Rapid clinical diagnosis and treatment of common, undetected, and uncultivable bloodstream infections using metagenomic sequencing from routine blood cultures with Oxford Nanopore

Kumeren N. Govender,[1] David W. Eyre,[1,2]

1. Nuffield Department of Medicine, John Radcliffe Hospital, University of Oxford, Oxford, UK

2. Big Data Institute, Nuffield Department of Population Health, University of Oxford, UK

Correspondence to: Dr Kumeren Govender, Nuffield Department of Medicine, John Radcliffe Hospital, University of Oxford, Oxford, UK, **kumeren.govender@ndm.ox.ac.uk.**

Keywords: metagenomics, taxonomic classification, infection, diagnosis, whole-genome sequencing

Abstract

Metagenomic sequencing could potentially revolutionise clinical microbiology, including identifying pathogens and predicting antimicrobial resistance (AMR) in seriously unwell patients with bloodstream infections. We present a direct-from-blood culture sequencing method that delivers results rapidly and accurately. In 273 randomly-selected blood cultures (211 positive, 62 negative), we achieved 97% sensitivity and 94% specificity (improving to 100% accounting for plausible additional infections) for species identification compared to established diagnostic methods. We detected 18 additional infections—13 polymicrobial and 5 previously unidentifiable—and delivered findings in 3.5 hours, nearly a third of time taken by routine methods. For the top ten pathogens, our method produced AMR results 20 hours quicker, with 88% sensitivity and 93% specificity. For *Staphylococcus aureus* and *Escherichia coli*, AMR prediction sensitivity was 100% and 91%, and specificity 99% and 94% respectively. These findings suggest the potential of integrating clinical metagenomics into standard diagnostics for faster and more comprehensive pathogen detection.

Introduction

Bloodstream infections (BSIs) represent a significant global health concern, affecting over 49 million individuals annually and leading to approximately 11 million sepsis-related deaths.[1] Early diagnosis and rapid administration of appropriate antibiotic therapy is crucial; delayed or inactive antibiotics lead to significant increases in mortality rates.[2] However, empirical antibiotic treatment can be ineffective in 20-34% of cases, with emerging antimicrobial resistance (AMR) exacerbating this problem.[3–5] Blood culture, first described in the mid 19th century, and widely automated, but largely unchanged, since the 1970s, is recognised as the standard of care for accurately identifying the causative pathogen and its antimicrobial susceptibility, essential for ensuring targeted and effective antibiotic treatment. However, blood cultures are negative in up to two thirds of probable sepsis cases due to a variety of factors including low numbers of organisms in the bloodstream, pre-exposure to antibiotics, non-infectious inflammatory disorders, or fastidious or atypical organisms.[6,7] Moreover, blood culture is slow relative to the pace of illness in severe infections; it requires 24-72 hours to yield results, whereas in severe sepsis each hour delay in appropriate antibiotic administration may increase in mortality by 8%.[8,9]

Molecular methods offer a promising solution to the limitations of traditional culture techniques.[10] Although multiplex PCR panels are fast, achieving results within 1-4 hours from positive blood cultures, this technology is limited by using a finite array of gene markers limiting the breath of species and resistance markers identifiable, which must also be updated as new resistance emerges.[11–14] Most platforms start with positive blood cultures; with only one FDA-approved platform that detects infection directly from patient blood samples (T2 Biosystems), detecting 6 bacterial pathogens in 3-5 hours.[15]

In contrast, clinical metagenomics, an unbiased, sequencing-based method capable of detecting nearly all organisms within a sample, potentially offers a more comprehensive diagnostic tool.[16] It overcomes the limitations of multiplex PCR by not relying on predefined gene markers, thereby broadening the spectrum of detectable pathogens and resistance mechanisms.[17]

Methods for sequencing bacterial DNA direct from positive blood cultures on Illumina MiSeq and Nanopore MinION systems have been previously described, identifying specific Gram-negative and *Staphylococcus sp*. species and limited AMR-encoding genes within 4 hours in one study[18,19] and accurate monomicrobial species prediction within 9-17 hours in another.[20] However, such studies are relatively underpowered, limited to spiked samples or enriched for specific organisms, or have reported only AMR gene detection without assessment against phenotypic resistance. Another approach, using microbial cell-free DNA (cfDNA) sequencing, performed directly from blood samples without needing any culture, can identify a diverse array of species but needs costly high-throughput sequencing per sample, often yielding <0.01% of pathogen reads and has limited AMR predictivity.[21,22]

Here, we present an optimized, unified sample-to-result metagenomic pipeline employing Nanopore sequencing directly from positive blood cultures, using an unbiased sample selection to evaluate species identification, resistance detection, and time-to-result against a routine clinical microbiology laboratory in Oxford, UK, in real-time.

Results

Sample characteristics

We performed metagenomic sequencing using the Oxford Nanopore platform in 273 randomly-selected blood culture bottles (211 positive, 62 negative) (Figure 1). Of 211 positive blood cultures, 205 contained 216 species identified by the routine laboratory using MALDI-TOF (Bruker, USA), including 11 polymicrobial infections (species and antimicrobial susceptibilities listed in Table 1, Table S1). Routine methods failed to identify any species in the remaining 6 positive blood cultures. Details on negative, internal, and positive controls, and barcode demultiplexing performance are provided in the Supplementary Results.

Species identification

Unadjusted species performance

Samples were randomly split into training (60%) and testing (40%) sets for independent validation of species prediction performance. To identify species, three bioinformatic methods using Bracken with a standard database adjusted for contamination (see Methods; thresholds in Table S2) were tested. Method 1 excluded plasmid sequences, Method 2 added genome mapping and heuristic thresholds (number of species reads and percentage of bacterial reads from the species), and Method 3 used mapping with a random forest model-derived thresholds. Metagenomics species identification was compared to routine MALDI-TOF analysis, representing "naïve" performance. Sensitivity was defined on a per species basis, i.e., the proportion of all species found by MALDI-TOF across all samples also detected by sequencing. Specificity was calculated within the positive samples, as the proportion of samples with ≥1 extra species detected by sequencing. Sensitivity and specificity in the testing datasets for Methods 1, 2 and 3 were 92% (95%CI, 85-97) and 95% (88-99), 94% (87-98) and 94% (86-98), and 92% (84-97) and 91% (83-96) respectively (Table 2, thresholds available in Table S7). Method 2 was used for subsequent analyses.

False-negative results

Using Method 2, within both training and test datasets, there were 6 false-negative results where sequencing failed to detect a species found by MALDI-TOF. In four, sequencing identified a species similar to MALDI-TOF that was potentially a more accurate representation of the species (Table S3, Methods). The other two missed species (samples 143 and 147) contained reads from the species, but not sufficient to pass thresholds set for calling these samples positive; of note both samples were polymicrobial by culture.

We therefore investigated relaxing the stringency of filtering to optimise sensitivity, selecting new species identification thresholds (50 species reads, 0.5% of all bacterial reads) that achieved 100% sensitivity in training data while also maximising specificity (Figure S6). Using these thresholds, sensitivity and specificity in the test dataset were 100% (95% CI, 97-100%) [128/128] and 94% (88-97%) [115/123] respectively, and 98% (95-99%) [212/216] and 92% (88-95%) [189/205] in the whole dataset. The four remaining false-negative results were explained by MALDI-TOF/sequencing discrepancies described above.

False-positive results

Additional species detected by sequencing but not by routine culture and MALDI-TOF, were assessed for bioinformatic plausibility (see Methods). 16 samples had false-positive species on sequencing, including the four MALDI-TOF/sequencing discrepancies described earlier as also false-negatives. In the 12 other samples, 13 additional species were found using sensitivity-optimised thresholds (Table 3). These were considered biologically plausible and likely true infections (Table 3; Table S4). Mapping was used to confirm the taxonomic classification, 11/13 species had >75% genome coverage and depths between >3x and 143x.

The other two species, *Collinsella aerofaciens* (48% coverage, 3x depth) and *Bacteroides nordii* (68% coverage, 1.5x depth alongside *Bacteroides fragilis* and *Bacteroides thetaiotaomicron*) had lower sequence coverage, but could plausibly have been part of polymicrobial infections although could also represent difficulties classifying genetically similar species for *Bacteroides*.[23] Anaerobes were over-represented compared to their prevalence in cultures (5/13 species), suggesting limitations in current microbiological methods for detecting such organisms, e.g. *Bacteroides fragilis* was linked to liver abscess in sample 78.

None of the additional species had high ANIs (>95%) with those detected phenotypically, indicating their distinct nature. In samples 33, 39, and 129, low levels of *Pseudomonas aeruginosa*, *Collinsella aerofaciens*, and *Staphylococcus haemolyticus* were identified respectively, which routine methods missed in initial cultures but detected in subsequent tests. This suggests some organisms present at low levels in blood cultures may evade standard microbiological detection.

Reviewing thresholds accounting for plausible additional species

As we considered most or all of the additional species identified as plausible, we also tested further lowering thresholds for species detection (50 reads, 0.1%), selecting the lowest level that still accounted for most contamination in simulations. This adjustment resulted in 100% sensitivity and 99.5% specificity until a 0.4% cut-off, where both improved to 100% (Figure S7). A single false positive, likely due to *Staphylococcus aureus* bioinformatic misclassification with poor genome coverage, impacted lower-threshold performance (Table S6). Using these final filtering thresholds, sensitivity and specificity were both 100% in both the training and testing datasets (95%CI 97-100% and 96-100% respectively) (Table 2).

False negative and false positive Method 3 results

Results using Method 3 were similar. Before adjustment for plausible species, the seven false negative results included the 4 MALDI-TOF misclassifications previously described (Table S3), and 3 other missed species (Table S5). Within the seven false positive results in the testing data, four were likely due to MALDI-TOF inaccuracy (Table S3) and three had been previously assessed as being plausible. This approach failed to identify a result that was considered plausible in the threshold-based analysis (Table S5). The total number of reads was the most important feature in the random forest model trained with adjusted plausible species (Figure S5B). After retraining models accounting for plausible additional species, sensitivity and specificity were both 100% in both the training and testing datasets (95%CI 97-100% and 96-100% respectively), matching Method 2.

Culture negative samples

In all methods, overall unadjusted species-level specificity in culture-negative samples was 61/62 (98%; 95%CI 91-100%), and after adjusting for a plausible infection was 61/61 (100%; 94-100%) (Table 2). One sample, Sample N47 was negative by the routine laboratory, however, sequencing confirmed the presence of *Pseudomonas aeruginosa*, which was also present in repeat blood culture obtained 24 hours later from the same patient. The number of bacterial reads before filtering was otherwise consistently low (Figure S8).

Positive blood cultures samples unidentifiable by the routine lab

Six blood cultures flagged positive, indicating potential bacterial growth, but species identification was not achieved using routine techniques (Table 3). Sample 44 likely represents a false positive ‘flag’ due to the patient's acute myeloid leukemia, where highly

active white blood cells may have triggered the flag. MALDI-TOF provided only genus-level identification for sample 91, but sequencing identified *Acinetobacter variabilis*. Notably, sequencing uncovered organisms missed by conventional methods in samples 2, 122, 131, and 181, identifying them as *Turicibacter sanguinis*, *Faecalibacterium prausnitzii*, *Corynebacterium kroppenstedtii*, and *Capnocytophaga canimorsus*, respectively.

Time to species identification

We applied the threshold filtering method derived from the training dataset after sequencing ended (24 hours) to earlier sequencing time intervals, for instance, analysing data available at 10 minutes after sequencing started resulted in a sensitivity and specificity of 78% (95CI% 72-83) [178/229] and 100% (98-100) [205/205] respectively, while sensitivity improved to sequencing end (Figure 2A). Using dynamic filtering thresholds based on mapping and the time since sequencing started (Table S8), sensitivity and specificity were 98% (95-99) [224/229] and 100% (98-100) [205/205] respectively 10 minutes after starting sequencing. This improved to 100% (98-100) [229/229] and >99% (97->99%) [204/205] after 120 minutes, even in the presence of polymicrobial infections with a low read abundance of 2% (Figure 2B).

Antimicrobial resistance prediction

Amongst the ten most frequently isolated species, phenotypic resistance was present in 18% (224 resistant, 1029 sensitive) of drug-organism combinations in monomicrobial samples (Table 1). The predictive capacity of drug resistance was evaluated through a comparative analysis with four software and database tools in these top ten species.

Using assembled genomes from data generated by 24hrs of sequencing, ResFinder had the highest overall sensitivity and specificity of 84% (95%CI 78-88%, 188/224) and 96% (94-97%, 983/1029) respectively, while mCARD had the lowest, 76% (70-81%, 170/224) and 76% (74-79%, 786/1029) respectively (Table 4).

Using ResFinder and raw fastq reads available after 24 hours of sequencing as input yielded a higher sensitivity and specificity, 91% (95% CI: 87-94%, 204/224) and 90% (95% CI: 88-91%, 923/1029) respectively, compared to using assembled genomes. We also investigated ResFinder sensitivity and specificity, using raw reads available after 1 hour of sequencing, 82% (95% CI: 76-87%, 184/224) and 94% (95% CI:92-95%, 963/1029), and 2 hours of sequencing, 88% (82-92%, 196/224) and 93% (91-94%, 953/1029) respectively.

Performance by species

For *Escherichia coli*, sensitivity and specificity with ResFinder [Input fastq at 2hrs] was 91% [95%CI 80-97%] (53/58 drug-organism combinations) and 94% [92-96%] (505/539) respectively, and for *Staphylococcus aureus* was 100% [77-100%] (12/12) and 99% [96-100%] (138/139) respectively (see Figure 2C for other species).

Across species, ResFinder-based tools generally outperformed others, but RGI (CARD) showed better performance for *Klebsiella pneumoniae* and *Enterobacter cloacae*, with 100% sensitivity and 95-100% specificity, compared to 88% and 69% for ResFinder. AMRFinder also performed similarly well for these species. Sensitivity was low for *Proteus mirabilis* (17% across tools) and for coagulase-negative staphylococci (58-65% with AMRFinder and RGI). Specificity issues were present for *Pseudomonas aeruginosa* (15-74%). Sensitivity

could not be assessed for *Streptococcus pneumoniae* and *Pseudomonas aeruginosa* due to no resistant cases.

Performance by drug and drug class

ResFinder [fastq reads at 24 hrs] had the highest sensitivity across drug classes, achieving 100% for glycopeptides (12/12), amphenicols (1/1), macrolides (33/33), and steroid antibacterials (13/13), but lower for tetracyclines (83%, 5/6) and quinolones (71%, 5/7). AMRFinder and RGI (CARD) outperformed ResFinder for beta-lactams and folate pathway antagonists (Figure S9, S10, S11, see supplement for details of genes implicated in true and false calls).

Polymicrobial Performance

The overall sensitivity and specificity at the drug level using ResFinder [Input: fastq reads at 24hrs] for clinically relevant polymicrobial isolates was 96% (95% CI, 79-100%, 26/27) and 74% (56-89%, 26/35) respectively (Table S11). The polymicrobial sample containing *Enterococcus faecium* and *Staphylococcus haemolyticus* had the highest sensitivity and specificity performance of 100% (10/10) and 100% (4/4) respectively, while the sample containing *Escherichia coli* and *Providencia rettgeri* had the lowest, 0% (0/5) and 100% (9/9).

Time-point analysis vs routine laboratory

Our metagenomic workflow, reported species results within 3hrs 20min (IQR, 3hrs7min-3hrs27min) [Range, 2hrs45min-3hrs48min] from obtaining positive blood cultures, i.e. including time required for DNA extraction and library preparation as well as the sequencing

time. At this time point, after 10 minutes of sequencing, we achieved a sensitivity and specificity of 98% (95CI%, 95-99) [224/229] and 100% (98-100) [205/205], outperforming routine methods by ≥10 hours (Figure 2D). Additionally, our method offered actionable AMR results 20 hours earlier for the top ten pathogens, with 88% sensitivity and 93% specificity.

Cost analysis

The total consumable cost estimated per sample processed was $105 (Table S12). Multiplexing 22 samples, which would still result in a >20x coverage for a species identified, would cost $40 per sample, while multiplexing by 94, that would provide a species only result, would cost $20 (Table S13).

Discussion

For over a century, microscopy, culture, and immunodiagnostics have been the gold standards for infection diagnosis.[24] In this study, we present a direct-from-blood culture metagenomic workflow using Oxford Nanopore, enabling rapid pathogen identification and AMR prediction in only 3 hours and 30 minutes in a clinical setting using randomly selected samples. This approach significantly shortened the time to species identification by at least 10 hours and AMR prediction by 20 hours, while detecting 18 missed infections, including 13 undetected polymicrobial cases and 5 that were unidentifiable by routine methods.

A key finding of this study was the 100% sensitivity and specificity achieved in detecting plausible infections using metagenomic sequencing compared to standard microbiological techniques. Additional infections identified were supported by clinical, microbial, and bioinformatic evidence. This suggests metagenomic sequencing could reveal a broader

spectrum of pathogens involved in bloodstream infections than currently recognised by conventional diagnostic methods, including better detecting fastidious organisms, closely-related species causing co-infections, and minority infections that are hidden behind dominant growth in cultures. For example, sequencing identified *Bacteroides fragilis* in a patient with a liver abscess that the routine laboratory missed, *Enterococcus faecium* was detected as a co-infection by sequencing but only cultured from samples taken a day later, and in another sample *Pseudomonas aeruginosa* was detected with sequencing, but only cultured later in with the in a subsequent blood culture from the same patient. Notably, five out of the 13 additional species identified were anaerobes, suggesting that current diagnostic methods may not be fully optimised to grow and detect these organisms, and they may be a more common cause of infection than currently thought. However, whether additional information provided by metagenomics will consistently improve clinical outcomes remains to be determined. For instance, anaerobes found in bloodstream infections related to intra-abdominal infections may already be covered by empirical therapies. However, detected species such as *Capnocytophaga canimorsus* did lead to a change in antibiotic treatment during the study, underscoring the potential clinical relevance of these findings. Ultimately, randomised trials will be needed to determine whether metagenomic sequencing truly impacts patient outcomes.

Early detection of antimicrobial resistance through metagenomic sequencing, as demonstrated by the resistance patterns identified in our study, provides a critical advantage in managing bloodstream infections. Our method delivered AMR results 20 hours faster, with 88% sensitivity and 93% specificity for the top 10 pathogens. Specifically, for the common pathogens *Staphylococcus aureus* and *Escherichia coli*, sensitivity reached 100% and 91%, with specificities of 99% and 94%, respectively. Sequencing could play a key role identifying

or excluding resistance, enabling faster switches to targeted antibiotics, and reducing reliance on empirical broad-spectrum agents.

Our results are consistent with previous studies demonstrating the feasibility of nanopore-based sequencing for rapid pathogen identification and resistance profiling.[18,20] Our study builds upon this by extending the application of metagenomics to polymicrobial infections and incorporating a bioinformatic approach that significantly improves species identification and AMR prediction, even in complex clinical samples. By optimising filtering and species identification thresholds, we could enhance the detection of low-abundance species, which are often missed by traditional methods.

Despite these promising results, several challenges remain in integrating metagenomic sequencing into routine clinical practice. Contamination control, particularly in highly multiplexed workflows, remains a critical issue. Although our internal controls demonstrated minimal cross-contamination between samples, stringent filtering thresholds are required to avoid misclassification of species, especially in low-density infections. Additionally, the cost of sequencing, while rapidly decreasing, remains higher than that of traditional culture methods, particularly when performed at scale. However, multiplexing strategies, as used in this study, can reducing per-sample costs. Limitations in reference databases for species identification and resistance gene prediction can restrict performance, studies to improve their comprehensiveness are required. If sufficient performance can be achieved, further large-scale studies are also required to demonstrate performance for specific pathogens and antimicrobials in sufficient numbers to meet regulatory requirements.

In conclusion, metagenomic sequencing with Oxford Nanopore offers a potentially transformative approach to diagnosing bloodstream infections. Our findings demonstrate that this technology can provide rapid, accurate, and comprehensive pathogen identification and resistance profiling, significantly outperforming traditional culture-based methods in speed and sensitivity. With further validation in larger, multicentre clinical studies, metagenomics has the potential to become a cornerstone of precision diagnostics in infectious diseases, which ultimately may lead to improved patient outcomes by enabling timely, targeted therapies.

Methods

A workflow was developed to deliver direct-from-blood culture species identification and antimicrobial resistance prediction using Nanopore metagenomic sequencing. Detailed sample processing time was recorded to be compared to the conventional microbiology workflow.

Sample collection and processing

We processed a total of 211 positive blood cultures and 62 negative blood cultures (de-duplicated to one BSI isolate per 30-day period per patient) between November 9, 2020, and November 3, 2022. Samples were received by the clinical microbiology laboratory of Oxford University Hospitals (OUH) at the John Radcliffe Hospital, Oxford, UK, and processed by the laboratory using standard operating procedures. Six consecutively samples were selected to be processed in our workflow each day from those received by the OUH microbiology laboratory. Blood cultures were taken as part of routine clinical care. We processed samples with no pre-selection according to, e.g. speciation, in vitro antibiotic susceptibility, patient factors and hence we refer to this as a completely “unselected” or “random” sample population.

We did not process clinical samples that were collected using the BD BACTEC™ Peds Plus™ media. Samples that were due to be discarded were taken immediately after routine tests were completed. Therefore, individual patient consent was not required. This study was done with approval from the London – Queen Square Research Ethics Committee (REC reference 17/LO/1420). Although aerobic and anaerobic cultures are drawn together, only one was processed as this was the bottle that flagged positive during the sample collection

period. Collected blood culture bottles included both BD BACTEC™ Plus Aerobic and Lytic Anaerobic bottles (BD, USA). The former contains resin, and the latter lytic agents for the release of intracellular pathogens. The BD BACTEC™ FX blood culture system stored liquid blood culture bottles at 35 °C until positivity or negativity (defined as not flagging positive after five days). 1.5ml of liquid blood culture was removed and immediately processed.

DNA extraction from blood culture

Samples were extracted using the BiOstic Bacteraemia kit (MoBio, Qiagen, USA) according to the manufacturers' protocol with minor adjustments and no pre-treatment steps (Figure S1). Prior to extraction, two additional negative controls, and one positive control was added to each batch of 6 samples for processing, totalling 9 samples being processed in each batch. As a positive control for each batch, we used a single 5ul loop of *Arcanobacterium haemolyticum* grown from blood agar plate cultures. The culture broth from uninoculated BD BACTEC™ Plus Aerobic, and Lytic Anaerobic bottles were used as negative controls. Extracted DNA from clinical samples were normalised prior to library preparation, and DNA from positive controls were normalised to one-fourth of this concentration to ensure greater sequencing coverage depth of clinical samples. A spiked internal control of *Thermus thermophilus* DNA at a concentration of 2% of the normalised DNA concentration (with an absolute minimum amount of 0.5ng) was added to every clinical sample, the positive and one negative control, leaving the other negative control completely free from any DNA input. CFU counts were determined for every sample at the point of extraction using chocolate, blood and chromagenic agar plates.

Library preparation and Nanopore sequencing

Samples were prepared for GridION sequencing using the Rapid Sequencing Kit (SQK-RBK004) following the manufacturer's protocol. Sequencing was performed using R9.4.1 flow cells for 24h and were live basecalled with Guppy (version 3.2.6).

Bioinformatic analysis

Species prediction

GridION sequencing data was pre-processed using Guppy (version 6.4.8) to demultiplex barcodes and prinseq-lite (v0.20.4) to remove reads with low complexity and lengths <1000bp (thresholds based on our previous work).[25] Three different methods were used for bioinformatic species identification, all initially using the classification software Kraken2[26] (V2.0.7) and Bracken[27] (V2.5; specific settings available in Figure S2) using a full standard database built from archaeal, bacterial, plasmid, viral and human reference genomes downloaded from NCBIRefSeq on the February 13th, 2022. Reports were adjusted for bioinformatic contamination using thresholds set from simulation studies (contamination thresholds available in Table S2; method as previously described).[25]

Heuristic cut-off threshold models (Method 1 and 2) and random forest model (Method 3) were derived to identify the presence of true infection. We used 60% of the culture bottles processed as training/derivation data and the remaining 40% as the testing set (selected at random using the sample_frac() function in RStudio (V 2023.03.0+386)). The same split was used throughout all methods tested to ensure comparable results. For Methods 1 and 2, optimal thresholds for determining if samples contained low-level contamination or true infection from sequencing were determined by numerical optimization in the training set, choosing thresholds that maximised the Youden's Index (Sensitivity + specificity - 1). In Method 3 we trained a random forest classifier.

The first method removed plasmid reads prior to classification using a plasmid only Kraken2 database that was built (sequences downloaded from NCBI on the March 16th, 2023), as plasmids, due to horizontal gene transfer, may be classified into multiple species erroneously. To derive cut-off thresholds, the following filters were used for Method 1: read counts and percentage cut-off, that is, the total number of reads classified as a specific species, and the percentage of all bacterial reads classified as a specific species.

The second method did not remove plasmid reads, but instead used Minimap2 (v2.24-r1122; using the map-ont function) to map sequencing reads to each species identified by Bracken to a species-specific reference genome (downloaded randomly from NCBI) after minimal filtering (<0.1% percentage of reads) that excluded only the least plausible species before mapping. Samtools (v 1.17) was used to determine 3 additional metrics that were used for filtering: coverage breadth, depth of coverage, and percentage of reads mapped from extracted species reads. Heuristic thresholds were derived for all 5 filtering metrics from the training set and the applied to the test set. The latter three, coverage breadth, depth of coverage and percentage of reads were set to the lowest values for a true species identified in the training set and the remaining thresholds set optimising the Youden index as above. Following an initial evaluation, thresholds were retrained using a "corrected" training dataset, correcting for plausible false-positive and false-negative errors that were identified.

The third method, used the same 5 metrics as inputs, but applied a machine learning approach using a random forest model to classify Bracken results and mapping results derived from Minimap2 and Samtools. The randomForest package in R developed by A. Liaw et al was used.[28] Hyperparameter selection was made using a repeated cross-validation approach

within the training data with 5 folds and 40 repeats. A random forest model was trained using the training set, and its variable importance was assessed. The model was evaluated on the testing set.

Assessing false negative Method 2 results for plausibility (i.e., potential inaccurate sequencing results or inaccurate MALDI-TOF results)

To interpret MALDI-TOF results on cultured isolates, probability scores are used: a value of >2.3 indicates "highly probable species identification", a score >2 and <2.3 indicates "secure genus identification, probable species identification", a score >1.7 and <2 indicates "probable genus identification", and a score <1.7 indicates "unreliable identification". [29] In samples 22, 23, 46 and 49, MALDI-TOF predicted *Corynebacterium lipophiloflavum, Staphylococcus epidermidis, Klebsiella pneumoniae* and *Klebsiella pneumoniae (MALDI scores of 2.08, 2.07, 2.35, and 2.35 respectively)* while sequencing results predicted *Corynebacterium sanguinis, Staphylococcus hominis, Klebsiella quasipneumoniae* and *Klebsiella quasipneumoniae* respectively. Samples were repeat tested grown from saved stock cultures and yielded the same MALDI-TOF results as described. In all four samples, there was higher coverage breadth of the sequencing reference genome species than the MALDI-TOF reference genome species (93% vs 92.7%, 95% vs 63%, 62% vs 59% and 68% vs 62%, Table S3). Furthermore, almost all blast hits were classified to the species identified by sequencing (99.2%, 97%, 96% and 96% respectively) rather than the MALDI-TOF species. Prior studies have shown that *Corynebacterium sanguinis* and *Corynebacterium lipophiloflavum,* and *Staphylococcus epidermidis* and *Staphylococcus hominis* are genotypically similar and are often misidentified by MALDI-TOF.[30–32] Furthermore, the MALDI-TOF probability scores for these two organisms were 2.08 and 2.07 respectively, suggesting a reliable genus identification, but only probable species identification.

Klebsiella quasipneumoniae is a newly defined species, not readily distinguished from *Klebsiella pneumoniae* with current MALDI-TOF identification techniques, which has resulted in its true prevalence being unknown.[33] This is likely due to a lack of discrimination between spectra patterns being produced by these species. *Klebsiella quasipneumoniae* has been shown to be capable of nosocomial transmission and acquiring and maintaining relevant resistance elements, placing an increased necessity for accurate identification of this organism.[33] Biomerieux API (Analytical Profile Index) 20E testing was performed on samples 46 and 49 which were both sequenced as *Klebsiella quasipneumoniae,* as well as sample 60 which was sequenced as *Klebsiella pneumoniae* as a control for phenotype. Sample 60 was identified by the API test to have a high probability (i.e., % ID: 82) as being *Klebsiella pneumoniae*, while both samples 46 and 49 were identified with a much lower probability (i.e., % ID: 21) as being identified as *Klebsiella pneumoniae*. This difference in probability score was due to the two samples sequenced as *Klebsiella quasipneumoniae* having a negative LDC result, which tests for decarboxylations of the amino acid lysine by lysine decarboxylase. Furthermore, *Klebsiella quasipneumoniae* is not in the API 20E database, and therefore a much lower probability is expected.

Assessing false positive Method 2 results for plausibility (i.e., potential inaccurate sequencing results or inaccurate MALDI-TOF results)

False positive results, that is, additional species detected from sequencing not found on routine culture, were assessed for bioinformatic plausibility. Firstly, 1000 reads were extracted from additional species detected and queried using NCBI BLAST (v2.13.0+) with the ref_prok_rep database (downloaded March 12th, 2023) to see if the same species was found. Hits were filtered using percent identity>=95, coverage >=80, E-value <1e-50 and the

-max_target_seqs option set to 1. Secondly, for each additional species identified by Bracken, extracted reads were mapped using Minimap2 (v2.24-r1122; using the map-ont function) to a randomly selected refseq genome of the same species (downloaded from NCBI April 29th, 2023). Samtools (v1.17) was used to calculate genome coverage breadth and depth. To investigate the extent to which bioinformatic misclassification might arise between genomically similar species, the Average Nucleotide Index (ANI) was calculated for additional bioinformatic species found against the MALDI result using ANI.pl (downloaded reference genomes from NCBI April 29th, 2023).[34] Additional microbiology data was captured from other blood culture samples, if any were available, within 5 days of the initial collected blood sample to look for evidence of the additional species in these samples. Laboratory requests were also reviewed for relevance to additional infections found by sequencing. Provision for this is included in the ethical approvals cited above.

Statistical analysis

The unadjusted performance of metagenomic sequencing was assessed by comparing the species identified from sequencing data with the species from pure culture isolates grown from blood culture samples [species from Bruker microflex MALDI-TOF MS (Bruker, USA)], while the adjusted performance additionally considered plausible infections identified bioinformatically. Sensitivity was calculated taking each species identified from each culture-positive sample as a separate data point contributing to the numerator and denominator; therefore culture-positive samples with a single species contributed once, and culture-positive samples with two or more species contributed two or more times. Specificity was determined both by considering the proportion of samples deemed negative by blood culture that were identified as negative by metagenomic sequencing, and in culture-positive samples. In the latter samples false-positive results are still possible where an additional species not found by

culture was reported. In culture positive results, specificity was defined on a per sample basis, where any sample with >1 additional species found from sequencing was considered to contribute as 1 to the false negative count.

Antimicrobial resistance prediction

As a reference standard, we compared phenotypic antimicrobial susceptibility results from pure culture isolates performed in the OUH microbiology laboratory during each patient's admission [using the BD Phoenix microbroth dilution (BD, USA)] against genotypic resistance predictions. Results that were determined to be "intermediate" was excluded from the analysis. To ensure higher quality predictions and to keep plasmid sequences, sequencing data was pre-processed by filtering out human and *Thermus thermophilus* reads prior to resistance prediction using the Kraken Software Suite, rather than extracting only identified species reads.[35] Genotypic resistance prediction was performed with fasta assemblies [Flye v. 2.9-b1768 with '-meta' and '-nano-raw'] and raw reads when possible using ResFinder 4.0 (ResFinder & PointFinder db downloaded July 3rd, 2022), Resistance Gene Identifier (CARD downloaded August 18th, 2022), and NCBI AMRFinderPlus 3.10.40 (db downloaded July 30th, 2022). Given that ResFinder was the only tool capable of predicting resistance at the individual drug level, we adopted the approach of interpreting resistance at the class level. This approach assumes that if a gene within a particular class is identified as resistant, all drugs within that same class are presumed to exhibit resistance. Additionally, a curated mapping database (that maps genes found to genotypic resistance) from ResPipe, an open-source software pipeline, was also applied to RGI (CARD) hits to determine genotypic resistance predictions which is referred to as ResPipe (modified CARD, mCARD).[36]

RGI(CARD) was configured to excluded hits associated with the "antibiotic efflux" mechanism and the "protein variant model". The former, a nonspecific mechanism influenced by factors such as efflux pump expression levels, and the latter, a context-dependent model susceptible to variability among different bacteria and conditions, were filtered out to focus on more specific and reliable resistance mechanisms. We set stringent thresholds for best identities (>60), best hit bitscore (>33), and percentage length of reference sequence (>99) to ensure significant sequence similarity, high-quality alignment, and comprehensive coverage of the reference sequence, respectively.[37,38] For AMRFinder, we set a percentage coverage of the reference sequence to >85%, percentage identity to the reference sequence >95% and minimum target length >120bp to ensure that the identified sequences likely represent substantial portions or complete genes, rather than fragments, reducing the risk of false positives from partial gene matches.[37,38] We also filtered for the "AMR" (Antimicrobial Resistance) subtype, focusing the analysis on elements specifically associated with antimicrobial resistance and excluding irrelevant elements.

Of note, we used the *Staphylococcus aureus* rules across all *Staphylococcus* species to ensure a broader coverage and attempt detection of resistance elements. A sample was considered to be genotypically sensitive if no resistant genes were found, in samples with a genome sequencing depth of at least 10x at the end of sequencing. We presented categorical agreement (CA; classified in the same susceptibility category), very major error (VME; predicted susceptible but phenotypically resistant) and major error (ME; predicted resistant but phenotypically susceptible) rates. Monomicrobial and polymicrobial infections were analysed separately. In cases where discordant phenotypic susceptibility results existed in polymicrobial species, a genotypic result of resistant was required to correctly predict this phenotype, counting as a single drug prediction.

Barcode contamination

To allow for sequencing multiple samples in the same sequencing run, short recognisable sequences (barcodes) are added to DNA fragments during library preparation to tag samples separately. These barcodes are read and used to separate out the resulting sequencing data. However within the multiplex sequencing data, reads were assigned to barcodes not included in the experiments (defined as mis-assigned reads), and reads were assigned to the incorrect barcode bins in-use (referred to as cross-assigned).[39] In this study, we assessed both mis-assigned reads by Guppy demultiplexer and cross-assigned reads using the positive control *Arcanobacterium haemolyticum* as a proxy. *A. haemolyticum* was used as this was added only to a single positive control sample per run sequenced and this species has not been present in any of the MALDI-TOF results from this dataset.

Mis-assigned reads

As there are 12 barcodes available, and 11 are in use with barcode number 3 not in use, $number(barcode\ 3\ reads) = \left(1 - p(t)\right) \times \left[\frac{1}{11}\right] \times z$, where p(t): probability of a read being classified to the correct barcode and z: total number of reads. The term 1/11 reflects that when mis-assigned there are 11 possible barcodes that a read can be mis-assigned to. Therefore,

$$p(t) = 1 - [\frac{number(barcode\ 3\ reads)}{z} \times 11].$$

Now assuming there are more barcodes not in use (a1, a2, …). The equation then becomes:

$$p(t) = 1 - [\frac{number(bar.\ a1\ reads)}{z} + \frac{number(bar.\ a2\ reads)}{z} + \ldots]$$
$$\times \frac{(number\ of\ total\ barcodes\ in\ use - 1)}{(number\ of\ barcodes\ not\ in\ use)}.$$

Cross-assigned reads

The probability of reads cross-assigned was calculated by the number of *Arcanobacterium haemolyticum* reads not in the positive control barcode, but assigned to a barcode in use, divided by the total *Arcanobacterium haemolyticum* reads present in a run assigned to any barcode in use.

Data availability

All clinical sample sequence data are available via European Nucleotide Archive (ENA) under study accession number PRJEB81413.

Authors and contributors

All authors contributed to conceptualization, methodology, software, formal analysis, investigation, resources, data curation, and visualisation. Writing - original draft preparation: KG; writing - review and editing, supervision and funding: DWE.

Conflicts of interest

No author has a conflict of interest to declare.

Financial support

The study was supported by the National Institute for Health Research (NIHR) Biomedical Research Centre, Oxford. The views expressed in this publication are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health. DWE is a Robertson Foundation Fellow. KG is a Rhodes Scholar.

Figure legends

Figure 1. Routine microbiology laboratory vs. metagenomic study workflow. Blood cultures are drawn from patients with suspected blood stream infections and are incubated until culture positivity or negativity (defined as two weeks of no positivity). **1.** In the routine laboratory positive blood cultures undergo immediate gram-staining and are re-cultured and then often sub-cultured on a petri dish to obtain a pure culture. MALDI-ToF and antimicrobial susceptivity testing is then performed. **2.** In the metagenomic workflow, DNA is extracted directly from positive blood cultures for processing using the QIAmp BiOstic Bacteremia kit. In this study, 211 positive and 62 negative blood cultures were selected at random and processed with added positive, negative and internal controls. Sequencing was performed in batches of 6 samples per flow cell using the Oxford Nanopore GridION platform. Three bespoke methods to determine metagenomic species and three resistance prediction software tools were evaluated against routine phenotypic results.

Figure 2. Time to species identification and antimicrobial resistance prediction from metagenomics vs. routine microbiology lab. A. Species identification performance by time with a static threshold applied. **B.** Species identification performance by time with a dynamic threshold applied (Table S8). **C.** Top AMR performing tools at 2- and 24-hour sequencing intervals for top 10 species. AMR tools predicting resistance to isolates were categorically evaluated against phenotypic antimicrobial susceptibility testing (AST) results obtained from the BD Phoenix microbroth dilution (BD, USA) using pure isolates. Results that were determined to be "intermediate" were excluded from the analysis. **D.** Time to metagenomic species and AMR results vs. routine microbiology laboratory.

Tables

Table 1. Species and resistance phenotypic summary for top 10 species. The top 10 species are displayed with phenotypic resistance data shown by number of resistant isolates/(resistant + sensitive isolates) grouped by drug. See Table S1 for further details.

Table 2. Species performance results across all models. This table summarises results of models across all classification approaches used. Samples were randomly split into training (60%) and testing (40%) sets to ensure independent validation of species prediction approaches. Three different methods were used for bioinformatic species identification, all using the classification software Bracken with a standard database and adjusted for bioinformatic contamination using thresholds set from simulation studies.

Table 3. Species missed by routine phenotypic methods identified by sequencing pipeline. Metagenomic sequencing identifies missed co-infections, traditionally unidentifiable, low-level infections in culture-negative samples and confirms false positives.

Table 4. Genotypic performance of top 10 clinically relevant species by tool. Four AMR prediction tools were benchmarked including ResFinder, AMRFinderPlus, Resistance Gene Identifier (RGI), and RGI with a modified database. Resistance was called from genome assembly, raw reads from the end of the sequencing run and raw reads at earlier sequencing time intervals.

References

1. Rudd, K. E. *et al.* Global, regional, and national sepsis incidence and mortality, 1990–2017: analysis for the Global Burden of Disease Study. *The Lancet* **395**, 200–211 (2020).
2. Rhodes, A. *et al.* Surviving Sepsis Campaign: International Guidelines for Management of Sepsis and Septic Shock: 2016. *Intensive Care Med* **43**, 304–377 (2017).
3. Peters, C. *et al.* Initiation of inappropriate antimicrobial therapy results in a fivefold reduction of survival in human septic shock. *Elsevier* (2009) doi:10.1378/chest.09-0087.
4. Fitzpatrick, J. M. *et al.* Gram-negative bacteraemia; a multi-centre prospective evaluation of empiric antibiotic therapy and outcome in English acute hospitals. *Clinical Microbiology and Infection* **22**, 244–251 (2016).
5. Kadri, S. S. *et al.* Inappropriate empirical antibiotic therapy for bloodstream infections based on discordant in-vitro susceptibilities: a retrospective cohort analysis of prevalence, predictors, and mortality risk in US hospitals. *Lancet Infect Dis* **21**, 241–251 (2021).
6. Duncan, C. F., Youngstein, T., Kirrane, M. D. & Lonsdale, D. O. Diagnostic challenges in sepsis. *Curr Infect Dis Rep* **23**, 1–14 (2021).
7. Yang, L. *et al.* Comparison of clinical characteristics and outcomes between positive and negative blood culture septic patients: a retrospective cohort study. *Infect Drug Resist* 4191–4205 (2021).
8. Kumar, A. *et al.* Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock. *Crit Care Med* **34**, 1589–1596 (2006).
9. Tabak, Y. P. *et al.* Blood culture turnaround time in US acute care hospitals and implications for laboratory process optimization. *J Clin Microbiol* **56**, 10–1128 (2018).
10. Courtenay, M. *et al.* Tackling antimicrobial resistance 2019-2024–the UK's five-year national action plan. *J Hosp Infect* **101**, 426–427 (2019).
11. Enne, V. I. *et al.* Multicentre evaluation of two multiplex PCR platforms for the rapid microbiological investigation of nosocomial pneumonia in UK ICUs: the INHALE WP1 study. *Thorax* **77**, 1220–1228 (2022).

12. Payne, M. *et al.* Evaluation of the FilmArray Blood Culture Identification Panel compared to direct MALDI-TOF MS identification for rapid identification of pathogens. *J Med Microbiol* **67**, 1253–1256 (2018).
13. Yang, S. & Rothman, R. E. PCR-based diagnostics for infectious diseases: uses, limitations, and future applications in acute-care settings. *Lancet Infect Dis* **4**, 337–348 (2004).
14. Caméléna, F. *et al.* Multicenter evaluation of the FilmArray blood culture identification 2 panel for pathogen detection in bloodstream infections. *Microbiol Spectr* **11**, e02547-22 (2023).
15. Samuel, L. Direct-from-blood detection of pathogens: a review of Technology and Challenges. *J Clin Microbiol* **61**, e00231-21 (2023).
16. Govender, K. N., Street, T. L., Sanderson, N. D. & Eyre, D. W. Metagenomic sequencing as a pathogen-agnostic clinical diagnostic tool for infectious diseases: a systematic review and meta-analysis of diagnostic test accuracy studies. *J Clin Microbiol* **59**, e02916-20 (2021).
17. Govender, K. N. Precision pandemic preparedness: improving diagnostics with metagenomics. *J Clin Microbiol* **59**, 10–1128 (2021).
18. Anson, L. W. *et al.* DNA extraction from primary liquid blood cultures for bloodstream infection diagnosis using whole genome sequencing. *J Med Microbiol* **67**, 347–357 (2018).
19. Taxt, A. M., Avershina, E., Frye, S. A., Naseer, U. & Ahmad, R. Rapid identification of pathogens, antibiotic resistance genes and plasmids in blood cultures by nanopore sequencing. *Sci Rep* **10**, 7622 (2020).
20. Harris, P. N. A. *et al.* Rapid nanopore sequencing and predictive susceptibility testing of positive blood cultures from intensive care patients with sepsis. *Microbiol Spectr* **12**, (2024).
21. Blauwkamp, T. A. *et al.* Analytical and clinical validation of a microbial cell-free DNA sequencing test for infectious disease. *Nat Microbiol* **4**, 663–674 (2019).
22. Rossoff, J. *et al.* Noninvasive Diagnosis of Infection Using Plasma Next-Generation Sequencing: A Single-Center Experience. *Open Forum Infect Dis* **6**, (2019).
23. Wexler, H. M. Bacteroides: the good, the bad, and the nitty-gritty. *Clin Microbiol Rev* **20**, 593–621 (2007).
24. Schmitz, J. E., Stratton, C. W., Persing, D. H. & Tang, Y.-W. Forty years of molecular diagnostics for infectious diseases. *J Clin Microbiol* **60**, e02446-21 (2022).
25. Govender, K. N. & Eyre, D. W. Benchmarking taxonomic classifiers with Illumina and Nanopore sequence data for clinical metagenomic diagnostic applications. *Microb Genom* **8**, 000886 (2022).
26. Wood, D. E., Lu, J. & Langmead, B. Improved metagenomic analysis with Kraken 2. *Genome Biol* **20**, 1–13 (2019).
27. Lu, J., Breitwieser, F. P., Thielen, P. & Salzberg, S. L. Bracken: estimating species abundance in metagenomics data. *PeerJ Comput Sci* **3**, e104 (2017).
28. Liaw, A. & Wiener, M. Classification and regression by randomForest. *R news* **2**, 18–22 (2002).
29. Normand, A.-C. *et al.* Decision criteria for MALDI-TOF MS-based identification of filamentous fungi using commercial and in-house reference databases. *BMC Microbiol* **17**, 1–17 (2017).
30. Rosa, N. M., Penati, M., Fusar-Poli, S., Addis, M. F. & Tola, S. Species identification by MALDI-TOF MS and gap PCR–RFLP of non-aureus Staphylococcus, Mammaliicoccus, and Streptococcus spp. associated with sheep and goat mastitis. *Vet Res* **53**, 84 (2022).

31. Spanu, T. *et al.* Evaluation of matrix-assisted laser desorption ionization-time-of-flight mass spectrometry in comparison to rpoB gene sequencing for species identification of bloodstream infection staphylococcal isolates. *Clin. Microbiol. Infect.* **17**, 44–49 (2011).
32. Jaen-Luchoro, D. *et al.* Corynebacterium sanguinis sp. nov., a clinical and environmental associated corynebacterium. *Syst Appl Microbiol* **43**, 126039 (2020).
33. Mathers, A. J. *et al.* Klebsiella quasipneumoniae provides a window into carbapenemase gene transfer, plasmid rearrangements, and patient interactions with the hospital environment. *Antimicrob Agents Chemother* **63**, e02513-18 (2019).
34. Richter, M. & Rosselló-Móra, R. Shifting the genomic gold standard for the prokaryotic species definition. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 19126–19131 (2009).
35. Lu, J. *et al.* Metagenome analysis using the Kraken software suite. *Nat Protoc* **17**, 1–25 (2022).
36. Gweon, H. S. *et al.* The impact of sequencing depth on the inferred taxonomic composition and AMR gene content of metagenomic samples. *Environ Microbiome* **14**, 7 (2019).
37. Neuert, S. *et al.* Prediction of phenotypic antimicrobial resistance profiles from whole genome sequences of non-typhoidal Salmonella enterica. *Front Microbiol* **9**, 592 (2018).
38. Day, M. R. *et al.* Comparison of phenotypic and WGS-derived antimicrobial resistance profiles of Salmonella enterica serovars Typhi and Paratyphi. *J. Antimicrob. Chemother*. **73**, 365–372 (2018).
39. Xu, Y. *et al.* Detection of viral pathogens with multiplex nanopore MinION sequencing: Be careful with cross-Talk. *Front Microbiol* **9**, (2018).

Rapid clinical diagnosis and treatment of common, undetected, and uncultivable bloodstream infections using metagenomic sequencing from routine blood cultures with Oxford Nanopore: supplementary results

Controls

Negative control: No spike

Two negative controls were added to each sequencing flow cell, 30 in total. In all negative control samples with no spike added and no bioinformatic filtering, total reads did not exceed 125 (Figure S16).

Negative control: *Thermus thermophilus* spike

When *Thermus thermophilus* DNA spikes were added to negative controls (targeting 2% of the normalised DNA concentration [with an absolute minimum amount of 0.5ng]) in each sample, >96% of the classified reads were from this species, except for one sample (DFB_R_22) which had a low total read number of 150 (Figure S17). After applying filtering thresholds (reads>50, percent reads>0.4%), all samples were classified as having >98% of the spike added, while filtering by reads>150 and percent reads>0.4% resulted in 100% of the spiked in species in all samples.

Negative control: Increasing *Thermus thermophilus* spike levels

In samples 15 and 16, an additional negative control of *Thermus thermophilus* spiked at 20% of the individual sample DNA concentration was added to assess the effect of concentration on contamination. Prior to any filtering, increasing the concentration of *T. thermophilus* DNA from 0% to 0.5% and to 20% improved contamination control and resulted in 66.7%, 99.87% and 99.92% and 100%, 99.89% and 99.97% of reads classified as *T. thermophilus* in both samples respectively (Figure S18).

Positive control

Positive controls which consisted of *Arcanobacterium haemolyticum* were present in every run. Unfiltered positive controls had >99.8% of reads across all samples classified as *A. haemolyticum*. After applying filtering thresholds (percent>0.4, reads>1000, coverage_breadth>32, reads_mapped>75, mean_depth>0.4), all positive controls had 100% of reads classified to *A. haemolyticum* in all samples (Figure S19).

Internal spike control present in positive controls

The internal spike of *Thermus thermophilus* DNA was added to all positive controls (2% of the normalised DNA concentration (with an absolute minimum amount of 0.5ng)). The median percent of *T. thermophilus* reads was 2.4% (1.7-4) [0.9-14] in positive control samples (Figure S20).

Internal spike control present in culture positive and culture negative clinical samples

The internal spike of *Thermus thermophilus* was present in all culture positive and culture negative clinical samples. The median percentage of *T. thermophilus* reads was 3.6% (2-8) [0.7-58] in culture positive clinical samples (Figure S21), and 100% (100-100) [4.5-100] in culture negative samples (Figure S22).

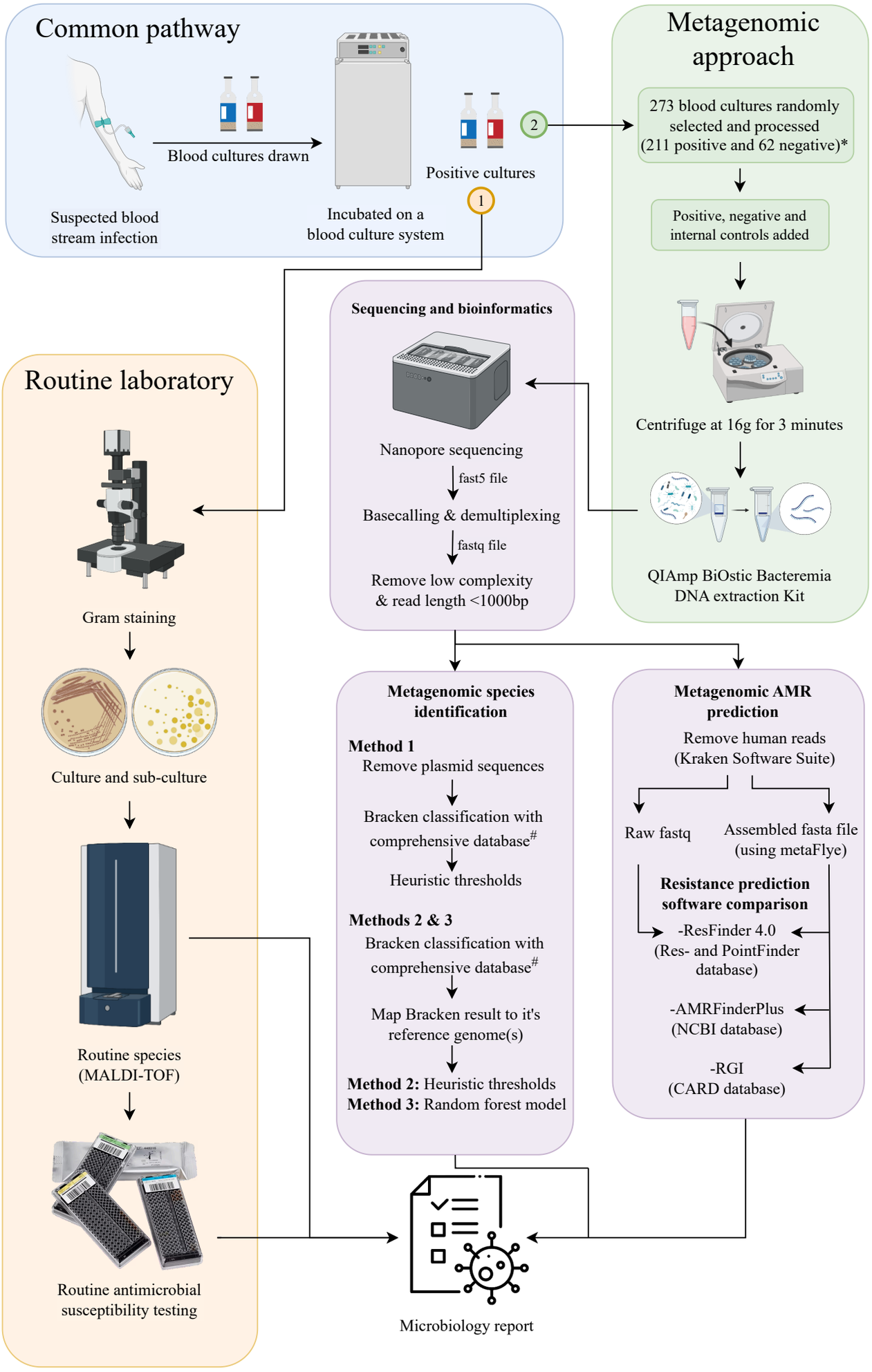
Barcode contamination

Within each flow cell we estimated rates of mis-assigned reads (assigned to unused barcodes) and cross-assigned reads (assigned to incorrect in-use barcodes). The median probability for a read to be mis-assigned was 0.000054 (IQR, 0.000008-0.000129) [Range, 0-0.000379] for unfiltered reads and 0.000050 (0.000008-0.000119) [0-0.000372] for filtered reads, equating to 1 misassigned read per ~18,519 unfiltered and ~20,000 filtered reads (Figure S23). For cross-assigned reads, the median probability was 0.000028 (0.000019-0.000039) [0.000009-0.000199] (Figure S24), and in negative samples, it was 0.000021 (0.000010-0.000026) [0.000008-0.000052] (Figure S25). This translates to 1 cross-assigned read per ~35,714 in positive samples and ~47,619 in negative samples. Barcode 08 showed a significant association with off-target reads (p=0.004) (Figure S26).

Genes implicated in true and false AMR calls

ResFinder, using fastq reads at 24 hours, demonstrated superior performance in identifying resistance genes (Figure S12). The most frequently detected genes in true calls were *bla*TEM

(n=324), *bla*Z (n=91), and *bla*CTX (n=90). However, *bla*TEM also caused the most false resistance (major error) calls, with 295 incorrect gene calls (IGC), followed by *bla*SHV (52x IGC), *bla*PAO (19x IGC), and others, particularly for ceftazidime, ceftriaxone, and piperacillin+tazobactam (Figure S13). Despite ResFinder's overall accuracy, AMRFinder and RGI (CARD) identified genes like *bla*EC and *bla*ACT that ResFinder missed, highlighting areas to reduce Very Major Errors (VMEs), while no tool was able to predict phenotypic resistance in 9 species-drug combinations (Table S9 and 10).



#Bioinformatic contamination accounted for using simulation studies

* Randomly selected from the Oxford University Hospital's microbiology laboratory (Dec. 2020 to Sep. 2021)

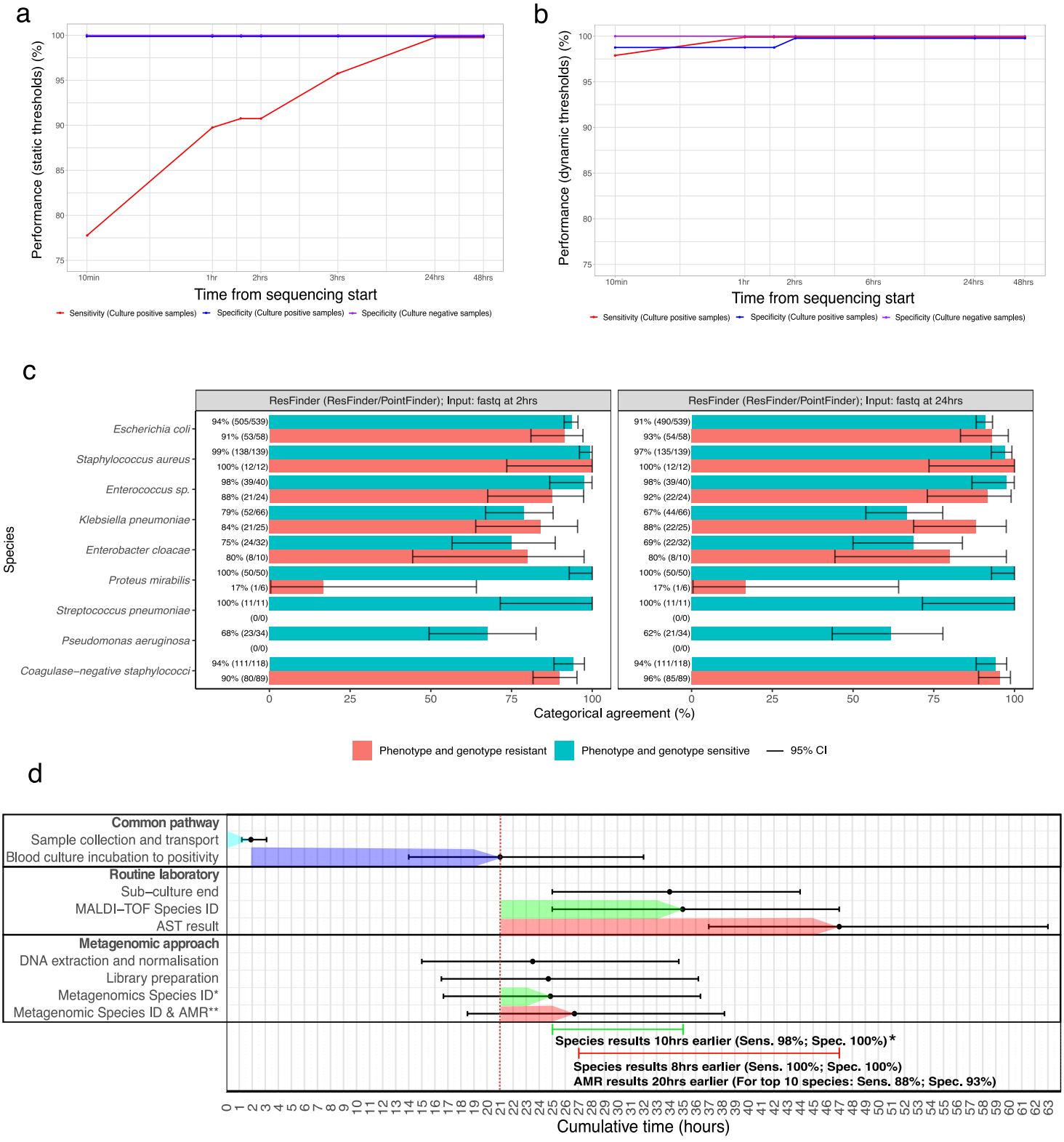

a
Performance (static thresholds) (%)
Time from sequencing start
10min
1hr
2hrs
3hrs
24hrs
48hrs
Sensitivity (Culture positive samples)
Specificity (Culture positive samples)
Specificity (Culture negative samples)
b
Performance (dynamic thresholds) (%)
Time from sequencing start
10min
1hr
2hrs
6hrs
24hrs
48hrs
Sensitivity (Culture positive samples)
Specificity (Culture positive samples)
Specificity (Culture negative samples)
c
ResFinder (ResFinder/PointFinder); Input: fastq at 2hrs
ResFinder (ResFinder/PointFinder); Input: fastq at 24hrs
Species
Escherichia coli
Staphylococcus aureus
Enterococcus sp.
Klebsiella pneumoniae
Enterobacter cloacae
Proteus mirabilis
Streptococcus pneumoniae
Pseudomonas aeruginosa
Coagulase−negative staphylococci
94% (505/539)
91% (53/58)
99% (138/139)
100% (12/12)
98% (39/40)
88% (21/24)
79% (52/66)
84% (21/25)
75% (24/32)
80% (8/10)
100% (50/50)
17% (1/6)
100% (11/11)
(0/0)
68% (23/34)
(0/0)
94% (111/118)
90% (80/89)
91% (490/539)
93% (54/58)
97% (135/139)
100% (12/12)
98% (39/40)
92% (22/24)
67% (44/66)
88% (22/25)
69% (22/32)
80% (8/10)
100% (50/50)
17% (1/6)
100% (11/11)
(0/0)
62% (21/34)
(0/0)
94% (111/118)
96% (85/89)
Categorical agreement (%)
Phenotype and genotype resistant
Phenotype and genotype sensitive
95% CI
d
Common pathway
Sample collection and transport
Blood culture incubation to positivity
Routine laboratory
Sub−culture end
MALDI−TOF Species ID
AST result
Metagenomic approach
DNA extraction and normalisation
Library preparation
Metagenomics Species ID*
Metagenomic Species ID & AMR**
Species results 10hrs earlier (Sens. 98%; Spec. 100%)*
Species results 8hrs earlier (Sens. 100%; Spec. 100%)
AMR results 20hrs earlier (For top 10 species: Sens. 88%; Spec. 93%)
Cumulative time (hours)

Table 1. Species and resistance phenotypic summary for top 10 species. The top 10 species are displayed with phenotypic resistance data shown by number of resistant isolates/(resistant + sensitive isolates) grouped by drug. See Supplementary Table 1 for further details.

	Species count	Amikacin	Amoxicillin	Azteonam	Cefalexin	Ceftazidime	Ceftriaxone	Ciprofloxacin	Clindamycin	Co-amoxiclav	Daptomycin	Ertapenem	Erythromycin	Flucloxacillin	Fosfomycin	Fusidic Acid	Gentamycin	Levofloxacin	Linezolid	Meropenem	Moxifloxacin	Penicillin	Piptaz	Pivmecillinam	Teicoplanin	Temocillin	Tetracydine	Tobramycin	Co-trimoxazole	Trimethoprim	Vancomycin
Escherichia coli	**46**	0/46 0%	19/45 42%	2/46 4%	3/28 11%	1/44 2%	1/45 2%	3/46 7%	-	17/45 38%	-	0/46 0%	-	-	0/46 0%	-	3/46 7%	3/45 7%	-	0/46 0%	-	-	3/45 7%	7/46 15%	-	2/45 4%	-	3/45 7%	10/46 22%	10/46 22%	-
Enterococcus faecium	**15**	-	15/15 100%	-	-	-	-	-	-	-	0/16 0%	-	-	-	-	-	13/14 93%	-	1/16 6%	-	-	-	-	-	8/15 53%	-	-	-	1/1 100%	-	8/15 53%
Staphylococcus aureus	**12**	-	-	-	-	-	-	0/12 0%	2/12 17%	-	0/12 0%	-	2/12 17%	0/12 0%	0/12 0%	1/12 8%	0/12 0%	0/12 0%	0/12 0%	-	0/12 0%	8/9 89%	-	-	0/12 0%	-	1/12 8%	-	0/12 0%	-	0/12 0%
Klebsiella pneumoniae	**10**	0/10 0%	10/10 100%	2/10 20%	0/1 0%	1/9 11%	2/9 22%	2/9 22%	-	4/10 40%	-	0/10 0%	-	-	0/10 0%	-	2/10 20%	2/10 20%	-	0/10 0%	-	-	2/10 20%	2/2 100%	-	2/10 20%	-	2/10 20%	4/10 40%	4/9 44%	-
Pseudomonas aeruginosa	**7**	0/8 0%	-	0/8 0%	-	0/8 0%	-	0/8 0%	-	-	-	-	-	-	-	-	0/8 0%	0/8 0%	-	0/6 0%	-	-	0/8 0%	-	-	-	-	0/8 0%	-	-	-
Enterobacter cloacae	**6**	0/4 0%	4/4 100%	1/3 33%	-	2/4 50%	2/4 50%	0/4 0%	-	4/4 100%	-	0/4 0%	-	-	0/4 0%	-	0/4 0%	0/4 0%	-	0/4 0%	-	-	2/4 50%	-	-	2/4 50%	-	0/4 0%	0/4 0%	0/4 0%	-
Enterococcus faecalis	**5**	-	0/5 0%	-	-	-	-	-	-	-	0/5 0%	-	-	-	-	-	1/5 20%	-	0/5 0%	-	-	-	-	-	0/5 0%	-	-	-	-	-	0/5 0%
Proteus mirabilis	**5**	0/5 0%	1/5 20%	0/5 0%	-	0/5 0%	0/5 0%	0/5 0%	-	0/5 0%	-	0/5 0%	-	-	1/5 20%	-	4/5 80%	0/5 0%	-	0/5 0%	-	-	0/5 0%	-	-	0/5 0%	-	1/5 20%	0/5 0%	1/5 20%	-
Streptococcus dysgalactiae	**5**	-	0/5 0%	-	-	-	-	-	-	-	0/5 0%	-	3/5 60%	-	-	-	0/5 0%	-	0/5 0%	-	0/5 0%	0/4 0%	-	-	0/5 0%	-	2/4 50%	-	-	-	0/5 0%
Klebsiella oxytoca	**4**	0/4 0%	4/4 100%	0/4 0%	-	0/4 0%	0/4 0%	0/3 0%	-	0/4 0%	-	0/4 0%	-	-	0/4 0%	-	0/4 0%	0/4 0%	-	0/4 0%	-	-	0/4 0%	-	-	0/4 0%	-	0/4 0%	0/4 0%	0/4 0%	-
Coagulase-negative staphylococci	**63**	0/1 0%	24/24 100%	-	-	-	-	18/27 67%	16/25 64%	-	0/26 0%	-	20/27 74%	22/27 81%	5/25 20%	15/27 56%	20/27 74%	17/26 65%	0/27 0%	-	17/26 65%	8/8 100%	-	-	0/25 0%	-	8/20 40%	1/1 100%	13/22 59%	1/1 100%	0/27 0%
Other	**38**	0/9 0%	4/12 33%	0/7 0%	-	0/7 0%	0/13 0%	3/12 25%	0/8 0%	3/10 30%	0/1 0%	0/7 0%	0/4 0%	-	2/7 29%	1/1 100%	3/18 17%	0/7 0%	0/11 0%	0/12 0%	1/2 50%	6/20 30%	0/7 0%	-	0/6 0%	0/6 0%	2/7 29%	1/7 14%	0/8 0%	1/7 14%	0/13 0%
Total	**216**	0/87 0%	81/129 63%	5/83 6%	3/29 10%	4/81 5%	5/80 6%	26/126 21%	18/45 40%	28/78 36%	0/65 0%	0/76 0%	25/48 52%	22/39 56%	8/113 7%	17/40 42%	46/158 29%	22/121 18%	1/76 1%	0/87 0%	18/45 40%	22/41 54%	7/83 8%	9/48 19%	8/68 12%	6/74 8%	13/43 30%	8/84 10%	28/112 25%	17/76 22%	8/77 10%

Table 2. Species performance results across all models. This table summarises models results across all classification approaches used. This table summarises results of models across all classification approaches used. Samples were randomly split into training (60%) and testing (40%) sets to ensure independent validation of species prediction approaches. Three different methods were used for bioinformatic species identification, all using the classification software Bracken with a standard database and adjusted for bioinformatic contamination using thresholds set from simulation studies.

	Training		**Testing**		**Training & testing dataset**		
	Culture positive		Culture positive		Culture positive		Culture negative
	Sensitivity (95% CI) [sample #]	Specificity (95% CI) [sample #]	Sensitivity (95% CI) [sample #]	Specificity (95% CI) [sample #]	Sensitivity (95% CI) [sample #]	Specificity (95% CI) [sample #]	Specificity (95% CI) [sample #]
Plasmid removal (Method 1): naive	97% (92-99) [124/128]	98% (93-99) [120/123]	92% (85-97) [81/88]	95% (88-99) [78/82]	94% (91-97 [201/216]	97% (93-99) [198/205]	98% (91-100%) [61/62]
Mapping and heuristic thresholds (Method 2): naive	99% (96-100) [127/128]	96% (91-99) [118/123]	94% (87-98) [83/88]	94% (86-98) [77/82]	97% (94-99) [210/216]	94% (90-97) [193/205]	98% (91-100%) [61/62]
Mapping and random forest model (Method 3): naive	100% (97-100) [128/128]	100% (97-100) [123/123]	92% (84-97%) [81/88]	91% (83-96%) [75/82]	97% (93-99%) [209/216]	97% (93-99%) [198/205]	98% (91-100%) [61/62]
Mapping and heuristic thresholds (Method 2): adjusted	100% (97-100) [137/137]	100% (97-100) [123/123]	100% (96-100) [92/92]	100% (97-100) [82/82]	100% (95 CI, 98-100) [229/229]	100% (95 CI, 98-100) [205/205]	100% (94-100%) [61/61]
Mapping and random forest model (Method 3): adjusted	100% (97-100) [137/137]	100% (97-100) [123/123]	100% (96-100) [92/92]	100% (96-100) [82/82]	100% (95 CI, 98-100) [229/229]	100% (95 CI, 98-100) [205/205]	100% (94-100%) [61/61]

*Specificity was calculated on a per sample basis, where any sample with >1 additional species found from sequencing was considered to contribute as 1 to the false negative count.

Table 3. Species missed by routine phenotypic methods identified by metagenomic sequencing approach. Metagenomic sequencing identifies missed co-infections, traditionally unidentifiable, low-level infections in culture-negative samples and confirms false positives.

Sample no.	Phenotypic results (Routine laboratory MALDI species)	Additional species identified though sequencing (Bracken %)	If applicable, clinical or microbiological factors
Metagenomics identifies co-infection(s) missed by routine laboratory			
5	*Klebsiella pneumoniae*	*Escherichia coli* (15%)	
16	*Clostridium perfringens*	*Paeniclostridium sordellii* (25%)	
33	*Lacticaseibacillus rhamnosus* *Klebsiella pneumoniae*	*Pseudomonas aeruginosa* (19%)	Old central venous catheter left subclavian.
39	*Bacteroides caccae*	*Collinsella aerofaciens* (3%)	*Collinsella aerofaciens* found in other blood culture bottle taken.
78	*Streptococcus anginosus*	*Bacteroides fragilis* (92%)	Patient was found to have a liver abscess
93	*Staphylococcus haemolyticus*	*Staphylococcus hominis* (43%)	
121	*Bacteroides fragilis*	*Bacteroides thetaiotaomicron* (9%) *Bacteroides nordii* (2%)	
129	*Enterococcus faecium*	*Staphylococcus haemolyticus* (2%)	*Staphylococcus haemolyticus* grew in the other culture bottle.
134	*Staphylococcus capitis*	*Staphylococcus epidermidis* (30%)	
135	*Enterobacter cloacae complex*	*Enterococcus faecium* (2%)	*Enterococcus faecium* isolated from additional blood culture taken a day later.
154	*Staphylococcus hominis*	*Haemophilus influenzae* (82%)	Patient found to have squamous cell carcinoma of the lung with a pleural effusion.
191	*Klebsiella variicola*	*Enterococcus faecium* (4%)	
Metagenomics identifies species undetectable by routine laboratory			
2	Gram-positive rods: No further ID.	*Turicibacter sanguinis* (98%)	
91	*Acinetobacter sp.* (Genus level only)	*Acinetobacter variabilis* (37%)	
122	No organism seen	*Faecalibacterium prausnitzii* (5%)	Clinically septic.
131	Gram-positive rods: No further ID.	*Corynebacterium kroppenstedtii* (91%)	
181	No organism seen.	*Capnocytophaga canimorsus* (98%)	This patient presented with a left swollen hand with evidence of a bite mark.
Metagenomics confirms false positive			
44	No organism seen	No organism found	Diagnosed with AML.
Metagenomics identifies low-level infection in culture-negative sample			
N47	Culture negative	*Pseudomonas aeruginosa (94%)*	*Pseudomonas aeruginosa* present in repeat blood culture obtained hours later.

Table 4. Genotypic performance of top 10 clinically relevant species by tool. Four AMR prediction tools were benchmarked including ResFinder, AMRFinderPlus, Resistance Gene Identifier (RGI), and RGI with a modified database. Resistance was called from genome assembly, raw reads at the end of the sequencing run and raw reads at earlier sequencing time intervals.

Database	Software	Input format	Sequencing start time to result	Resistant agreement	Sensitive agreement	Major Error rate	Very Major Error rate
Resistance called after genome assembly							
ResFinder and PointFinder database	ResFinder	fasta	24 hours	84% (78-88%, 188/224)	96% (94-97%, 983/1029)	4%	16%
NCBI's AMRFinderPlus database*	AMRFinderPlus	fasta	24 hours	79% (74-85%, 178/224)	79% (77-82%, 816/1029)	21%	21%
Comprehensive Antibiotic Resistance Database (CARD)*	Resistance Gene Identifier (RGI)	fasta	24 hours	79% (73-84%, 177/224)	77% (74-79%, 788/1029)	23%	21%
ResPipe: modified Comprehensive Antibiotic Resistance Database (CARD)*	Resistance Gene Identifier (RGI)	fasta	24 hours	76% (70-81%, 170/224)	76% (74-79%, 786/1029)	24%	24%
Resistance called using raw reads (no genome assembly)							
ResFinder and PointFinder database	ResFinder	fastq	24 hours	91% (87-94%, 204/224)	90% (88-91%, 923/1029)	10%	9%
Resistance called using raw reads at earlier time intervals (no genome assembly)							
ResFinder and PointFinder database	ResFinder	fastq	2 hours	88% (82-92%, 196/224)	93% (91-94%, 953/1029)	7%	12%
ResFinder and PointFinder database	ResFinder	fastq	1 hour	82% (76-87%, 184/224)	94% (92-95%, 963/1029)	6%	18%

*These tools predicted genotypic resistance at class level only, which was interpreted as resistance at all drug levels within that specific class.