White Rose Research Online

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk
Universities of Leeds, Sheffield and York

Deposited via The University of Sheffield.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/247762/

Version: Published Version

Article:
Saville, T.J., Dabrowska, M.B., Zhang, P. et al. (2026) Clinical accuracy and operational evaluation of a novel nanopore-based 16S sequencing workflow for diagnostic pathogen detection. Journal of Infection, 93 (5). 106876. ISSN: 0163-4453

https://doi.org/10.1016/j.jinf.2026.106876

Reuse
This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:
https://creativecommons.org/licenses/

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

UNIVERSITY OF LEEDS
University of Sheffield
UNIVERSITY of York

Journal of Infection 93 (2026) 106876

Contents lists available at ScienceDirect

Journal of Infection

journal homepage: www.sciencedirect.com/journal/journal-of-infection

Clinical accuracy and operational evaluation of a novel nanopore-based 16S sequencing workflow for diagnostic pathogen detection

Thomas J. Saville [a,b,*], Magdalena B. Dabrowska [a], Peijun Zhang [a], Joanne Bullivant [b], Amy State [c], Jack Goodall [a,b], Sharon Hsu [a], Katie Johnson [a,c], Christopher Lynch [c], Rob Townsend [c], Gayti Morris [c], Helena Parsons [b,c], David Partridge [b,c], Thushan I. de Silva [a,b]

[a] Division of Clinical Medicine, School of Medicine and Population Health, The University of Sheffield, United Kingdom
[b] The Florey Institute of Infection and Sheffield NIHR Biomedical Research Centre, The University of Sheffield, United Kingdom
[c] Sheffield Laboratory Medicine, Sheffield Teaching Hospital NHS Foundation Trust, Sheffield, United Kingdom

ARTICLE INFO

Keywords:
16S rRNA sequencing
Oxford Nanopore long-read sequencing
Pathogen detection
Culture-negative infections
Species-level identification
Clinical diagnostics

ABSTRACT

Culture-based methods remain standard practice for pathogenic bacterial detection in routine diagnostic laboratories but frequently fail to identify pathogens, particularly in low bacterial burden or antibiotic-exposed samples. While full-length 16S rRNA sequencing improves detection and taxonomic resolution, its adoption is hindered by bioinformatic complexity and operational feasibility.

We developed an in-house Oxford Nanopore Technologies (ONT)-based full-length 16S assay coupled with an open-access, automated analysis pipeline designed for non-specialist staff. In this pilot study, 116 culture-negative clinical samples underwent sequencing, with 86 tested in parallel against outsourced standard-of-care 16S or species-specific PCR. Medical microbiologists, blinded to comparative results, independently assessed clinical relevance. Turnaround time, workflow integration, and cost requirements were also evaluated.

Compared to outsourced 16S testing, the in-house workflow demonstrated a 100% positive percent agreement (12/12; 95% CI: 75.7–100.0%). It significantly expanded clinical diagnostic yield from 14.0% (12/86) to 37.2% (32/86) ($p<0.0001$), identifying clinically relevant organisms in an additional 20 samples reported as negative by outsourced testing. It enhanced species-level resolution and identified polymicrobial infections (n=8). The NanoPATH pipeline generated clinically interpretable outputs via a single-command workflow and implementation modelling demonstrated that twice-weekly runs would achieve comparable turnaround times to outsourced testing, with cost-neutrality reached at 9 samples per run.

An open-access ONT-based 16S workflow that is feasible for routine diagnostic laboratory implementation and operated without specialist bioinformaticians significantly improves pathogen detection from clinical samples. By overcoming analytical and operational barriers, this approach provides a practical framework for integrating sequencing-based diagnostics into standard clinical workflows and improving patient care.

Introduction

Culture-based methods remain standard practice for identifying pathogenic bacteria in routine clinical diagnostic laboratories, but frequently fail to identify infectious aetiologies, particularly in low-burden samples or patients with prior antibiotic exposure [1,2]. Molecular approaches such as targeted quantitative polymerase chain reaction (qPCR) improve sensitivity but are inherently limited to predefined organisms [3,4]. In contrast, sequencing-based methods offer a pathogen-agnostic approach improving diagnostic yield in culture-negative samples [4,5].

The 16S ribosomal RNA (rRNA) gene (~1500 bp) consists of nine hypervariable regions (V1–V9) interspersed with conserved sequences, enabling universal PCR amplification and species-level differentiation [5–7]. Traditional short-read assays target isolated regions such as V3–V4 [8,9], but long-read sequencing captures the full-length 16S gene

* Correspondence to: School of Medicine and Population Health, Room LU112, L Floor, Royal Hallamshire Hospital, Glossop Road, Sheffield S10 2JF, United Kingdom
E-mail address: t.j.saville@sheffield.ac.uk (T.J. Saville).

https://doi.org/10.1016/j.jinf.2026.106876
Accepted 9 October 2026
Available online 16 October 2026
0163-4453/© 2026 The Authors. Published by Elsevier Ltd on behalf of British Infection Association. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

providing enhanced taxonomic resolution and species-level identification [7].

Despite this, adoption of 16S testing in routine diagnostic laboratories is limited. It is often outsourced to specialist providers, resulting in high costs, delays, and reduced integration into clinical workflows. A key barrier to local implementation is the reliance on specialist bioinformatics expertise to process and interpret sequencing data. Additionally, few studies address how to operationalise and integrate these assays logistically and financially within routine diagnostic laboratories.

To address these barriers, we developed an in-house Oxford Nanopore Technologies (ONT)-based workflow for full-length 16S rRNA sequencing, coupled with an automated analysis pipeline, NanoPATH [10], executable without specialist bioinformatics expertise. We evaluated its diagnostic performance against the current outsourced standard-of-care and assessed its operational integration, turnaround time (TAT), and cost-effectiveness within a routine clinical diagnostic laboratory.

Methods

Standard-of-care and sample selection

Samples requiring 16S rRNA sequencing from Sheffield Teaching Hospitals NHS Foundation Trust (STH) are routinely outsourced to Micropathology Ltd. [8] for 16S PCR and subsequent Sanger sequencing if positive. To evaluate our in-house assay, we aimed to perform parallel testing at The University of Sheffield on a consecutive series of 80–120 residual culture-negative sterile-site samples collected prospectively between 06/03/2023 and 19/09/2023, a pragmatic target based on historical volumes. The full sampling and analytical process is shown in Fig. 1.

Nucleic acid extraction

DNA from 200 µl sample aliquots was extracted using a Roche MagNA Pure 96 Instrument (DNA and Viral NA Small Volume Kit), eluting in 50 µl nuclease-free water (NFW). Each run included a NFW negative extraction control and a positive control of pure *Moraxella osloensis* (Gram-negative) and *Bacillus subtilis* (Gram-positive) cultures, chosen for their clinical irrelevance in sterile samples. Cultured controls were diluted in 5 ml NFW to 0.5 McFarland, then diluted 1:100 for assay use. A secondary NFW PCR-negative control was included during amplification to isolate potential contamination steps.

16S rRNA amplification, sequencing, and assay optimisation

Amplification and library preparation used ONT 16S Barcoding Kits (SQK-RAB204, SQK-16S024, SQK-16S114.24). Initially, 10 ng of ZymoBIOMICS Microbial Community DNA Standard (D6311) was amplified (25 cycles) and sequenced to verify accurate species identification and relative abundance against the manufacturer's datasheet. To mimic low biomass clinical samples, total input was reduced to 1 ng and PCR cycles increased to 35. An NFW negative control assessed whether increased cycling amplified background contaminants. To determine the assay's detection threshold, the ZymoBIOMICS standard was serially diluted (1 ng to 10 fg total input) and run using the optimised 35-cycle protocol. Additionally, comparative runs were conducted using twelve 16S positive clinical samples to evaluate performance across different ONT sequencing kits and flow cell chemistries (see web-only Supplementary Material S1, S2).

For clinical testing, extracted DNA input was maximised (24 µl or 15 µl, depending on the kit). Full-length 16S genes were amplified using 27F (AGAGTTTGATCMTGGCTCAG) and 1492R (CGGTTACCTTGTTACGACTT) primers with barcoded tails (see web-only Supplementary Material S3). PCR reactions contained 25 µl LongAmp Taq 2X Master Mix (NEB) with either 24 µl extract and 1 µl 10 µM barcode primer

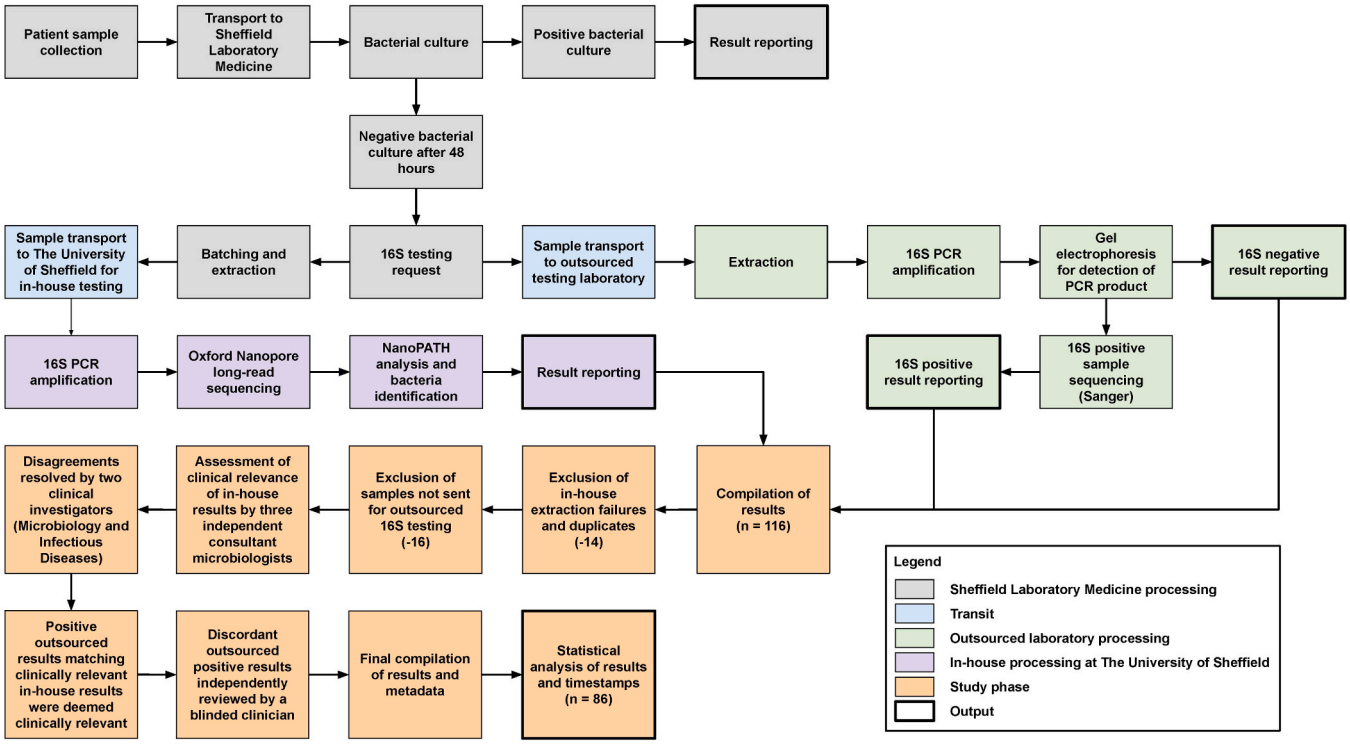


Fig. 1. Study and analytical workflow comparing 16S testing pathways. Diagram tracking sample flow across clinical and research environments: Sheffield Laboratory Medicine (grey), transit (blue), outsourced laboratory processing (green), and in-house processing (purple). The study phase (orange) outlines the multi-stage clinical review mechanism for resolving discordant data and defines the final sample cohort size (n=86) following data-cleaning exclusions.

(SQK-RAB204), or 15 µl extract and 10 µl 1 µM barcode primer (SQK-16S024/SQK-16S114.24). Thermocycling conditions were 95°C for 1 min; 35 cycles of 95°C for 20 s, 55°C for 30 s, 65°C for 2 min; and a 5 min final extension at 65°C. Amplicons were purified using 30 µl AMPure XP (Beckman Coulter), washed with 80% ethanol, and eluted in 10 µl Nanopore Elution Buffer. Following quantification by Qubit fluorometer (HS dsDNA) (Invitrogen), up to 24 samples were pooled (~10–15 µl total volume; ~115 ng DNA). 1 µl ONT Rapid Adaptor was added, incubating at room temperature for 15 min.

Sequencing was performed on a GridION Mk1 using ONT MinION Flow Cells (R9.4.1, R10.4.1). Flow cell priming, library loading, and sequencing followed manufacturer protocols. Raw signal data were basecalled, quality-filtered, and demultiplexed using Guppy (v6.4.8) [9] with the high-accuracy model. Reads <400 bp, with a Phred score <8 or without barcodes detected at both ends were excluded.

NanoPATH analysis pipeline and reporting interface

To overcome the limitations of individual read classification which can lead to inflated species counts in ONT's wf-metagenomics (v2.13.0) workflow [11], we developed NanoPATH. Both workflows were benchmarked in parallel using the ZymoBIOMICS Standard (see web-only Supplementary Material S4).

NanoPATH was adapted from NanoCLUST [12], with both tools sharing a core analytical framework of UMAP dimensionality reduction, HDBSCAN clustering, Canu error correction and consensus polishing. While NanoCLUST was designed as a single-sample research tool, classifying exclusively with BLAST, NanoPATH includes key modifications to meet clinical diagnostic and nf-core portability standards. These include automated quality control, primary read classification, and a multi-classifier read assignment. The pipeline processes demultiplexed FASTQ reads through the following sequential stages:

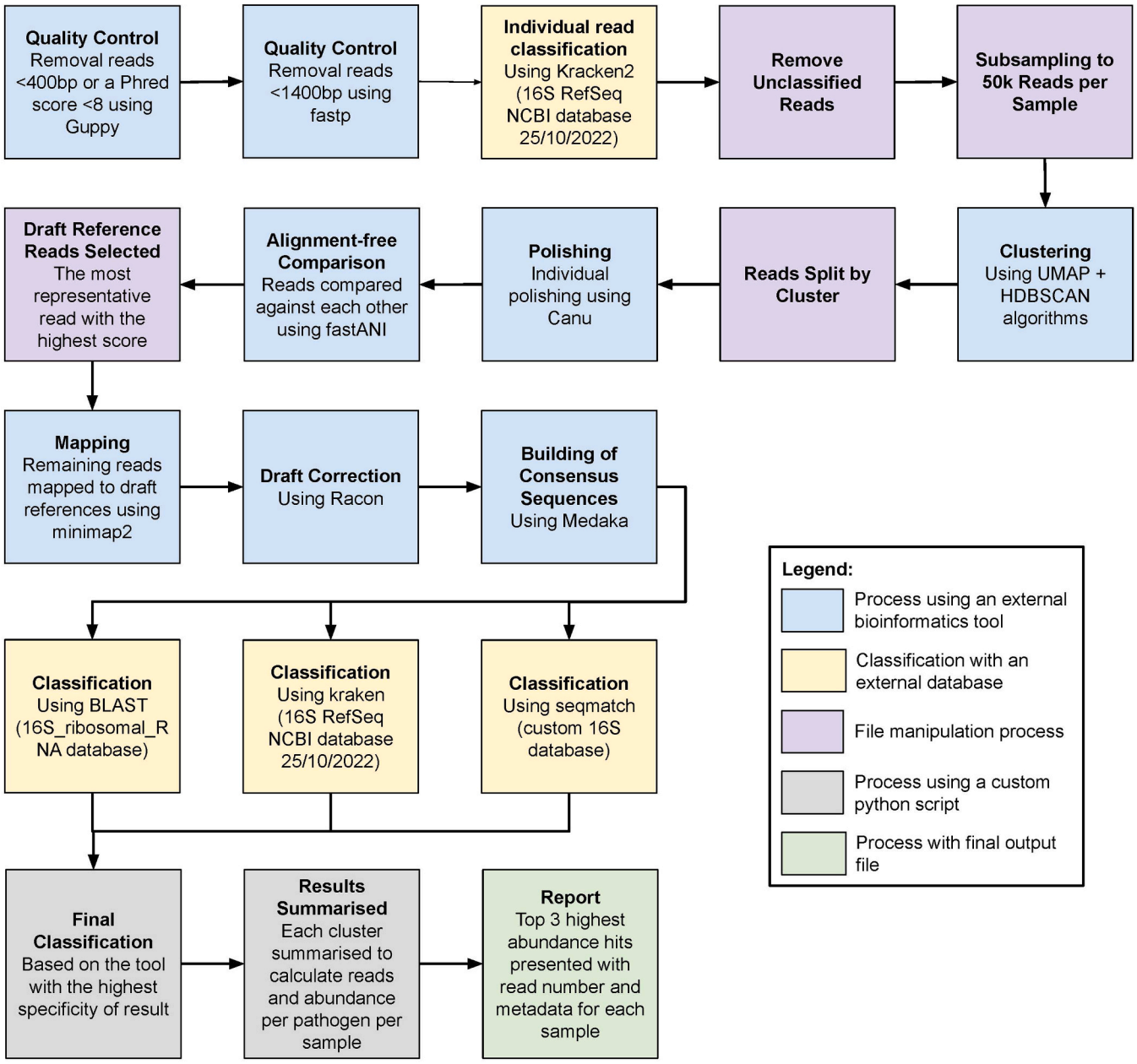


Fig. 2. Schematic representation of the NanoPATH analysis pipeline. Processing stages are color-coded to differentiate external bioinformatics software (blue), database classification steps (yellow), file manipulation processes (purple), custom Python scripting (grey), and final reporting (green). NanoPATH analysis pipeline development and parameters details are provided in web-only Supplementary Material S5.

- Quality Control & Pre-filtering: Reads shorter than 1400 bp are removed using fastp [13]. An optional Kraken2 [14] step removes non-16S reads.
- Primary Classification: Remaining reads are classified against the NCBI 16S RefSeq database (25/10/2022) [15] using Kraken2; unclassified reads are discarded.
- Downsampling & Clustering: Each sample is downsampled to 50,000 reads, converted to 5-mer frequency profiles, and clustered using UMAP [16] and HDBSCAN [17], then error corrected by Canu [18].
- Consensus Generation: The highest-scoring representative read is selected as a draft reference using fastANI [19]. Reads are mapped to this reference using minimap2 [20], corrected with Racon [21], and polished with Medaka [22] to generate a high-accuracy consensus sequence per cluster.
- Multi-Classifier Taxonomic Assignment: Consensus sequences are evaluated in parallel using Kraken2, seqmatch [23], and BLAST [24] against the NCBI 16S RefSeq (25/10/2022), Ribosomal Database Project 16S [25], and NCBI 16S_ribosomal_RNA databases [26], respectively. The result providing the highest taxonomic specificity is automatically selected for reporting.

The pipeline is illustrated in Fig. 2 with additional pipeline details including default clustering parameters provided in web-only Supplementary Material S5.

NanoPATH integrates run metadata and tracks patient identifiers via a user-input sample sheet. Run-level quality control metrics (FastQC [27] and MultiQC [28]) and metadata are automatically compiled into standardised, easily interpretable HTML reports highlighting the top three detected species including their relative abundances and absolute read counts. The pipeline also supports ITS2 fungal amplicon sequencing. NanoPATH is implementable within EPI2ME Labs [29] and executed either via the command line or a user-friendly graphical interface, with a total execution time of approximately 30 min per sequencing run on the ONT GridION.

Clinical review and interpretation of results

Three consultant medical microbiologists (RT, CL, GM), blinded to outsourced data, independently assessed the clinical relevance of the in-house results. When making their assessments, they were blinded to the study hypotheses and based their responses on the organism's known pathogenicity, the sample type and the clinical details provided on the request form from the clinical team. Disagreements were resolved by consensus between two clinical investigators (TdS, DP). Positive outsourced results matching clinically relevant in-house results were automatically deemed clinically relevant. Discordant outsourced positive results were reviewed by a clinician (JG). For these samples, clinical significance was classified according to how the result had been interpreted and managed in routine clinical practice, specifically whether the authorising microbiologist and treating clinical team regarded it as a clinically significant result or as a contaminant.

Data analysis

Concordance of diagnostic positivity between the in-house assay and the outsourced laboratory was assessed using a two-tailed McNemar's test, with a significance threshold set at alpha=0.05. Post-pipeline data analysis, statistical testing, and generation of graphical figures were performed in R version 4.3.1.

Results

Assay validation

The manufacturer's (10 ng input, 25 PCR cycles) and low input (1 ng input, 35 PCR cycles) conditions yielded similar results, detecting the 3 most abundant bacteria in the Zymo standard while missing organisms at or below 0.0089% relative abundance. Measured abundances remained within 2% of expected values (see web-only Supplementary Material S6), with negligible contaminant reads in negative controls (see web-only Supplementary Material S7).

The Zymo standard serial dilution assay detected bacteria down to a total input of 0.00001 ng (10 fg). Total reads remained stable (>32,000) down to 0.001 ng. A distinct threshold was observed at 0.0001 ng (100 fg), where read depth declined sharply and the relative proportion of contaminants rose, reaching 26.7% at 0.00001 ng (10 fg) (see web-only Supplementary Material S8). Detection depended heavily on total input and target relative abundance. The highest-abundance target (*L. monocytogenes*, 89.1% relative abundance), achieved robust detection (4/4 replicates) across the entire tested range down to 0.00001 ng (10 fg) total input (8.91 fg species-specific input). For low-abundance targets within complex microbial environments, *P. aeruginosa* (8.9%) and *B. subtilis* (0.89%) maintained robust detection (4/4 replicates) down to 0.001 ng total input (89.0 fg and 8.9 fg species-specific input respectively). Detection became inconsistent at 0.089% relative abundance and failed entirely at 0.0089%, or when species-specific input fell below 0.89 fg (Fig. 3).

Thus, the low-abundance detection threshold in polymicrobial samples is a total input of 0.001 ng and a target relative abundance of 0.89% (8.9 fg). This 0.001 ng input equates to a minimum post-PCR sample concentration of 0.47 ng/μl and pool input of 0.8 ng within a 15.33 ng total (5.2%). While dominant targets (≥89.1%) remain detectable below 0.001 ng/8.9 fg input, these extreme dilutions result in post-PCR concentrations below the Qubit fluorometer's limit of detection (0.01 ng/μl). This prevents accurate pooling and library preparation while causing a sharp rise in contaminant reads; thus, a true formal LOD could not be tested.

The twelve 16S positive clinical samples used to evaluate performance across different ONT sequencing kits and flow cell chemistries showed excellent result concordance with no significant organism compositional shift between conditions (see web-only Supplementary Material S1, S2).

Clinical sample validation

Of 116 clinical samples collected, 30 were excluded (3 duplicates, 11 in-house extraction failures, 16 not sent for outsourced 16S testing). The remaining 86 parallel-tested samples consisted of 30 cerebrospinal fluid, 23 pleural fluid, 15 joint aspirate, 8 pus, 5 biopsy, and 5 other fluid samples. Using NanoPATH, the in-house method detected a 16S gene and identified an organism in 76/86 samples, compared to 13/86 by the outsourced laboratory (Table 1 A).

During clinical review, inter-observer agreement was high, with disagreements occurring within only 11.7% of species, which then required further review as described above. Following clinical review, clinically relevant organisms were identified in 32/86 (37.2%) samples using the in-house workflow, compared to 12/86 (14.0%) by the outsourced laboratory. This represents a statistically significant 2.7-fold increase in clinical diagnostic yield, with an absolute increase of 23.3% (95% CI: 14.3–32.3%; McNemar's test, $p<0.0001$; Table 1B). The in-house assay demonstrated 100% positive percent agreement (12/12; 95% CI: 75.7–100.0%), detecting all clinically relevant pathogens identified by the outsourced laboratory. Among these 12 cases, 6 were fully concordant to the species level and 3 to the genus level. The remaining 3 were semi-concordant, where the in-house assay detected additional, clinically relevant pathogens missed by the outsourced laboratory (Fig. 4). The full distribution of organisms identified is shown in web-only supplemental material S9.

In fifty-four samples, both labs detected no clinically relevant organisms (excluded from Fig. 4). The in-house method identified eight clinically relevant polymicrobial infections. The outsourced laboratory detected a single organism in three of these cases and did not detect

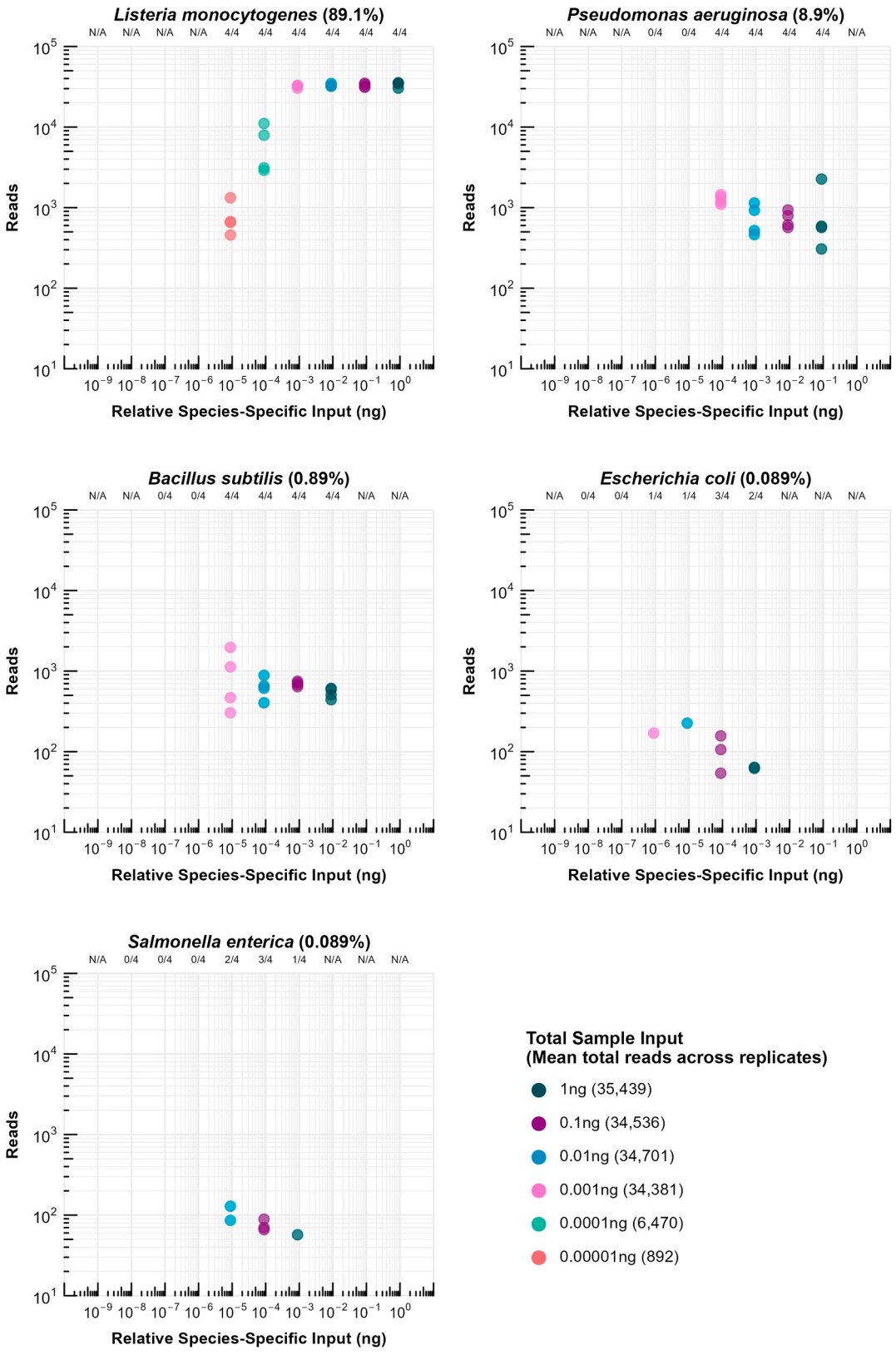


Fig. 3. Sequencing reads by relative species-specific input. Observed species specific reads of ZymoBIOMICS Community DNA Standard II (Log Distribution) (D6311) at total sample inputs 1–0.00001 ng showing species specific relative input and reads. Panels are faceted by species with theoretical community percentages. Points are coloured by total sample input mass, with the mean total read throughput for each dilution level indicated in the legend. Fractional annotations (x/4) represent the frequency of successful detection across four technical replicates. *Lactobacillus fermentum* (0.0089% component), *Enterococcus faecalis* (0.00089% component), and *Staphylococcus aureus* (0.000089% component) were not detected at any input level and thus not shown in the plot. Samples sequenced using R10.4.1 flow cells and analysis performed using NanoPATH v1.0dev.

Table 1
Comparison of in-house assay and outsourced laboratory performance for 16S gene detection and clinically relevant isolate detection (n=86). (A) Concordance of 16S gene detection and organism identification between methods. (B) Concordance of clinically relevant isolate identification via 16S sequencing between methods. Statistical significance was determined using a two-tailed McNemar's test (*p*-value of <0.0001 for both panels).

A		Outsourced laboratory	
	n=86 $p<0.0001$	**16S gene detected**	**16S gene NOT detected**
In-house assay	**16S gene detected**	13	63
	16S gene NOT detected	0	10

B		Outsourced laboratory	
	n=86 $p<0.0001$	**Clinically relevant isolate**	**Clinically negative**
In-house assay	**Clinically relevant isolate**	12	20
	Clinically negative	0	54

anything in the remaining five (Table 2). Additionally, six samples that yielded negative results through outsourced 16S testing were subsequently identified as positive via species-specific qPCR. Compared to these qPCR results, the in-house assay showed full concordance in two samples, partial concordance in one (detecting three species where qPCR detected one genus), and discrepant results in three (outsourced qPCR positive; in-house 16S detected but clinically negative (n=2) and 16S not detected (n=1)) (see web-only Supplementary Fig. S10).

Assay implementation

Based on study timestamps, the median pre-analytical duration from sample collection to a 16S test request was 63.2 h (IQR: 42.6–94.9, n=86). Once requested, the outsourced 16S testing pathway demonstrated a median turnaround time (TAT) of 78.4 h (IQR: 51.4–101.0, n=86) to reporting, which rose to 121.7 h (IQR: 94.7–131.5, n=12) for positive samples (Fig. 5 A). To provide a meaningful comparison of the assay's potential, a routine clinical workflow was modelled (see web-only Supplementary Material S11) based on a proposed twice-weekly schedule (Fig. 5B), yielding a projected total median TAT of 91.5 h (IQR: 67.1–99.8, n=86) from request to reporting. Notably, while the outsourced laboratory maintained a shorter overall TAT, it was considerably slower to report positive results. Furthermore, 23.3% (20/86) of samples in this cohort were reported as false negatives missing clinically relevant isolates detected by the in-house assay (Table 1B).

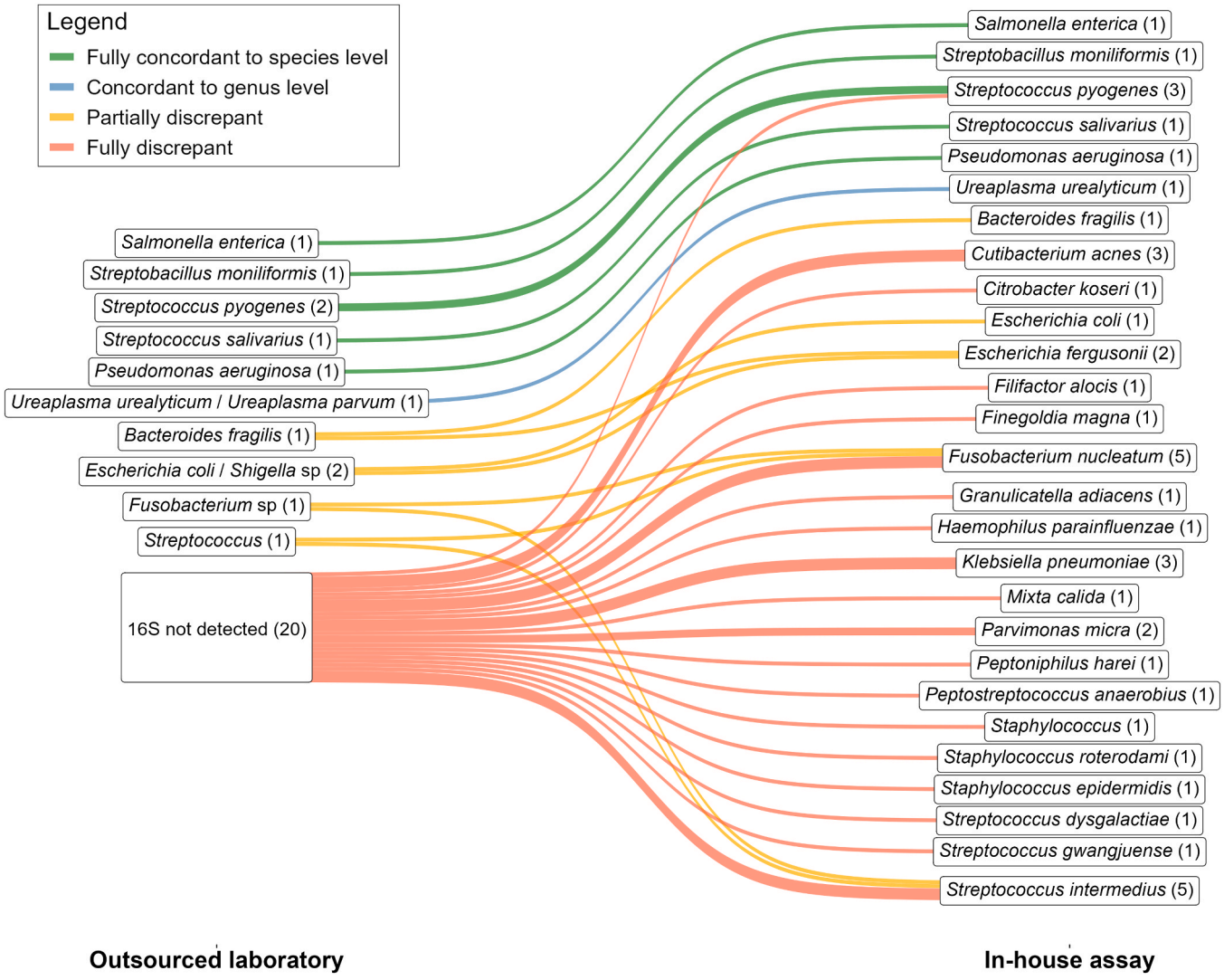


Fig. 4. Concordance of clinically relevant species identification between the in-house method and the outsourced laboratory. Colours represent the level of concordance between the in-house assay and outsourced laboratory results: Fully concordant to species level (green), Concordant to genus level (blue), Partially discrepant (orange), and Fully discrepant (red). Fifty-four fully concordant clinically negative samples are excluded for clarity.

Table 2
Polymicrobial samples detected by the in-house 16S assay. Summary of eight samples containing two or more clinically relevant organisms identified by the in-house assay alongside outsourced laboratory results. In-house species relative abundance is shown in brackets.

Sample	Sample type	Outsourced laboratory result	In-house assay result species 1	In-house assay result species 2	In-house assay result species 3
1	Liver abscess pus	*Bacteroides fragilis*	*Escherichia fergusonii* (76%)	*Bacteroides fragilis* (24%)	-
2	Cerebral abscess pus	*Fusobacterium sp.*	*Fusobacterium nucleatum* (70%)	*Streptococcus intermedius* (30%)	-
3	Pleural fluid	*Streptococcus*	*Streptococcus intermedius* (94%)	*Fusobacterium nucleatum* (6%)	-
4	Cerebrospinal fluid	16S not detected	*Cutibacterium acnes* (59%)	*Streptococcus epidermidis* (26%)	-
5	Pleural fluid	16S not detected	*Streptococcus intermedius* (54%)	*Fusobacterium nucleatum* (38%)	-
6	Pleural fluid	16S not detected	*Fusobacterium nucleatum* (45%)	*Parvimonas micra* (19%)	*Filifactor alocis* (17%)
7	Extradural collection	16S not detected	*Staphylococcus roterodami* (43%)	*Peptostreptococcus anaerobius* (24%)	*Peptoniphilus harei* (12%)
8	Pleural debris	16S not detected	*Streptococcus intermedius* (67%)	*Fusobacterium nucleatum* (31%)	*Parvimonas micra* (2%)

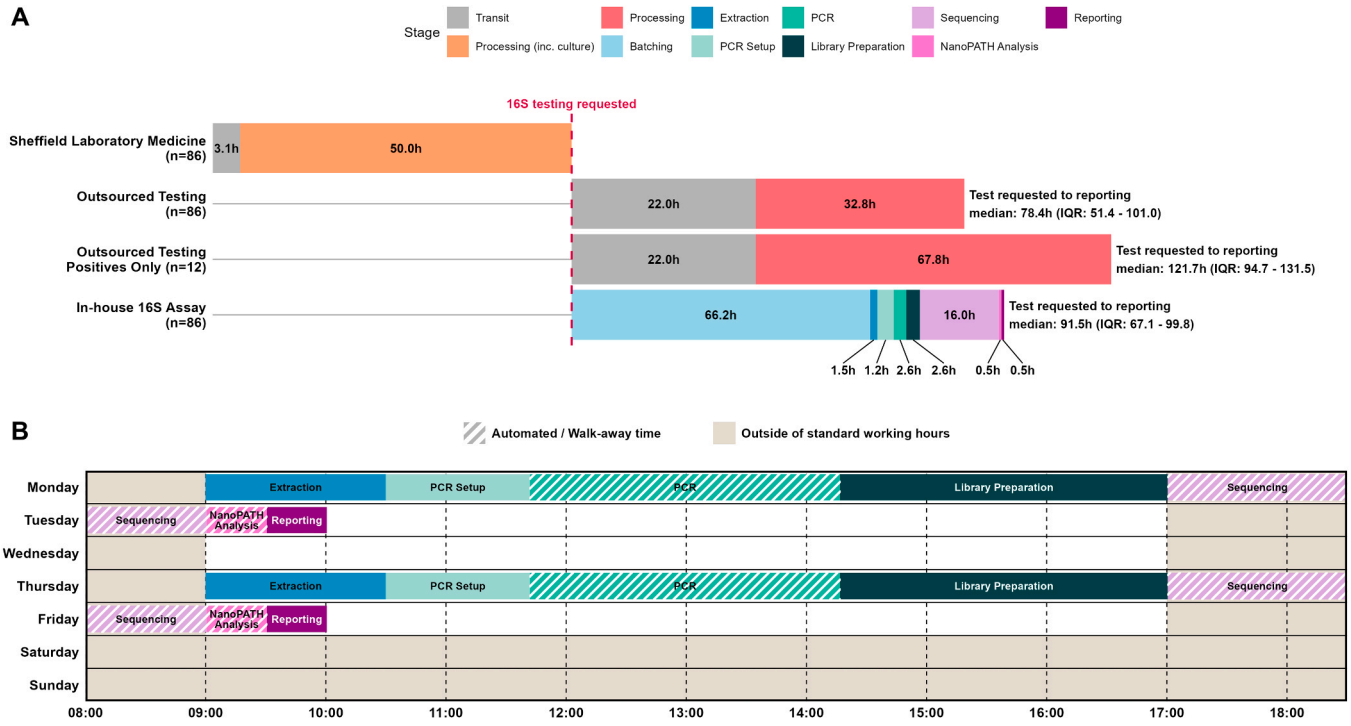


Fig. 5. Sample processing timelines and proposed in-house operational schedule. (A) Sample Processing Timeline: Comparison of workflows and turnaround times (TAT) across different testing modalities. The vertical pink dashed line indicates the point at which 16S testing is requested following initial sample transit and processing (including culture) at Sheffield Laboratory Medicine. Timelines compare outsourced testing (all samples vs. positives only) against a modelled in-house 16S assay workflow broken down by operational stage (batching, extraction, PCR setup, PCR, library preparation, sequencing, and NanoPATH analysis/reporting). Text callouts indicate the "test requested to reporting" median duration and interquartile range (IQR) for each pathway. Note: Medians were calculated for each category individually and are non-additive. Turnaround time (TAT) modelling calculations and details are provided in web-only Supplementary Material S11. **(B)** In-house 16S Assay Proposed Weekly Operational Schedule: A theoretical bi-weekly batching workflow mapping out laboratory tasks across a standard work week. Coloured blocks define specific assay stages, while hatched patterns denote automated/walk-away time (e.g., PCR and sequencing). Shaded tan regions indicate periods outside of standard working hours (before 09:00 and after 17:00).

Cost analysis revealed that the high fixed cost of sequencing coupled with a low incremental cost of £6.88 per additional sample results in cost per sample decreasing non-linearly with increased multiplexing throughput. The in-house workflow achieves cost-efficiency relative to outsourced testing (£69.96 per sample) at 9 samples per run (£69.72 per sample). Scaling up to the maximum kit capacity of 21 samples further reduced costs to £33.87 per sample (Fig. 6 and see web-only Supplementary File S12). While multiplexing capacity is currently constrained by available barcodes and total flow cell output, scaling could be achieved by custom barcode expansion and transitioning to higher-capacity ONT PromethION flow cells. These models demonstrate the workflow is operationally viable and financially advantageous within a routine NHS diagnostic laboratory setting.

Discussion

This study demonstrates that a full-length 16S sequencing workflow can be both analytically effective and practically implementable within a routine diagnostic laboratory. While sequencing-based diagnostics can outperform culture, particularly for fastidious or low-biomass infections and in resolving complex infections, adoption has been limited by reliance on bioinformatics expertise and a lack of operational validation. By combining an automated analysis pipeline executable by non-specialist staff with a clinically validated ONT-based assay, we address both barriers.

The in-house workflow delivered a significant 2.7-fold increase in diagnostic yield over the current standard-of-care, identifying clinically relevant pathogens in 20 samples reported as negative by the outsourced laboratory. It improved species-level resolution and resolved

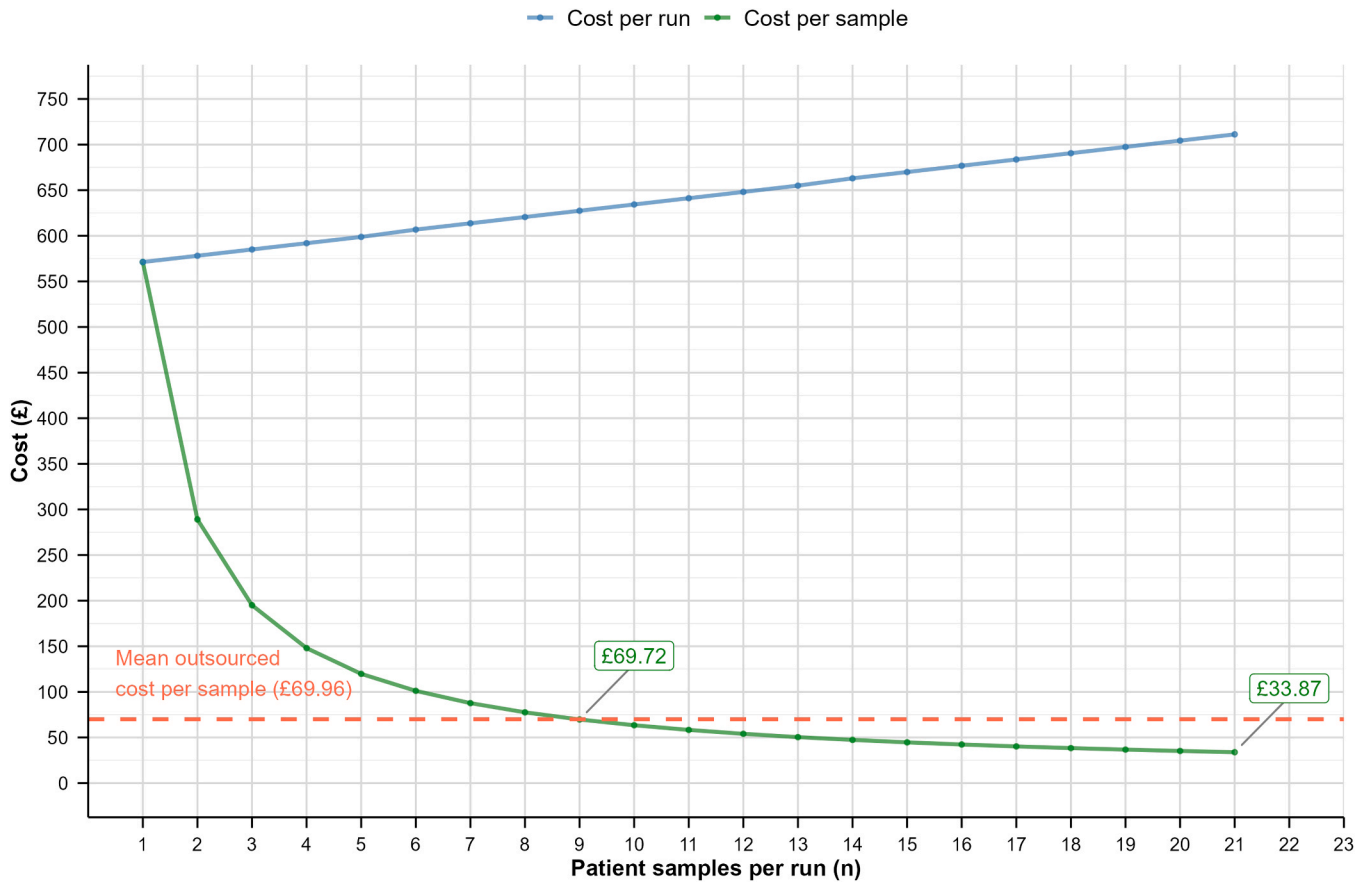


Fig. 6. Sequencing economy of scale for the in-house 16S assay. The overall cost per run (blue) increases linearly with sample scale. The cost per sample (green) decreases as the number of patient samples per run increases, crossing the mean outsourced equivalence threshold (£69.96; dashed red line) between 8 and 9 samples per run. Detailed cost breakdowns and underlying calculations are provided in web-only Supplementary File S12.

polymicrobial infections that are frequently missed by traditional methods, reinforcing its diagnostic value.

While broad metagenomic sequencing offers an entirely agnostic approach, its high cost and computational requirements restrict its use in cost-conscious settings. Conversely, targeted qPCR, inherently limited to predefined targets, is unsuitable as a primary diagnostic approach for unknown infections. Full-length 16S provides a pragmatic balance of diagnostic breadth, cost, and operational simplicity.

As with any highly sensitive amplification assay, background DNA contamination ("kitome") remains a potential limitation, requiring stringent control tracking and appropriate bioinformatic filtering. Although this study was conducted at a single centre, the use of widely available ONT platforms and an open-access pipeline supports broad reproducibility. The NanoPATH analysis pipeline removes the need for specialist bioinformatics support, enabling routine laboratory staff to process raw sequencing files into clear, standardised HTML reports. While this structured reporting format is designed to streamline clinical review, result interpretation still requires dedicated clinician time. This represents an indirect operational cost that was not quantified in this study but could become substantial as sequencing-based assays scale across routine diagnostics. Additionally, during this clinical evaluation, we were unable to blind the clinical microbiologists to whether a report came from the commercial provider or the in-house assay. The in-house assay provided additional details and metrics not offered by the outsourced result, which would have reduced the report's utility if removed. While this prevented true blinding to the report origin, the assessing microbiologists were blinded to the study's hypotheses and had not been involved in any prior development stages.

Operational modelling showed that a twice-weekly run schedule provides clinically actionable timelines, outperforming outsourced reporting times for positive samples while returning local testing control to the hospital. Economically, the workflow becomes cost-advantageous at 9 samples per run, delivering greater cost reductions at higher multiplex capacities. While molecular workflows face scrutiny in a cost-conscious healthcare environment, particularly as laboratories vary in their approach to culture-negative samples, the ability to identify a pathogen allows for tailored antibiotic selection and optimised treatment duration. In the context of rising antimicrobial resistance, maximising diagnostic yield to optimise therapy remains paramount; the judicious implementation of such molecular methods in routine diagnostics is therefore essential. This validated model offers a sustainable roadmap for integrating sequencing into standard clinical workflows, with an established infrastructure capable of expanding to fungal ITS2 amplicon testing [30].

Conclusion

An ONT-based full-length 16S sequencing workflow can be successfully translated into routine diagnostic practice. By successfully overcoming analytical and operational barriers, this study combines improved pathogen detection with an easily operable analysis pipeline, NanoPATH. The resulting feasible implementation model provides a practical pathway for integrating sequencing-based diagnostics into standard clinical diagnostic workflows and improving patient care.

CRediT authorship contribution statement

TJS - Methodology, Validation, Formal analysis, Investigation, Resources, Data curation, Writing - Original draft, Visualization, Project Administration, MBD - Methodology, Software, Validation, Resources,

Data curation, Writing – Review & Editing, PZ - Methodology, Investigation, Validation, Resources, Data curation, Writing – Review & Editing, Project Administration, JB - Methodology, Investigation, Resources, Data curation, Writing – Review & Editing, AS - Methodology, Investigation, Resources, Data curation, Writing – Review & Editing, JG - Methodology, Formal analysis, Writing – Review & Editing, SH - Methodology, Software, Validation, Resources, Data curation, Writing – Review & Editing, KJ - Methodology, Validation, Investigation, Data curation, Writing – Review & Editing, CL - Investigation, Formal analysis, Writing – Review & Editing, RT - Investigation, Formal analysis, Writing – Review & Editing, GM - Investigation, Formal analysis, Writing – Review & Editing, HP - Conceptualisation, Investigation, Writing – Review & Editing, DP - Conceptualisation, Formal analysis, Resources, Writing – Review & Editing, Project Administration, TIdS - Conceptualisation, Formal analysis, Resources, Supervision, Writing – Review & Editing, Project Administration, Funding acquisition.

Ethics

The authors have not obtained informed consent from participants or their legal representatives. The samples used in this study were unlinked anonymised residual clinical samples submitted for routine diagnostic testing within an NHS laboratory. NHS ethical approval was granted on the basis that individual consent was not required.

This study was performed in compliance with relevant laws, regulatory frameworks, and guidelines where the research took place. This study was approved by the London - Central Research Ethics Committee. (Approval No. 21/PR/0675)

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the author(s) used Gemini by Google to improve the readability and grammar of the text. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the published article.

Funding

This study was funded by the National Institute for Health and Care Research (NIHR) Sheffield Biomedical Research Centre (BRC) (Grant Reference Number NIHR203321). The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank the virology staff at Sheffield Laboratory Medicine for sample processing and extraction.

Data availability

The datasets used are available from the corresponding author on reasonable request.

NanoPATH is openly available on GitHub at: https://github.com/Clinical-Infection-Research-UoSheffield/nf-core-nanopatht

The R code required to reproduce analyses and figures in this manuscript is openly available on GitHub at: https://github.com/Clinical-Infection-Research-UoSheffield/16S-Analysis

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.jinf.2026.106876.

References

[1] Harris AM, Bramley AM, Jain S, Arnold SR, Ampofo K, Self WH, et al. Influence of antibiotics on the detection of bacteria by culture-based and culture-independent diagnostic tests in patients hospitalized with community-acquired pneumonia. January 1 Open Forum Infect Dis 2017;4(1). https://doi.org/10.1093/ofid/ofx014.

[2] Lampejo T, Ciesielczuk H, Lambourne J. Clinical utility of 16S rRNA PCR in pleural infection. January 1 J Med Microbiol 2021;70(5). https://doi.org/10.1099/jmm.0.001366.

[3] Morel AS, Dubourg G, Prudent E, Edouard S, Gouriet F, Casalta JP, et al. Complementarity between targeted real-time specific PCR and conventional broad-range 16S rDNA PCR in the syndrome-driven diagnosis of infectious diseases. March 1 Eur J Clin Microbiol Infect Dis 2015;34(3):561–70. https://doi.org/10.1007/s10096-014-2263-z.

[4] Aggarwal D, Kanitkar T, Narouz M, Azadian BS, Moore LSP, Mughal N. Clinical utility and cost-effectiveness of bacterial 16S rRNA and targeted PCR based diagnostic testing in a UK microbiology laboratory network. May 14 Sci Rep 2020;10(1). https://doi.org/10.1038/s41598-020-64739-1.

[5] Fida M, Khalil S, Abu Saleh O, Challener DW, Sohail MR, Yang JN, et al. Diagnostic value of 16S ribosomal RNA gene polymerase chain reaction/sanger sequencing in clinical practice. September 15 Clin Infect Dis 2021;73(6):961–8. https://doi.org/10.1093/cid/ciab167.

[6] Aftab H, Schouw CH, Dargis R, Vejrum LK, Johansen RL, Møller JT, et al. Next generation sequencing improves diagnostic 16S rRNA amplicon-based microbiota analyses of clinical samples compared to sanger sequencing. APMIS Acta Pathol Microbiol Et Immunol Scand September 2025;133(9). https://doi.org/10.1111/apm.70067.

[7] Johnson JS, Spakowicz DJ, Hong BY, Petersen LM, Demkowicz P, Chen L, et al. Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. November 6 Nat Commun 2019;10(1). https://doi.org/10.1038/s41467-019-13036-1.

[8] Micropathology Ltd. Laboratory User Handbook [Internet]. Coventry: Micropathology Ltd,; 2026. Mar 27]. Available from: https://micropathology.com.

[9] Oxford Nanopore Technologies. Guppy [software]. Version 6.4.2. Oxford, UK: Oxford Nanopore Technologies,; 2023 (Available from: nanoporetech.com).

[10] Dabrowska M. nf-Core-nanopath [Internet]. Version 1.0.0. Sheffield: Clinical Infection Research UoSheffield,; 2024 [updated 2024 Jul 3; cited 2026 Mar 27]. Available from: https://github.com/Clinical-Infection-Research-UoSheffield/nf-core-nanopath.

[11] EPI2ME Labs. wf-metagenomics [Internet]. Version 2.14.2. London: Oxford Nanopore Technologies,; 2026 [updated 2026 Jan 22; cited 2026 Mar 27]. Available from: https://github.com/epi2me-labs/wf-metagenomics.

[12] Rodríguez-Pérez H, Ciuffreda L, Flores C. NanoCLUST: a species-level analysis of 16S rRNA nanopore sequencing data. July 12 Bioinforma (Oxf Engl) 2021;37(11):1600–1. https://doi.org/10.1093/bioinformatics/btaa900.

[13] Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. September 1 Bioinformatics 2018;34(17):i884–90. https://doi.org/10.1093/bioinformatics/bty560.

[14] Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. November 28 Genome Biol 2019;20(1). https://doi.org/10.1186/s13059-019-1891-0.

[15] O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. November 8 Nucleic Acids Res 2015;44(D1):D733–45. https://doi.org/10.1093/nar/gkv1189.

[16] McInnes L, Healy J, Saul N, Großberger L. UMAP: uniform manifold approximation and projection. September 2 J Open Source Softw 2018;3(29):861. https://doi.org/10.21105/joss.00861.

[17] McInnes L, Healy J, Astels S. hdbscan: Hierarchical density based clustering. March 21 J Open Source Softw 2017;2(11):205. https://doi.org/10.21105/joss.00205.

[18] Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: scalable and accurate long-read assembly via adaptive k -mer weighting and repeat separation. Genome Res May 2017;27(5):722–36. https://doi.org/10.1101/gr.215087.116.

[19] Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. November 30 Nat Commun 2018;9(1). https://doi.org/10.1038/s41467-018-07641-9.

[20] Li H. Minimap2: pairwise alignment for nucleotide sequences. September 15 Bioinformatics 2018;34(18):3094–100. https://doi.org/10.1093/bioinformatics/bty191.

[21] Vaser R, Sović I, Nagarajan N, Šikić M. Fast and accurate de novo genome assembly from long uncorrected reads. May 1 Genome Res 2017;27(5):737–46. https://doi.org/10.1101/gr.214270.116.

[22] Oxford Nanopore Technologies. medaka: Sequence correction provided by ONT Research [Internet]. Version 1.4.4. 2021 [cited 2026 Apr 16]. Available from: https://github.com/nanoporetech/medaka.

[23] RDP Staff. SequenceMatch [Internet]. East Lansing (MI): Center for Microbial Ecology, Michigan State University,; 2023. Sep 20; cited 2024 May 22]. Available from: https://github.com/rdpstaff/SequenceMatch.
[24] Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST +: architecture and applications. December 15 BMC Bioinforma 2009;10(1). https://doi.org/10.1186/1471-2105-10-421.
[25] Cole JR, Chai B, Marsh TL, Farris RJ, Wang Q, Kulam SA, et al. The ribosomal database project (RDP-II): previewing a new autoaligner that allows regular updates and the new prokaryotic taxonomy. January 1 Nucleic Acids Res 2003;31 (1):442–3. https://doi.org/10.1093/nar/gkg039.
[26] Pruitt KD, Tatusova T, Brown GR, Maglott DR. NCBI reference sequences (RefSeq): current status, new features and genome annotation policy. Nucleic Acids Res January 2012;40(D1):D130–5. https://doi.org/10.1093/nar/gkr1079.
[27] Andrews S. FastQC: A Quality Control Tool for High Throughput Sequence Data [Internet]. Cambridge, UK: Babraham Institute,; 2010 [cited 2026 Apr 16]. Available from: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.
[28] Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. October 1 BIOINFORMATICS 2016;32(19):3047–8. https://doi.org/10.1093/bioinformatics/btw354.
[29] Oxford Nanopore Technologies. EPI2ME [Internet]. Oxford (UK): Oxford Nanopore Technologies,; 2025. Mar 27]. Available from: https://epi2me.nanoporetech.com/.
[30] Ohta A, Nishi K, Hirota K, Matsuo Y. Using nanopore sequencing to identify fungi from clinical samples with high phylogenetic resolution. June 16 Sci Rep 2023;13 (1). https://doi.org/10.1038/s41598-023-37016-0.