

RESEARCH

Open Access



Limited value of Nanopore adaptive sampling in a long-read metagenomic profiling workflow of clinical sputum samples

Weizhen Xu^{1†}, Janetta Top^{1*†}, Mattheus C. Viveen¹, Andrii Slyzkyi², Noud Hermans^{2,3}, Sarah van Erp⁴, Dafna Eiloz⁴, Richard Anthony³, Kristin Kremer² and Anita C. Schürch¹

Abstract

Background Oxford Nanopore adaptive sampling (NAS) is a method by which the long-read sequencing flowcell accepts or rejects DNA molecules that are actively being sequenced based on their initial ~ 500 bp sequences, selectively increasing target data output. NAS promises up to 5–10× enrichment of target sequencing yield without additional sample preparation, but this optimal performance is dependent on ideal sample parameters which may be difficult to achieve under many real-world use-cases. We evaluated the use of NAS for profiling clinical sputum metagenomes.

Methods We sequenced DNA extracted from clinical sputa and spike-in controls of a mock community of bacterial respiratory pathogens, using the current R10.4.1 MinION flowcell chemistry.

Results We achieved at best 3.1× enrichment of bacterial sequence output with NAS due to the shorter read lengths (~ 2.5 kb) from the PCR amplification necessary to compensate for low DNA extraction yields. More critically, we encountered rapid pore loss during our runs that reduced total sequencing yield by an estimated 80%. We were unable to mitigate the pore loss despite extensive attempts to reduce contaminant carry-over, and we could not determine its cause but ruled out NAS and pore underloading as contributing factors.

Conclusions We conclude that the utility of NAS is often limited by the characteristics of the metagenomic sample studied, and that the factors contributing to pore loss need to be resolved before ONT sequencing can be reliably applied to long-read metagenomics.

Keywords Long-read sequencing, Metagenomics, Pathogen surveillance

[†]Weizhen Xu, and Janetta Top contributed equally to this work.

*Correspondence:

Janetta Top

j.top@umcutrecht.nl

¹Department of Medical Microbiology, University Medical Center Utrecht, Utrecht, the Netherlands

²Diagnostics Team, Division of TB Elimination & Health Systems Innovations, KNCV Tuberculosis Foundation, The Hague, the Netherlands

³Department Bacterial Surveillance and Response, Center for Infectious Disease Control, Diagnostics and Laboratory Surveillance, National Institute for Public Health and the Environment (RIVM), Bilthoven, the Netherlands

⁴Mozand B.V, Eindhoven, the Netherlands



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Introduction

Culture-based isolate sequencing followed by phenotypic testing of drug susceptibility is generally regarded to be the gold-standard practice for the epidemiological surveillance of infectious diseases and the associated spread of antimicrobial resistance and virulence genes. However, this process is slow and complex, especially for difficult-to-culture or slow-growing pathogens such as *Mycobacterium tuberculosis*. Furthermore, it relies on well-staffed centralised laboratory infrastructures that are more difficult to establish in lower-middle income country settings where the burden of infectious disease is the greatest [1–4]. Direct culture-free metagenomic sequencing of clinical samples would avoid the aforementioned difficulties with microbial culture and offers a comprehensive genomic overview of multiple species present at the site of infection [1, 3–5], but is complicated by high levels of off-target host or commensal DNA (up to 99.99% in respiratory samples) that often hinder the detection of low-abundance species and genes of interest [1, 3, 4, 6, 7]. Chemical host-depletion methods can improve metagenomic sensitivity and performance [3, 4, 6] but impose additional costs and workflow complexity that might be challenging in low-resource settings.

Nanopore adaptive sampling (NAS) is a methodology within Oxford Nanopore Technologies (ONT) sequencing that enriches for target sequences without additional wet lab preparation. NAS utilises the real-time sequencing capabilities of ONT to read the first 400–500 bp of a DNA molecule, comparing it against a database of target or non-target genomic sequences and rejecting off-target molecules accordingly, preferentially sequencing target molecules at full length [8, 9]. NAS enrichment performance is dependent on multiple sample parameters, in particular longer (> 10 kb) read length [8]. While the ONT NAS best-practices guide [10] promises up to 5–10× target enrichment, occasionally achievable when higher-quality DNA samples can be obtained [8, 11, 12], enrichment performance was lower (1.8–4x) in many other NAS metagenomic studies [9, 13–20], likely reflective of real-world samples deviating from optimal parameters. These studies were done using the earlier generation R9 ONT flowcell chemistry in which NAS was first developed and optimised [9]. We thus evaluated the use of the current generation of R10.4.1 flowcells in a multiplexed NAS-based workflow for the metagenomic and antimicrobial resistance (AMR) gene profiling of actual clinical sputum samples in a low-resource laboratory setting. While easily collectible in a minimally invasive manner, the higher host DNA composition and low extraction yields of sputum present a greater challenge for metagenomic profiling.

We encountered rapid pore loss in the ONT flowcells during this evaluation, severely limiting metagenomic

performance. Pore degradation either from electric potential [21] or pore blockage [22] is a major factor affecting ONT sequencing yield and inter-run variability [8], and while accelerated pore loss is a commonly-cited cause of run failure on the ONT community forums (closed access to registered ONT customers, 2023–2025), its specific causes and solutions are not well-documented in scientific literature. We thus present our efforts to troubleshoot pore loss based on ONTs' recommendations of improving pore occupancy and reducing sample contamination.

Materials and methods

Sputum collection

The sputum samples used in this study included residual and completely anonymized material from routine diagnostic protocols at University Medical Centre Utrecht (UMCU). Ethical approval for the use of the sputum samples for this study was obtained from the Biobank Research Ethics Committee (TCBio) from the University Medical Centre Utrecht with approval number: TCBio 25–063. The need for informed consent was waived by the Biobank Research Ethics Committee (TCBio) from the University Medical Centre Utrecht.

Mock respiratory pathogen community

A mock community of respiratory pathogens was prepared from clinical isolates of *Pseudomonas aeruginosa* and *Klebsiella pneumoniae* from our hospital and type cultures of *Staphylococcus aureus* USA300 and *Streptococcus pneumoniae* ATCC49619. Glycerol stocks of these isolates were initially plated and incubated overnight on tryptone-soy-agar (TSA) (BD, Alphen aan den Rijn, The Netherlands) at 37 °C, and in the case of *S. pneumoniae* at 37 °C in a 5% CO₂ atmosphere incubator. Colonies of the first three pathogens were inoculated into 10 ml Luria Broth (LB) medium (Sigma Aldrich, Zwijndrecht, The Netherlands) and cultured overnight at 37 °C, while *S. pneumoniae* was subcultured on four TSA plates at 37 °C in a 5% CO₂ atmosphere overnight and resuspended in 10 ml LB medium. Bacterial load from these cultures was estimated based on OD₆₀₀ absorbance with the assumption that OD₆₀₀ of 1.0 ≈ 1 × 10¹⁰ cfu, pooled at equal culture densities and aliquoted into 50 µl volumes of 3.1 × 10⁸ cfu total bacteria each. In the pilot run, ten-fold serial dilutions (3.1 × 10⁷–10⁴ cfu) of this mock community stock were added to each sample from two clinical sputa for subsequent spike-in experiments.

High molecular weight DNA extraction from sputum

Approximately 200 µl of each sputum sample was mixed with an equal volume of SPUTOLYSIN® reagent (Merck), resuspended with a Vortex Genie-2 laboratory vortex mixer (Scientific Industries) and incubated at 37 °C

with shaking at 800 rpm for 15 min to obtain homogeneous suspensions. The samples were then centrifuged at 16,060 g for 5 min, after which the supernatant was discarded and the pellets resuspended in 300 µl sterile 1×Dulbecco's phosphate-buffered saline (DPBS, Capricorn Scientific) each.

Bacteria lysis was performed by adding an enzymatic mix of 5 µl achromopeptidase (50000U/ml, Merck), 1.5 µl lysostaphin (10 mg/ml, Sigma), 20 µl lysozyme (50 mg/ml, Merck) and 6 µl mutanolysin (2000U/ml, Sigma) to each sample, and incubating the samples at 37 °C with shaking at 800 rpm for 2 h. Further mechanical lysis and DNA extraction was performed using the DNeasy Ultra-Clean Microbial kit (QIAGEN). We adapted the mechanical lysis step for easier implementation in low-resource lab settings, adding 300 µl of PowerBead solution from the DNeasy kit to each sample and vortexing at maximum speed with the same Vortex Genie-2 vortex mixer with a vortex adaptor for 24 1–1.5 ml tubes(QIAGEN). DNA was subsequently extracted according to standard kit protocol and dsDNA concentrations were quantified on a Qubit™ 2 fluorometer using the Qubit™ dsDNA high sensitivity assay kit (ThermoFisher). High molecular weight DNA size was assessed in earlier extractions by the Genomic DNA ScreenTape assay on the TapeStation electrophoresis system (Agilent).

A negative control consisting of 300 µl sterile 1×DPBS subjected to the same bacteria lysis treatment and DNA extraction procedure was included in each run.

Initial library preparation and sequencing

For the first two ONT sequencing runs, multiplexed libraries were prepared with the Ligation Sequencing Kit V14 (SQK-LSK114) with PCR Expansion (EXP-PCA001) according to the ONT ligation sequencing of low input gDNA protocol (version LWP_9183_v114_revG_07Mar2023), using a PCR template input of up to 100 ng depending on sample yield and 10×PCR extension cycles of 8 min each, using the LongAMP Taq 2×Master Mix (NEB). Barcoded samples were quantified by Qubit™ fluorometer and combined into a equimolar pooled library, with the volume of the negative control sample matching that of the lowest concentration sample. Completely new ONT R10.4.1 flowcells (FLO-MIN114) with included light shields were used during sequencing for a duration of 72 h. For the first pilot run, 150 ng (~75fmol) of pooled library was loaded, with no subsequent need for sample reloading or flowcell washing. For the second run, 125 ng (~64fmol) of library was loaded without the nuclease wash step for the first flowcell (2A), and with a nuclease wash/reload of 125 ng library pool for the second (2B) at 24 h.

Flowcells were run on a ONT GridION×5 Mk1 platform using the Dorado 7.3.11 high-accuracy basecalling

model (400bps). The pilot run was performed using MinKNOW software v23.11.7, and the second run using MinKNOW v24.02.16. NAS was implemented through the MinKNOW interface to deplete host reads aligning to the human T2T-CHM13v2 genome reference (GCF_009914755.1). In the pilot run, channels 1–256 were designated as NAS channels and channels 257–512 as control non-NAS channels. All channels were used for NAS in the second run.

Troubleshooting run library preparation and sequencing

We revised the earlier library preparation protocol to address potential contaminant carryover, flowcell underloading and to troubleshoot possible pore blockage resulting from NAS. Multiplexed libraries were prepared using the Rapid PCR Barcoding Kit 24 V14 (SQK-RPB114.24) according to ONT protocol (version RPB_9191_v114_revB_28Jun2023), using a PCR template input of up to 4 ng DNA and 30×PCR extension cycles of 6 min each. Samples were combined into an equimolar pool as in the pilot runs.

To troubleshoot if there were any effects of NAS on pore health, we loaded 400 ng (~200fmol) each of the same library pool onto four separate ONT R10.4.1 flowcells to maximize pore occupancy, with two flowcells running NAS on half the channels similar to the pilot run, and two control flowcells not running NAS. In the second set of replicates for the control/NAS runs, we added 0.25 µl of RNase A (10 µg/ml, QIAGEN, RP14) to the library pool for 15 min at room temperature before loading the flowcell to troubleshoot possible RNA contamination degrading flowcell performance.

Flowcells were run for 72 h with a nuclease wash-reload of 200 ng of the library pool (~100fmol) at 24 h on the GridION platform using the same software versions as the second run (MinKNOW v23.11.7 and Dorado 7.3.11 high-accuracy basecalling).

Bioinformatic analysis

We initially demultiplexed basecalled.fastq reads by barcode and read quality in the MinKNOW interface. Total flowcell data output was calculated from successfully barcoded reads with a minimum Q score of 9, inclusive of short (~500 bp) NAS-rejected reads. NAS-level read demultiplexing was performed using custom in-house scripts available at <https://gitlab.com/weizhenxu/pas4amr-tools>, with read output demultiplexed based on whether the read was obtained from a control or NAS channel (channel_demultiplex.py), or if it was rejected ("reject") or accepted ("keep") by NAS (unblock_demultiplex.py) based on the MinKNOW generated adaptive sampling report.

Reads were taxonomically classified using Kraken 2 (v2.1.3) [23] using the standard-16 database available

at <https://benlangmead.github.io/aws-indexes/k2> (last accessed 9 Jan 2024). To reduce noise and filter for low abundant or spurious species, only species ≥ 100 reads were included in the sample report. In samples where reads were assigned to multiple closely-related species in the same genus, identified species with $< 10\%$ of the read count of the most abundant species in that genus were considered to be cross-mapped and removed in the final analysis.

Acquired genes and chromosomal point mutations linked to AMR were detected by ResFinder v4.5.0 [24] using the ResFinder 2.02 and PointFinder 4.1.0 databases, with identity and alignment coverage thresholds of $> 95\%$ each.

Results

Reduced bacterial enrichment by NAS of PCR-barcoded libraries

DNA extraction yields from sputa ranged from 7.4 to 46.2 ng/ μ l and were often below the recommended concentration (< 40 ng/ μ l) for a ligation-based native barcoding library preparation workflow. We thus opted for the PCR barcoding expansion kit (EXP-PCA001)-based protocol to accommodate low-yield samples and to improve equimolar loading of multiplexed samples.

The aim of our pilot run was to assess the practical detection limits of pathogen genomes in actual clinical sputa with NAS and to accordingly determine the target read yields necessary for robust metagenomic profiling. We sequenced a mock community of four equally-pooled respiratory bacterial pathogens (Gram-negative: *Klebsiella pneumoniae* and *Pseudomonas aeruginosa*, Gram-positive: *Staphylococcus aureus* and *Streptococcus pneumoniae*) spiked at ten-fold dilutions of $\approx 3.1 \times 10^7$ – 3.1×10^4 cfu into two clinical sputa. This spike series, unspiked sputa and positive and negative controls were sequenced in multiplex on a single MinION flowcell divided in half into NAS channels (channel 1–256) and control channels (channel 257–512) to evaluate the

relative enrichment benefit of NAS (Table S1). The total sequencing output (including short ~ 500 bp rejected reads by NAS) was 9.62 M reads/13.7 Gb passing quality thresholds ($Q > 9$) (Table 1, Fig. S1A). The PCR amplification resulted in shorter reads compared to the input template (previously determined to 15–30 kb by TapeStation sizing) despite long 8 min extension cycles, indicated by a ~ 2.5 kb modal peak corresponding to the non-rejected reads (Fig. S1B).

The total sequencing output (including rejected reads) of the control and NAS channels were 6.9 Gb/6.8 Gb respectively, suggesting that NAS did not affect output and pore health in NAS channels in this setup. Similar to the previous metagenomic study by Marquet et al. [14], NAS had a minor effect on the proportional count of bacterial reads sequenced (5.5% NAS reads vs 5.2% control reads, or a $1.06 \times$ increase) – NAS enrichment was a result of increased read turnover in NAS channels due to the earlier rejection of host reads, resulting in a $3.1 \times$ increase in bacterial reads and bacterial base output (Table 1), comparable to other PCR-based NAS metagenomic studies [14, 19, 25] but less than the 5–10 $\times$ optimal enrichment suggested in the ONT best practices guide and observed in studies using longer non-PCR reads [8, 11, 12].

We analysed the taxonomic composition of the “keep” and “reject” fractions of the NAS channel output to determine NAS decision accuracy (Table S2). The overall accuracy of NAS was 94.9%, with similar proportions of falsely-accepted human reads (5.2%) and falsely-rejected bacterial reads (5.1%). Given the high proportion of host reads in the input (94.8% based on the control channels), non-rejected human reads accounted for 48% of the “keep” fraction. In agreement with Marquet et al. [14], NAS alone was insufficient to comprehensively remove human reads from the final sequencing output, necessitating the use of more stringent software host-removal tools [25] to avoid potential patient privacy issues.

Table 1 Summary statistics of first two nanopore adaptive sampling (NAS) runs

	Pilot run			Run 2 (100% NAS run)		
	Total	Control channels	NAS channels	Total pool	Flowcell 2A	Flowcell 2B
Total read count (M)	9.62	2.43	7.19	4.18	1.89	2.19
Total base output (Gb)	13.7	6.9	6.8	5.0	2.2	2.8
Human reads (M)	9.19	2.30	6.79	3.59	1.67	1.92
Bacterial reads (M)	0.52	0.13	0.39	0.49	0.22	0.27
Bacterial output (Gb)	1.58	0.39	1.19	1.67	0.76	0.91
% bacterial reads	5.4	5.2	5.5	11.9	11.6	12.4
% bacterial output	11.5	5.7	17.5	33.4	34.6	32.5
Target enrichment (reads)	3.1x			N.A. (no control channels)		
Target enrichment (output)	3.1x					
Available pores (initial)	1305			1076	1021	
Available pores (end run)	28			21	144	

Analysis of the sequencing output from the positive control consisting of the mock community alone revealed a disproportionate read yield for the two Gram-negative species accounting for >90% of the total reads (Table S3). This possibly reflects a bias in DNA extraction efficiency for Gram-positive and -negative species, or discrepancies in estimation of the bacterial loads of the mock community from OD₆₀₀ absorbance. Dilution of the mock community spike in sputum A and B resulted in a decrease of reads for *K. pneumoniae* and *S. pneumoniae* (Fig. 1). However, in unspiked sputum A we also observed up to $\sim 10^3$ reads per million of *S. aureus*, and sputum B contained similar amounts of *P. aeruginosa*. This provides an example of expected pathogenic read yields in clinical sputum but impacted the ability to access the lower spike concentrations of these species. Baseline levels of bacteria that were also part of the mock community were observed in unspiked sputa, with up to $\sim 10^3$ reads per million of *S. aureus* in sputum A and *P. aeruginosa* in sputum B. These reads likely represent the presence of these mock community species already in the clinical sputa. While hindering the quantification of lower spike dilutions, they also provide an example of expected pathogenic read yields in clinical sputa.

In addition, while we observed a general trend of decreasing bacterial reads with reduced spike input,

read counts often did not correlate directly with the ten-fold diluted spike series, likely indicative of difficulties in consistent sample homogenization or DNA extraction from viscous sputum samples. We tested if AMR genotypic profiles from previous in-house assemblies of the *K. pneumoniae* and *P. aeruginosa* mock community isolates could be reproduced from our NAS sequencing data (Fig. 2). Most of the expected AMR genes from the assemblies could be recovered in the positive control sample (*K. pneumoniae*: 16/17 acquired AMR genes and 4/4 AMR-linked chromosomal point mutations, *P. aeruginosa*: 10/11 acquired AMR genes), but only partial profiles were detected in 3.1×10^7 cfu spiked sputum B (*K. pneumoniae*: 13/17 acquired genes and 2/4 chromosomal point mutations, *P. aeruginosa*: 5/11 acquired AMR genes), 1/17 *K. pneumoniae* acquired genes detected in the 3.1×10^7 cfu spike of sputum A, and no genes detected at the lower spike concentrations/species read counts.

Sequencing of the negative control indicated two sources of contamination that need to be accounted for. *Lysobacter enzymogenes* accounted for 1405 reads, possibly originating from the achromopeptidase preparation used during DNA extraction, whereas 515 reads were taxonomically assigned to *Homo sapiens*, which could represent barcode leakage or trace cross-contamination

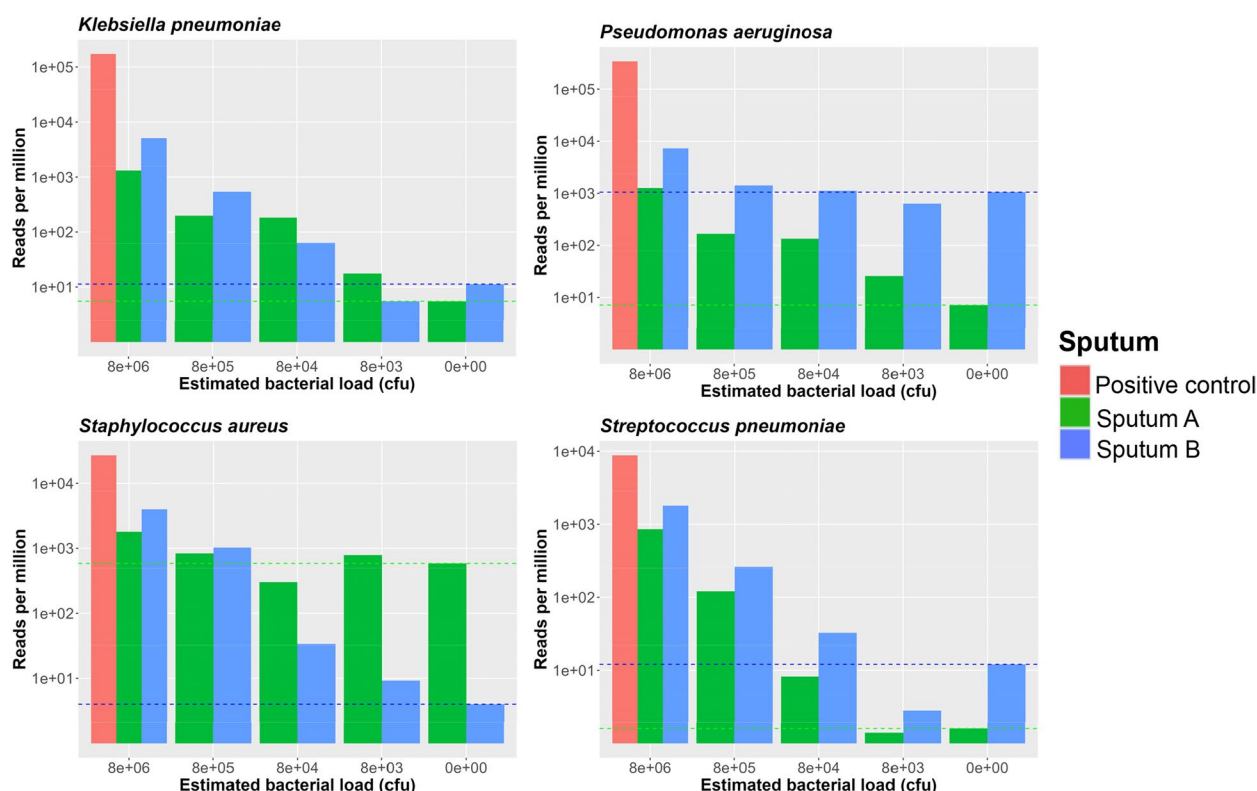


Fig. 1 Detection of mock respiratory pathogen community reads in spiked sputa. Bars indicate reads from the mock community positive control without sputum (pink) and sputa A (green) and B (blue). Dotted lines represent baseline counts from DNA already present in unspiked sputa

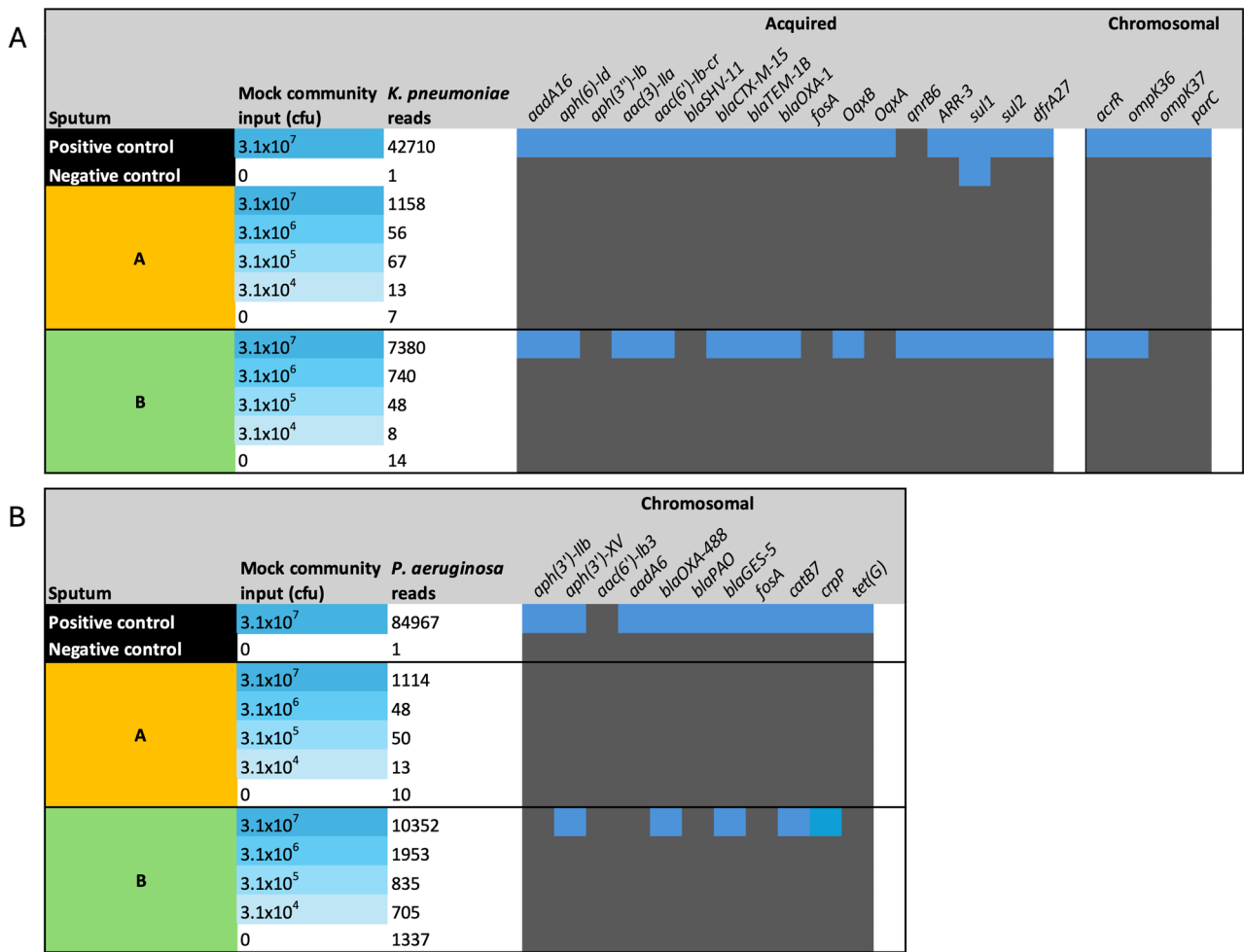


Fig. 2 Detection of AMR gene profiles in two sputum samples. Genes in the header row represent AMR genes or AMR-linked chromosomal mutations previously detected in in-house isolate assemblies of **A** *K. pneumoniae* and **B** *Pseudomonas aeruginosa* used in the mock community

from other high-host proportion samples multiplexed on the same flowcell (Table S2). Taxonomic classification of reads from the unspiked sputa also suggested that their innate bacterial loads were comparable to those of the spikes, with 33,413 reads taxonomically assigned to *Staphylococcus haemolyticus* in sputum A, and >10³ reads assigned to species including *Enterococcus faecium*, *Lactacaseibacillus rhamnosus* and *Streptococcus dysgalactiae* in sputum A and *Streptococcus oralis*, *Streptococcus dysgalactiae* and *Pseudomonas aeruginosa* in sputum B.

In summary, we could detect >100 reads per million for each mock community species at a bacterial load of approximately 8 × 10⁵ cfu per species. Higher read counts of >10⁴ reads were required for comprehensive AMR profile detection, and we estimated that we would need >10⁶ total reads per sample assuming a pathogenic read abundance of ~1%. As the total read counts per sample from the NAS half of the pilot flowcell ranged from 0.21–1.07 × 10⁶ (Table S1), we reasoned that this target read yield was attainable from multiplexing eight samples

on one or two flowcells fully running NAS, assuming that the pilot run performance could be reproduced.

Rapid pore loss greatly reduces sequencing output and metagenomic profiling performance

We next sequenced eight different clinical sputa (C1–C8) on two new MinION flowcells running NAS on all channels to evaluate metagenomic profiling performance across samples of highly variable taxonomic composition. Contrary to our expectation that whole-flowcell NAS would improve read turnover and sequencing performance, we observed a reduction in sequencing output of ~80% relative to the pilot sequencing run, obtaining 1.89 M/2.19 M reads and 2.2/2.8 Gb output from both flowcells (Table 1). The pore activity reports indicated that sequencing pores were lost at an accelerated rate relative to the pilot run (Fig. S1A), with ~90% of the initially available pores displaying the “inactive” status 24 h into the run (Fig. S2). Visual inspection of the flowcells indicated that this was not due to the introduction of air

bubbles. Slightly fewer pores were available for sequencing at the start of the runs relative to the pilot flowcell (~1000 vs 1300) – this number was still higher than the ONT guaranteed minimum of 800 and presumed to be a result of manufacturing variance. Pores could be partially and temporarily rescued with the nuclease wash and sample reloading in flowcell 2B, but could only improve sequencing output by ~25% relative to flowcell 2A (Fig. S2). This drastically reduced yield necessitated the pooling of sequencing output from both flowcells, reaching only 73.5% (5 Gb) of the base output and 58.1% (4.18 M) of the read count of the NAS half-flowcell in the pilot sequencing run.

Taxonomic composition varied greatly between samples, with total bacterial read counts ranging from 1.4×10^3 – 4.2×10^5 reads (median: 5.0×10^3) and bacterial read proportion ranging from 0.27–85.8% (median: 1.7%) (Fig. 3A&B). These distributions were heavily skewed by sample C5, which accounted for 87.2% of the bacterial reads sequenced across both flowcells (Table 2). Bacterial read count and proportion was generally higher in less viscous samples, which likely contained less host material (Fig. 3B). AMR genes were only detected in 2/8 of the samples, supporting our estimation from the pilot run that $>10^5$ reads from the relevant species were needed for AMR detection (Fig. 3C, Table 2). Most of the isolate species identified from sputum culture could be detected from the samples at >100 reads (6/8 samples), and some species previously unidentified from the culture were detected at a similar level or higher. However, read counts of species identified from metagenomic

sequencing generally fell short of our 10^5 target required for comprehensive metagenomic profiling. Similar to the pilot run, negative control sequencing indicated read contamination or barcode cross-talk from host DNA and *Lysobacter enzymogenes*, and also *Serratia marcescens* which was highly abundant in C5 and *S. dysgalactiae*. These negative control reads represent species that would likely need to be subtracted from the output in actual clinical diagnostic and surveillance on a specific run-by-run basis. In summary, rapid pore loss in the flowcells reduced total sequencing output by 5x, overriding the estimated $3 \times$ target enrichment from NAS and limiting metagenomic functionality to just species detection.

Rapid pore loss cannot be mitigated by extensive sample decontamination measures

We consequently focused on troubleshooting pore loss in an attempt to improve metagenomic functionality. While rapid pore loss has frequently been anecdotally reported on the ONT community forums, especially in the context of the current generation R10.4.1 chemistry flowcells, its causes are still to-date poorly defined and documented in both the official ONT best practices guides and scientific literature. Consultation with the ONT technical support team suggested that 1. pore loss could be accelerated by insufficient sample loading and incomplete pore occupancy during the run or 2. caused by the presence of unknown contaminants from the sample. We also investigated 3. any potential effects of NAS on flowcell health, due to previous reports [8, 9, 12] and the NAS best practices guide [10] indicating that NAS read rejection could

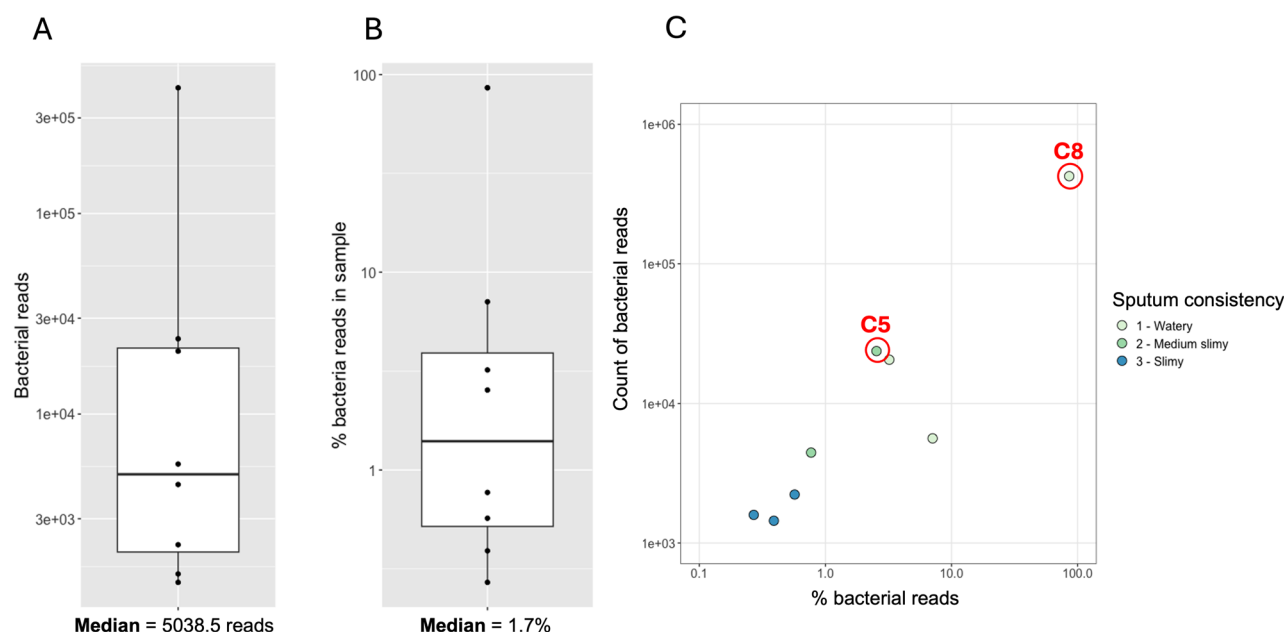


Fig. 3 Variability in bacterial composition between different clinical sputa. **A** Bacterial read count per sample. **B** Proportion of bacterial reads per sample. **C** Scatterplot summary of bacterial composition of eight clinical sputa. Sputa with detected AMR genes are circled in red

Table 2 Taxonomic composition of NAS-sequenced clinical sputa

Sample	Total reads	Bacterial reads	Bacterial %	AMR genes	Detected species (> 100 reads)	Species read counts
C1	373,196	1444	0.39%	-	<i>Serratia marcescens</i> †	554
					<i>Streptococcus dysgalactiae</i> †	498
					<i>Pseudomonas aeruginosa</i> **	95
C2	575,202	4449	0.77%	-	<i>Streptococcus dysgalactiae</i> †	933
					<i>Serratia marcescens</i> †	545
					<i>Streptococcus parasanguinis</i>	297
					<i>Streptococcus oralis</i>	290
					<i>Streptococcus mitis</i>	275
					<i>Streptococcus salivarius</i>	204
					<i>Rothia mucilaginosa</i>	129
					<i>Gemella haemolysans</i>	126
					<i>Streptococcus anginosus</i>	123
					<i>Veillonella atypica</i>	101
					<i>Staphylococcus aureus</i> **	16
C3	393,058	2226	0.57%	-	<i>Haemophilus influenzae</i> *†	733
					<i>Serratia marcescens</i> †	495
					<i>Streptococcus dysgalactiae</i> †	367
					<i>Streptococcus salivarius</i>	114
C4	77,016	5628	7.31%	-	<i>Streptococcus dysgalactiae</i> †	3265
					<i>Serratia marcescens</i> †	821
					<i>Lysobacter enzymogenes</i> †	759
					<i>Escherichia coli</i> *	123
C5	493,704	424,278	85.94%	<i>aac(6′)-Ic</i> , <i>blaSST-1</i>	<i>Serratia marcescens</i> *†	361,086
					<i>Citrobacter freundii</i> *	11,192
					<i>Streptococcus dysgalactiae</i> †	1887
					<i>Lysobacter enzymogenes</i> †	922
					<i>Serratia ureilytica</i>	520
					<i>Streptococcus pyogenes</i>	427
					<i>Rothia mucilaginosa</i>	279
					<i>Klebsiella pneumoniae</i>	252
C6	639,859	20,616	3.22%	-	<i>Lysobacter enzymogenes</i> †	10,897
					<i>Streptococcus dysgalactiae</i> †	2941
					<i>Serratia marcescens</i> †	2720
					<i>Streptococcus pneumoniae</i> *†	2405
					<i>Haemophilus influenzae</i>	432
C7	584,192	1591	0.27%	-	<i>Serratia marcescens</i> *†	657
					<i>Streptococcus dysgalactiae</i> †	592
C8	931,635	23,769	2.55%	<i>cfXA</i> , <i>erm(C)</i> , <i>tetQ</i>	<i>Streptococcus dysgalactiae</i> †	5425
					<i>Streptococcus anginosus</i>	2833
					<i>Hoylella buccalis</i>	2083
					<i>Bacteroides fragilis</i>	1565
					<i>Serratia marcescens</i> †	1456
					<i>Escherichia coli</i>	1233
					<i>Segatella oris</i>	956
					<i>Staphylococcus epidermidis</i>	638
					<i>Prevotella denticola</i>	396
					<i>Prevotella corporis</i>	318
					<i>Burkholderia cenocepacia</i> *	100
Negative control	6121	2468	40.32%	-	<i>Lysobacter enzymogenes</i> †	1022
					<i>Streptococcus dysgalactiae</i> †	802
					<i>Serratia marcescens</i> †	474

*species identified from sputum culture

**species identified from culture below detection threshold

†likely contaminant species also detected in negative control

result in reduced sequencing output, especially in the case of longer reads (> 10 kb).

We first addressed the likely sources of contamination from our samples. For this purpose, we used DNA extracted from four clinical sputa for our troubleshooting run, and switched from the earlier ligation sequencing kit with PCR-barcoding extension to a rapid barcoding kit protocol as it requires less input DNA per sample (1–4 ng), allowing us to use a smaller volume of the extracted sputum DNA and hence reduce contaminant carryover from the sputum. We additionally increased the number of PCR cycles to 30× to further amplify sample DNA concentration relative to potential contaminants, and the pooled libraries were washed thrice using AMPure beads as recommended in the protocols. This resulted in a superficially high-purity DNA sample based on Nanodrop UV absorbance spectroscopy, with a 260/230 ratio of 1.92 and a 260/280 ratio of 1.93. In order to address point 1, we loaded an increased amount of pooled library per flowcell (150fmol, 3× higher than the recommended 50fmol minimum) to maximize pore occupancy. The same library pool was loaded on four flowcells, with two flowcells running NAS on 50% of the channels to assess if NAS accelerated pore loss on the NAS half, and also two control flowcells without NAS to investigate there was accelerated global pore loss across the NAS flowcells relative to non-NAS setups.

Rapid pore loss was once again observed on both NAS and control flowcells despite all the troubleshooting measures taken, resulting in a 75% reduced sequencing output relative to the pilot run (Fig. S3, Table 3). As before, ~ 90% of initially available pores were inactive 24 h into the run and could not be adequately recovered by a nuclease wash/reload step. Pore occupancy was at ~ 90% throughout the runs, suggesting that low pore occupancy was not the cause. In response to the poor performance of the first set of replicates, we treated the pooled libraries with RNase A to rule out pore blockage by RNA molecules from the PCR reaction or the sample. This did not improve run performance in the second NAS cell, and

while there was a modest increase in sequencing output in the second control flowcell (4.6 Gb vs 3.3 Gb, 39%) it was difficult to tell whether this could also be due to inter-flowcell variability with a single replicate. Total sequencing output from the NAS flowcells was similar to that of the control flowcells, suggesting that NAS was not a major cause of pore loss in our situation.

NAS enrichment of bacterial reads was significantly lower than that of the pilot run, with an estimated 1.2× on-cell enrichment comparing NAS and control pores on the NAS flow cells, and a relative 1.1× enrichment in bacterial read yield in NAS flowcells relative to the control flowcells (Table 3). This limited enrichment was likely due one of the four clinical samples being 95% bacterial by composition, resulting in 35% target bacterial reads in the multiplexed library pool, which is higher than the recommendation of previous studies [8] and the ONT best practices guide of a 1–10% target composition to maximize potential enrichment gain. This highlights how a single deviant sample in a multiplexed setup can affect NAS enrichment across the entire flowcell. In summary, while we were able to rule out NAS and pore underloading as likely causes for the observed rapid pore loss, we could not eliminate the problem despite rigorous measures to reduce potential contaminants.

Discussion

We evaluated the use of NAS for the enrichment of target bacterial genomic sequences from sputum metagenomic samples observing an enrichment of at best 3.1× by read yield, similar to previous metagenomic applications of NAS [9, 13–20]. Whether this constitutes a functionally significant enrichment is dependent on the metagenomic problem at hand—under applications where the scientific question can already be answered without NAS, a three-fold enrichment introduced by NAS could indeed triple multiplexing capacity or improve the time-to-result in the context of rapid diagnosis. However, in our intended use-case of culture-free metagenomic surveillance of respiratory pathogens, this level of enrichment

Table 3 Summary statistics of troubleshooting runs

	Control flowcells		NAS flowcells					
			NAS #1		NAS #2			
	Control #1	Control #2	Total	Control channels	NAS channels	Total	Control channels	NAS channels
Total read count (M)	1.08	1.44	1.43	0.66	0.77	1.43	0.64	0.79
Total base output (Gb)	3.3	4.6	3.3	2.1	1.3	3.4	2.0	1.3
Human reads (M)	0.70	0.94	0.94	0.43	0.50	0.93	0.41	0.51
Bacterial reads (M)	0.38	0.50	0.50	0.23	0.26	0.50	0.22	0.28
% bacterial reads	35.1	34.6	34.6	34.7	34.5	35.2	35.0	35.3
On-cell target enrichment	N.A. (no control channels)		1.15x			1.25x		
Inter-cell target enrichment			1.10x					
Available pores (initial)	990	1025	1021			1113		
Available pores (end run)	79	61	91			72		

did not yield sufficient reads for the near-complete genomes required for AMR profiling in most of the clinical samples tested, due to the overwhelming proportion of contaminant host DNA of up to >99.7%. We suggest that NAS should thus be treated more as a low-cost polishing step in already-feasible workflows, rather than as a reliable replacement for wet-lab target enrichment protocols. Existing approaches for enriching low abundance respiratory pathogens currently outperform NAS – culture-based isolate sequencing remains the gold-standard in delivering pure high-coverage genomes, whereas established chemical sputum host-depletion protocols and target-capture using biotinylated baits yield superior target enrichment of up to 256x [4] and 300x [2] respectively, with the caveat of increased sample processing time and costs.

We multiplexed up to eight samples on the same flowcell to improve sample throughput and reduce overall sequencing costs, but in the process introduced two major complications to the workflow. Firstly, the presence of a few high bacterial load samples such as those in the troubleshooting run could raise the overall target composition on the flowcell above the 1–10% best-practices ideal and reduce the overall enrichment benefit across all samples loaded on the flowcell. Secondly, barcode cross-talk was observed in the sequence reads, with high-abundance species from one sample appearing at lower counts in the other samples and the negative control. This needs to be accounted for by manual curation, automatic bioinformatic methods or the use of dual barcoding systems with reduced barcode crosstalk. Given these limitations and the low target enrichment by NAS, we agree with the recommendation of a prior ONT respiratory metagenomics study to multiplex metagenomic samples for optimal data yield and performance [1], with the major caveat of increased sequencing costs.

We attempted to troubleshoot the rapid pore loss issue with the aim of improving metagenomic profiling with the potential 5–10× recovery in sequencing output, but were unable to resolve it or conclusively determine its cause. We ruled out pore underloading and NAS as probable causes, but could not prevent pore loss despite extensive sample decontamination measures. Two possible explanations remain, the first being that sputum samples might be inherently difficult to sequence with ONT, containing persistent inhibitors of pore activity that remain potent despite rigorous sample dilution and washing, and the second being that the undetermined pore inhibitor may have been introduced during library preparation. As the Nanodrop UV absorbance spectra of our pooled libraries conformed to what would be expected of a high-purity DNA sample, this unknown factor either has inconveniently similar absorbance characteristics to DNA, does not register on the absorbance spectrum, or is DNA itself.

We speculate that sputum-origin inhibitors could include covalently-modified DNA, secondary DNA structures (e.g. hairpins), or sputum polysaccharides co-purified with the sample. PCR-based library preparation might also systematically introduce short barcoding primer-dimers that might affect flowcell performance. This could be exacerbated by the inclusion of a negative control and other low-concentration samples, which could promote primer-dimer formation due to insufficient PCR template DNA. Further research is needed to investigate if stringent DNA size selection or the omission of low-concentration samples can mitigate the problem.

Conclusion

While we were able to modestly enrich target bacterial sequences by up to 3.1× with NAS, read counts were generally insufficient for our surveillance purposes due to pore loss. Rapid pore loss is a common cause of ONT run failure but remains undocumented in peer-reviewed literature. Our study indicates that rapid pore loss significantly reduces data output and limits metagenomic application, and that its idiosyncratic nature poses a challenge to our original goal of deployment in lower-middle-income countries, which often lack the necessary resources and logistical support to troubleshoot on-the-fly. We thus document our troubleshooting efforts to stimulate systematic discussion of the problem and further development, in the hope that long-read sequencing of clinical sputum samples will be attainable for surveillance and diagnostic applications in the future.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12920-025-02272-8>.

Supplementary Material 1.

Supplementary Material 2.

Acknowledgements

This study was co-funded by the PPP allowance made available by Health-Holland, Top Sector Life Sciences & Health (project number: LSHM22031) as part of the “AMR-Global: Reducing inappropriate exposure to antibiotics” (GLORIA-2) program. In addition, we acknowledge KNCV’s Dream Fund project “No more pandemics”, financially supported by the Dutch Postcode Lottery, for enabling KNCV’s contribution to this project. We also acknowledge the Utrecht Sequencing Facility (USEQ) for providing sequencing service and data. USEQ is subsidized by the University Medical Center Utrecht and The Netherlands X-omics Initiative (NWO project 184.034.019). Adri van der Zanden, founder and previous director of Mozand B.V., is very much thanked for his support to the start-up of this project.

Authors’ contributions

A.C.S. and K.K. conceived and supervised the study. A.C.S., J.T., M.C.V. and W.X. designed experiments and optimised study protocols. J.T., M.C.V. and W.X. acquired experimental data. J.T. and W.X. analysed the results. A.C.S., J.T. and W.X. initially drafted the manuscript, figures and tables. All authors (A.C.S., J.T., M.C.V., W.X., A.S., N.H., S.v.E., D.E., R.A. and K.K.) were involved in experimental troubleshooting, critical review of the manuscript and have read and approved of this final version.

Funding

This study was co-funded by the PPP allowance made available by Health ~ Holland, Top Sector Life Sciences & Health (project number: LSHM22031) as part of the "AMR-Global: Reducing inappropriate exposure to antibiotics" (GLORIA-2) program.

Data availability

Raw sequencing data was processed using Hostile v2.0.0 with the human-t2t-hla-argos985 index [26] to remove human origin reads, and the processed. fastq files were uploaded and made available to the European Nucleotide Archive (ENA) project PRJEB86674.

Declarations

Ethics approval and consent to participate

This study is compliant with the Declaration of Helsinki. The sputum samples used in this study included residual and completely anonymized material from routine diagnostic protocols at University Medical Centre Utrecht (UMCU). Ethical approval for the use of the sputum samples for this study was obtained from the UMCU Biobank Research Ethics Committee (TCBio) with approval number: TCBio 25–063. This approval waives the need for patient consent to participation according to TCBio regulations.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 25 March 2025 / Accepted: 11 November 2025

Published online: 01 December 2025

References

1. Nilgiriwala K, Rabodoarivelo MS, Hall MB, Patel G, Mandal A, Mishra S, et al. Genomic sequencing from sputum for tuberculosis disease diagnosis, lineage determination, and drug susceptibility prediction. *J Clin Microbiol*. 2023;61(3):e0157822.
2. Lanza VF, Baquero F, Martinez JL, Ramos-Ruiz R, Gonzalez-Zorn B, Andreumont A, et al. In-depth resistome analysis by targeted metagenomics. *Microbiome*. 2018;6(1):11.
3. 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.
4. Alcolea-Medina A, Alder C, Snell LB, Charalampous T, Aydin A, Nebbia G, et al. Unified metagenomic method for rapid detection of microorganisms in clinical samples. *Commun Med*. 2024;4(1):135.
5. Cookson W, Cox MJ, Moffatt MF. New opportunities for managing acute and chronic lung infections. *Nat Rev Microbiol*. 2018;16(2):111–20.
6. Kok NA, Paker N, Schuele L, de Beer JL, Rossen JWA, Sinha B, et al. Host DNA depletion can increase the sensitivity of *Mycobacterium* spp. detection through shotgun metagenomics in sputum. *Front Microbiol*. 2022;13:949328.
7. Kim M, Parrish RC 2nd, Tisza MJ, Shah VS, Tran T, Ross M, et al. Host DNA depletion on frozen human respiratory samples enables successful metagenomic sequencing for microbiome studies. *Commun Biol*. 2024;7(1):1590.
8. Martin S, Heavens D, Lan Y, Horsfield S, Clark MD, Leggett RM. Nanopore adaptive sampling: a tool for enrichment of low abundance species in metagenomic samples. *Genome Biol*. 2022;23(1):11.
9. Payne A, Holmes N, Clarke T, Munro R, Debebe BJ, Loose M. Readfish enables targeted nanopore sequencing of gigabase-sized genomes. *Nat Biotechnol*. 2021;39(4):442–50.
10. Oxford Nanopore Technologies. Adaptive Sampling 2024 [(ADS_S1016_v1_revM_11Dec2024)] [Available from: <https://nanoporetech.com/document/adaptive-sampling>].
11. Ulrich JU, Epping L, Pilz T, Walther B, Stingl K, Semmler T, et al. Nanopore adaptive sampling effectively enriches bacterial plasmids. *mSystems*. 2024;9(3):e0094523.
12. Cheng H, Sun Y, Yang Q, Deng M, Yu Z, Zhu G, et al. A rapid bacterial pathogen and antimicrobial resistance diagnosis workflow using Oxford nanopore adaptive sequencing method. *Brief Bioinform*. 2022. <https://doi.org/10.1093/bib/bbac453>.
13. Weilguny L, De Maio N, Munro R, Manser C, Birney E, Loose M, et al. Dynamic, adaptive sampling during nanopore sequencing using Bayesian experimental design. *Nat Biotechnol*. 2023;41(7):1018–25.
14. Marquet M, Zollkau J, Pastuschek J, Viehweger A, Schleussner E, Makarewicz O, et al. Evaluation of microbiome enrichment and host DNA depletion in human vaginal samples using Oxford Nanopore's adaptive sequencing. *Sci Rep*. 2022;12(1):4000.
15. Su J, Lui WW, Lee Y, Zheng Z, Siu GK, Ng TT, et al. Evaluation of *Mycobacterium tuberculosis* enrichment in metagenomic samples using ONT adaptive sequencing and amplicon sequencing for identification and variant calling. *Sci Rep*. 2023;13(1):5237.
16. Sun Y, Cheng Z, Li X, Yang Q, Zhao B, Wu Z, et al. Genome enrichment of rare and unknown species from complicated microbiomes by nanopore selective sequencing. *Genome Res*. 2023;33(4):612–21.
17. Kipp EJ, Lindsey LL, Khoo B, Faulk C, Oliver JD, Larsen PA. Metagenomic surveillance for bacterial tick-borne pathogens using nanopore adaptive sampling. *Sci Rep*. 2023;13(1):10991.
18. Viehweger A, Marquet M, Holzer M, Dietze N, Pletz MW, Brandt C. Nanopore-based enrichment of antimicrobial resistance genes - a case-based study. *GigaByte*. 2023;2023:gigabyte75.
19. Lin Y, Dai Y, Liu Y, Ren Z, Guo H, Li Z, et al. Rapid PCR-based nanopore adaptive sequencing improves sensitivity and timeliness of viral clinical detection and genome surveillance. *Front Microbiol*. 2022;13:929241.
20. Buytaers FE, Verhaegen B, Van Nieuwenhuysen T, Roosens NHC, Vanneste K, Marchal K, et al. Strain-level characterization of foodborne pathogens without culture enrichment for outbreak investigation using shotgun metagenomics facilitated with nanopore adaptive sampling. *Front Microbiol*. 2024;15:1330814.
21. Laszlo AH, Derrington IM, Gundlach JH. MspA nanopore as a single-molecule tool: from sequencing to SPRNT. *Methods*. 2016;105:75–89.
22. Kovaka S, Fan Y, Ni B, Timp W, Schatz MC. Targeted nanopore sequencing by real-time mapping of raw electrical signal with UNCALLED. *Nat Biotechnol*. 2021;39(4):431–41.
23. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. *Genome Biol*. 2019;20(1):257.
24. Bortolaia V, Kaas RS, Ruppe E, Roberts MC, Schwarz S, Cattori V, et al. ResFinder 4.0 for predictions of phenotypes from genotypes. *J Antimicrob Chemother*. 2020;75(12):3491–500.
25. Munro R, Holmes N, Moore C, Carlile M, Payne A, Tyson JR, et al. A framework for real-time monitoring, analysis and adaptive sampling of viral amplicon nanopore sequencing. *Front Genet*. 2023;14:1138582.
26. Constantinides B, Hunt M, Crook DW. Hostile: accurate decontamination of microbial host sequences. *Bioinformatics*. 2023. <https://doi.org/10.1093/bioinformatics/btad728>.

Publisher's Note

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