



Article scientifique

Article

2023

Published version

Open Access

This is the published version of the publication, made available in accordance with the publisher's policy.

A metagenomics method for the quantitative detection of bacterial pathogens causing hospital-associated and ventilator-associated pneumonia

Hauser, S; Lazarevic, Vladimir; Tournoud, M; Ruppe, Etienne; Santiago Allexant, E; Guigon, G; Schicklin, S; Lanet, V; Girard, Myriam; Mirande, C; Gervasi, G; Schrenzel, Jacques

How to cite

HAUSER, S et al. A metagenomics method for the quantitative detection of bacterial pathogens causing hospital-associated and ventilator-associated pneumonia. In: Microbiology spectrum, 2023, vol. 11, n° 6, p. e0129423. doi: 10.1128/spectrum.01294-23

This publication URL: <https://archive-ouverte.unige.ch/unige:175404>

Publication DOI: [10.1128/spectrum.01294-23](https://doi.org/10.1128/spectrum.01294-23)

A metagenomics method for the quantitative detection of bacterial pathogens causing hospital-associated and ventilator-associated pneumonia

S. Hauser,¹ V. Lazarevic,² M. Tournoud,¹ E. Ruppé,² E. Santiago Allexant,³ G. Guigon,³ S. Schicklin,³ V. Lanet,³ M. Girard,² C. Mirande,⁴ G. Gervasi,³ J. Schrenzel²

AUTHOR AFFILIATIONS See affiliation list on p. 13.

ABSTRACT The management of ventilator-associated and hospital-acquired pneumonia requires rapid and accurate quantitative detection of the infecting pathogen(s). To achieve this, we propose a metagenomics next-generation sequencing (mNGS) assay that includes the use of an internal sample processing control (SPC) for the quantitative detection of 20 relevant bacterial species of interest (SOI) from bronchoalveolar lavage (BAL) samples. To avoid very major errors in the identification of respiratory pathogens due to “false-negative” cases, each sample was spiked with *Bacillus subtilis*, at a precisely defined concentration, using rehydrated BioBall. This SPC ensured the detection and quantification of the pathogen(s) at defined minimum concentrations. In the presented mNGS workflow, absolute quantification of *Staphylococcus aureus* was as accurate as quantitative PCR. We defined a metagenomics threshold at 5.3×10^3 genome equivalent unit per milliliter of the sample for each SOI, to distinguish colonization from higher amounts of pathogens that may be associated with infection. Complete mNGS process and metrics were assessed on 40 clinical samples, showing >99.9% sensitivity compared to microbial culture. However, 19 out of the 29 (66%) SOI detections above the metagenomics threshold were not associated with bacterial growth above classical culture-based clinical thresholds. Taxonomic classification of 7 (37%) of the “false-positive” detections was confirmed by finding specific 16S/MetaPhlAn2 markers, the 12 other “false-positive” detections did not yield enough reads to check their taxonomic classification. Our SPC design and analytical workflow allowed efficient detection and absolute quantification of pathogens from BAL samples, even when the bacterial DNA quantity was largely below the manufacturer’s recommendations for NGS. The frequent “false-positive” detections suggested the presence of nonculturable cells within the tested BAL samples. Furthermore, mNGS detected mixed infections, including bacterial species not reported by routine cultures.

IMPORTANCE The management of ventilator-associated pneumonia and hospital-acquired pneumonia requires rapid and accurate quantitative detection of the infecting pathogen. To this end, we propose a metagenomic sequencing assay that includes the use of an internal sample processing control for the quantitative detection of 20 relevant bacterial species from bronchoalveolar lavage samples.

KEYWORDS clinical metagenomics, diagnosis, standardization, sample processing control, spike-in bacteria, calibrator, absolute quantification, metagenomics threshold, HAP-VAP

Hospital-acquired pneumonia (HAP) and ventilator-associated pneumonia (VAP) are common nosocomial infections, predominantly of bacterial origin, causing high

Editor Anne Jamet, Institut Necker Enfants Malades, Paris, France

Address correspondence to S. Hauser, sebastien.hauser@biomerieux.com.

Some of the authors are employees of bioMérieux, a company creating and developing infectious disease diagnostics. No other potential conflicts of interest relevant to this article are reported.

Received 9 May 2023

Accepted 29 September 2023

Published 27 October 2023

Copyright © 2023 Hauser et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

morbidity rates (1). Optimal care for HAP/VAP requires prompt detection and identification of the causative pathogen.

Today, the “gold standard” method for the microbiological diagnosis of HAP/VAP is culture-based (2) and requires at least 24 h to identify the causative pathogen (3). Molecular methods that are independent of cultivability may improve speed and analytical sensitivity (4) over conventional tests (5–7). Clinical applications of metagenomic next-generation sequencing (mNGS) are currently drawing the attention of infectious diseases specialists (8, 9) and guidelines for NGS-based diagnostics have emerged (10–12). However, the technical requirements for NGS diagnostics remain to be defined. To control for the integrity of the reagents, equipment functionality, and the potential presence of inhibitors, the US Food and Drug Administration (FDA) (12) recommends the use of an internal control (IC), typically a “foreign sequence” co-extracted and co-analyzed with the sequences of the sample. Spiked nucleic acids (13) or exogenous bacteria (14) can be used as IC. Moreover, metagenomics analyses require control of the quantity and quality of the DNA extracted from the sample as well as the concentration, purity, and size distribution of the sequencing library (15). These control steps waste sample volume and are time-consuming, expensive, and tedious to make.

We propose the utilization of a sample processing control (SPC) that is spiked into the clinical sample and follows the entire analytical process. A positive SPC detection indicates that all steps of the sample processing were successful, and that sequencing data can be safely interpreted. This concept is widely used for integrated PCR cartridge systems such as GeneXpert (Cepheid), which contains *Bacillus globigii* spores or FilmArray (bioMérieux), which includes *Schizosaccharomyces pombe* cells. However, these SPC cannot be directly transposed from PCR to mNGS assays. As mNGS relies on random sequencing of DNA fragments, the detection of SPC competes with the detection of pathogens, commensal microbiota (flora), and the DNA of the patient. Therefore, the efficacy of the detection of SPC and pathogens depends also on the biological composition of each sample (relative abundance of microorganisms and the ratio of microbial to human cells) (16–18). In addition, clinical samples may contain large quantities of dead bacteria and extracellular bacterial DNA (19). The depletion of DNA from human cells and extracellular (human and bacterial) DNA is possible using selective lysis of host cells followed by endonuclease digestion. The SPC detection can then establish whether the mNGS is able to detect pathogens at a defined minimal concentration.

Clinically defined thresholds of pathogen concentrations are used to distinguish infection from asymptomatic colonization (20). For HAP/VAP diagnosis, culture-based thresholds are currently defined at 10^3 colony-forming units (CFU)/mL for mini-BAL and 10^4 CFU/mL for BAL samples (21–23). Therefore, mNGS should also provide absolute quantification of the detected pathogens. Various experimental approaches have been proposed including cell counting by flow cytometry (24), normalization of bacterial relative abundance based on defined cell numbers that are spiked into the samples before nucleic acid isolation (25), or use of spiked nucleic acids (26). However, these designs cannot provide an absolute quantification of pathogens.

To control all processing steps and to quantify the abundance of the detected pathogen(s), we spiked samples with rehydrated BioBall (bioMérieux) as an SPC. Diagnostic capabilities of the mNGS workflow were assessed in bronchoalveolar lavage (BAL) samples using a panel of bacterial species that commonly cause HAP/VAP infection.

MATERIALS AND METHODS

Selection of 20 bacterial pathogens for HAP/VAP diagnosis

We defined a panel of bacterial pathogens frequently involved in HAP/VAP in immunocompetent patients. This panel, composed of 20 bacterial species of interest (SOI), is listed in Table 1.

TABLE 1 List of the 20 pathogens composing the HAP/VAP panel

Enterobacterales	Gram-negative bacilli	Gram-positive cocci
<i>Escherichia coli</i>	<i>Pseudomonas aeruginosa</i>	<i>Staphylococcus aureus</i>
<i>klebsiella oxytoca</i>	<i>Stenotrophomonas maltophilia</i>	<i>Streptococcus pneumoniae</i>
<i>klebsiella pneumoniae</i>	<i>Acinetobacter baumannii</i>	
<i>klebsiella aerogenes</i>	<i>Legionella pneumophila</i>	
<i>Enterobacter cloacae</i>	<i>Haemophilus influenzae</i>	
<i>Serratia marcescens</i>		
<i>Proteus mirabilis</i>		
<i>Proteus vulgaris</i>		
<i>Hafnia alvei</i>		
<i>Citrobacter freundii</i>		
<i>Citrobacter koseri</i>		
<i>Morganella morganii</i>		
<i>Providencia stuartii</i>		

Collection of fresh BAL and mini-BAL samples

BAL or mini-BAL fluids were obtained at Geneva University Hospitals (HUG, Switzerland) from patients suspected of pneumonia, with no other specific clinical or demographic inclusion criteria. The samples were sent to the bacteriology laboratory for routine culture analysis on solid media (blood agar, chocolate agar, Mac Conkey agar, and Columbia colistin-nalidixic acid agar). The 20 SOI may grow on these media used for culturing the respiratory specimens except for *Legionella pneumophila*. Exceptionally, *L. pneumophila* can be grown on chocolate agar, albeit with extended incubation time, and possibly when using a nutrient-rich specimen (e.g., a biopsy) (27). In our study, solid media were incubated for 48 h; hence, *L. pneumophila* would not be isolated, even if nutrients were putatively brought by the specimen. Following plating and a 48-h incubation at 35°C in a 5% CO₂ enriched atmosphere, CFU were counted.

Strain identification of the predominant colony morphotypes was performed using matrix-assisted laser desorption ionization-time of flight mass spectrometry (MALDI Biotyper, Bruker Daltonics, Bremen, Germany) according to the manufacturer's instructions. For each sample, a list of the SOI detected is provided along with an estimate of their concentration rounded to the nearest log₁₀ value.

Two sets of samples were collected, the first set named "training set" included 45 samples (22 BAL and 23 mini-BAL), and the second set, named "validation set," included 40 samples (33 BAL and 7 mini-BAL).

Sample processing control

To control the whole mNGS process, we spiked the samples with spores of the Gram-positive bacterium *Bacillus subtilis* as an SPC. The SPC sequences are detected by our pipeline with an accuracy of 99.95%, without cross-identification of the SOI (see Fig. S1).

The *B. subtilis* strain ATCC 19659 BioBall MultiShot 10E8 (bioMérieux, Catalog #416721) was re-hydrated in 1.1 mL of phosphate buffered saline (PBS) to provide a solution of 10⁸ CFU/mL ± 20%. After a 100-fold dilution in PBS, 10 µL of the suspension was added to 600 µL of the (mini-)BAL fluid. The final SPC concentration within a sample was 1.7 × 10⁴ CFU/mL, close to the culture-based diagnostic threshold (10⁴ CFU/mL) for HAP/VAP.

Sample preparation and DNA extraction

Six hundred microliters of BAL supplemented with SPC was mixed with 6 mL of Tris-HCl 50 mM (pH 8) containing polyethylene glycol (PEG) 4% and saponin 0.08% and incubated at room temperature for 10 min for the selective lysis of human cells. After saponin treatment, the non-lysed cells were pelleted (12,000 g for 10 min) and treated with DNase I (5,000 U, 37°C, for 15 min). The DNase was inactivated by heating (80°C for 10 min).

and the addition of EDTA (10 mM final concentration). The sample was then added to a tube containing a mix of 1 mm glass beads (600 mg per 1.5 mL tube) and 0.1 mm zirconia/silica beads (150 mg per 1.5 mL tube) and bacteria were disrupted by shaking for 20 min on a vortex with a horizontal tube holder. Nucleic acids were extracted from the lysate on an easyMAG platform (bioMérieux) using the generic protocol (V2.0.1). Elution was carried out in a volume of 25 μ L, and the extracts were stored at -20°C . Human and bacterial DNA load were determined as described previously (18).

Library preparation and MiSeq sequencing

Libraries were prepared for 2×250 paired-end sequencing with a modified protocol of the Nextera XT DNA Library Preparation Kit (Illumina).

To allow efficient preparation of DNA libraries for all extracts, the protocol was adapted to run with only a few picograms of DNA (28, 29). The following steps were added to the standard protocol: (i) when the concentrations of extracted DNA were above 0.2 ng/ μ L, 1 ng was used for the preparation of the library; otherwise, a maximum volume of 5 μ L of DNA extract was used, regardless of its concentration; (ii) after tagmentation, 16 amplification cycles were applied to the DNA library in order to obtain enough material for sequencing; and (iii) the library was purified using 25 μ L of AMPure XP beads (Beckman Coulter). Indexed libraries prepared from two samples were pooled at equal quantities before sequencing on the MiSeq platform (see Fig. 1) with the MiSeq Reagent kit V3 (Illumina) following the manufacturer's instructions.

Bioinformatics pipeline

For the identification of pathogens, sequence reads were analyzed using the metagenomics pipeline described by Jaillard et al. (30) and Tournoud et al. [awaiting peer review (31)]. The detection of antibiotic resistance genes was not assessed in this study.

Briefly, the bioinformatics pipeline consisted of the following steps: quality control of the reads, trimming and filtering of poor quality reads, and taxonomic read binning with Kraken (32) using an internal reference database including more than 10,000 genomes from the 20 SOIs, *B. subtilis*, bacteria found in the lung and oral cavity, and the human genome (*Homo sapiens* genome assembly GRCh37). Bacterial genomes are both public [e.g., PATRIC (33), RefSeq (34), FDA ARGOS (35)] and private (strains sequenced from BAL samples). The median number of genomes per species was 42 with inter-quartile-range equal to (19–297), and the total number of genomes belonging to the flora was 7,593. The performance of taxonomic classification of simulated sequence reads of SOI, *B. subtilis*, and Human from the internal reference database is reported in Fig. S1.

Pathogen detection and quantification relied on the number of reads attributed to each SOI and the SPC. To avoid spurious pathogen detection due to erroneous taxonomic classification, reads associated with an SOI with an average genome sequencing depth $>1\times$ were assembled. Assembly was performed using the idba_ud500 assembler (IDBA-UD 1.1.1) (36), with the following parameters: —mink 40 —maxk 250 —min_pairs 2. To confirm species identification, the assemblies were aligned with BLAST (37) against a pathogen marker database built by selecting clade-specific MetaPhlAn2 (38) and 16S markers (for *Hafnia alvei*, *Proteus vulgaris*, and *Morganella morganii*, for which no MetaPhlAn2 markers were available) to allow unambiguous taxonomic assignment to the 20 SOIs. The BLAST algorithm was run with 75% coverage and 97% identity. When at least one marker of the tested SOI was detected, the taxonomic assignment of the sequences to the SOI was confirmed. When a marker from another species was detected and not the one of the tested SOI, the tested SOI was invalidated to avoid false-positive detection induced by misclassification of the reads. When no pathogen marker was detected, mainly because of low-coverage assemblies, no confirmation or invalidation of taxonomic classification to SOI was reported.

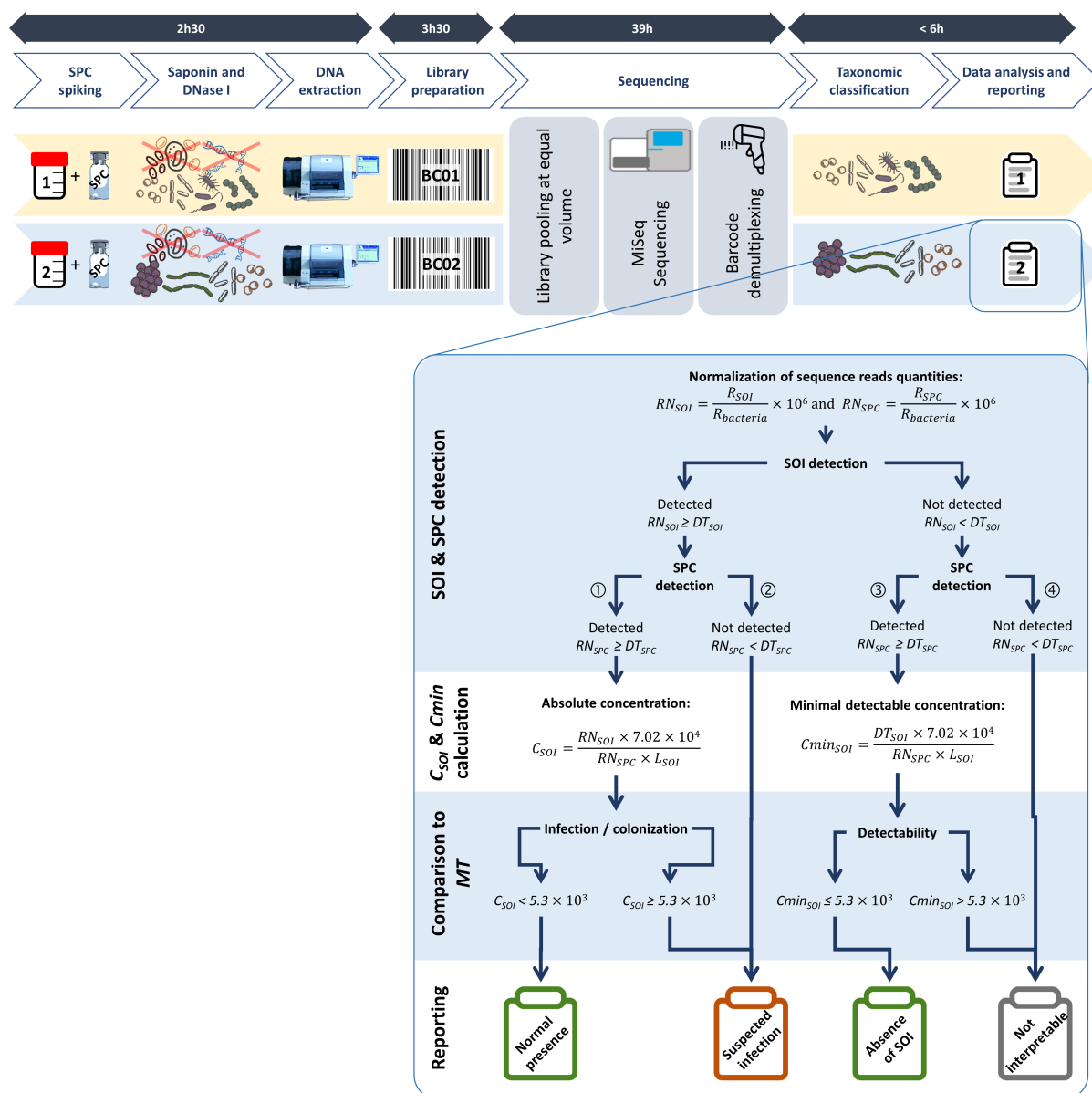


FIG 1 Complete workflow for quantitative detection of pathogen causing HAP/VAP by metagenomics. The duration of the key steps is shown at the top of the figure. RN_{SOI} and RN_{SPC} are normalized quantities in reads per million of bacterial reads classified as SOI and SPC, respectively. DT_{SOI} and DT_{SPC} are thresholds of detection of the normalized quantity of reads associated with SOI and SPC, respectively. L_{SOI} is the genome size of SOI (in Mb). C_{SOI} is the calculated absolute concentration of SOI in a sample in genome equivalent per milliliter (GEq/mL). $C_{min_{SOI}}$ is the minimal detectable concentration of SOI in a sample (in GEq/mL) for a given mNGS run.

Sequence data analysis and interpretation

The complete mNGS workflow is presented in Fig. 1. For each sample, all SOIs were analyzed separately as independent detection assays. Detailed procedures and calculations are provided in the supplemental material. Briefly, data interpretation consists of four steps.

Step 1: SOI and SPC detection

The numbers of reads classified to a bacterial species of interest (R_{SOI}) or to the SPC (R_{SPC}) are normalized to the number of millions of reads associated with all bacterial

species in our database (R_{bacteria}). The normalized values (RN_{SOI} and RN_{SPC}) are reported in reads per million of bacterial reads (RPMB).

We defined detection thresholds (DTs) to differentiate background from “true” taxonomic classification of sequence reads as SPC or SOI within sequenced sample (see supplemental material). These DTs represent the minimal quantity of classified sequence reads normalized per million of bacterial reads required to report SPC or SOI as detected.

Step 2: Calculation of absolute concentration or minimal detectable concentration of SOI

As SPC is added at a defined concentration ($C_{\text{SPC}} = 1.7 \times 10^4$ CFU/mL) within each sample, it can be used as a calibrator.

- When SOI and SPC are detected (Fig. 1), the absolute concentration of SOI (C_{SOI}) can be calculated as

$$C_{\text{SOI}} = \frac{RN_{\text{SOI}} \times 7.02 \times 10^4}{RN_{\text{SPC}} \times L_{\text{SOI}}}$$

- When a SOI is not detected (Fig. 1), its minimal detectable concentration (C_{minSOI}) can be calculated if SPC has been detected

$$C_{\text{minSOI}} = \frac{DT_{\text{SOI}} \times 7.02 \times 10^4}{RN_{\text{SPC}} \times L_{\text{SOI}}}$$

RN_{SOI} and RN_{SPC} correspond to the normalized quantity of sequence reads associated to SOI and SPC, respectively. L_{SOI} corresponds to the genome size of SOI in megabases. 7.02×10^4 is the product of SPC spiked concentration (1.07×10^4 GEq/mL) and its genome size in megabases (4.13 Mb).

Step 3: Comparison to the metagenomics threshold

A metagenomics threshold (MT) was defined (see supplementary material) as the concentration in genome equivalent (GEq) of SOI above which infection can be suspected. When a SOI is not detected (Fig. 1), the comparison of MT to the calculated C_{minSOI} makes it possible to differentiate the absence of SOI in the sample from the inability to detect this SOI at a concentration defined by the MT.

Step 4: Reporting

In each sample, the result of the detection of each SOI is reported separately. The rules of interpretation (see Fig. 1) define four possible reporting outcomes:

- Suspected infection, when an SOI is detected and quantified at or above MT or when the SOI is detected, but the SPC is not.
- Suspected colonization, when an SOI is detected and quantified below the MT.
- Absence of detection, when the SOI is not detected and the calculated C_{minSOI} is <MT.
- Not interpretable, when both SOI and SPC are not detected or when the calculated C_{minSOI} is >MT.

PCR quantification of *Staphylococcus aureus*

Quantitative PCR targeting the *spa* gene was used to quantify *S. aureus*. Master mixes were prepared by mixing 0.2 μM of forward primer 1914F (CAGCAAACCA TGCAGATGCT AA), 0.2 μM of reverse primer 1992R (ACAGTTGTAC CGATGAATGG ATTTT), and 0.1 μM of probe 1945T (AGCATTACCA GAAACT) (39) in the ABsolute QPCR Mix, no ROX (Thermo Scientific). One microliter of DNA extract was added in a final reaction volume of 50 μL . The amplification and the real-time reading of the fluorescent signal were carried

out on the CFX96 Touch Real-Time PCR Detection System (Bio-Rad) using the following program: 95°C for 15 min; 42 cycles (95°C, 15 s; 60°C, 1 min).

A standard curve was produced from the genomic DNA of *S. aureus* strain MW2 at concentrations ranging between 1 pg and 1 ng and corresponding to 3.1×10^2 to 3.1×10^5 GEq.

RESULTS

Absolute quantification of *S. aureus*

We compared the quantity of *S. aureus* obtained by our mNGS approach to the results of qPCR and culture quantifications (Table 2) on 11 samples from the “training set.” These samples were either culture-positive for *S. aureus* ($n = 10$) or culture-negative but with *S. aureus* detection by mNGS (sample T07).

The comparison of the nearest log₁₀ of concentrations determined by mNGS (in GEq/mL) and microbial culture (in CFU/mL) (Fig. 2A) showed moderate correlation (slope = 0.6087) with a weak linear association ($R^2 = 0.3043$), which was statistically insignificant ($P = 0.156$). The comparison of the log₁₀ values of concentrations determined by mNGS with those obtained from quantitative PCR (Fig. 2B) revealed a strong correlation (slope ≈ 1), strong linear association ($R^2 = 0.8392$), and statistical significance ($P < 0.01$).

Moreover, for sample T07, which was culture-negative, both qPCR and mNGS detected and quantified *S. aureus* above 10^4 GEq/mL.

Validation of quantitative detection of SOI

We evaluated the complete mNGS workflow (Fig. 1) with a “validation set” of 40 samples, comparing the quantitative detection of SOIs by culture and mNGS (Table 3). Twenty

TABLE 2 Comparison of *S. aureus* quantification by microbial culture, qPCR, and mNGS

ID	Sample type	Bacterial species	Quantification (log ₁₀) ^a			16S MetaPhlAn2 markers
			Microbial culture	mNGS	qPCR (<i>S. aureus</i>)	
T01	Mini-BAL	<i>S. aureus</i>	>5	7.4	7.2	+
T03	Mini-BAL	<i>S. aureus</i>	3	6.2	6.3	+
		<i>S. pneumoniae</i>	3	4.9		
T07	Mini-BAL	<i>S. aureus</i>	NEG	4.0	5.0	+
T17	BAL	<i>K. pneumoniae</i>	NEG	7.5		+
		<i>S. aureus</i>	2	5.1	4.3	+
T18	BAL	<i>E. coli</i>	>5	7.0		+
		<i>S. aureus</i>	>5	6.3	6.3	+
T19	BAL	<i>S. aureus</i>	5	6.3	5.5	+
T26	BAL	<i>E. coli</i>	>5	7.7		+
		<i>P. mirabilis</i>	>5	3.9		
		<i>S. aureus</i>	>5	5.6	5.0	+
T28	BAL	<i>H. influenzae</i>	>5	7.1		+
		<i>P. vulgaris</i>	4	4.0		
		<i>S. aureus</i>	5	5.0	4.5	+
T31	BAL	<i>H. influenzae</i>		3.8		+
		<i>S. aureus</i>	5	6.0	5.2	+
		<i>S. maltophilia</i>	5	7.9		+
T43	BAL	<i>H. influenzae</i>		5.7		+
		<i>P. mirabilis</i>	>5	7.7		+
		<i>S. aureus</i>	2	2.4	3.0	
T44	BAL	<i>S. aureus</i>	4	6.1	5.9	+
T45	BAL	<i>S. aureus</i>	4	3.7	3.6	+

^aLog₁₀ of quantified pathogen concentrations in CFU/mL (for microbial culture) and GEq/mL (for mNGS and qPCR) is presented in the corresponding columns.

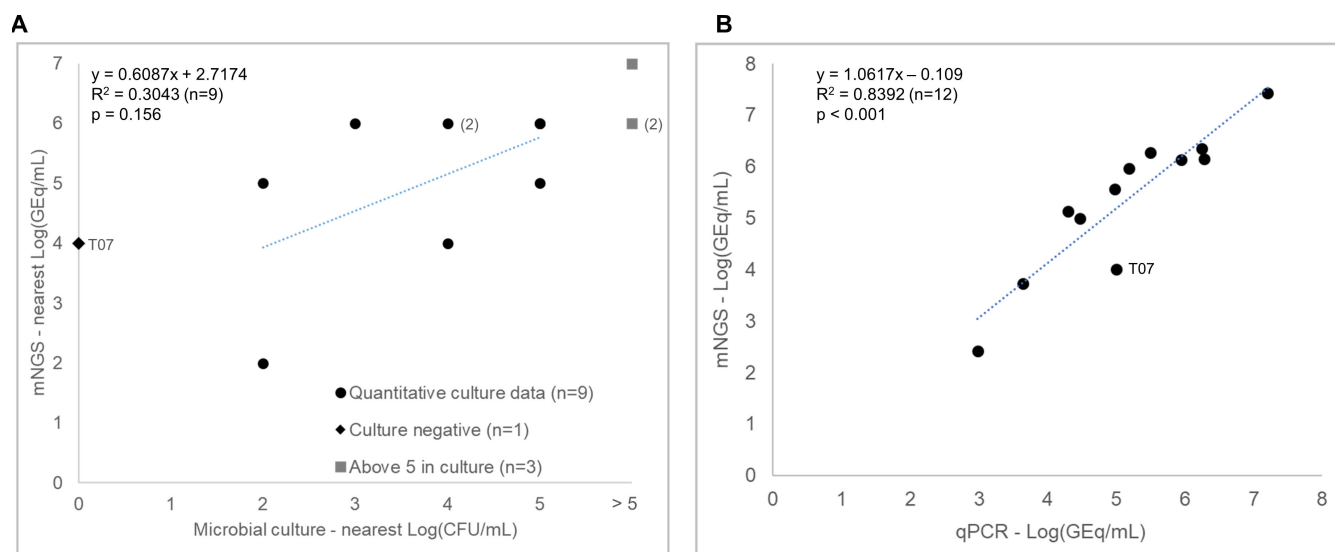


FIG 2 Comparison of *S. aureus* quantification. (A) Comparison of quantification in GEq/mL obtained through mNGS to CFU/mL determined by microbial culture. (B) Comparison of quantification by mNGS to quantitative PCR, both in GEq/mL.

samples (50%) tested negative for all SOIs by culture. Bacterial culture and mNGS identified other species outside the study panel. However, as we did not define the detection threshold for these species and their MetaPhlAn markers were not considered, our data analysis focused on the 20 SOIs.

Control of the sample processing

The SPC detection failed ($RN_{SPC} < DT_{SPC}$) in four samples (10%) in which at least one SOI was detected as a probable infecting agent (samples 3, 7, 23, and 35 in Table 3). In these samples, non-detected SOIs are reported as “not interpretable” (Fig. 1), as the minimal detectable concentration may be above the MT.

The SPC detection ($RN_{SPC} \geq DT_{SPC}$) in 36 samples (90%) provided conditions for mNGS results' validation. The presented mNGS workflow was able to process culture-negative samples, e.g., sample 36 (Table 3) that contained as low as 0.5 pg/μL DNA in the extract and generated only 4,850 reads. The detection of SPC and estimated C_{minSOI} below MT allowed for validating the sequencing results and reporting a negative detection for these samples. However, we were not able to validate 10.8% (75/697) of negative mNGS SOI detections because the calculated C_{minSOI} exceeded the MT (Fig. 1). It mainly concerned *E. coli*, *K. pneumoniae*, *P. vulgaris*, *S. aureus*, and *S. pneumoniae* (Table 3), which had high DT_{SOI} values (see Fig. S4).

Quantitative detection of the SOI panel

mNGS results were compared to the current “gold standard” for HAP/VAP diagnosis, i.e., the microbial culture (Fig. 3). Ten “true positive” detections of probable infection by SOIs and no “false-negative” results (Table 3; Fig. 3) revealed a test sensitivity above 99.9%. With 19 “false-positives” (Table 3; Fig. 3), the test specificity was 96.6% but the “false discovery rate” reached 65.5%. Importantly, we confirmed proper taxonomic classification of 26.3% (5/19) of the “false-positive” detections by finding specific 16S/MetaPhlAn2 markers. The 14 other “false-positive” detections did not yield enough reads to check their taxonomic classification by 16S/MetaPhlAn2 marker search.

Sample 27, which was culture positive for *E. cloacae* complex above 10^5 CFU/mL, led to seven “false-positive” detections (Table 3) corresponding mostly to Enterobacteriales. The presence of *K. pneumoniae*, detected at a concentration above 10^6 CFU/mL, was confirmed by the 16S/MetaPhlAn2 markers search. But the other “false-positive”

TABLE 3 Detection of SOI(s) above clinical threshold by microbial culture or above MT by mNGS in the “validation set”^a

N°	Sample type	Bacterial species	Culture	mNGS	16S/MetaPhlan2 markers	SPC detected	Comments
1	BAL	<i>S. pneumoniae</i>	> 5	4.8	+	+	
2	Mini-BAL					+	
3	BAL	<i>P. aeruginosa</i>	> 5	>MT	+	–	Negative detections of other SOI cannot be validated.
4	BAL					+	Cmin > MT for <i>E. aerogenes</i> , <i>E. coli</i> , <i>H. influenzae</i> , <i>K. pneumoniae</i> , <i>P. vulgaris</i> , <i>S. aureus</i> , and <i>S. pneumoniae</i> .
5	BAL					+	
6	BAL	<i>P. aeruginosa</i>	2	4.7		+	Cmin > MT for <i>E. aerogenes</i> , <i>E. coli</i> , <i>H. influenzae</i> , <i>K. pneumoniae</i> , <i>P. vulgaris</i> , <i>S. aureus</i> , and <i>S. pneumoniae</i> .
		<i>S. marcescens</i>	2	5.5	+		
7	Mini-BAL	<i>S. marcescens</i>	2	>MT	+	–	Negative detections of other SOI cannot be validated.
8	Mini-BAL					+	Cmin > MT for <i>A. baumannii</i> , <i>C. koseri</i> , <i>E. aerogenes</i> , <i>E. coli</i> , <i>H. influenzae</i> , <i>K. oxytoca</i> , <i>K. pneumoniae</i> , <i>P. vulgaris</i> , <i>S. aureus</i> , <i>S. maltophilia</i> , and <i>S. pneumoniae</i> .
9	BAL	<i>S. pneumoniae</i>	5	6.0	+	+	Cmin > MT for <i>E. coli</i> and <i>P. vulgaris</i> .
10	BAL					+	
11	BAL	<i>L. pneumophila</i>	+	1.4		+	
12	BAL					+	
13	BAL	<i>S. aureus</i>	5	5.9	+	+	Cmin > MT for <i>P. vulgaris</i> .
14	BAL					+	
15	BAL					+	
16	Mini-BAL					+	
17	BAL					+	
18	BAL					+	
19	BAL	<i>S. pneumoniae</i>		4.1		+	Cmin > MT for <i>E. coli</i> and <i>P. vulgaris</i> .
20	BAL	<i>P. aeruginosa</i>	5	5.6	+	+	Cmin > MT for <i>P. vulgaris</i> .
21	Mini-BAL					+	
22	BAL	<i>S. maltophilia</i>	5	7.3	+	+	Cmin > MT for <i>A. baumannii</i> , <i>C. koseri</i> , <i>E. aerogenes</i> , <i>E. coli</i> , <i>H. influenzae</i> , <i>K. oxytoca</i> , <i>K. pneumoniae</i> , <i>M. morganii</i> , <i>P. vulgaris</i> , <i>P. stuartii</i> , <i>P. aeruginosa</i> , <i>S. aureus</i> , and <i>S. pneumoniae</i> .
23	Mini-BAL	<i>P. aeruginosa</i>	> 5	>MT	+	–	Negative detections of other SOI cannot be validated.
		<i>S. aureus</i>	2	Invalid			
24	BAL					+	
25	BAL					+	
26	BAL					+	
27	BAL	<i>C. freundii</i>		4.5		+	Cmin > MT for <i>A. baumannii</i> , <i>C. koseri</i> , <i>E. aerogenes</i> , <i>E. coli</i> , <i>H. influenzae</i> , <i>M. morganii</i> , <i>P. vulgaris</i> , <i>P. stuartii</i> , <i>S. aureus</i> , and <i>S. pneumoniae</i> .
		<i>H. alvei</i>		4.8			
		<i>K. oxytoca</i>		4.8			
		<i>K. pneumoniae</i>		6.1	+		
		<i>P. aeruginosa</i>		4.5			
		<i>S. marcescens</i>		5.0			
		<i>S. maltophilia</i>		4.4			
		<i>E. cloacae</i> complex	> 5	7.1	+		
28	BAL	<i>S. marcescens</i>	2	4.2	+	+	
29	BAL	<i>E. aerogenes</i>	3	6.0	+	+	Cmin > MT for <i>E. coli</i> , <i>H. influenzae</i> , <i>K. pneumoniae</i> , <i>P. vulgaris</i> , and <i>S. pneumoniae</i> .
		<i>S. aureus</i>	2	5.5	+		
		<i>K. oxytoca</i>		3.8			
30	Mini-BAL					+	
31	BAL	<i>H. influenzae</i>	> 5	4.8	+	+	
32	BAL					+	
33	BAL					+	

(Continued on next page)

TABLE 3 Detection of SOI(s) above clinical threshold by microbial culture or above MT by mNGS in the “validation set”^a (Continued)

N°	Sample type	Bacterial species	Culture	mNGS	16S/MetaPhlAn2 markers	SPC detected	Comments
34	BAL	<i>S. marcescens</i>	3	4.8	+	+	Cmin > MT for <i>E. coli</i> , <i>K. pneumoniae</i> , <i>P. vulgaris</i> , and <i>S. pneumoniae</i> .
35	BAL	<i>P. aeruginosa</i>	5	>MT	+	–	Negative detections of other SOI cannot be validated.
36	BAL					+	
37	BAL					+	Cmin > MT for <i>E. coli</i> and <i>P. vulgaris</i> are above MT.
38	BAL	<i>S. maltophilia</i>	3	4.4		+	
39	BAL	<i>P. aeruginosa</i>	> 5	7.0	+	+	Cmin > MT for <i>A. baumannii</i> , <i>C. koseri</i> , <i>E. aerogenes</i> , <i>E. coli</i> , <i>H. influenzae</i> , <i>K. oxytoca</i> , <i>K. pneumoniae</i> , <i>P. vulgaris</i> , <i>S. aureus</i> , and <i>S. pneumoniae</i> .
		<i>S. maltophilia</i>		4.2			
40	BAL	<i>P. aeruginosa</i>	2	3.8		+	

^aLog10 of quantified pathogen concentrations in CFU/mL (for microbial culture) and GEq/mL (for mNGS) is presented in the corresponding columns. >MT means that SOI was detected but not the SPC, suggesting that detected SOI were likely present at a concentration above MT (Fig. 1B). (+) in 16S/MetaPhlAn2 markers column means that taxonomic classification of sequence reads to a bacterial species is confirmed by 16S/MetaPhlAn2 markers search. Values in bold correspond to quantified pathogen concentrations exceeding the clinical threshold (in Culture column) or metagenomics threshold (in mNGS column).

SOI detections were quantified at 10- to 100-fold lower concentrations relative to *K. pneumoniae* and had too few reads for 16S/MetaPhlAn2 confirmation.

Detection of co-infections

The presence of a single infecting agent at a high concentration may preclude the detection of other SOIs and the SPC and may limit the detection of co-infections by mNGS. However, we were able to detect two (samples 6 and 39) or three (sample 29) co-infecting pathogens with absolute concentrations differing by up to three orders of magnitude (Table 3). None of these mixed infections were reported by microbial cultures.

DISCUSSION

Here, we propose spiking BioBall into (mini-)BAL samples to control processing and provide absolute SOI quantification in the mNGS workflow. We selected *B. subtilis* as SPC because of its rare natural presence in BAL samples and the ability of mNGS to distinguish its sequence reads from those of SOIs, commensal flora, and the human genome (see Fig. S1). The quantitative metagenomics assay and metrics (Fig. 1) were developed using a “one system” approach. This means that all individual steps, from sample preparation to results reporting, are controlled by the SPC, thus eliminating the need for the fastidious steps required to control and quantify extracted DNA and sequencing libraries.

In a mNGS run, the detection limit of a SOI can be impacted by intrinsic factors including genome length and efficiency of its DNA extraction as well as extrinsic factors such as the accuracy of taxonomic classification and sample composition (host cell load, relative abundance of microorganisms, peculiarities of genome sequences for certain microorganisms) (19). To avoid reporting false-negative results, the detection of a SOI should only be reported as negative when $C_{min} \leq MT$ (Fig. 1). This was especially useful for the analysis of culture-negative samples in which, after patient’s DNA removal, remaining DNA quantities were significantly below those recommended for library preparation and sequencing. While modifications of the library preparation protocol allowed the sequencing of these samples, the number of reads remained very low. By detecting the SPC and calculating the Cmin for each SOI, we validated the ability of the test to detect an SOI at a concentration at least equal to the metagenomics threshold. Unfortunately, we were unable to validate 20.7% of negative SOI detections: in the four samples where SPC were not detected, we reported a probable infection by the detected SOI, but the 72 negative detections could not be validated. In the samples where SPC were detected, 75 negative SOI detections had an estimated Cmin > MT. High Cmin values resulted mainly from DT_{SOI} above 1,000 RPMB. To reduce DT_{SOI} , it may be

		mNGS detection			
		Above or equal to <i>MT</i>	No detection or below <i>MT</i>	No detection with no SPC detection or <i>Cmin_{SOI}</i> above <i>MT</i>	
Culture based detection	Above or equal to clinical threshold	True Positive (TP) 10	False negative (FN) 0	Not interpretable 0	Sensitivity $\frac{TP}{TP + FN}$ 100.0 %
	Negative or below clinical threshold	False positive (FP) 19	True negative (TN) 548	Not interpretable 143	Specificity $\frac{TN}{TN + FP}$ 96.6 %
		False discovery rate $\frac{FP}{FP + TP}$ 65.5 %	False omission rate $\frac{FN}{FN + TN}$ 0 %		

FIG 3 Comparison of mNGS to culture detections of 18 SOIs from all of the 40 samples from validation set.

necessary to decrease the level of false taxonomic classification of reads by improving the specificity of classification algorithms and by using a curated reference sequence database. The example of *E. cloacae* illustrates difficulties in interpreting results due to a high detection threshold (>1,000,000 RPMB) likely caused by a contaminated reference database and, as a consequence, a high rate of misclassifications.

It is interesting to note that despite the competition in the detection of SOI, our mNGS process allowed us to detect co-infections by two or three SOIs with concentrations ranging over three orders of magnitude. These co-infections were not detected by the routine cultures. Our results are consistent with previous observations that mNGS assays could be more effective in characterizing polymicrobial infections when compared to traditional culture-based methods (40).

Qualitative detection of microorganisms by mNGS can reflect resident microbiota, transient colonization, sample contamination, and/or infection. To differentiate the asymptomatic presence of bacteria from probable infection, absolute quantification has to be performed and compared to defined clinical decision thresholds (5, 41). For that purpose, we used the counts of reads assigned to SPC as a calibrator for the quantification of SOIs (Fig. 2). Using *S. aureus* as an example, the results of absolute quantification by mNGS (Fig. 2B) were comparable to those of qPCR (42).

Clinical microbiology laboratories have defined clinical decision thresholds for the HAP/VAP causative pathogen(s) in CFU/mL (mini-BAL: 10^3 CFU/mL and BAL: 10^4 CFU/mL)

(20). We did not find an obvious correlation between the number of genomes quantified by mNGS or qPCR and CFU counts from culture plates (Fig. 2A). This may have different causes such as a lack of precision of the culture report that provides the concentration of CFU at the nearest log₁₀ level. However, even when reducing the precision of mNGS quantification by rounding to the nearest log₁₀ value, the correlation between microbial culture and mNGS remains moderate and statistically insignificant (Fig. 2A). Other contributing factors such as the presence of viable but non-culturable (VBNC) (43, 44) cells in the sample, or the clustering of bacterial cells, can lead to an underestimation of the culture results relative to mNGS. Therefore, we defined the metagenomics threshold at $5.3E \times 10^3$ GEq/mL to differentiate, similar to the clinical decision thresholds, the asymptomatic presence of bacteria from infection (5, 41, 45–47).

Assessment of our mNGS process and defined metrics on the HAP/VAP panel showed good diagnostic capabilities (specificity: 96.6%; sensitivity: >99.9%), albeit with a “false discovery rate” of 65.5%. To avoid “false-positive” detections that may result from a lack of accuracy in the taxonomic classification, additional control methods should be considered. In this study, we confirmed correct sequence classification to the SOI for at least 26% of “false-positive” detections using 16S/MetaPhlAn2 markers detection by BLAST. For the other “false-positive” results, the number of reads was insufficient for the sequence assembly required for the BLAST-based 16S/MetaPhlAn2 markers search. Therefore, other controls should be considered such as the removal of reads that are stacked on a single location and share identity with the human genome or with commensal flora, as suggested by Uprety et al. (48). Nevertheless, in the presented mNGS workflow, no “false-positive” tests seemed to result from (k-mer based) misclassification of sequence reads, as they were not invalidated by finding 16S/MetaPhlAn2 markers from a different species. As the DNase I treatment of samples before the bacterial lysis step may remove extracellular DNA and genomic DNA from dead bacteria (6, 49), the new detections might reflect the presence of VBNC (6, 7) or antibiotic persister bacterial cells (44) within BAL samples. Currently, we have defined an MT that we apply to all bacterial species of our HAP/VAP panel, by analogy with cultures where a single clinical decision threshold is applied to all bacterial pathogens. It was suggested by Jahn et al. (41) that specific MT should be established for each bacterium depending on its pathogenicity. The presence of VBNC and antibiotic persisters could also be taken into account for the setting of specific MT. However, the evaluation of specific thresholds for each SOI would require large numbers of samples and clinical data, which were not available to us at the time of this study.

Conclusions and perspectives

We present a new clinical metagenomics workflow for the detection of the causative pathogen(s) of HAP/VAP (Fig. 1). It includes the use of an SPC to ensure that all steps, from sample preparation to data reporting, are performed properly. Detection of the SPC spiked at concentrations slightly above the MT allows to determine whether pathogens are present at or above the MT. SPC was especially useful for validating the analysis of culture-negative samples that yielded only a few reads. However, we could not validate 20.7% of individual negative SOI detections, mainly because of the competition effect by reads generated when other SOI(s) were detected at high concentrations in the same sample. This should have a limited impact on the diagnosis as it would not affect the detection and identification of the major pathogen causing the infection.

We have also demonstrated that SPC can be used as a calibrator in mNGS, allowing absolute quantification of *S. aureus* GEq as efficiently as quantitative PCR (Fig. 2B). This allowed us to define a metagenomics threshold to differentiate colonization from suspected infection. Our quantitative mNGS process showed good diagnostic capabilities (specificity: 96.6%; sensitivity: >99.9%). The “false discovery rate” of 65.5% could be due, at least in part, to a unique MT defined for all bacterial species instead of individual MT (41). However, some “false-positive” mNGS results might also reflect the presence of VBNC and antibiotic persister cells and their potential for infection recurrences (44).

Additional studies, including clinical assessments, will be necessary to evaluate the diagnostic values of such “false positives” or to set specific MT to each species.

Before implementing mNGS in routine clinical diagnosis, it is imperative to address several limitations. Since metagenomics can detect any organism present in a sample, it is technically challenging to validate metagenomics assays for compliance with regulatory guidelines, rules, and standards, considering the vast array of organisms potentially detectable by NGS. In this study, we have chosen not to report the results of out-of-panel detections for which no DT has been defined and no MetaPhlAn markers have been considered. To enable the clinical use of metagenomics, *in vitro* diagnostic regulation bodies need to anticipate the modalities for reporting positive results outside the validation panel, which could have clinical value.

Furthermore, in our experimental design, which relies on SPC-based pathogen quantification, we did not use typical negative controls obtained by omitting the addition of clinical samples in the mNGS workflow. Nonetheless, it is worth considering the use of controls consisting of human cells spiked with SPC as an additional validation measure in the mNGS process.

Additionally, variations in lysis efficiencies between SPC and SOIs may lead to distorted quantification of SOIs. These differences can be empirically tested for each SPC/SOI combination and taken into account to enhance the accuracy of SOI quantification. In the present study, we employed a combination of mechanical and chemical lysis, an approach that has recently demonstrated the ability to yield reasonably accurate estimates of bacterial absolute quantifications (50).

Other limitations of mNGS are the extended sample processing time, the necessity for skilled technicians and trained bioinformaticians, as well as the associated costs for NGS system installation, maintenance, and operation, which pose challenges for most routine diagnostic laboratories in terms of affordability. In our study, the mNGS time-to-result was approximately 50 hours, mostly due to the time required for Illumina MiSeq sequencing. While this timeframe is comparable to that of standard culture-based methods, it offers the advantage of detecting a wider range of bacterial species. Our SPC-based mNGS approach, one of the pioneering applications for clinical samples, streamlined the process related to the control and quantification of extracted DNA and sequencing library, resulting in a reduction of hands-on time by 4–5 hours. The use of an SPC is also compatible with the development of integrated platforms capable of performing all steps, from sample preparation to reporting, without additional intervention. Current sequencing technologies now deliver results more rapidly than those reported in our study, making it feasible to identify pathogens within a timeframe suitable for clinical practice and thus improving patient care.

ACKNOWLEDGMENTS

The authors would like to thank Alex van Belkum, Scientific Director of Microbiology Research at bioMérieux, for constructive comments on the manuscript.

This study was investigator-initiated and sponsored by Geneva University Hospitals. An unrestricted grant from bioMérieux (Marcy l'Étoile, France) was used to financially support this study, which was independently designed, executed, analyzed, and reported by the authors.

M.T., E.R., S.S., and G. Guigon developed the bioinformatics. M.T., V. Lazarevic, S.H., E.S.A., V. Lanet, M.G., C.M., G. Gervasi, and J.S. conceived and designed the study, and performed the analysis. M.T., S.H., E.S.A., and S.S. wrote the article and all the authors reviewed it.

AUTHOR AFFILIATIONS

¹bioMérieux, Grenoble, France

²Genomic Research Laboratory, Service of Infectious Diseases, Geneva University Hospitals, Geneva, Switzerland

³bioMérieux, Marcy-l'Étoile, France

⁴bioMérieux, La Balme-les-Grottes, France

AUTHOR ORCIDs

S. Hauser  <http://orcid.org/0000-0001-5076-2730>

AUTHOR CONTRIBUTIONS

S. Hauser, Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Writing – original draft | V. Lazarevic, Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Validation, Writing – original draft, Writing – review and editing | M. Tournoud, Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Writing – review and editing | E. Ruppé, Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Writing – original draft, Writing – review and editing | E. Santiago Allexant, Formal analysis, Methodology, Supervision, Validation, Writing – original draft, Writing – review and editing | G. Guigon, Project administration, Software, Supervision, Writing – review and editing | S. Schicklin, Data curation, Investigation, Software, Writing – review and editing | V. Lanet, Data curation, Formal analysis, Investigation, Methodology, Writing – review and editing | M. Girard, Formal analysis, Writing – review and editing | C. Mirande, Investigation, Supervision, Writing – review and editing | G. Gervasi, Investigation, Project administration, Supervision, Writing – review and editing | J. Schrenzel, Conceptualization, Supervision, Validation, Writing – original draft, Writing – review and editing

DATA AVAILABILITY

The data sets supporting the conclusions of this article are available in the NCBI Sequence Read Archive (SRA) repository under study number [PRJNA771365](https://www.ncbi.nlm.nih.gov/sra/PRJNA771365). The sequence reads were deposited after filtering out the read pairs that matched the human genome.

ETHICS APPROVAL

BAL samples were collected by the Geneva University Hospitals (Switzerland) for diagnostic purposes. Well-characterized left-over BAL samples were then used for research activities. Because of the difficulty of obtaining consent in the short time frame required for the right execution of the DNA extraction (no sample freezing possible) and because there was no interference with the normal care of the patient (no intervention, no collection of demographic or clinical data, no additional sample), an exemption of the necessity of informed consent was obtained in accordance with Art. 34 from the Swiss Federal Law on Research on Human Beings (LRH) and Art 37-40 from the Swiss Ordinance on research on humans with the exception of clinical trials (ORH).

After receiving the samples at the Genomic Research Laboratory of the Geneva University Hospitals, the DNA was promptly extracted, and any remaining materials were completely discarded.

ADDITIONAL FILES

The following material is available [online](#).

Supplemental Material

Supplemental material (Spectrum01294-23-s0001.docx). Additional method information with supporting figures and tables.

Open Peer Review

PEER REVIEW HISTORY (review-history.pdf). An accounting of the reviewer comments and feedback.

REFERENCES

1. Rea-Neto A, Youssef NCM, Tuche F, Brunkhorst F, Ranieri VM, Reinhart K, Sakr Y. 2008. Diagnosis of ventilator-associated pneumonia: a systematic review of the literature. *Crit Care* 12:R56. <https://doi.org/10.1186/cc6877>
2. IDSA/ATS. 2005. Guidelines for the management of adults with hospital-acquired, ventilator-associated, and healthcare-associated pneumonia. *Am J Respir Crit Care Med* 171:388–416. <https://doi.org/10.1164/rccm.200405-6445T>
3. Ruppé E, Baud D, Schicklin S, Guigon G, Schrenzel J. 2016. Clinical metagenomics for the management of hospital- and healthcare-acquired pneumonia. *Future Microbiol* 11:427–439. <https://doi.org/10.2217/fmb.15.144>
4. Lung M, Codina G. 2012. Molecular diagnosis in HAP/VAP. *Curr Opin Crit Care* 18:487–494. <https://doi.org/10.1097/MCC.0b013e3283577d37>
5. Gadsby NJ, Russell CD, McHugh MP, Mark H, Conway Morris A, Laurenson IF, Hill AT, Templeton KE. 2016. Comprehensive molecular testing for respiratory pathogens in community-acquired pneumonia. *Clin Infect Dis* 62:817–823. <https://doi.org/10.1093/cid/civ1214>
6. Pezzullo AA, Kelly PH, Nassar BS, Rutland CJ, Gansemer ND, Dohrn CL, Costello AJ, Stoltz DA, Zabner J. 2013. Abundant DNase I-sensitive bacterial DNA in healthy porcine lungs and its implications for the lung microbiome. *Appl Environ Microbiol* 79:5936–5941. <https://doi.org/10.1128/AEM.01752-13>
7. Potgieter M, Bester J, Kell DB, Pretorius E. 2015. The dormant blood microbiome in chronic, inflammatory diseases. *FEMS Microbiol Rev* 39:567–591. <https://doi.org/10.1093/femsre/fuv013>
8. Ruppé E, Greub G, Schrenzel J. 2017. Messages from the first International Conference on Clinical Metagenomics (ICCMg). *Microbes Infect* 19:223–228. <https://doi.org/10.1016/j.micinf.2017.01.005>
9. Chiu CY, Miller SA. 2019. Clinical metagenomics. *Nat Rev Genet* 20:341–355. <https://doi.org/10.1038/s41576-019-0113-7>
10. Rehm HL, Bale SJ, Bayrak-Toydemir P, Berg JS, Brown KK, Deignan JL, Friez MJ, Funke BH, Hegde MR, Lyon E, Working Group of the American College of Medical Genetics and Genomics Laboratory Quality Assurance Committee. 2013. ACMG clinical laboratory standards for next-generation sequencing. *Genet Med* 15:733–747. <https://doi.org/10.1038/gim.2013.92>
11. Gargis AS, Kalman L, Berry MW, Bick DP, Dimmock DP, Hambuch T, Lu F, Lyon E, Voelkerding KV, Zehnbauser BA, et al. 2012. Assuring the quality of next-generation sequencing in clinical laboratory practice. *Nat Biotechnol* 30:1033–1036. <https://doi.org/10.1038/nbt.2403>
12. FDA. 2016. Draft guidance for infectious disease next generation sequencing based diagnostic devices: microbial identification and detection of antimicrobial resistance and virulence markers
13. Tourlousse DM, Yoshiike S, Ohashi A, Matsukura S, Noda N, Sekiguchi Y. 2017. Synthetic spike-in standards for high-throughput 16S rRNA gene amplicon sequencing. *Nucleic Acids Res* 45:e23. <https://doi.org/10.1093/nar/gkw984>
14. Jones MB, Highlander SK, Anderson EL, Li W, Dayrit M, Klitgord N, Fabani MM, Seguritan V, Green J, Pride DT, Yooshep S, Biggs W, Nelson KE, Venter JC. 2015. Library preparation methodology can influence genomic and functional predictions in human microbiome research. *Proc Natl Acad Sci U S A* 112:14024–14029. <https://doi.org/10.1073/pnas.1519288112>
15. Kozyreva VK, Truong C-L, Greninger AL, Crandall J, Mukhopadhyay R, Chaturvedi V. 2017. Validation and implementation of clinical laboratory improvements act-compliant whole-genome sequencing in the public health microbiology laboratory. *J Clin Microbiol* 55:2502–2520. <https://doi.org/10.1128/JCM.00361-17>
16. Schlager R, Chiu CY, Miller S, Procop GW, Weinstock G. 2017. Validation of metagenomic next-generation sequencing tests for universal pathogen detection. *Arch Pathol Lab Med* 141:776–786. <https://doi.org/10.5858/arpa.2016-0539-RA>
17. Leo S, Gaia N, Ruppé E, Emonet S, Girard M, Lazarevic V, Schrenzel J. 2017. Detection of bacterial pathogens from broncho-alveolar lavage by next-generation sequencing. *Int J Mol Sci* 18:2011. <https://doi.org/10.3390/ijms18092011>
18. Lazarevic V, Gaia N, Emonet S, Girard M, Renzi G, Despres L, Wozniak H, Yugueros Marcos J, Veyrieras J-B, Chatellier S, van Belkum A, Pugin J, Schrenzel J. 2014. Challenges in the culture-independent analysis of oral and respiratory samples from intubated patients. *Front Cell Infect Microbiol* 4:65. <https://doi.org/10.3389/fcimb.2014.00065>
19. Nelson MT, Pope CE, Marsh RL, Wolter DJ, Weiss EJ, Hager KR, Vo AT, Brittnacher MJ, Radey MC, Hayden HS, Eng A, Miller SI, Borenstein E, Hoffman LR. 2019. Human and extracellular DNA depletion for metagenomic analysis of complex clinical infection samples yields optimized viable microbiome profiles. *Cell Rep* 26:2227–2240. <https://doi.org/10.1016/j.celrep.2019.01.091>
20. Chastre J, Fagon JY, Bornet-Lecro M, Calvat S, Dombret MC, al Khani R, Basset F, Gibert C. 1995. Evaluation of bronchoscopic techniques for the diagnosis of nosocomial pneumonia. *Am J Respir Crit Care Med* 152:231–240. <https://doi.org/10.1164/ajrccm.152.1.7599829>
21. Rotstein C, Evans G, Born A, Grossman R, Light RB, Magder S, McTaggart B, Weiss K, Zhanel GG. 2008. Clinical practice guidelines for hospital-acquired pneumonia and ventilator-associated pneumonia in adults. *Can J Infect Dis Med Microbiol* 19:19–53. <https://doi.org/10.1155/2008/593289>
22. Chastre J, Fagon J-Y. 2002. Ventilator-associated pneumonia. *Am J Respir Crit Care Med* 165:867–903. <https://doi.org/10.1164/ajrccm.165.7.2105078>
23. Kollef MH, Bock KR, Richards RD, Hearn ML. 1995. The safety and diagnostic accuracy of minibronchoalveolar lavage in patients with suspected ventilator-associated pneumonia. *Ann Intern Med* 122:743–748. <https://doi.org/10.7326/0003-4819-122-10-199505150-00002>
24. Vandeputte D, Kathagen G, D'hoel K, Vieira-Silva S, Valles-Colomer M, Sabino J, Wang J, Tito RY, De Commer L, Darzi Y, Vermeire S, Falony G, Raes J. 2017. Quantitative microbiome profiling links gut community variation to microbial load. *Nature* 551:507–511. <https://doi.org/10.1038/nature24460>
25. Stämmler F, Gläser J, Hiergeist A, Holler E, Weber D, Oefner PJ, Gessner A, Spang R. 2016. Adjusting microbiome profiles for differences in microbial load by spike-in bacteria. *Microbiome* 4:28. <https://doi.org/10.1186/s40168-016-0175-0>
26. Venkataraman A, Parlov M, Hu P, Schnell D, Wei X, Tiesman JP. 2018. Spike-in genomic DNA for validating performance of metagenomics workflows. *Biotechniques* 65:315–321. <https://doi.org/10.2144/btn-2018-0089>
27. Barigou M, Cavalie L, Daviller B, Dubois D, Mantion B, Delobel P, Debarb A, Prere M-F, Marchou B, Martin-Blondel G. 2015. Isolation on chocolate agar culture of *Legionella pneumophila* isolates from subcutaneous abscesses in an immunocompromised patient. *J Clin Microbiol* 53:3683–3685. <https://doi.org/10.1128/JCM.01116-15>
28. Parkinson NJ, Maslau S, Ferneyhough B, Zhang G, Gregory L, Buck D, Ragoussis J, Ponting CP, Fischer MD. 2012. Preparation of high-quality next-generation sequencing libraries from picogram quantities of target DNA. *Genome Res* 22:125–133. <https://doi.org/10.1101/gr.124016.111>
29. Adey A, Morrison HG, AsanXun X, Kitzman JO, Turner EH, Stackhouse B, MacKenzie AP, Caruccio NC, Zhang X, Shendure J. 2010. Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density *in vitro* transposition. *Genome Biol* 11:R119. <https://doi.org/10.1186/gb-2010-11-12-r119>
30. Jaillard M, Lima L, Tournoud M, Mahé P, van Belkum A, Lacroix V, Jacob L. 2018. A fast and agnostic method for bacterial genome-wide association studies: bridging the gap between k-mers and genetic events. *PLoS Genet* 14:e1007758. <https://doi.org/10.1371/journal.pgen.1007758>

31. Tournoud M, Ruppé E, Perrin G, Schicklin S, Guigon G, Mahé P, Lazarevic V, Hauser S, Mirande C, Levrat A, Louis K, Gervasi G, Schrenzel J. 2019. Clinical metagenomics bioinformatics pipeline for identification of hospital-acquired pneumonia pathogens and resistance determinants from bronchoalveolar lavage samples. *bioRxiv*. <https://doi.org/10.1101/2020.02.26.966309>
32. Wood DE, Salzberg SL. 2014. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol* 15:R46. <https://doi.org/10.1186/gb-2014-15-3-r46>
33. Davis JJ, Wattam AR, Aziz RK, Brettin T, Butler R, Butler RM, Chlenski P, Conrad N, Dickerman A, Dietrich EM, et al. 2020. The PATRIC bioinformatics resource center: expanding data and analysis capabilities. *Nucleic Acids Res* 48:D606–D612. <https://doi.org/10.1093/nar/gkz943>
34. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res* 44:6614–6624. <https://doi.org/10.1093/nar/gkw569>
35. Sichtig H, Minogue T, Yan Y, Stefan C, Hall A, Tallon L, Sadzewicz L, Nadendla S, Klimke W, Hatcher E, Shumway M, Aldea DL, Allen J, Koehler J, Slezak T, Lovell S, Schoepp R, Scherf U. 2019. FDA-ARGOS is a database with public quality-controlled reference genomes for diagnostic use and regulatory science. *Nat Commun* 10:3313. <https://doi.org/10.1038/s41467-019-11306-6>
36. Peng Y, Leung HCM, Yiu SM, Chin FYL. 2012. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* 28:1420–1428. <https://doi.org/10.1093/bioinformatics/bts174>
37. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol* 215:403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2)
38. Truong DT, Franzosa EA, Tickle TL, Scholz M, Weingart G, Pasoli E, Tett A, Huttenhower C, Segata N. 2015. MetaPhlAn2 for enhanced metagenomic taxonomic profiling. *Nat Methods* 12:902–903. <https://doi.org/10.1038/nmeth.3589>
39. Renzoni A, Francois P, Li D, Kelley WL, Lew DP, Vaudaux P, Schrenzel J. 2004. Modulation of fibronectin adhesins and other virulence factors in a teicoplanin-resistant derivative of methicillin-resistant *Staphylococcus aureus*. *Antimicrob Agents Chemother* 48:2958–2965. <https://doi.org/10.1128/AAC.48.8.2958-2965.2004>
40. Cummings LA, Kurosawa K, Hoogestraat DR, SenGupta DJ, Candra F, Doyle M, Thielges S, Land TA, Rosenthal CA, Hoffman NG, Salipante SJ, Cookson BT. 2016. Clinical next generation sequencing outperforms standard microbiological culture for characterizing polymicrobial samples. *Clin Chem* 62:1465–1473. <https://doi.org/10.1373/clinchem.2016.258806>
41. Jahn K, Kuisma M, Mäki M, Grendelmeier P, Hirsch HH, Tamm M, Papakonstantinou E, Stolz D. 2015. Molecular diagnostics for bacterial infections in bronchoalveolar lavage—a case-control, pilot study. *Swiss Med Wkly* 145:w14193. <https://doi.org/10.4414/smww.2015.14193>
42. Kralik P, Ricchi M. 2017. A basic guide to real time PCR in microbial diagnostics: definitions, parameters, and everything. *Front Microbiol* 8:108. <https://doi.org/10.3389/fmicb.2017.00108>
43. Li L, Mendis N, Trigui H, Oliver JD, Faucher SP. 2014. The importance of the viable but non-culturable state in human bacterial pathogens. *Front Microbiol* 5:258. <https://doi.org/10.3389/fmicb.2014.00258>
44. Ayrappetyan M, Williams T, Oliver JD. 2018. Relationship between the viable but nonculturable state and antibiotic persister cells. *J Bacteriol* 200:e00249–18. <https://doi.org/10.1128/JB.00249-18>
45. Albrich WC, Madhi SA, Adrian PV, van Niekerk N, Marelets T, Cutland C, Wong M, Khoosal M, Karstaedt A, Zhao P, Deatly A, Sidhu M, Jansen KU, Klugman KP. 2012. Use of a rapid test of pneumococcal colonization density to diagnose pneumococcal pneumonia. *Clin Infect Dis* 54:601–609. <https://doi.org/10.1093/cid/cir859>
46. Gadsby NJ, McHugh MP, Russell CD, Mark H, Conway Morris A, Laurenson IF, Hill AT, Templeton KE. 2015. Development of two real-time multiplex PCR assays for the detection and quantification of eight key bacterial pathogens in lower respiratory tract infections. *Clin Microbiol Infect* 21:788. <https://doi.org/10.1016/j.cmi.2015.05.004>
47. Johansson N, Kalin M, Tiveljung-Lindell A, Giske CG, Hedlund J. 2010. Etiology of community-acquired pneumonia: increased microbiological yield with new diagnostic methods. *Clin Infect Dis* 50:202–209. <https://doi.org/10.1086/648678>
48. Uprety P, Mitchell SL, Krock BL, Graf EH. 2018. The current state of metagenomics in infectious disease. *Adv Mol Pathol* 1:81–89. <https://doi.org/10.1016/j.yamp.2018.07.001>
49. Willis AL, Calton JB, Carr TF, Chiu AG, Chang EH. 2016. Dead or alive: deoxyribonuclease I sensitive bacteria and implications for the sinus microbiome. *Am J Rhinol Allergy* 30:94–98. <https://doi.org/10.2500/ajra.2016.30.4278>
50. Lazarevic V, Gaia N, Girard M, Mauffrey F, Ruppé E, Schrenzel J. 2022. Effect of bacterial DNA enrichment on detection and quantification of bacteria in an infected tissue model by metagenomic next-generation sequencing. *ISME Commun* 2:122. <https://doi.org/10.1038/s43705-022-00208-2>