of the IR Biotyper for Klebsiella pneumoniae typing and its potentials in hospital hygiene management. *Microbial Biotechnology* 2021;14(4):1343-1352.
17. **Martak D, Valot B, Sauget M, Cholley P, Thouverez M et al.** Fourier-Transform InfraRed Spectroscopy Can Quickly Type Gram-Negative Bacilli Responsible for Hospital Outbreaks. *Frontiers in microbiology*, Original Research 2019;10.
18. **Rakovitsky N, Frenk S, Kon H, Schwartz D, Temkin E et al.** Fourier Transform Infrared Spectroscopy Is a New Option for Outbreak Investigation: a Retrospective Analysis of an Extended-Spectrum-Beta-Lactamase-Producing Klebsiella pneumoniae Outbreak in a Neonatal Intensive Care Unit. *J Clin Microbiol* 2020;58(5).
19. **Wang-Wang JH, Bordoy AE, Martró E, Quesada MD, Pérez-Vázquez M et al.** Evaluation of Fourier Transform Infrared Spectroscopy as a First-Line Typing Tool for the Identification of Extended-Spectrum β-Lactamase-Producing Klebsiella pneumoniae Outbreaks in the Hospital Setting. *Frontiers in microbiology* 2022;13:897161.
20. **Vogt S, Löffler K, Dinkelacker AG, Bader B, Autenrieth IB et al.** Fourier-Transform Infrared (FTIR) Spectroscopy for Typing of Clinical Enterobacter cloacae Complex Isolates. *Frontiers in microbiology* 2019;10:2582.

21. **Wendel AF, Peter D, Mattner F, Weiss M, Hoppenz M et al.** Surveillance of Enterobacter cloacae complex colonization and comparative analysis of different typing methods on a neonatal intensive care unit in Germany. *Antimicrobial resistance and infection control* 2022;11(1):54.
22. **Oho M, Nagasawa Z, Funashima Y, Ueda O, Watamabe S et al.** Correlation of Strain Classification with IR Biotyper and Molecular Epidemiological Method of Pseudomonas aeruginosa. *Rinsho Biseibutsu Jinsoku Shindan Kenkyukai shi = JARMAM : Journal of the Association for Rapid Method and Automation in Microbiology* 2021;31(1):29-40.
23. **Matuschek E, Brown DFJ, Kahlmeter G**. Development of the EUCAST disk diffusion antimicrobial susceptibility testing method and its implementation in routine microbiology laboratories. *Clinical Microbiology and Infection* 2014;20(4):O255-O266.
24. **R Core Team**. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/. 2019.
25. **Wickham H**. *ggplot2: Elegant Graphics for Data Analysis. https://ggplot2.tidyverse.org* 2016.
26. **Lombardo D, Cordovana M, Deidda F, Pane M, Ambretti S**. Application of Fourier transform infrared spectroscopy for real-time typing of Acinetobacter baumannii outbreak in intensive care unit. *Future Microbiol* 2021;16:1239-1250.
27. **Kizny Gordon AE, Mathers AJ, Cheong EYL, Gottlieb T, Kotay S et al.** The Hospital Water Environment as a Reservoir for Carbapenem-Resistant Organisms Causing Hospital-Acquired Infections-A Systematic Review of the Literature. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 2017;64(10):1435-1444.
28. **Volling C, Ahangari N, Bartoszko JJ, Coleman BL, Garcia-Jeldes F et al.** Are Sink Drainage Systems a Reservoir for Hospital-Acquired Gammaproteobacteria Colonization and Infection? A Systematic Review. *Open Forum Infect Dis* 2021;8(2):ofaa590.
29. **Dieckmann R, Hammerl JA, Hahmann H, Wicke A, Kleta S et al.** Rapid characterisation of Klebsiella oxytoca isolates from contaminated liquid hand soap using mass spectrometry, FTIR and Raman spectroscopy. *Faraday Discussions*, 10.1039/C5FD00165J 2016;187(0):353-375.