**Rapid metagenomic sequencing for diagnosis and antimicrobial sensitivity prediction of canine bacterial infections**

## Author names

Natalie Ring1,2 (ORCiD: 0000-0002-8971-8507)  
Alison S. Low1 (ORCiD: 0000-0002-8360-1316)  
Bryan Wee1 (ORCiD: 0000-0001-6135-5903)  
Gavin K. Paterson1,2 (ORCiD: 0000-0002-1880-0095)  
Tim Nuttall2 (ORCiD: 0000-0002-7412-9398)  
David Gally1 (ORCiD: 0000-0002-2566-0844)  
Richard Mellanby2 (ORCiD: 0000-0002-3467-7007)  
J. Ross Fitzgerald1 (Corresponding, ORCiD: 0000-0002-9233-8468)

## Affiliation

1 The Roslin Institute, University of Edinburgh, UK  
2 Royal (Dick) School of Veterinary Studies, University of Edinburgh, UK

## Corresponding author and email address

J. Ross Fitzgerald, [ross.fitzgerald@roslin.ed.ac.uk](mailto:n.a.ring@bath.ac.uk)

## Keywords

nanopore sequencing, whole genome sequencing, rapid diagnostics, AMR, infection

## Repositories:

All raw sequencing data (nanopore) produced during this work have been deposited under NCBI BioProject Accession PRJNA925092 (<https://www.ncbi.nlm.nih.gov/bioproject/PRJNA925092>).

# Abstract

Antimicrobial resistance is a major threat to human and animal health. There is an urgent need to ensure that antimicrobials are used appropriately to limit the emergence and impact of resistance. In the human and veterinary healthcare setting, traditional culture and antimicrobial sensitivity testing typically requires 48-72 h to identify appropriate antibiotics for treatment. In the meantime, broad-spectrum antimicrobials are often used, which may be ineffective or impact non-target commensal bacteria. Here, we present a rapid, culture-free, diagnostics pipeline, involving metagenomic nanopore sequencing directly from clinical urine and skin samples of dogs. We have planned this pipeline to be versatile and easily implementable in a clinical setting, with the potential for future adaptation to different sample types and animals. Using our approach, we can identify the bacterial pathogen present within 5 hours, in some cases detecting species which are difficult to culture. For urine samples, we can predict antibiotic sensitivity with up to 95% accuracy. Skin swabs usually have lower bacterial abundance and higher host DNA, confounding antibiotic sensitivity prediction; an additional host depletion step will likely be required during the processing of these, and other types of samples with high levels of host cell contamination. In summary, our pipeline represents an important step towards the design of individually tailored veterinary treatment plans on the same day as presentation, facilitating the effective use of antibiotics and promoting better antimicrobial stewardship.

# Impact statement

Antimicrobial resistance (AMR) is a major threat to veterinary and human healthcare. It is a One Health problem, as humans and dogs are in close contact, require similar antibiotics, and share bacterial pathogens and AMR genes. Limited treatment options due to AMR would have a catastrophic effect. The risk of infection would render much of modern healthcare (including critical care, orthopaedic and complex surgeries, implants and oncology) impossible. In addition, routine infections could become life threatening. It is therefore critical to preserve the efficacy of these drugs for the future. Inappropriate antimicrobial use is the single biggest factor driving AMR. Antimicrobial stewardship involves reducing antimicrobial use, using first-line narrow-spectrum drugs, and avoiding overly long treatment. Delays in culture-based diagnosis lead clinicians to speculatively use broad-spectrum antibiotics and prolong courses of treatment beyond clinical cure. Sequencing-based rapid diagnostic approaches will have a major impact in reducing, refining and replacing antibiotic use. Such techniques will advance antimicrobial stewardship in veterinary and human healthcare.

# Data summary

All sequencing data mentioned in this work is available from NCBI, BioProject PRJNA925092, Biosamples SAMN32880396 to SAMN32880438, run accessions SRR23195371 to SRR23195413.   
This includes 35 samples sequenced during the development of the protocol (full details in Supplementary Table S2, “Protocol Development”) and 8 samples sequenced during the real-time processing phase with the finalised protocol (full details in Supplementary Table S7, “Realtime clinical samples”). Samples from which no DNA could be extracted were not sequenced during either phase, therefore no data is available for them from the NCBI.

Source code and full commands used for mentioned tools are available from Github:   
<https://github.com/nataliering/Dogstails>

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

# Introduction

Antimicrobial resistance (AMR) levels among human and veterinary bacterial pathogens are escalating globally, to the extent that the World Health Organization now classifies AMR as one of the biggest threats to global health, food security and development (1, 2). The current gold-standard diagnostic methods used in human and veterinary practice are culture-based or involve remote service providers for PCR, requiring several days to yield results. Broad-spectrum or inappropriate antibiotics are often started while waiting for the results; this inappropriate antimicrobial use is a major driver for AMR (3). Furthermore, animals with suspected highly contagious or potentially zoonotic infections like leptospirosis may require quarantining while waiting for a diagnosis, leading to escalated costs, wasted resources, and unnecessary stress for the quarantined animal (4).

Novel methods that enable sensitive, same-day rapid diagnosis and prediction of antibiotic resistance across different infection types are required. Targeted molecular tests, including PCR panels of selected AMR genes, are one option already in use (5, 6). However, targeted tests often require prior knowledge (or suspicion) of the causative pathogen, whilst gene panels may risk missing AMR caused by unexpected or novel mechanisms. Likewise, strategies involving metabarcoding for the identification of bacteria in human clinical samples, such as 16S rRNA gene amplification and sequencing, have been thoroughly tested and reviewed elsewhere (7-13). These studies have often used Oxford Nanopore Technologies’ (ONT) long read sequencing technology, which allows rapid sequencing and real-time data analysis, and have shown the ability to accurately identify bacteria, usually to the species level, despite the historically higher error rates produced by the technology. However, although full-length 16S sequencing with long reads may enable accurate species identification, the 16S gene gives no information about potential AMR phenotypes present in the present species.

An alternative method is therefore metagenomic whole genome sequencing (mWGS), involving the extraction of all genomic DNA (gDNA) present in a sample and subsequent identification of pathogens by unbiased DNA sequencing. Such methods would also use ONT sequencers, due to their rapid turnaround time. The ONT platforms most suited for this kind of rapid diagnosis are the MinION and GridION, which have already been tested with mWGS for human clinical sputum, endotracheal aspirate, blood, urine and other sample types (14-23). The MinION’s small size (similar to a small smartphone for the Mk1B or a handheld games console for the Mk1C) and relatively low start-up costs could enable its usage in a wide variety of clinical settings. In contrast, the GridION is larger (requiring around as much bench space as two desktop computers plus their monitors) and more expensive but may be suited to a larger hospital environment where more samples are processed. Indeed, the MinION and GridION are already both in use in some NHS hospitals (24, 25).

With both the MinION and the GridION, sequence data is accessible in real-time as it is produced, potentially reducing the time required to identify pathogens and their AMR genotypes to as little as 10 minutes (26, 27). Furthermore, the Flongle is an adaptor which fits into MinION and GridION sequencing devices, and enables the use of lower throughput, single-use Flongle flow cells, which may offer a more cost-effective solution which is suited to a clinical environment (14, 28). The long reads produced by nanopore sequencing also readily facilitate whole genome assembly which potentially allows the linkage of AMR genes to specific bacterial strains (29). Whilst many previous studies have focussed on human clinical applications, Nanopore sequencing (isolate, metagenomic and 16S) has also already been used in a veterinary context, for example to examine the spread of AMR genes within a veterinary hospital, and to investigate a wide variety of veterinary pathogens (30-37)

Our primary objective here was to develop a rapid, culture-free mWGS pipeline to identify pathogens and predict AMR in canine samples in a veterinary hospital setting. This included selecting and validating an extraction protocol, determining what additional clean-up steps might be required for clinical samples, determining which sequencing library preparation method produced the most reliable results, determining the most cost-effective way to use sequencing flow cells in a clinical setting where sample numbers might be inconsistent, and testing how the finalised protocol performed when compared to the current gold standard culture-based clinical microbiology techniques, taking into account the accuracy of the sequencing-based predictions as well as the time taken to produce them.

To initially develop our pipeline, we focus here on two common canine infections: urinary tract infections (UTIs) and skin infections (pyoderma). Antibiotic therapy is often the first line of treatment in these infections, but AMR, including multi-drug resistance, is frequently observed, and increasing in prevalence (38-42). Furthermore, recurrent infections are common, leading to frequent return visits to the clinic and further courses of antibiotics (43, 44). A rapid, sensitive method for diagnosing the bacterial pathogens in urine and skin swab samples, and predicting their antimicrobial sensitivity, could therefore prevent the use of inappropriate antibiotics, and limit the amount of clinical care required.

# Methods

Source code and full commands used for mentioned data analysis tools are available from Github:   
<https://github.com/nataliering/Dogstails>

***Optimisation of DNA extraction***

Our first objective was to develop and test a DNA extraction protocol suitable for both urine and skin swab samples, and ideally adaptable to additional sample types in the future, ending with DNA of suitable quality for sequencing library preparation. We examined the records from the University of Edinburgh’s Hospital for Small Animals covering a two year period (2018 and 2019) to determine which species were most commonly seen in urinary tract and skin infections, and aimed to develop a protocol which would be capable of producing sequencing-quality DNA from them all. This objective included several steps: selection of a DNA extraction kit, selection and optimisation of a metagenomic lysis step to incorporate into the selected kit, validating that the selected lysis method and kit could extract DNA from the most commonly seen species in the HfSA urine and skin records, and determining whether additional post-extraction steps would be required.

**Selection of DNA extraction kit**

Three different kits recommended in the literature were tested: DNeasy Blood + Tissue (Qiagen, Hilden, Germany), DNeasy Powersoil (Qiagen) and MagAttract HMW DNA (Qiagen). To test each kit, overnight *Escherichia coli* CAN-50 growth in Luria-Bertani (LB, ThermoFisher, Massachusetts, USA) broth equivalent to 109 CFUs was spun down at 16,000 *xg*, and the cell pellet was resuspended in 1 ml healthy dog urine. The urine-cell suspension was then processed using each kit, according to the relevant manufacturer’s instructions for each, finishing with an elution into 50 µl nuclease-free water. For each extraction, two negative controls were used: 3ml of uncultured broth and 160 µl 50 mM Tris, 10 mM EDTA, pH 8.0 (buffer P1 as per the MagAttract HMW DNA protocol. The resulting extractions were used to compare the kits in terms of (i) gDNA yield in 50 µl nuclease-free water (quantified by Qubit dsDNA HS kit), (ii) lysis method for Gram-positive species (enzymatic vs bead-beating), (iii) speed, and (iv) cost.

**Optimisation of metagenomic bacterial lysis**

Records for canine urine and skin swab samples processed at the Royal (Dick) School of Veterinary Studies Hospital for Small Animals (HfSA) in Edinburgh between 2018 and 2019 were analysed to determine which species were most commonly detected, so the broad efficacy of our extraction protocol for the most relevant pathogens could be tested (**Supplementary Table S1**). Metapolyzyme (Sigma-Aldrich, Missouri, USA), an enzymatic lysis cocktail containing six lysis enzymes (achromopeptidase, chitinase, lyticase, lysostaphin, lysozyme and mutanolysin), was trialled for our extraction protocol. An isolate of the Gram-positive species *Staphylococcus pseudintermedius*, ED99 (45), was used to trial four enzymatic lysis options: lysostaphin, lysozyme, and two different concentrations of metapolyzyme. Cells were grown overnight on tryptic soy agar (TSA, Oxoid ThermoFisher, Massachusetts, USA) plates at 37oC, then a single colony was transferred into tryptic soy broth (TSB, Oxoid) media and cultured overnight at 37oC with shaking. 2.5 ml of overnight culture was pelleted by centrifuging for 3 min at 16,000 *xg*, then resuspended in 3 ml healthy dog urine. Two negative control samples were used: 3ml of the same healthy dog urine, and 160 µl 50 mM Tris, 10 mM EDTA, pH 8.0 (buffer P1 as per the MagAttract HMW DNA protocol.

For each of the four lysis options tested, the resulting urine-and-cell suspension was centrifuged for 3 min at 16,000 *xg*, then resuspended in 160 µl 50 mM Tris, 10 mM EDTA, pH 8.0 (buffer P1 for the MagAttract HMW DNA kit). 20 µl lysozyme (100 mg ml-1), lysostaphin (10 mg ml-1), metapolyzyme (6.6 mg ml-1) or metapolyzyme (3.3 mg ml-1) was added, and the solution mixed by flicking. The solution was then incubated on a thermomixer for 1h at 37oC with 900 RPM shaking. After 1h, 20 µl proteinase K was added, and the solution was incubated for a further 30 min at 56oC with 900 RPM shaking. The rest of the MagAttract HMW DNA Gram-positive protocol was then followed as per the manufacturer’s instructions (starting from step 8 on page 26 in the MagAttract HMW DNA Handbook 03/2020), eluting into 50 µl nuclease-free water as the final step. DNA concentrations were quantified using the Qubit dsDNA HS kit according to the manufacturer’s instructions.

**Testing our extraction protocol on different species**

The optimised MagAttract + Metapolyzyme extraction protocol was tested on the most commonly isolated species identified in the HfSA’s records: *E. coli*, *S. pseudintermedius*, *Streptococcus canis*, *Enterococcus faecalis*, *Pseudomonas aeruginosa*, *Proteus mirabilis*, *Pasteurella canis*, *Klebsiella pneumoniae*, *Kocuria kristinae/Kocuria rosea*, and *Clostridium perfringens*. For aerobic species, cells were grown on TSA plates at 37oC for 24h or 72h (*Kocuria/Streptococcus*), then in TSB medium at 37oC with shaking for 24h or 72h (*Kocuria*/*Streptococcus*). For two anaerobic or facultative anaerobic species (*C. perfringens* and *P. canis*), cells were grown on TSA plates at 37oC in a sealed box with Anaerogen sachets (Oxoid ThermoFisher) for 24h, then in TSB medium in growth flasks in a sealed box with Anaerogen sachets at 37oC with shaking for 24h. For all species, 3 ml of broth culture was pelleted by centrifuging at 16,000 *xg* for 3 min, the pellet was resuspended in 3 ml healthy dog urine, then pelleted again in the same way. Again, 3ml the healthy dog urine and 160 µl buffer P1 were also used as a negative control. Cell pellets was then processed as per the optimised protocol, starting with resuspension in 160 µl buffer P1 and 60 minutes lysis with 20 µl metapolyzyme. DNA was eluted into 50 µl nuclease-free water, and quantified using Qubit’s dsDNA HS kit.

**Determining whether additional post-extraction DNA clean-up was required**

The purity of the extracted DNA was assessed using a Nanodrop spectrophotometer (ThermoFisher) to measure the 260/280nm and 260/230nm absorbance ratios of 21 clinical samples (a selection of the total samples processed, indicated in **Supplementary Table S2**). The samples were then cleaned up using the ProNex Size-Selective Purification System (Promega, Wisconsin, USA) according to manufacturer’s instructions, starting with 50 µl DNA and 200 µl ProNex beads, and eluting into 20 µl nuclease-free water at the end of the protocol. The purity of the cleaned-up DNA was then measured again by Nanodrop. The differences in purity of the pre- and post-clean-up samples were assessed for significance using the Wilcoxon test for paired samples. After these 21 samples, the ProNex clean-up was incorporated as standard, and for reasons of speed, Nanodrop readings were no longer taken.

**Final optimised protocols for metagenomic DNA extraction and clean-up**

The final optimised protocols, including post-extraction clean-up with ProNex beads, are available from protocols.io:

[dx.doi.org/10.17504/protocols.io.n2bvj8o5bgk5/v2](file:///\\cmvm.datastore.ed.ac.uk\cmvm\eb\groups\fitzgerald_grp\Dogstails\Paper\dx.doi.org\10.17504\protocols.io.n2bvj8o5bgk5\v2) (urine)

[dx.doi.org/10.17504/protocols.io.q26g7yr19gwz/v2](file:///\\cmvm.datastore.ed.ac.uk\cmvm\eb\groups\fitzgerald_grp\Dogstails\Paper\dx.doi.org\10.17504\protocols.io.q26g7yr19gwz\v2) (skin swabs)

***Optimisation of the sequencing process***

After developing our metagenomic DNA extraction protocol for clinical samples, our next objective was to optimise the sequencing process. Throughout the development of this process, we needed to assess the taxonomic composition of our sequenced DNA, hence we first selected two commonly used bioinformatics taxonomic classification tools to use throughout this phase. The remaining steps in the sequencing optimisation phase included: selection of a library preparation kit, testing the lower limits of the selected kit, comparison and selection of flow cell type (MinION vs Flongle), and developing the optimal usage strategy for the selected flow cell type.

**Species identification strategies**

For each experiment detailed here, species were identified from sequencing reads using either EPI2ME workflows “Fastq WIMP (What’s in my pot?)” and “Fastq Antimicrobial Resistance”   
(v3.0.1-7052513, <https://epi2me.nanoporetech.com/>, ONT account required), Kraken2 (v2.1.1, 46) with one of two custom databases, “pathogens\_plus” or “bacteria\_plus”, or a combination thereof. EPI2ME is ONT’s own software, designed to be used with uncorrected nanopore long reads. The WIMP workflow classifies reads using the Centrifuge standard database, whilst the Antimicrobial Resistance workflow uses minimap2 to align input reads to the AMR CARD database to identify AMR genes.

The custom Kraken2 bacteria\_plus database was constructed from all bacterial representative genomes present in the NCBI RefSeq database in November 2022 (16,113 species) plus eight mammalian genomes (*Canis lupus familiaris*, *Homo sapiens*, *Felis catus*, *Equus caballus*, *Oryctolagus cuniculus*, *Sus domesticus*, *Bos taurus* and *Ursus arctos*). The “pathogens\_plus” database contains 668 genomes of various bacterial, viral, protozoan and fungal pathogens, including the 100 common human and animal pathogens identified in a 2014 study (47), plus selected other pathogens known to be important in veterinary samples, such as *Leptospira* and *Mycobacterium spp* (a full list of species included can be seen in **Supplementary Table S3**)*.* The pathogens\_plus database also contains the same eight mammalian genomes included in the bacteria\_plus database.The two Kraken2 databases were tested using two ZymoBIOMICS Microbial Community Standards datasets (log and even distributions of species) sequenced on R9.4.1 flow cells by Nicholls et al. in 2019 (48), as well as single isolate datasets for the same ten most commonly isolated species tested in the “Testing our extraction on different species” section above. Full details of these tests and their results are given in **Supplementary Results SR1**.

The tool and database used is noted for each step of the protocol development below, with further details as necessary. Negative control samples (unless otherwise stated, nuclease-free water which was subjected to the entire same protocol as the test samples) were sequenced to characterise any background contamination and any barcode-crosstalk (when sequencing multiple samples simultaneously). When testing clinical samples, a threshold of ≥1% of classified bacterial reads was used to determine true positives and minimise false positives arising from any remaining background contaminants and low quality database alignments.

**Selection of a library preparation kit**

Two rapid library preparation kits were tested: rapid PCR barcoding (SQK-RPB004, Oxford Nanopore Technologies (ONT), Oxford, UK) and rapid barcoding (SQK-RBK004, ONT). The SQK-RPB004 PCR reaction was carried out on 22 clinical samples (**Supplementary Table S2**), according to the manufacturer’s instructions, and the concentration of each was measured by Qubit’s dsDNA HS kit before and after the reaction.

All 22 samples prepared with the rapid PCR barcoding kit were subsequently sequenced on MinION R9.4.1 flow cells (ONT) according to the SQK-RPB004 protocol, although ProNex beads were substituted into steps which required AMPure XP beads. Four of these samples, two urines (DTU09 and DTU16) and two skin swabs (SkSw08A and SkSw14), were re-sequenced with the rapid barcoding kit, to compare the results of sequencing with and without the PCR amplification. The DNA was prepared according to the manufacturer’s instructions for SQK-RBK004, except for the substitution of ProNex in place of AMPure XP again. For both SQK-RPB004 and SQK-RBK004, the sequencing was performed using a MinION Mk1C device running either MinKNOW v19.12.12 or v21.05.12 (see **Supplementary Table S2**) and basecalling was performed after sequencing using the University of Edinburgh’s high-performance computing cluster (Eddie) to run GPU Guppy (v5.0.7) in super accuracy (SUP) mode. The results of the two library preparation methods were compared in terms of (i) run yield, and (ii) bacterial reads vs eukaryotic host contamination (according to ONT’s online EPI2ME classification “WIMP” tool).

**Determining the lower detection limits for the rapid barcoding (SQK-RBK004) kit**

To estimate the lowest DNA concentration at which species could still be identified, a sequencing run was conducted using serially diluted gDNA. The gDNA was a 50:50 mix of *S. pseudintermedius* and dog DNA. Both were extracted using the MagAttract HMW kit (*S. pseudintermedius* from overnight TSB culture, dog from a surplus diagnostic blood sample from the HfSA’s clinical biobank, as per the MagAttract protocol for blood), cleaned-up with ProNex beads as described above, quantified using Qubit’s dsDNA HS kit and adjusted to 75 ng µl-1 prior to mixing. The mixed sample was further diluted to a starting concentration of 10 ng µl-1 (confirmed by Qubit dsDNA HS kit), then the 10 ng µl-1 sample was serially 1:1 diluted in nuclease-free water, from around 10 ng µl-1 down to below 0.1 ng µl-1. Each dilution was prepared for sequencing with SQK-RBK004 as described above, with a different barcode used for each dilution and barcode01 used as a negative control (nuclease-free water). The dilution series plus negative control was then sequenced on a fresh MinION R9.4.1 flow cell for 24 h. We know from prior experience that 10 ng µl-1 sequences adequately using SQK-RBK004, hence this starting concentration acted as our positive control for this test. The sequencing data was analysed using Kraken2 with the bacteria\_plus database. The results for the negative control sample were used to determine levels of background contamination and barcode crosstalk. The number of reads for each expected species (*S. pseudintermedius* and *C. lupus familiaris*) seen in the negative control (representing barcode crosstalk) were subtracted from the other samples, and theresults for each dilution were compared in terms of (i) run yield, and (ii) reads from the two expected species.

**Comparing MinION flow cells with Flongle flow cells**

Two relatively highly concentrated gDNA samples were used to compare higher throughput MinION flow cells with single-use, low-throughput Flongle flow cells: a clinical skin swab sample (SkSw21, 55 ng µl-1) and a 1:1 mix of *E. coli* and *P. aeruginosa* isolate DNA, both extracted previously during the DNA extraction optimisation trials (80 ng µl-1). Each sample was prepared for sequencing on an R9.4.1 Flongle flow cell (FLO-FLG001) according to the Flongle-specific manufacturer’s instructions for SQK-RBK004 (using ProNex beads instead of AMPure XP). Each sample was then sequenced separately for a full 24h with the Flongle adapter on a MinION Mk1C running MINKOW (v21.02.2) and real-time basecalling with Guppy (v4.3.4) in “fast” mode in order to monitor sequence acquisition, and the total yield of the 24 h run was noted. More DNA from the same samples was then prepared for sequencing on an R9.4.1 MinION flow cell with SQK-RBK004 as above. Sequencing was started for the first (*E. coli + P. aeruginosa*) sample on a Mk1C MinION running MINKOW (v21.02.2) and real-time basecalling with Guppy (v4.3.4) in “fast” mode in order to monitor sequence acquisition. The sequencing was stopped when the run yield matched that of the 24 h Flongle run, and the time taken to reach that yield was recorded. The MinION flow cell was then washed using the flow cell wash kit (EXP-WSH004, ONT) according to the manufacturer’s instructions, and the second sample (SkSw21) was loaded and sequenced using the same conditions. Again, the sequencing was monitored closely and stopped once the run yield matched that of the 24 h Flongle run, and the time taken to reach that yield was compared between the MinION and the Flongle flow cells.

**Determining the optimal MinION flow cell usage strategy**

As the theoretical yield of each MinION flow cell is much higher than required for a single clinical sample, and to improve the cost-effectiveness of using MinION flow cells, two flow cells and a variety of DNA samples were used to test how many times a MinION flow cell can be washed and re-used. For the first flow cell, the DNA sample was a relatively high concentration previously extracted *E. coli* isolate (110 ng µl-1) which had both been stored at 4oC since extraction. An initial sample of 7.5 µl of DNA was prepared for sequencing with SQK-RBK004 as previously described, using barcode01. Sequencing commenced on a fresh flow cell (on a Mk1C MinION running MINKOW (v21.02.2) and real-time basecalling with Guppy (v4.3.4) in “fast” mode in order to monitor sequence acquisition) and was stopped when a target of 200 Mbp was reached (we assumed at this stage that most real clinical samples would need a maximum of 200 Mbp for species identification and AMR prediction). The time taken to reach 200 Mbp was recorded, and the flow cell was washed using EXP-WSH004 as previously described.

A second 7.5 µl was processed in the same way, with barcode02, followed by another wash with EXP-WSH004. This protocol continued until the time taken to reach 200 Mbp was longer than 2 h, using the next subsequent barcode (barcode03, barcode04, etc.). For each sequencing step, the starting number of available sequencing pores was recorded at the beginning of the run. After each run, the sequenced reads were quality controlled by EPI2ME and NanoStat (v1.6.0, 49), and re-basecalled using the University of Edinburgh’s high-performance computing cluster (Eddie) to run GPU Guppy (v5.0.7) in super accuracy (SUP) mode in order to accurately determine the percentage of DNA assigned to the wrong barcode.

For the second flow cell, the same protocol was followed, but freshly extracted DNA from clinical samples, ranging from 0.76 to >120 ng µl-1, was sequenced. Each sample was sequenced for up to 3 h, and some samples were sequenced simultaneously (with different barcodes). Full details of the samples sequenced are given in **Supplementary Table S4**.

***Testing the optimised extraction and sequencing process on real clinical samples***

After finalising our extraction and sequencing protocol, the objective of our final stage was to test it on real clinical samples, to determine how quickly we could produce results and how accurate our predictions were, compared to those produced by the current gold-standard culture-based techniques.

**Sequencing clinical urine samples in real-time using the finalised protocol**

To test the full optimised protocol in a realistic manner, a further nine “post-development” urine samples were processed in real-time in batches of one to four using the finalised protocol. The post-development samples were collected from the HfSA and processed immediately using the finalised extraction protocol (<https://www.protocols.io/view/magattract-metapolyzyme-metagenomic-gdna-extractio-chnrt5d6>). The extracted DNA was prepared for sequencing using the rapid barcoding kit (SQK-RBK004) as described above, using barcode 01 for a negative control (nuclease-free water, which was processed using the same extraction protocol as the real samples to control for contaminating DNA from the extraction and barcode cross-contamination during sequencing) and the remaining sequential barcodes for the real samples. Samples were sequenced for up to 3h, or until 100 Mbp had been sequenced, using an R9.4.1 MinION flow cell on a GridION device (MinKNOW version 22.10.5), with real-time Guppy super accurate basecalling (v6.3.8). Although we tested flow cell longevity by sequencing 200 Mb, we wanted to minimise sequencing time further here, hence testing the sequencing of only 100 Mb to determine if sufficient data was produced. In between sequencing runs, the flow cell was stopped, washed using the flow cell wash kit (EXP-WSH004) according to manufacturer’s instructions and stored with storage buffer at 4oC until the next use.

**Analysing the data produced by all of our samples, compared to culture-based results**

During sequencing, the data produced underwent preliminary analysis using EPI2ME’s Fastq Antimicrobial Resistance workflow. After sequencing, the data were further analysed, including more accurate species identification with Kraken2 (with both the pathogens\_plus and bacteria\_plus databases) with the --confidence 0.05 flag. The results of running Kraken2 with both databases were manually compared, and a final call on predicted species was made (see **Supplementary Results S1** for full details on this decision process).

After species prediction, reads were trimmed with Porechop (v0.4.2, 50), genomes were assembled with with Flye (v2.9-b1774, 51) and annotated using Prokka (v1.14.5, 52). Further AMR prediction was carried out using Abricate (v1.0.1, 53) with the NCBI database downloaded on 18 January 2022 and AMRFinderPlus (v3.10.45, 54) with the database version 2022-10-11.2. The Comprehensive Antibiotic Resistance Database (CARD, <https://card.mcmaster.ca/home>) was consulted to manually relate any identified genes and mutations with their likely resistance phenotypes. The results from all data analyses were collated to predict which species had been present in the original sample, and the AMR phenotypes expected from them. In addition, the length of time taken for each sample was recorded. These were then compared with results from the HfSA’s Veterinary Pathology Unit (VPU), who processed the same nine samples at the same time, using their usual gold-standard culture-based techniques (including species identification and antibiotic susceptibility testing (AST) using a using a VITEK® 2 (bioMérieux) instrument. The accuracy of our AMR predictions was compared to both “Resistant and Intermediate” AST results, and “Resistant” alone.

During the development of this protocol, 44 clinical samples were processed (in addition to the final nine “post-development” samples; 23 of these had paired results from the VPU. To more extensively test our current bioinformatics pipeline, these 23 samples with paired VPU results were also processed using the same data analysis strategy as the post-development samples (Kraken2, Flye, Prokka, ABRicate and AMRFinderPlus). For some of these samples volumes of data much larger than 100 Mb had been sequenced. These large datasets were randomly down-sampled to 100 Mbp (for reasons detailed above) using Rasusa (v0.6.1, 55) before data analysis. The results for these samples were then compared retrospectively to VPU results (see **Supplementary Table S6** for full sample details and VPU results).

# Results and Discussion

***Optimisation of the extraction process***

As mentioned in the methods section, the main objective for this stage of our project was to develop and test a DNA extraction protocol suitable for both urine and skin swab samples, which would be capable of producing sequencing-quality DNA from the vast majority of pathogens seen in urinary tract and skin infections. The development of this extraction protocol include: selection of a DNA extraction kit, optimisation of metagenomic lysis, validation that the selected lysis method and kit could extract DNA from the most commonly seen pathogens in the HfSA urine and skin records (**Figure 1**), and determination of required additional post-extraction clean-up steps.

**The MagAttract HMW kit can extract micrograms of DNA in around 30 minutes**

A wide variety of DNA extraction methods are available, and many of them have been used in previous studies aiming to develop rapid diagnostic protocols or to extract metagenomic DNA from urine samples (16, 18, 20, 26, 29, 56-60). We wanted to use a kit-based approach to increase consistency between samples, especially if the extraction was being performed by different individuals. We selected three commonly used kits to test (**Table 1**), based on their potential flexibility for application to other sample types in the future. A metagenomic approach using enzymatic lysis rather than mechanical cell disruption should result in high molecular weight (HMW) fragments, facilitating better genome assembly and more efficient species and AMR gene identification. Nonetheless, we tested one kit using mechanical disruption, to compare the efficiency of mechanical versus enzymatic lysis.

We tested three kits in the same way: attempting to extract cultured bacteria spiked into healthy dog urine, with the same healthy dog urine without spiked-in bacteria as a negative control. Although we tested the extractions with *E .coli*, we used the standard Gram positive protocol (with lysozyme) for each kit, as our eventual protocol would need to efficiently extract DNA from both Gram positive and Gram negative species. Of the three kits we trialled, the one with mechanical lysis (the DNeasy Powersoil) was the most expensive per-sample, and produced lowest yield of gDNA when tested (**Table 1**). The two kits that included optional enzymatic Gram positive lysis steps were the DNeasy blood + tissue kit and the MagAttract HMW kit, each producing similar yields of DNA in a similar time-frame when tested with *E. coli* (3.16 µg in 40 minutes vs 3.08 µg in 30 minutes, respectively). None of the kits we trialled produced levels of “kitome” DNA that were detectable in our negative controls by the Qubit dsDNA HS kit, hence we assumed them all to contain similar (low) levels of background contamination (61). Although the spin column-based DNeasy kit had a lower cost-per-sample, we ultimately selected the MagAttract for our protocol due to its optimisation towards extracting HMW DNA, slightly shorter extraction time, and its extensive washing steps that should limit the presence of inhibitors that affect library preparation.

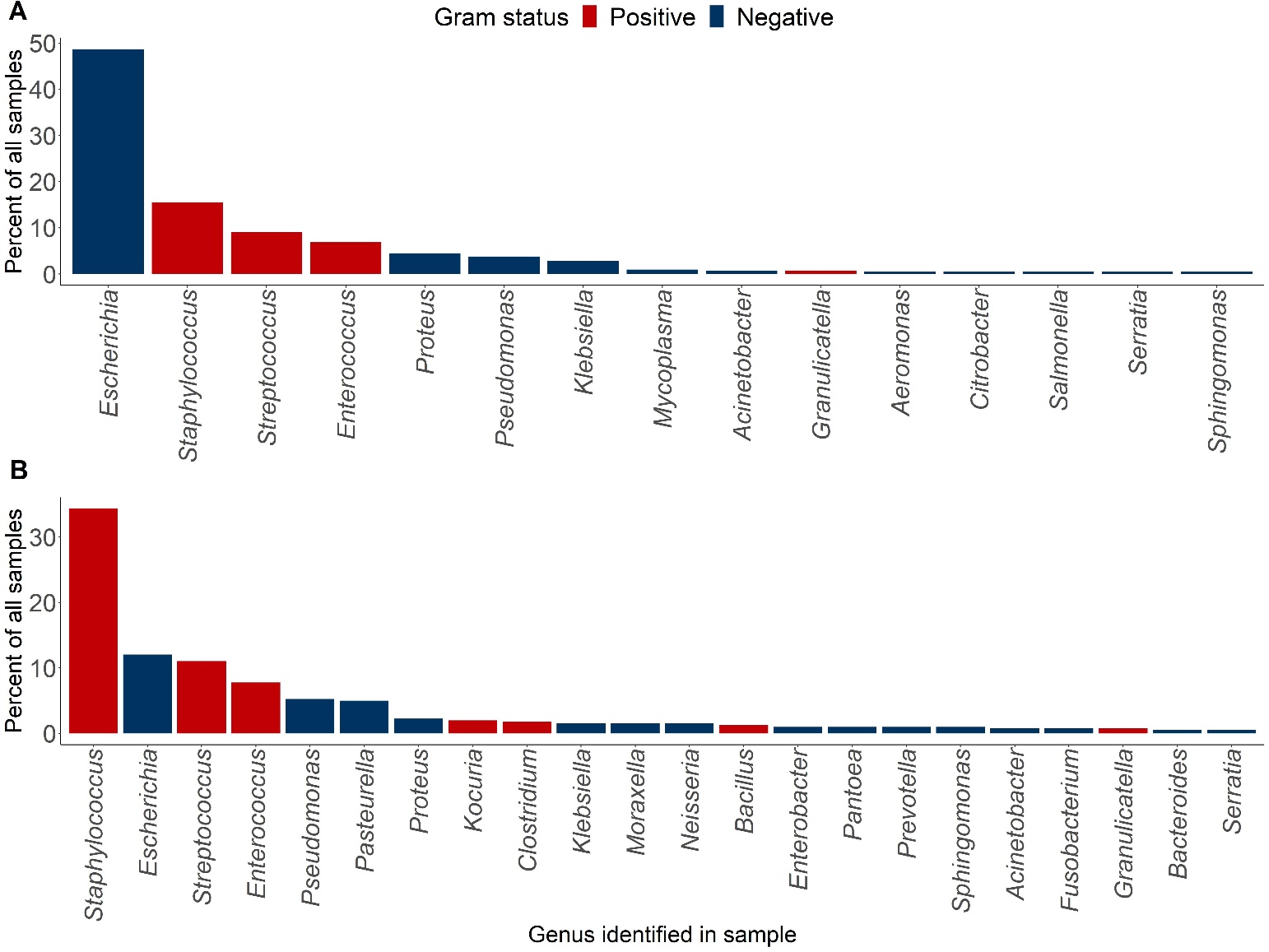
**Enzymatic lysis with 3.3 mg ml-1 metapolyzyme can produce micrograms of extracted DNA**

We examined records from samples seen at the HfSA in 2018 and 2019 to assess the variety of bacterial pathogens associated with clinical urine or skin swab samples from dogs (**Figure 1, Supplementary Figure S1**). Overall, around 90% of culture-positive cases were associated with ten different genera, with a total of 53 different genera in total, including 48% Gram positive and 52% Gram negative (**Supplementary Table S1**). We therefore required a lysis method that would allow the extraction of DNA from a wide variety of different species. As discussed above, we preferred enzymatic lysis over physical methods, due to the high levels of DNA shearing caused by physical methods. However, a single lysis enzyme such as lysozyme would not be capable of lysing the variety of Gram positive bacteria seen in urine and skin infections.

Metapolyzyme is a cocktail of six different enzymes for the lysis of bacterial and/or fungal cell walls. Of these six, lysostaphin, mutanolysin and lysozyme should lyse all of the Gram-positive species most commonly seen in the 2018-2019 HfSA data. We assessed the efficiency of two concentrations of metapolyzyme compared to lysozyme and lysostaphin by lysing a cultured isolate of *Staphylococcus pseudintermedius* (the most common Gram-positive species in skin swabs) for 60 minutes, followed by DNA extraction with the MagAttract HMW kit selected above. We found that 3.3 mg ml-1 of metapolyzyme resulted in better yield of extracted DNA than either lysozyme or Lysostaphin used at

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Table 1: gDNA extraction kits tested, and their pros and cons | | | | |
| **Kit** | **Pros** | **Cons** | **Yield from test *E. coli* extraction (µg)** | **Cost per sample (kit only)** |
| **Qiagen DNeasy Blood + Tissue** | * Quick extraction (<40 minutes after lysis) * Easy, column-based * No additional reagents or equipment required * No bead-beating | * Different Gram +ve protocol (enzymatic lysis step required) | 3.16 | £3.92 |
| **Qiagen DNeasy PowerSoil** | * Very quick extraction (30 minutes or less) * Same protocol for Gram +ve or –ve, optimised for metagenomics extractions | * Extremely low yield * Bead-beating, highly sheared DNA * Not optimised for liquid samples | 0.6 | £6.06 |
| **Qiagen MagAttract HMW DNA** | * Optimised for extraction of HMW DNA * No bead-beating * Magnetic beads and 6 wash steps, so theoretically very clean DNA even from dirty starting samples * Quick extraction (30 minutes after lysis) | * Different Gram +ve protocol (enzymatic lysis step required) * Magnetic rack required | 3.08 | £4.75 |

Fig. 1 Pathogens identified in HfSA (A) urine and (B) skin swab samples, 2018 & 2019.   
Species observed in only one sample (17 species for urine samples, 22 for skin swab samples) are not shown here, but can be seen in Supplementary Table S1.



their recommended concentrations (**Table 2**). Although the higher concentration of metapolyzyme (6.6 mg ml-1) resulted in a slightly higher mass of extracted DNA than the lower, the gain was minimal (5.5µg vs 5.2µg), and thus not thought to be worth the added cost of using twice the enzyme per reaction. We consequently incorporated a 60 minute lysis step with 3.3 mg ml-1 metapolyzyme before DNA extraction with the MagAttract kit.

|  |  |
| --- | --- |
| Table 2: Different lysis enzymes tested with *S. pseudintermedius* ED99 DNA extraction | |
| **Lysis enzyme** | **DNA yield (µg)** |
| Lysozyme (100 mg ml-1) | 0.46 |
| Lysostaphin (10 mg ml-1) | 2.75 |
| Metapolyzyme (6.6 mg ml-1) | 5.5 |
| Metapolyzyme (3.3 mg ml-1) | 5.2 |

**Our optimised protocol can efficiently extract DNA from the bacterial pathogens most commonly identified in canine urinary tract and skin infections**

To validate that it would work for a variety of species, the optimised MagAttract + Metapolyzyme extraction protocol was next tested on representatives of the top ten genera identified in urine and/or skin swab samples (**Figure 1**). Each species was grown in appropriate media and spiked into 3 ml healthy dog urine to simulate a clinical sample. As shown in **Table 3**,our protocol was successful in extracting micrograms of DNA from all ten species tested, including anaerobes, Gram positives, and slow-growing bacteria (e.g. *Kocuria*). Although it is technically feasible that the healthy dog urine used may contain host cells as well as the spiked-in bacterial cells, our negative controls (extraction from the healthy dog urine without spiked-in bacteria) did not contain levels of DNA detectable by the Qubit dsDNA HS kit. We therefore concluded that any host DNA in our samples was negligible, and the vast majority of the micrograms of DNA extracted from our test samples was bacterial.

According to the HfSA records, the ten genera tested here represented around 90% of urine and skin swab samples; the remaining 10% of samples included 43 additional genera, each of which was seen only once or twice over a two year period at the HfSA. Therefore, testing every genera seen in the records was unfeasible at this point. However, overall, these tests served as proof that our protocol is effective in extracting DNA from the vast majority, if not all, of the species encountered in clinical canine urine or skin swab samples.

**A post-extraction clean-up improves DNA purity for clinical samples**

As clinical samples may contain contaminants or inhibitors that impact library preparation or sequencing, we tested if a further clean-up step would enhance our protocol and help to prevent library preparation failure. Nanodrop 260/280nm and 260/230nm absorbance ratios were used to assess the purity of 21 clinical samples (14 urine, 7 skin swabs, **Supplementary Figure S2**) before and after a clean-up protocol using the ProNex Size-Selective Purification System. This kit was selected because it should maintain the high molecular weight of the DNA extracted by the MagAttract protocol, and for its affordability relative to the AMPure XP system.

Prior to the clean-up, neither the mean 260/280nm ratio nor the mean 260/230nm ratio across the 21 samples was ideal (mean=2.00 and mean=0.78, compared to the expected ratios for pure DNA of ~1.8 and 2.0-2.2, respectively (62)). This suggested that significant levels of contaminants remained after the DNA extraction. After the clean-up, the mean 260/280nm ratio across the 21 samples was close to ideal at 1.86, although the difference between pre- and post-clean-up ratios was not significant (*p*=0.8, *n*=21). In contrast, the 260/230nm ratios were significantly improved post-clean-up (*p*=0.007, *n*=21); the mean value of 1.36 was still not optimal, but it did enable efficient library preparation. Our findings support the inclusion of the additional clean-up step between DNA extraction and library preparation, despite adding extra time to the final length of the rapid diagnostics protocol. An additional benefit of the clean-up protocol is that by eluting into a final volume lower than the starting volume (20 µl vs 50 µl) the DNA is further concentrated.

|  |  |
| --- | --- |
| Table 3: gDNA concentrations extracted from ten most commonly encountered species, using our optimised lysis and extraction protocol | |
| **Species** | **Mean DNA yield (µg) *(standard deviation)*** |
| *Escherichia coli* | >6.00 *(±0)* |
| ***Staphylococcus pseudintermedius*** | 5.35 *(±0.15)* |
| *Proteus mirabilis* | 5.85 *(±0.15)* |
| *Pseudomonas aeruginosa* | 2.4 *(±0.72)* |
| ***Enterococcus faecalis*** | 2.74 *(±0.87)* |
| ***Streptococcus canis*** | 2.68 *(±0.18)* |
| ***Kocuria rosea/kristinae*** | 2.09 *(±0.21)* |
| *Pasteurella canis*1 | >6.00 *(±0)* |
| ***Clostridium perfringens*1** | 1.86 *(±0.31)* |
| *Klebsiella pneumoniae* | >6.00 *(±0)* |
| **Bold** indicates Gram positive species 1 Cultured in anaerobic conditions | |

**Final DNA extraction protocol**

Our final protocol for metagenomic DNA extraction from urine and skin swab samples is therefore a 60 minute lysis with 3.3 mg ml-1 of metapolyzyme, 30 minute DNA extraction with the MagAttract HMW kit, and 30 minute DNA clean-up with ProNex beads, eluting into a concentrated 20 µl nuclease-free water.

***Optimisation of the sequencing process***

After the development of our DNA extraction protocol, which outputs high quality DNA ready for sequencing, we next optimised the steps necessary to produce DNA sequencing reads. This included the selection of a sequencing library preparation kit, testing the limits of that library preparation kit, selecting a flow cell type, and determining the most cost-effective usage strategy for the selected flow cells.

**Rapid barcoding enables sequencing and identification of species from extracted concentrations of DNA as low as 0.08 ng µl-1**

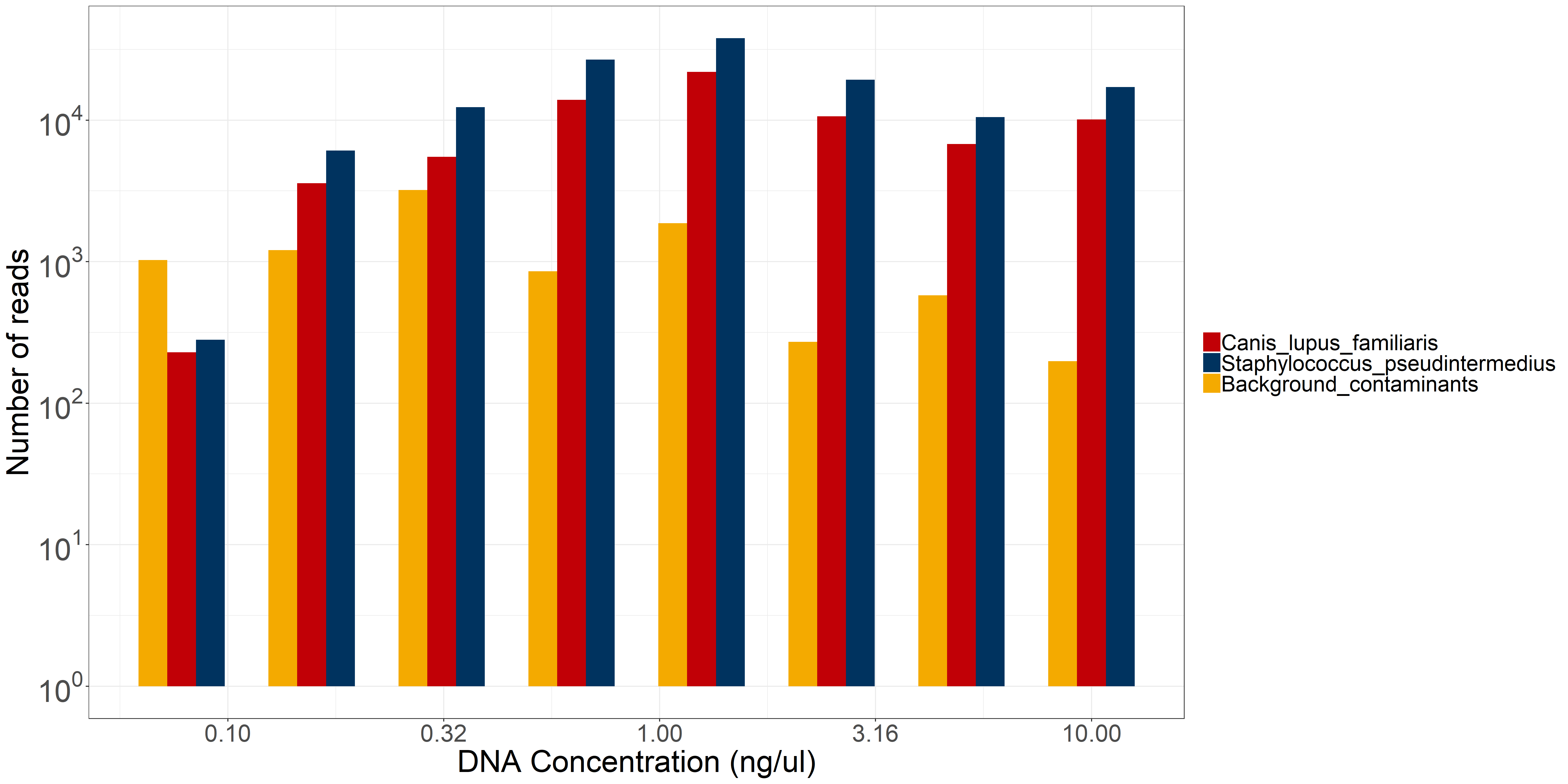
Our experience extracting DNA from the first few clinical samples (**Supplementary Table S2**) indicated that DNA concentrations may be relatively low in some samples (the mean concentration of the first five skin swab samples was just 4.86 ng µl-1, whilst the mean from the first five urine samples was 21.98 ng µl-1). Some previous protocols developed for rapid bacterial identification by whole genome nanopore sequencing have used ONT’s rapid PCR barcoding library preparation kit, SQK-RPB004 (16, 18, 63, 64). The major benefit of this kit is the amplification of DNA during the PCR step, which may be important for low abundance samples, though with the disadvantage of the additional time required. We therefore tested this kit first, using it to attempt library preparation of DNA extracted from 27 samples (15 urine, 12 skin swabs), with pre-PCR DNA masses ranging from 1.95 to 8.7 ng. However, we found that the PCR step resulted in unpredictable yields of DNA (independent of starting mass), and a tendency to amplify host DNA more than bacterial (**Supplementary figures S3 and S4).** The failure of the PCR reaction for some samples was likely due to PCR inhibitors remaining even after a DNA clean-up step; whether or not inhibitors still remain is unpredictable as it seems to vary between samples. A second DNA clean-up step might result in more consistent PCR success, but it would add at least 30 minutes to our protocol, and PCR might still fail.

We therefore trialled ONT’s rapid barcoding kit without the PCR step (SQK-RBK004) instead. We used a serial dilution series of combined *S. pseudintermedius* and *C. lupus familiaris* DNA to determine the lowest concentration which could be sequenced and still produce enough usable data to identify selected bacterial species from our samples. A 50:50 ratio of bacterial to canine DNA was chosen, as this represents the average level of host contamination seen in our urine samples throughout development, and therefore was a realistic approximation of a low concentration clinical sample.

A small number of *S. pseudintermedius* and *C. lupus familiaris* reads were seen in our negative control (12 and 8 reads, respectively). These represent barcode crosstalk between different samples; the spillage of barcode from a more highly concentrated sample into a less highly concentrated (that is, negative) sample. The level of barcode crosstalk seen here was, however, extremely low, especially compared to the number of background contaminant reads seen in the negative control. ~2,000 contaminant reads were seen in the negative control after 24h of sequencing; these likely came from the kits and tubes used throughout DNA extraction and library preparation, and were identified as “*Homo sapiens”* and “*Nocardioides alcanivorans*” by Kraken2 with our bacteria\_plus database (see **Supplementary Results SR1** for further discussion of our experiences with this database, including the identification of contaminants). After subtracting the crosstalk numbers from the other test samples, we found that the *S. pseudintermedius* DNA was still identifiable at concentrations much lower than our means of 4.86 ng µl-1 (skin swabs) and 21.98 ng µl-1 (urine). As seen in **Figure 2**, both *S. pseudintermedius* and *C. lupus familiaris* were seen at levels above background contamination with starting concentrations as low as 0.16 ng µl-1. In addition, even at 0.08 ng µl-1, where the *S. pseudintermedius* reads were outnumbered by the contaminants, *S. pseudintermedius* reads were still seen at an abundance far greater than 1%, hence would be called as a positive (especially as the background contaminant reads would be identified as such and excluded by analysis of the negative control sample). Although we only tested once (due to cost limitations of MinION flow cells), we concluded that we could use the rapid barcoding kit to sequence even very low concentration clinical samples without the need for the PCR amplification step.

Fig. 2 Sequencing serial dilutions of *S. pseudintermedius* and *C. lupus familiaris* DNA

1:1 dilution of gDNA in nuclease-free water, sequenced for 24 hours using SQK-RBK004 (rapid barcoding) and R9.4.1 MinION flow cells. A) shows the number of reads identified by EPI2ME’s WIMP tool as the correct species vs starting gDNA concentration. B), C) and D) show the number of reads identified as the correct species vs the number identified as the next most common species for *E. coli*, *S. pseudintermedius* and *S. canis*, respectively.



**MinION flow cells produce sequencing data significantly faster than Flongle flow cells, and washing MinION flow cells allows their repeated re-use**

The Flongle is an adapter which fits into MinION or GridION sequencers and allows the use of Flongle flow cells, which are a single-use, lower yield alternative to MinION flow cells. These characteristics of Flongle flow cells are desirable for clinical use: single-use means no potential cross-contamination between different samples, and the lower yield means the flow cells are correspondingly less expensive than MinION flow cells. Clinical applications for which Flongle flow cells are already being used include rapid sequencing of viruses such as SARS-CoV-2 and monkeypox (65, 66), HLA-typing (67, 68), and 16S metagenomics (7, 69). A number of previous studies have also investigated the use of Flongle flow cells for the rapid identification and typing of bacterial infections in some sample types (14, 25, 70, 71). For veterinary purposes, where even in a hospital sample numbers may be low, sequencing single samples presents a more practical and cost-effective solution than either waiting for enough samples for a full MinION flow cell run or using a whole MinION flow cell per sample.

The major downside of Flongle flow cells with regards to rapid sequencing is their reduced number of pores; with up to 126 pores available on a Flongle flow cell compared to up to 512 available at any one time to sequence on a MinION flow cell, DNA is sequenced correspondingly less rapidly on a Flongle flow cell. However, when sequencing low concentrations of DNA, pore occupancy is often low on a MinION flow cell (not every pore is sequencing and many are empty because of the low number of DNA strands loaded), hence it was possible that sequencing speeds would be similar on both types of flow cell.

We therefore sought to evaluate the utility of Flongle flow cells in our protocol compared to the MinION, sequencing two clinical samples on each, and comparing the length of time taken to produce the same volume of data on each (**Supplementary Table S4**). The volume of DNA sequence data produced by Flongle in a 24h period was matched by MinION within 1.5h for one sample, and 3h for the other, even for relatively low masses of loaded DNA. The MinION therefore produces data much faster than the Flongle, even for low concentration samples.

However, as mentioned above, although MinION flow cells offer greater sequencing speed, the price of each MinION flow cell precludes the use of one flow cell per sample. Barcoding with SQK-RBK004 allows multiplexing of up to 12 samples, whilst barcoding with more recent kits allows multiplexing of up to 96 (SQK-RBK110.96), but in a clinical setting there may not be sufficient samples to load a full 12+ sample flow cell and still produce results within the desired timeframe.

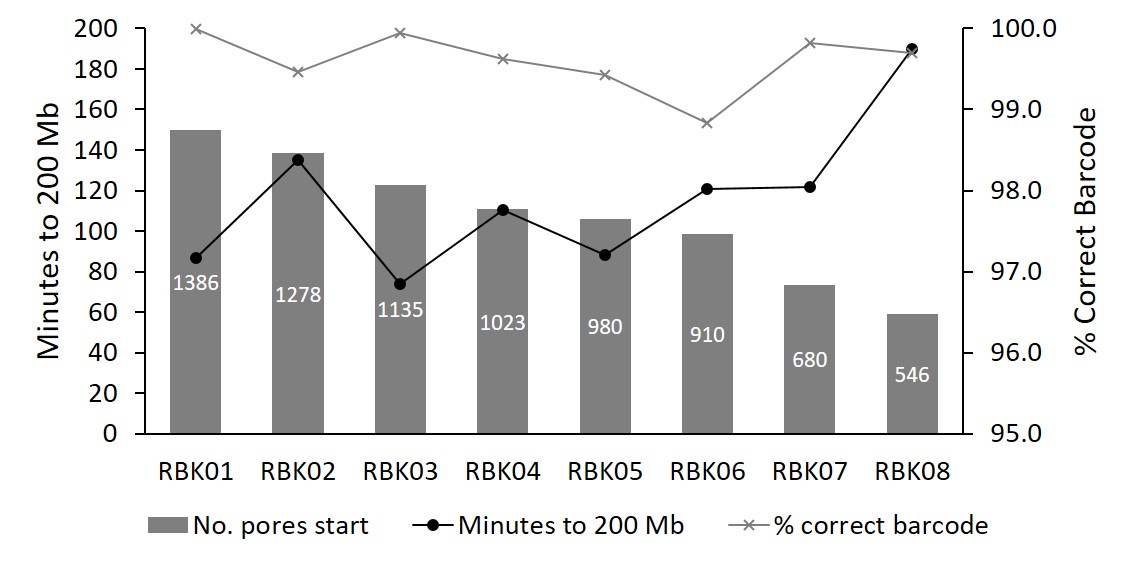
We therefore examined the possibility of re-using MinION flow cells. This would involve sequencing individual samples for as long as necessary to produce the sequence data required, stopping the run and performing a DNase wash of the flow cell, then storing the flow cell until another sample was received. In this way, the same flow cell could theoretically be used many times, thus reducing the cost-per-sample without the need for simultaneous sequencing of multiple samples. Accordingly, we aimed to establish i) how many times could we re-use a flow cell to produce sufficient sequence data in a timely manner, and ii) how much residual DNA from previous samples would remain in the flow cell after washing. Employing a cultured *E. coli* sample, we examined the capacity to repeatedly produce 200 Mbp of DNA within 2h, and what percentage of the reads had the correct barcode attached for each post-wash sequencing run.

We were able to use the same flow cell eight times before it was exhausted, and the percentage of reads with the correct barcode never fell below 98.8% (**Figure 3, Supplementary Table S5**). Although using different barcodes for each subsequent sample reduces the risk of cross-contamination between runs to negligible, especially when paired with a negative control sample each time to control for barcode misclassification, the low levels of DNA carry-over means that available pores were, in general, not being used to sequence previous samples, which would be inefficient and waste time. We also later postulated that only 100 Mbp of sequence may be sufficient for species identification and AMR prediction (representing up to 20x coverage of the average 5 Mbp bacterial genome (72), dependent on levels of host contamination), hence flow cells could potentially be used even more times.

Since the quality and longevity of the pore array likely varies from flow cell to flow cell, and likely also depend on the nature of the samples being sequenced, we also trialled a second flow cell, using real clinical urine samples that had been extracted on the same day they were sequenced (**Supplementary Table S7**). This second flow cell was able to produce 100 Mbp each for nine samples before being used to exhaustion, with samples being sequenced in batches of one to four per run, and over a period of roughly one month. Ultimately, we estimate that the average MinION flow cell can be reliably used and re-used at least eight times to sequence 100-200 Mbp per sample. The slight increase in cost of re-using MinION flow cells versus Flongle flow cells is balanced against producing the data at least eight times faster than it would be on a single-use Flongle flow cell, as well as producing less plastic waste (especially as MinION flow cells can be returned to ONT for recycling, whilst Flongle flow cells cannot).

Fig. 3 MinION flow cell progression when washing and re-using multiple times.

The same 108 ng µl-1 *E. coli* sample was sequenced up to 200Mbp in 8 consecutive sequencing runs on a single flow cell. In between runs, the sequencing was ceased, and the flow cell was washed with wash kit EXP-WSH004-XL according to manufacturer’s instructions. The flow cell was either then stored over one or two nights with storage buffer, or the next sequencing run was immediately commenced.



***Testing the optimised extraction and sequencing process on real clinical samples***

Having established an extraction and sequencing protocol, and a strategy for using and re-using MinION flow cells, we next wanted to test the performance of the finalised protocol for processing clinical samples in real-time, measuring the time taken from sample to result as well as establishing the accuracy of our results compared to those produced by the current gold-standard culture-based tests.

**A wide range of DNA concentrations were retrieved from clinical samples, and we were able to extract DNA from all but the lowest abundance samples**

During the development of this protocol, 44 clinical urine and skin swab samples (19 urine and 25 skin swab) were processed. Of these, 35 produced enough DNA to be sequenced, using either the rapid barcoding or rapid PCR barcoding library preparation kit (see library kit, Guppy version and MinKNOW version in **Supplementary Table S2**), and either MinION or Flongle flow cells. These samples were not usually processed on the same day they were received; instead, they were processed and sequenced in batches, and the raw sample and/or extracted DNA was stored at 4oC in the meantime. After finalising the protocol, we processed a further nine “post-development” urine samples in real-time, which involved collecting, processing and producing results all on the same day. This equalled a total of 53 clinical samples processed here (44 during development and nine post-development), of which 43 were sequenced (35 development and 8 post-development) (**Supplementary Tables S2, S6, and S7, and Table 4**).

A wide variety of DNA concentrations were retrieved from these samples, ranging from 0 to >120 ng µl-1 for the urine samples (mean=37.57, SD= 48.23) and 0 to 116 ng µl-1 for the skin swabs (mean=22.51, SD=32.45). As mentioned previously, the lower limit of detection for sequencing was determined to be < 1 ng µl-1 of cleaned-up DNA. In total, only four urine samples and five skin swabs produced DNA with lower concentrations than our lower limit, and colony forming units (CFU) per ml of sample indicated that those samples had either no growth of any kind, or contained fewer than 1x105 cells. We therefore estimate our lower limit for sufficient DNA from clinical samples to be between 1x105 and 1x106 CFU/ml. The exact lower limit is likely to depend on other sample characteristics, such as the presence of host cells, sediment or inhibitors; some of these factors may assist DNA extraction, whilst others may be detrimental.

**Our optimised protocol can accurately identify species within five hours**

In order to test relevance to a clinical scenario, nine post-development urine samples were processed in real-time. Quantifiable levels of DNA were extracted from eight samples; the remaining sample was classified as “negative” after the DNA extraction, at roughly 4 hours after the sample was received. The average time taken to process the remaining eight samples through to the commencement of sequencing was just under 5 hours (**Table 4 and Supplementary Table S7**). Using the online EPI2ME tool, species identification was carried out in real-time as soon as the first sequencing reads were produced. The speed of EPI2ME can be affected by the number of concurrent online users, but we were able to produce preliminary species identification calls within approximately 20 min of sequencing. In each case, we continued to sequence until 100 Mbp of reads had been produced, theoretically representing up to 20x coverage of the average bacterial genome, depending on levels of host contamination.

EPI2ME uses a database which includes species for which reference exitsts NCBI RefSeq database for viruses, bacteria, fungi and archaea, and does not include *C. lupus familiaris.* The tool can be prone to misclassifying missing species as either “*Homo sapiens”* or “*Escherichia coli*” if they are not present in its standard database. To reduce the likelihood of false positive calls, we therefore also performed species identification using Kraken2 with two custom databases after sequencing was complete. The “pathogens\_plus” database is smaller and therefore produces results more quickly, but could misclassify species if the pathogen being sequenced was missing from the database. Hence, we also ran Kraken2 with the “bacteria\_plus” database, which contains all RefSeq representative bacterial genomes and therefore can clarify any ambiguous results from “pathogens\_plus”. Notably,for our post-development samples, the initial rapid EPI2ME species calls differed from the Kraken2 calls for only one sample, in which EPI2ME detected both *P. mirabilis* and *E. coli*, whilst Kraken2 only identified *P. mirabilis* at an abundance greater than 1% of the bacterial reads, a commonly used threshold for accepting true positive results (18, 73).The culture-based diagnostic process also only identified *P. mirabilis*, indicating a false positive result from EPI2ME. Across all the clinical samples processed during the study, EPI2ME and Kraken2 agreed on the species present in all samples except this one, and one other (SkSw08A, where EPI2ME detected *Finegoldia magna* in addition to *E. coli* and *E. faecalis*, but Kraken2 did not, **Supplementary Table S2**).

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Table 4: Results from processing 9 clinical urine samples in real-time using our final optimised protocol | | | | | | |
| Sample name | **Processing time to sequencing (hours)** | **Processing time to full results (hours)** | **Species Kraken** | **Species VITEK® 2** | **AMR sequencing** | **AST VITEK® 2** |
| Post-dev-1 | 5 | 8 | *S. aureus* | *S. aureus* | Ampicillin/Penicillin, Tetracycline, Fosfomycin, Tigecycline | Penicillin |
| Post-dev-2 | 04:35 | 05:30 | *E. coli* | *E. coli* | Cephalosporins, Colistin | Cefalexin, Cefalotin |
| Post-dev-3 | 05:15 | 8 | *E. coli Streptococcus gallolyticus* | *E. coli (2) Streptococcus gallolyticus ssp pasteurianus* | Trimethoprim, Tetracycline | Cefalexin, Cefalotin Cefalexin, Cefalotin (I) |
| Post-dev-4 | 05:15 | 8 | NA | No growth | None | NA |
| Post-dev-5 | 4 | 4 | NA | No growth | NA | NA |
| Post-dev-6 | 05:15 | 8 | *Mycoplasmopsis canis* | No growth | None | NA |
| Post-dev-7 | 5 | 8 | *E. coli* | *E. coli (2)* | Cephalosporins, Class A beta-lactams, Colistin | Ampicillin, Cefalexin, Cefalotin, Chloramphenicol (I) |
| Post-dev-8 | 5 | 8 | *Proteus mirabilis* | *Proteus mirabilis* | Trimethoprim, Chloramphenicol, Tetracycline, Streptomycin, Sulfonamide, Class A beta-lactams, Streptothricin, Spectinomycin | Ampicillin, (I) Co-amoxiclav, (I) Cefalotin, Enrofloxacin, Marbofloxacin, Predofloxacin, Doxycycline, Tetracycline, Nitrofurantoin, Chloramphenicol, Trimethoprim/Sulfamethoxazole |
| Post-dev-9 | 5 | 8 | *E. coli* | *E. coli* | Cephalosporins, Class A beta-lactams, Colistin, Class C beta-lactams, Streptomycin, Streptothricin, Trimethoprim | Ampicillin, Co-amoxiclav, Cefalexin, Cefalotin, Chloramphenicol (I) |

Culture-based VITEK® 2 results were also available for 32 samples (23 development samples and the nine post-development samples). These were in almost 100% agreement with both EPI2ME and Kraken2 (**Table** **4 and Supplementary Table S7**). For two development samples (SkSk08A and SkSw10) EPI2ME identified additional species compared to the VITEK® 2 results (*Finegoldia magna* and *Escherichia coli*, respectively). As mentioned above, with our Kraken2 databases, *Finegoldia magna* did not occur at a greater than 1% abundance, thus was deemed to be a (non-pathogenic) member of the background skin microbiome. Likewise, interpreting the results of our Kraken2 SkSw10 results alongside our negative control revealed the “*Escherichia coli*” identified by EPI2ME was actually background contamination. This demonstrates the value of running additional thorough taxonomic classification with more extensive databases after EPI2ME real-time analysis, as well as the necessity of negative controls for every extraction and sequencing run.

We were able to extract DNA from two post-development urine samples for which no growth was seen via culture (post-dev-4 and post-dev-6). On species identification with Kraken2, the DNA from one (post-dev-4) was found to be due to contamination, and therefore a false positive. The DNA extracted from post-dev-6, however, was shown by both EPI2ME and Kraken2 to befrom *Mycoplasmopsis canis*. *Mycoplasma* and *Mycoplasmopsis* species are notoriously fastidious and difficult to culture (74); therefore, we believe that in this case, it is highly likely that the sequence-based approach was more sensitive than the current gold-standard culture-based techniques. It is worth noting that while culture-based species identification and ASTs are the “gold standard” they are themselves not 100% accurate, with problems including small sample sizes compared to the infecting population, results skewed to easily cultured and faster growing organisms, and laboratory error (e.g. single doubling dilution steps between sensitive, intermediate and resistant). A 2007 evaluation estimated VITEK® 2 accuracy for species identification to be 98.3%, while AST accuracy was estimated to be 97.7% (75).

**Metagenome-based AMR prediction can be highly accurate, but is dependent on bacterial sequence data volume**

The principle of predicting AMR phenotypes from sequencing data alone involves screening the data for known AMR-related genes or SNPs, based on curated databases like ResFinder or CARD (76, 77). The accuracy of using sequencing data alone for AMR identification varies from species to species, but previous studies have suggested >97% accuracy for *E. coli* and over 99% for *Staphylococcus aureus* (78, 79). This level of accuracy is, however, dependent on the volume of sequencing data available, the complexity of the samples being sequenced, and the amount of contaminating host DNA present. Here we interrogated DNA sequence produced from the 32 urine samples and skin swabs with paired VITEK® 2 results, using our metagenomic protocol for AMR detection and comparing the results to the phenotypic data produced by VITEK® 2 AST, looking at combined “Resistant” and “Intermediate” phenotypes as well as “Resistant” alone (**Supplementary Tables S2, S6 and S7, and Table 4**).

For urine samples, 71.7% of the combined “Resistant” and “Intermediate” resistance phenotypes identified by phenotypic AST were also predicted by our protocol. 50% of samples were in exact agreement for all AMR calls for each sample. Further investigation revealed that the vast majority of AST calls that did not correlate with the sequence-based predictions were defined as intermediate resistance, often to chloramphenicol, which may represent unknown genetic mechanisms. Excluding intermediate resistance calls from the VITEK® 2 results (i.e. looking only at the “Resistant” calls), we detected 83.7% of the same AMR using sequencing data alone. Excluding a single missed gene (likely an *oqx* gene conferring resistance to multiple antimicrobials) in one of our more complex post-optimisation samples (post-dev-8) increases the sensitivity of our AMR predictions to 95.3% (41 of 43 resistant phenotypes accurately predicted, excluding intermediates).

The remaining two AMR phenotypes not detected by our pipeline were from one sample (post-dev-3), which was determined by the VITEK® 2 to be a co-infection consisting of two different strains of *E. coli* and one strain of *Streptococcus gallolyticus ssp pasteurianus*. Though the type of resistance (multiple cephalosporins) identified in this sample was frequently detected accurately by our pipeline, the complexity of this sample may have hindered our AMR prediction. The selection of 100 Mbp as a target was based on a compromise between genome coverage and sequencing time; 100 Mbp represents 20x coverage of the average bacterial genome (72) and is usually easily sequenced in under 2 hours. However, we propose that complicated co-infection samples may require significantly more than 100 Mbp of sequencing data, particularly when host DNA is also present (Kraken2 identified 47.5% of the DNA in this sample as canine). A potential solution for this is the recently improved EPI2ME Labs software, which allows the real-time use of tools such as Kraken2 (with or without custom databases) whilst data is still being sequenced. Real-time analysis will indicate the complexity of a sample, as well as the levels of host contamination, thereby allowing a more informed decision about when “enough” data has been produced. In the meantime, our pipeline can predict the vast majority of resistant phenotypes in non-complex urine samples.

In contrast, for skin swab samples, of which five had paired VITEK® 2 phenotypes, effective sequence-based prediction of AMR was not possible due to the low abundance of the bacteria in the samples, (3x102 to 7.5x105 cfu/ml), almost all below our predicted lower limit of detection. In addition, four of the five samples were more than 95% dog DNA, so bacterial DNA comprised a very low proportion of the extracted DNA (reflecting 1 or 2x coverage). Although species can still easily and quickly be identified for these low abundance skin samples from 1 or 2x coverage, the amount of sequencing data required to accurately detect AMR genotypes is likely cost-prohibitive for veterinary applications, as it would reduce the number of uses of each flow cell. One of our urine samples (post-dev-9) contained over 83% dog DNA, as well as a multidrug-resistant *E*. *coli*, yet we were able to predict all four of its non-intermediate AMR phenotypes from just 100 Mbp. This suggests that even small increases in the relative levels of bacterial cells we are seeing in our low abundance skin swabs could greatly improve our ability to accurately detect AMR genotypes.

Subsequent to sequencing our clinical samples, we tested a feature of the GridION, adaptive sampling, which allows the rejection of DNA which maps to a reference genome during sequencing, and which can therefore potentially deplete some of the dog DNA seen in our samples. We have included our results in **Supplementary Results SR2*,*** which shows that adaptive sampling did greatly reduce the proportion of dog reads sequenced compared to bacterial.Nonetheless, removing host DNA during sequencing is costly in terms of time, as pore-time is still wasted sequencing host DNA until it is rejected, and the repeated rejection of host DNA by reversal of current across individual pores may reduce the overall life-span of a flow cell, resulting in fewer re-uses and therefore increased per-sample cost. In addition, adaptive sampling requires a powerful computer (or GridION) to map reads and make decisions in real-time, the cost of which may preclude its use in smaller clinics. Therefore, reducing the amount of host DNA present in a sample prior to sequencing is likely a better a strategy. Previous studies have shown promising results for pre-extraction host depletion using a range of techniques including saponin-and-DNase enzymatic methods, PMA plus UV-light-based chemical methods, and even methods as simple and cost-effective as physical filtering of host cells through a 22 µm filter prior to extraction (18, 80, 81). A kit also exists (New England Biolab’s NEBNext Microbiome Enrichment kit) which can reduce the levels of eukaryotic DNA in a sample after extraction, based on the predicted higher levels of methylation on eukaryotic DNA compared to bacterial. These methods will need to be explored further, and in the future probably incorporated into our protocol for all sample types which are likely to contain high levels of host cells (such as skin, blood and tissue).

**Concluding comments and future considerations**

We have developed and validated a protocol for the rapid, culture-free, agnostic identification of pathogenic species from clinical canine urine and skin swab samples, by cost-effective metagenomic sequencing. We have shown that this protocol is capable of detecting a wide array of species, representing over 90% of the urine and skin infections seen in the R(D)SVS HfSA. Although we did not test the remaining 10% of species due to their large number and relative rarity, our metagenomic extraction should effectively extract DNA from any species present in a sample. Further, we have shown that we can identify the same species identified by culture-based testing in 100% of the samples we sequenced, using sequencing data alone. False positive results can be ruled out by the use of negative controls and multiple different taxonomic classification methods, and we have also shown an ability to detect pathogenic species which are generally undetectable by culture. A small number of samples had pathogens which were detected by culture but for which we were unable to extract DNA, but these had CFUs/ml counts of 105 or less, indicating a lower limit of detection for our protocol of between 105 and 106 CFUs/ml. We intentionally developed a protocol that can also be adapted to a variety of other sample types, and the MagAttract protocol can be easily adapted for tissue, blood and other bodily fluids. In this way, one simple protocol can be deployed in a clinical setting to detect pathogens in a wide variety of infections, in different animals, in as little as 5 hours, compared to the 48 hours plus commonly seen in the current gold-standard diagnostics techniques.

Although the large amounts of host cell contamination we saw in skin swab samples was problematic with regards to the prediction of AMR, a number of depletion techniques exist which could be incorporated into our protocol, at the cost of time and money, but with the benefit of allowing accurate AMR prediction from a greater range of samples. Besides the samples with the highest levels of host DNA contamination, our ability to match “Resistant” VITEK® 2 calls was relatively accurate, approaching 95% in all but the most complicated co-infections. However, a number of samples displayed “Intermediate” VITEK® 2 resistance to certain antimicrobials, mainly chloramphenicol, which were not predicted from the sequencing data. This suggests that as-yet unknown mechanisms may be responsible for intermediate resistance, or that certain mechanisms are currently missing from the AMR databases used (NCBI, ResFinder and CARD). Our current AMR prediction pipeline includes combining the results of three different tools (EPI2ME, Abricate and AMRFinderPlus), which function in different ways, in order to capture all potential information from our sequencing data.

The methods validated herein focus primarily on the wet-lab; we have used bioinformatics tools which have already been previously well-described and well-validated, including for use with ONT long reads and in human and veterinary clinical metagenomics, and we did not seek here to benchmark them in any further detail than already available in the peer-reviewed literature (18, 32, 73, 82-86). Nonetheless, future optimisations of our pipeline will likely include testing alternative tools and databases, and producing a finalised easy-to-run data analysis workflow. Potential alterations to the methods used here include the addition of Bracken after Kraken2 for species identification, as a recent paper showed this approach to be more accurate and less prone to misclassification of clinical metagenomic samples than Kraken2 alone using long reads (84). Likewise, our current AMR prediction tools may be replaced by the newly released, ISO-certified, abritAMR, which combines the approaches of the tools used herein (87). Future developments to the wet-lab aspects of our protocol include, as described above, the potential inclusion of host depletion steps prior to sequencing. We also note that relatively few of the samples we tested here represented co-infections, although we were able to identify those that did (for example, sample post-dev-3). Future testing of larger numbers of clinical samples will help to validate our ability to accurately identify all pathogens in co-infection samples.

Although we focussed on urine and skin swab samples here, we aimed to design a protocol which could be adapted to an array of other sample types (e.g. blood) from infections in other animals including humans. Moreover, the approach could be used in a variety of clinical settings, from small practices to large hospitals. By comparing and optimising a number of different kit-based gDNA extractions and sequencing techniques, combined with community-built DNA analysis tools, we have developed a pipeline which can identify the bacterial species present in clinical samples in as little as 5 hours. It can also predict the antimicrobial resistance phenotype of those species with up to 95% accuracy, in around 8 hours.

Lastly, we note that the flow cells (R9.4.1) and kits (SQK-RBK004) used here throughout development are now being replaced with R10.4.1 and SQK-RBK114. We expect these new flow cells and kits to incorporate seamlessly into our existing protocol and, indeed, they will likely improve the accuracy of the sequencing data produced, which could in turn improve the accuracy of our species identification and prediction of SNP-based AMR.

# Author statements

## Authors and contributors

Conceptualization: NR, AL, BW, GP, TN, RM, DG, JRF; Data curation: NR, AL; Formal analysis: NR; Funding acquisition: BW, GP, TN, DG, JRF; Investigation: AL, BW, NR; Methodology: NR, AL, BW; Project administration: JRF; Resources: GP, TN; Supervision: RM, DG, JRF; Validation: NR, AL, BW; Visualization: NR, BW; Writing - original draft: NR, JRF; Writing - review & editing: NR, AL, BW, GP, TN, RM, DG, JRF.

## Conflicts of interest

The authors declare there are no conflicts of interest affecting this work.

## Funding information

This work was funded by the Dog’s Trust Canine Welfare Grant, Direct Diagnostic Genomics Towards Effective Antimicrobial Use (“Dogstails”).

## Ethical approval

This work involved the use of non-experimental pet animals only and followed established internationally recognised high standards (‘best practice’) of individual veterinary clinical patient care. Samples were non-invasive and/or were excess from those taken during ordinary clinical treatment and, as such, were covered by the R(D)SVS’s ethical approval for research.

## Acknowledgements

Many thanks to Jennifer Harris, Easter Bush Pathology, R(D)SVS, University of Edinburgh, for sharing clinical samples and the results thereof. Many thanks also to the Hospital for Small Animals’ Clinical Biobank for providing a single surplus blood sample for canine DNA extraction.

The authors are grateful for the fantastic bioinformatics resource, CLIMB-BIG-DATA (developed by the MRC, grant number MR/T030062/1), without which much of the data analysis undertaken here would not have been possible.

For the purpose of open access, the authors have applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising from this submission.

# References

1. Antimicrobial Resistance C. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. Lancet (London, England). 2022;399(10325):629-55.

2. WHO WHO. Antibiotic Resistance 2020 [Available from: <https://www.who.int/news-room/fact-sheets/detail/antibiotic-resistance>.

3. Nuttall T. Choosing the best antimicrobial for the job. Vet Rec. 2013;172(1):12-3.

4. Samrot AV, Sean TC, Bhavya KS, Sahithya CS, Chan-Drasekaran S, Palanisamy R, et al. Leptospiral Infection, Pathogenesis and Its Diagnosis-A Review. Pathogens. 2021;10(2):145.

5. Riedel S, Halls J, Dutta S, Toraskar N, Lemon J, Carter K, et al. Clinical evaluation of the acuitas® AMR gene panel for rapid detection of bacteria and genotypic antibiotic resistance determinants. Diagnostic microbiology and infectious disease. 2021;100(4):115383.

6. Simner PJ, Musser KA, Mitchell K, Wise MG, Lewis S, Yee R, et al. Multicenter Evaluation of the Acuitas AMR Gene Panel for Detection of an Extended Panel of Antimicrobial Resistance Genes among Bacterial Isolates. J Clin Microbiol. 2022;60(3):e0209821.

7. Baldan R, Cliff PR, Burns S, Medina A, Smith GC, Batra R, et al. Development and evaluation of a nanopore 16S rRNA gene sequencing service for same day targeted treatment of bacterial respiratory infection in the intensive care unit. Journal of Infection. 2021;83(2):167-74.

8. Fu Y, Chen Q, Xiong M, Zhao J, Shen S, Chen L, et al. Clinical Performance of Nanopore Targeted Sequencing for Diagnosing Infectious Diseases. Microbiol Spectr. 2022;10(2):e0027022.

9. Komiya S, Matsuo Y, Nakagawa S, Morimoto Y, Kryukov K, Okada H, et al. MinION, a portable long-read sequencer, enables rapid vaginal microbiota analysis in a clinical setting. BMC Medical Genomics. 2022;15(1):68.

10. Matsuo Y. Full-Length 16S rRNA Gene Analysis Using Long-Read Nanopore Sequencing for Rapid Identification of Bacteria from Clinical Specimens. In: Arakawa K, editor. Nanopore Sequencing: Methods and Protocols. New York, NY: Springer US; 2023. p. 193-213.

11. Matsuo Y, Komiya S, Yasumizu Y, Yasuoka Y, Mizushima K, Takagi T, et al. Full-length 16S rRNA gene amplicon analysis of human gut microbiota using MinION™ nanopore sequencing confers species-level resolution. BMC Microbiology. 2021;21(1):35.

12. Sheka D, Alabi N, Gordon PMK. Oxford nanopore sequencing in clinical microbiology and infection diagnostics. Briefings in bioinformatics. 2021;22(5):bbaa403.

13. Zhang LL, Zhang C, Peng JP. Application of Nanopore Sequencing Technology in the Clinical Diagnosis of Infectious Diseases. Biomedical and Environmental Sciences. 2022;35(5):381-92.

14. Avershina E, Frye SA, Ali J, Taxt AM, Ahmad R. Ultrafast and Cost-Effective Pathogen Identification and Resistance Gene Detection in a Clinical Setting Using Nanopore Flongle Sequencing. Frontiers in Microbiology. 2022;13.

15. Taxt AM, Avershina E, Frye SA, Naseer U, Ahmad R. Rapid identification of pathogens, antibiotic resistance genes and plasmids in blood cultures by nanopore sequencing. Scientific reports. 2020;10(1):7622.

16. Street TL, Barker L, Sanderson ND, Kavanagh J, Hoosdally S, Cole K, et al. Optimizing DNA Extraction Methods for Nanopore Sequencing of Neisseria gonorrhoeae Directly from Urine Samples. Journal of Clinical Microbiology. 2020;58(3):e01822-19.

17. Sanderson ND, Swann J, Barker L, Kavanagh J, Hoosdally S, Crook D, et al. High precision Neisseria gonorrhoeae variant and antimicrobial resistance calling from metagenomic Nanopore sequencing. Genome research. 2020;30(9):1354-63.

18. Charalampous T, Kay GL, Richardson H, Aydin A, Baldan R, Jeanes C, et al. Nanopore metagenomics enables rapid clinical diagnosis of bacterial lower respiratory infection. Nat Biotechnol. 2019;37(7):783-92.

19. Yang L, Haidar G, Zia H, Nettles R, Qin S, Wang X, et al. Metagenomic identification of severe pneumonia pathogens with rapid Nanopore sequencing in mechanically-ventilated patients. medRxiv. 2019:19002774.

20. Schmidt K, Mwaigwisya S, Crossman LC, Doumith M, Munroe D, Pires C, et al. Identification of bacterial pathogens and antimicrobial resistance directly from clinical urines by nanopore-based metagenomic sequencing. J Antimicrob Chemother. 2017;72(1):104-14.

21. Whittle E, Yonkus JA, Jeraldo P, Alva-Ruiz R, Nelson H, Kendrick ML, et al. Optimizing Nanopore Sequencing for Rapid Detection of Microbial Species and Antimicrobial Resistance in Patients at Risk of Surgical Site Infections. mSphere. 2022;7(1):e0096421.

22. Janes VA, Matamoros S, Munk P, Clausen P, Koekkoek SM, Koster LAM, et al. Metagenomic DNA sequencing for semi-quantitative pathogen detection from urine: a prospective, laboratory-based, proof-of-concept study. The Lancet Microbe. 2022;3(8):e588-e97.

23. Irwin AD, Coin LJM, Harris PNA, Cotta MO, Bauer MJ, Buckley C, et al. Optimising Treatment Outcomes for Children and Adults Through Rapid Genome Sequencing of Sepsis Pathogens. A Study Protocol for a Prospective, Multi-Centre Trial (DIRECT). Frontiers in Cellular and Infection Microbiology. 2021;11.

24. Lycett SJ, Hughes J, McHugh MP, Filipe AdS, Dewar R, Lu L, et al. Epidemic waves of COVID-19 in Scotland: a genomic perspective on the impact of the introduction and relaxation of lockdown on SARS-CoV-2. medRxiv. 2021:2021.01.08.20248677.

25. Charalampous T, Alcolea-Medina A, Snell LB, Alder C, Tan M, Williams TGS, et al. Routine respiratory metagenomics service for intensive care unit patients. medRxiv. 2023:2023.05.15.23289731.

26. Břinda K, Callendrello A, Ma KC, MacFadden DR, Charalampous T, Lee RS, et al. Rapid inference of antibiotic resistance and susceptibility by genomic neighbour typing. Nature Microbiology. 2020;5(3):455-64.

27. Steinig E, Pitt M, Aglua I, Suttie A, Greenhill A, Heather C, et al. Genomic neighbor typing for bacterial outbreak surveillance. bioRxiv. 2022:2022.02.05.479210.

28. Jeck WR, Iafrate AJ, Nardi V. Nanopore Flongle Sequencing as a Rapid, Single-Specimen Clinical Test for Fusion Detection. The Journal of Molecular Diagnostics. 2021;23(5):630-6.

29. Moss EL, Maghini DG, Bhatt AS. Complete, closed bacterial genomes from microbiomes using nanopore sequencing. Nat Biotechnol. 2020.

30. Huggins LG, Colella V, Atapattu U, Koehler AV, Traub RJ. Nanopore Sequencing Using the Full-Length 16S rRNA Gene for Detection of Blood-Borne Bacteria in Dogs Reveals a Novel Species of Hemotropic Mycoplasma. Microbiol Spectr. 2022;10(6):e0308822.

31. Bokma J, Vereecke N, Pas ML, Chantillon L, Vahl M, Weesendorp E, et al. Evaluation of Nanopore Sequencing as a Diagnostic Tool for the Rapid Identification of Mycoplasma bovis from Individual and Pooled Respiratory Tract Samples. J Clin Microbiol. 2021;59(12):e0111021.

32. Kamathewatta KI, Bushell RN, Young ND, Stevenson MA, Billman-Jacobe H, Browning GF, et al. Exploration of antibiotic resistance risks in a veterinary teaching hospital with Oxford Nanopore long read sequencing. PloS one. 2019;14(5):e0217600.

33. Lanszki Z, Tóth GE, Schütz É, Zeghbib S, Rusvai M, Jakab F, et al. Complete genomic sequencing of canine distemper virus with nanopore technology during an epizootic event. Scientific reports. 2022;12(1):4116.

34. Brown E, Freimanis G, Shaw AE, Horton DL, Gubbins S, King D. Characterising Foot-and-Mouth Disease Virus in Clinical Samples Using Nanopore Sequencing. Front Vet Sci. 2021;8:656256.

35. Vanmechelen B, Bertelsen MF, Rector A, Van den Oord JJ, Laenen L, Vergote V, et al. Identification of a novel species of papillomavirus in giraffe lesions using nanopore sequencing. Veterinary Microbiology. 2017;201:26-31.

36. D’Andreano S, Cuscó A, Francino O. Rapid and real-time identification of fungi up to species level with long amplicon nanopore sequencing from clinical samples. Biology Methods and Protocols. 2021;6(1):bpaa026.

37. Ferrer L, García-Fonticoba R, Pérez D, Viñes J, Fàbregas N, Madroñero S, et al. Whole genome sequencing and de novo assembly of Staphylococcus pseudintermedius: a pangenome approach to unravelling pathogenesis of canine pyoderma. Veterinary dermatology. 2021;32(6):654-63.

38. Loeffler A, Lloyd DH. What has changed in canine pyoderma? A narrative review. Veterinary journal (London, England : 1997). 2018;235:73-82.

39. Wong C, Epstein SE, Westropp JL. Antimicrobial Susceptibility Patterns in Urinary Tract Infections in Dogs (2010-2013). Journal of veterinary internal medicine. 2015;29(4):1045-52.

40. White SD. Systemic treatment of bacterial skin infections of dogs and cats. Veterinary dermatology. 1996;7(3):133-43.

41. Paradis M, Lemay S, Scott DW, Miller WH, Wellington J, Panich R. Efficacy of Enrofloxacin in the Treatment of Canine Bacterial Pyoderma. Veterinary dermatology. 1990;1(3):123-7.

42. Byron JK. Urinary Tract Infection. Veterinary Clinics of North America: Small Animal Practice. 2019;49(2):211-21.

43. Norris CR, Williams BJ, Ling GV, Franti CE, Johnson, Ruby AL. Recurrent and persistent urinary tract infections in dogs: 383 cases (1969-1995). Journal of the American Animal Hospital Association. 2000;36(6):484-92.

44. Reddy BS, Kumari KN, Rao VV, Rayulu VC, Sivajothi S. Efficacy of Enrofloxacin in the Treatment of Recurrent Pyoderma in Dogs. Journal of Advanced Veterinary Research. 2014;4(3):108-12.

45. Bannoehr J, Ben Zakour NL, Waller AS, Guardabassi L, Thoday KL, van den Broek AH, et al. Population genetic structure of the Staphylococcus intermedius group: insights into agr diversification and the emergence of methicillin-resistant strains. Journal of bacteriology. 2007;189(23):8685-92.

46. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. Genome Biology. 2019;20(1):257.

47. McIntyre KM, Setzkorn C, Hepworth PJ, Morand S, Morse AP, Baylis M. A Quantitative Prioritisation of Human and Domestic Animal Pathogens in Europe. PloS one. 2014;9(8):e103529.

48. Nicholls SM, Quick JC, Tang S, Loman NJ. Ultra-deep, long-read nanopore sequencing of mock microbial community standards. GigaScience. 2019;8(5):giz043.

49. De Coster W, D’Hert S, Schultz DT, Cruts M, Van Broeckhoven C. NanoPack: visualizing and processing long-read sequencing data. Bioinformatics (Oxford, England). 2018;34(15):2666-9.

50. Wick RR. Porechop: Adapter trimmer for Oxford Nanopore reads 2017 [Available from: <https://github.com/rrwick/Porechop>.

51. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol. 2019;37(5):540-6.

52. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics (Oxford, England). 2014;30(14):2068-9.

53. Seemann T. ABRicate 2022 [Available from: <https://github.com/tseemann/abricate>.

54. Feldgarden M, Brover V, Gonzalez-Escalona N, Frye JG, Haendiges J, Haft DH, et al. AMRFinderPlus and the Reference Gene Catalog facilitate examination of the genomic links among antimicrobial resistance, stress response, and virulence. Scientific reports. 2021;11(1):12728.

55. Hall MB. Rasusa: Randomly subsample sequencing reads to a specified coverage. Journal of Open Source Software. 2022;7(69):3941.

56. Ackerman AL, Anger JT, Khalique MU, Ackerman JE, Tang J, Kim J, et al. Optimization of DNA extraction from human urinary samples for mycobiome community profiling. PloS one. 2019;14(4):e0210306.

57. Bialasiewicz S, Duarte TPS, Nguyen SH, Sukumaran V, Stewart A, Appleton S, et al. Rapid diagnosis of Capnocytophaga canimorsus septic shock in an immunocompetent individual using real-time Nanopore sequencing: a case report. BMC Infectious Diseases. 2019;19(1):660.

58. Yang L, Haidar G, Zia H, Nettles R, Qin S, Wang X, et al. Metagenomic identification of severe pneumonia pathogens in mechanically-ventilated patients: a feasibility and clinical validity study. Respiratory Research. 2019;20(1):265.

59. Ferreira FA, Helmersen K, Visnovska T, Jørgensen SB, Aamot HV. Rapid nanopore-based DNA sequencing protocol of antibiotic-resistant bacteria for use in surveillance and outbreak investigation. Microb Genom. 2021;7(4).

60. Dalla-Costa LM, Morello LG, Conte D, Pereira LA, Palmeiro JK, Ambrosio A, et al. Comparison of DNA extraction methods used to detect bacterial and yeast DNA from spiked whole blood by real-time PCR. Journal of microbiological methods. 2017;140:61-6.

61. Kim D, Hofstaedter CE, Zhao C, Mattei L, Tanes C, Clarke E, et al. Optimizing methods and dodging pitfalls in microbiome research. Microbiome. 2017;5(1):52.

62. Lucena-Aguilar G, Sánchez-López AM, Barberán-Aceituno C, Carrillo-Ávila JA, López-Guerrero JA, Aguilar-Quesada R. DNA Source Selection for Downstream Applications Based on DNA Quality Indicators Analysis. Biopreservation and biobanking. 2016;14(4):264-70.

63. Helmersen K, Aamot HV. DNA extraction of microbial DNA directly from infected tissue: an optimized protocol for use in nanopore sequencing. Scientific reports. 2020;10(1):2985.

64. Wang CX, Huang Z, Fang W, Zhang Z, Fang X, Li W, et al. Preliminary Assessment of Nanopore-based Metagenomic Sequencing for the Diagnosis of Prosthetic Joint Infection. International journal of infectious diseases : IJID : official publication of the International Society for Infectious Diseases. 2020.

65. Israeli O, Guedj-Dana Y, Shifman O, Lazar S, Cohen-Gihon I, Amit S, et al. Rapid Amplicon Nanopore Sequencing (RANS) for the Differential Diagnosis of Monkeypox Virus and Other Vesicle-Forming Pathogens. Viruses. 2022;14(8):1817.

66. Chan WS, Au CH, Lam HY, Wang CLN, Ho DN-Y, Lam YM, et al. Evaluation on the use of Nanopore sequencing for direct characterization of coronaviruses from respiratory specimens, and a study on emerging missense mutations in partial RdRP gene of SARS-CoV-2. Virology Journal. 2020;17(1):183.

67. Mosbruger TL, Dinou A, Duke JL, Ferriola D, Mehler H, Pagkrati I, et al. Utilizing nanopore sequencing technology for the rapid and comprehensive characterization of eleven HLA loci; addressing the need for deceased donor expedited HLA typing. Human Immunology. 2020;81(8):413-22.

68. De Santis D, Truong L, Martinez P, D'Orsogna L. Rapid high-resolution HLA genotyping by MinION Oxford nanopore sequencing for deceased donor organ allocation. HLA. 2020;96(2):141-62.

69. de Siqueira GMV, Pereira-dos-Santos FM, Silva-Rocha R, Guazzaroni M-E. Nanopore Sequencing Provides Rapid and Reliable Insight Into Microbial Profiles of Intensive Care Units. Frontiers in Public Health. 2021;9.

70. Grädel C, Terrazos Miani MA, Barbani MT, Leib SL, Suter-Riniker F, Ramette A. Rapid and Cost-Efficient Enterovirus Genotyping from Clinical Samples Using Flongle Flow Cells. Genes (Basel). 2019;10(9).

71. Leggett RM, Alcon-Giner C, Heavens D, Caim S, Brook TC, Kujawska M, et al. Rapid MinION profiling of preterm microbiota and antimicrobial-resistant pathogens. Nature microbiology. 2020;5(3):430-42.

72. Land M, Hauser L, Jun SR, Nookaew I, Leuze MR, Ahn TH, et al. Insights from 20 years of bacterial genome sequencing. Functional & integrative genomics. 2015;15(2):141-61.

73. Charalampous T, Alcolea-Medina A, Snell LB, Williams TGS, Batra R, Alder C, et al. Evaluating the potential for respiratory metagenomics to improve treatment of secondary infection and detection of nosocomial transmission on expanded COVID-19 intensive care units. Genome Medicine. 2021;13(1):182.

74. Jansson E. Isolation of fastidious mycoplasma from human sources. Journal of clinical pathology. 1971;24(1):53-6.

75. Nakasone I, Kinjo T, Yamane N, Kisanuki K, Shiohira CM. Laboratory-based evaluation of the colorimetric VITEK-2 Compact system for species identification and of the Advanced Expert System for detection of antimicrobial resistances: VITEK-2 Compact system identification and antimicrobial susceptibility testing. Diagnostic microbiology and infectious disease. 2007;58(2):191-8.

76. Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, et al. Identification of acquired antimicrobial resistance genes. J Antimicrob Chemother. 2012;67(11):2640-4.

77. Alcock BP, Raphenya AR, Lau TTY, Tsang KK, Bouchard M, Edalatmand A, et al. CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. Nucleic Acids Research. 2020;48(D1):D517-D25.

78. Tyson GH, McDermott PF, Li C, Chen Y, Tadesse DA, Mukherjee S, et al. WGS accurately predicts antimicrobial resistance in Escherichia coli. J Antimicrob Chemother. 2015;70(10):2763-9.

79. Holden MT, Hsu LY, Kurt K, Weinert LA, Mather AE, Harris SR, et al. A genomic portrait of the emergence, evolution, and global spread of a methicillin-resistant Staphylococcus aureus pandemic. Genome Res. 2013;23(4):653-64.

80. Marotz CA, Sanders JG, Zuniga C, Zaramela LS, Knight R, Zengler K. Improving saliva shotgun metagenomics by chemical host DNA depletion. Microbiome. 2018;6(42).

81. Shi Y, Wang G, Lau HC, Yu J. Metagenomic Sequencing for Microbial DNA in Human Samples: Emerging Technological Advances. International journal of molecular sciences. 2022;23(4).

82. Ye SH, Siddle KJ, Park DJ, Sabeti PC. Benchmarking Metagenomics Tools for Taxonomic Classification. Cell. 2019;178(4):779-94.

83. Pearman WS, Freed NE, Silander OK. Testing the advantages and disadvantages of short- and long- read eukaryotic metagenomics using simulated reads. BMC Bioinformatics. 2020;21(1):220.

84. Govender KN, Eyre DW. Benchmarking taxonomic classifiers with Illumina and Nanopore sequence data for clinical metagenomic diagnostic applications. Microbial Genomics. 2022;8(10).

85. Dahl LG, Joensen KG, Østerlund MT, Kiil K, Nielsen EM. Prediction of antimicrobial resistance in clinical Campylobacter jejuni isolates from whole-genome sequencing data. European journal of clinical microbiology & infectious diseases : official publication of the European Society of Clinical Microbiology. 2021;40(4):673-82.

86. Sydenham TV, Overballe-Petersen S, Hasman H, Wexler H, Kemp M, Justesen US. Complete hybrid genome assembly of clinical multidrug-resistant Bacteroides fragilis isolates enables comprehensive identification of antimicrobial-resistance genes and plasmids. Microb Genom. 2019;5(11).

87. Sherry NL, Horan KA, Ballard SA, Gonҫalves da Silva A, Gorrie CL, Schultz MB, et al. An ISO-certified genomics workflow for identification and surveillance of antimicrobial resistance. Nature communications. 2023;14(1):60.