

A Metagenomics Method for the Quantitative Detection of Pathogens Causing Ventilator-Associated Pneumonia

Sébastien HAUSER (✉ sebastien.hauser@biomerieux.com)

bioMérieux <https://orcid.org/0000-0001-5076-2730>

Vladimir Lazarevic

Genomic Research Laboratory

Maud Tournoud

bioMérieux: bioMérieux SA

Etienne Ruppé

AP-HP: Assistance Publique - Hopitaux de Paris

Emmanuelle Santiago Alexant

bioMérieux: bioMérieux SA

Ghislaine Guigon

bioMérieux: bioMérieux SA

Stephane Schicklin

bioMérieux SA

Veronique LANET

bioMérieux: bioMérieux SA

Myriam Girard

Genomic Research Laboratory

Caroline Mirande-Meunier

bioMérieux: bioMérieux SA

Gaspard Gervasi

bioMérieux

Jacques Schrenzel

Genomic Research Laboratory

Methodology

Keywords: Clinical metagenomics, Diagnosis, Standardization, Sample Processing Control, Spike-in bacteria, Calibrator, Absolute quantification, Metagenomics threshold, HAP/VAP.

Posted Date: November 9th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-1031465/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Microbiology Spectrum on October 27th, 2023. See the published version at <https://doi.org/10.1128/spectrum.01294-23>.

Abstract

Background

The management of ventilator-associated and hospital-acquired pneumonia requires rapid and accurate quantitative detection of the infecting pathogen(s). To do it, we propose a metagenomics next-generation sequencing (mNGS) assay that includes an internal sample processing control (SPC) for the quantitative detection of 20 relevant bacterial species of interest (SOI) from bronchoalveolar lavage (BAL) samples.

Results

To avoid very major errors in 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 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 $5.3E+3$ 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 100 % 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 these “false positive” detections were confirmed by finding specific 16S/MetaPhlAn2 markers, the 12 other (63 %) “false positive” detections did not yield enough reads to check their taxonomic classification.

Conclusions

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 manufacturer’s recommendations for NGS. The frequent “false positive” detection suggested the presence of non culturable cells within the tested BAL samples. Finally, mNGS detected mixed infections including bacterial species that were not reported by routine cultures.

Background

Hospital-acquired pneumonia (HAP) and ventilator-associated pneumonia (VAP) are common nosocomial infections causing high 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 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 of 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 here 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) that contains *Bacillus globigii* spores or FilmArray® (bioMérieux) that 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 1.0E+4 colony-forming units (CFU) per mL for bronchoalveolar lavage (BAL) samples and 1.0E+3 CFU/mL for mini-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 a SPC. Diagnostic capabilities of the mNGS workflow were assessed in BAL samples using a panel of bacterial species that are commonly causing HAP/VAP infection.

Methods

Selection of 20 pathogens for HAP/VAP diagnosis

We defined a panel of 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
Listing 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.

Following plating on solid media (blood agar, chocolate agar, Mac Conkey agar, Columbia colistin-nalidixic acid agar) and a 48 h incubation at 35°C in 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-TOF MS; MALDI Biotyper, Bruker Daltonics, Bremen, Germany) according to the manufacturer's instructions.

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 (SPC)

To control the whole mNGS process, we spiked the samples with the Gram-positive bacterium *Bacillus subtilis* as a SPC. The SPC sequences are detected by our pipeline with an accuracy of 99.95%, without cross-identification of the SOI (see Additional figure 1).

The *B. subtilis* strain ATCC 19659 BioBall® MultiShot 10E8 (bioMérieux, Catalog #416721) was re-hydrated in 1.1 mL of PBS to provide a solution of $1.0E+8$ CFU/mL \pm 20 %. After a 100-fold dilution in PBS, 10 μ L of the suspension were added to 600 μ L of the (mini-)BAL fluid. The final SPC concentration within a sample was $1.7E+4$ CFU/mL, close to the culture-based diagnostic threshold ($1.0E+4$ CFU/mL) for HAP/VAP.

Sample preparation and DNA extraction

Six hundred microliters of (mini-)BAL fluid supplemented with SPC were mixed with 6 mL of Tris-HCl 50 mM (pH 8) containing 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 during 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 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 minutes on a vortex with 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 x 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 [27, 28]. The following modifications were made 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, iii) the library was purified using 25 μ L of AMPure XP beads (Beckman Coulter). Indexed libraries prepared from 2 samples were pooled at equal quantities before sequencing on the MiSeq® platform (see Figure 1) with the MiSeq Reagent kit V3 (Illumina) following manufacturer's instructions.

Bioinformatics pipeline

For identification of pathogens, sequence reads were analyzed using the metagenomics pipeline described by Tournoud *et al.* [29, 30]. The detection of antibiotic resistance genes (ARG) 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 [31] 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 genome sequences were from both public (e.g. PATRIC [32], RefSeq [33], FDA ARGOS [34]) and private (strains sequenced from BAL samples) databases. The median number of genomes per species was 42 with interquartile-range equal to [19; 297] and the total number of genomes belonging to the flora was 7593. The performance of taxonomic classification of simulated sequence reads of SOI, *B. subtilis* and Human from the internal reference database is reported in Additional figure 1.

Pathogen detection and quantification relied on the number of reads attributed to each SOI. To avoid spurious pathogen detection due to erroneous taxonomic classification, reads associated to a SOI with an average genome coverage depth $> 1x$ were assembled. Assembly was performed using the idba_ud500 assembler (IDBA-UD 1.1.1) [35], with the following parameters: `-mink 40 -maxk 250 -min_pairs 2`. To confirm species identification, the assemblies were aligned with BLAST [36] against a pathogen marker database built by selecting clade-specific MetaPhlan2 [37] and 16S markers (for *H. alvei*, *P. vulgaris*, and *M. 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 is detected, 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 miss-classification of the reads. When no pathogen marker was detected, mainly because of low-coverage assemblies, no confirmation or invalidation of taxonomic classification to SOI were reported.

Sequence data analysis and interpretation

The complete mNGS workflow is presented in Figure 1. For each sample, all SOIs were analyzed separately as independent detection assays. Detailed procedure and calculations are provided in the Additional file. Briefly, data interpretation consisted of 4 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 to all bacterial species including *B. subtilis* ($R_{bacteria}$). 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 Additional file). 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 ($C_{SPC} = 1.7E+4$ CFU/mL) within each sample, it can be used as a calibrator.

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

$$C_{SOI} = \frac{RN_{SOI} \times 7.02E + 04}{RN_{SPC} \times L_{SOI}}$$

- When a SOI is not detected (Figure 1), its minimal detectable concentration ($C_{min_{SOI}}$) can be calculated if SPC has been detected:

$$C_{min_{SOI}} = \frac{DT_{SOI} \times 7.02E + 04}{RN_{SPC} \times L_{SOI}}$$

RN_{SOI} and RN_{SPC} correspond to the normalized quantity of sequence reads associated to SOI and SPC, respectively. L_{SOI} correspond to the genome size of SOI.

Step 3: Comparison to the metagenomics threshold

A metagenomics threshold (MT) was defined (see Additional file) as the concentration in genome equivalent (GEq) of SOI above which infection can be suspected. When a SOI is not detected (Figure 1), the comparison of MT to the calculated $C_{min_{SOI}}$ 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 detection of each SOI is reported separately. The rules of interpretation (see Figure 1 and Additional figure 2) define 4 possible reports:

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

PCR quantification of Staphylococcus aureus

Quantitative PCR targeting the *spa* gene was used to quantify *S. aureus*. Master mixes contained 0.2 µM forward primer 1914F (CAGCAAACCA TGCAGATGCT AA), 0.2 µM reverse primer 1992R (ACAGTTGTAC CGATGAATGG ATTTT) and 0.1 µM probe 1945T (AGCATTACCA GAAACT) [38] in the ABsolute™ QPCR Mix, no ROX (Thermo Scientific). One µL 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 genomic DNA of *S. aureus* strain MW2 at concentrations ranging between 1 pg and 1 ng and corresponding to 3.1E+2 to 3.1E+5 GEq.

Results

Absolute quantification of *S. aureus*

We compared the quantity of *S. aureus* obtained by our mNGS process 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).

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

ID	Sample type	Bacterial species	Quantification (log10)			16S/MetaPhlan2 markers
			Microbial culture	mNGS	qPCR (<i>S. aureus</i>)	
T01	mini-BAL	<i>S. aureus</i>	6	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	+

The quantities determined by mNGS (in GEq) and by microbial culture (in CFU) showed poor ($R^2 = 0.2929$) correlation (Figure 2A). We observed a better correlation ($R^2 = 0.9904$) when we compared quantification

results of mNGS and qPCR (Figure 2B). Moreover, for sample T07 that was culture-negative, both qPCR and mNGS detected and quantified *S. aureus* above 1.0E+4 GEq/mL.

Validation of quantitative detection of SOI

We assessed the complete mNGS workflow (Figure 1) with a “validation set” of 40 samples by comparing quantitative detection of SOI by culture and mNGS (Table 3).

Table 3 : Detection of SOI(s) above clinical threshold by microbial culture or above *MT* by mNGS in the “validation set”. Log10 of quantified pathogen concentrations in CFU/mL (for microbial culture) and GEq/mL (for mNGS) are 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* (Figure 1). (+) in 16S/MetaPhlan2 markers column means that taxonomic classification of sequence reads to a bacterial species is confirmed by 16S/MetaPhlan2 markers search.

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 cannot be validated
4	BAL					+	<i>Cmin</i> >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		+	<i>Cmin</i> >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 cannot be validated

8	mini-BAL						+	<i>Cmin>MT</i> 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	+		+	<i>Cmin > MT</i> for <i>E. coli</i> and <i>P. vulgaris</i>
10	BAL						+	
11	BAL	<i>L. pneumophila</i>	Presence	1,4			+	
12	BAL						+	
13	BAL	<i>S. aureus</i>	5	5,9	+		+	<i>Cmin>MT</i> for <i>P. vulgaris</i>
14	BAL						+	
15	BAL						+	
16	mini-BAL						+	
17	BAL						+	
18	BAL						+	
19	BAL	<i>S. pneumoniae</i>		4,1			+	<i>Cmin>MT</i> for <i>E. coli</i> and <i>P. vulgaris</i>
20	BAL	<i>P. aeruginosa</i>	5	5,6	+		+	<i>Cmin>MT</i> for <i>P. vulgaris</i>
21	mini-BAL						+	

22	BAL	<i>S. maltophilia</i>	5	7,3	+		+	<i>Cmin>MT</i> 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. morgani</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 cannot be validated
		<i>S. aureus</i>	2	Invalid				
24	BAL						+	
25	BAL						+	
26	BAL						+	
27	BAL	<i>C. freundii</i>		4,5			+	<i>Cmin>MT</i> for <i>A. baumannii</i> , <i>C. koseri</i> , <i>E. aerogenes</i> , <i>E. coli</i> , <i>H. influenzae</i> , <i>M. morgani</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> Cplx	> 5	7,1	+			
28	BAL	<i>S. marcescens</i>	2	4,2	+		+	
29	BAL	<i>E. aerogenes</i>	3	6,0	+		+	<i>Cmin>MT</i> for <i>E. coli</i> , <i>H. influenzae</i> , <i>K. pneumoniae</i> ,
		<i>S. aureus</i>	2	5,5	+			

		<i>K. oxytoca</i>		3,8			<i>P. vulgaris</i> and <i>S. pneumoniae</i> .
30	mini-BAL					+	
31	BAL	<i>H. influenzae</i>	> 5	4,8	+	+	
32	BAL					+	
33	BAL					+	
34	BAL	<i>S. marcescens</i>	3	4,8	+	+	<i>Cmin>MT</i> 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 cannot be validated
36	BAL					+	
37	BAL					+	<i>Cmin>MT</i> 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	+	+	<i>Cmin>MT</i> 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		+	

Control of the sample processing

The SPC detection failed ($RN_{SPC} < DT_{SPC}$) in 4 samples (10 %) in which at least one SOI was detected as probable infecting agent (samples 3, 7, 23 and 35 in Table 3). In these samples, non-detected SOIs are reported as “not interpretable” (Figure 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/ μ L DNA in the extract and generated only 4,850 reads. The detection of SPC and estimated $Cmin_{SOI}$ below MT allowed 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 detections because the calculated $Cmin_{SOI}$ exceeded the MT (Figure 1). It mainly concerned *E. coli*, *K. pneumoniae*, *P. vulgaris*, *S. aureus* and *S. pneumoniae* (Table 3) that had high DT_{SOI} values (See Additional figure 4).

Quantitative detection of the SOI panel

mNGS results were compared to the current “gold standard” for HAP/VAP diagnosis, i.e. the microbial culture (Table 4). Ten “true positive” detections of probable infection by SOIs and no “false negative” results (Table 3 and Table 4) revealed a test sensitivity of 100 %. With 19 “false positives” (Table 3 and Table 4), the test specificity was 96.8 % but the “false discovery rate” reached 65.5 %. Importantly, we confirmed proper taxonomic classification of 37 % (7/19) of the “false positive” detections by finding specific 16S/MetaPhlan2 markers. The 12 other “false positive” detections did not yield enough reads to check their taxonomic classification by 16S/MetaPhlan2 marker search.

Table 4

Comparison of mNGS to culture detections of all SOI(s) from all of the 40 samples from validation set.

		mNGS detection			
		Above or equal to <i>MT</i>	No detection or below <i>MT</i>	No detection with no SPC detection or $C_{min_{SOI}}$ 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 %
	Negative or below clinical threshold	False positive (FP) 19	True negative (TN) 584	Not interpretable 147	Specificity $\frac{TN}{TN+FP}$ 96.8 %
		False discovery rate $\frac{FP}{FP+TP}$ 65.5 %	False omission rate $\frac{FN}{FN+TN}$ 0 %		

Sample 27, which was culture positive for *E. cloacae* complex above 1.0E+5 CFU/mL, led to 7 “false positive” detections (Table 3) corresponding mostly to Enterobacterales. The presence of *K. pneumoniae*, detected at a concentration above 1.0E+6 CFU/mL, was confirmed by the 16S/MetaPhlan2 markers search. But the other “false positive” SOI detections were quantified at 10 to 100-fold lower concentrations relative to *K. pneumoniae* and likely had too few reads for 16S/MetaPhlan2 confirmation.

Detection of co-infections

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 3 orders of magnitude (Table 3). None of these mixed infections were reported by microbial cultures.

Discussion

Here we propose to spike BioBall® to control the processing of (mini-)BAL samples and to provide absolute quantification of detected SOI 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 human genome (see Additional figure 1). The quantitative metagenomics assay and metrics (Figure 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 $C_{min} \leq MT$ (Figure 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 sequencing of these samples, the number of reads remained very low. Thanks to the detection of SPC and calculation of C_{min} for each SOI, it was possible to validate the ability of the test to detect a SOI at a concentration at least equal to the metagenomics threshold. Unfortunately, we were unable to validate 21.1 % of negative detections: in the 4 samples where SPC were not detected, we reported a probable infection by the detected SOI but the 72 negative SOI detections could not be validated. In the samples where SPC were detected, 75 negative SOI detections had an estimated $C_{min} > MT$. High C_{min} values resulted mainly from DT_{SOI} above 1,000 RPMB. To reduce DT_{SOI} it may be 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 effect in detecting SOIs, our mNGS assay allowed 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 [39].

Qualitative detection of microorganisms by mNGS can reflect resident microbiota, transient colonization, sample contamination, and/or infection. To differentiate asymptomatic presence of bacteria from probable infection, the absolute quantity of pathogens has to be determined and compared to defined clinical decision thresholds [5, 40]. For that purpose, we used the counts of reads assigned to SPC as calibrator for quantification of SOIs (Figure 2). Using *S. aureus* as an example, the results of absolute quantification by mNGS (Figure 2B) were comparable to those of qPCR [41].

Clinical microbiology laboratories have defined clinical decision thresholds for the HAP/VAP causative pathogen(s) in CFU/mL (mini-BAL: 1.0E+3 CFU/mL and BAL: 1.0E+4 CFU/mL) [20]. We did not find obvious correlation between the number of genomes quantified by mNGS or qPCR and CFU counts from culture plates (Figure 2A). This may have different causes such as a lack of precision of the culture report that provides concentration of CFU at the nearest log level. A second cause could be the presence of

viable but non-culturable (VBNC) cells in the sample from which genomic DNA remained detectable [42, 43]. Therefore, we defined the metagenomics threshold (*MT*) at 5.3E+3 GEq/mL to differentiate, similarly to the clinical decision thresholds, asymptomatic presence of bacteria from infection [5, 40, 44–46].

Assessment of our mNGS process and defined metrics on the HAP/VAP panel showed good diagnostic capabilities (specificity: 96.8 %; sensitivity: 100 %), 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 quantity of reads was insufficient for sequence assembly required for 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 human genome or with commensal flora, as suggested by Uprety *et al.* [47]. 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 bacterial lysis step may remove extracellular DNA and genomic DNA from dead bacteria [6, 48], the new detections might reflect the presence of VBNC [6, 7] or antibiotic persister bacterial cells [43] within BAL samples. Currently, we have defined a *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.* [40] that specific *MT* should be established for each bacterium depending on its pathogenicity. 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 that were not available to us at the time of this study.

Conclusions

We present a new clinical metagenomics workflow for the detection of causative pathogen(s) of HAP/VAP (Figure 1). It includes the use of a SPC to control 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 few reads. However, we could not validate 21.1 % 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 (Figure 2B). This allowed us to define a metagenomics threshold to differentiate colonization from suspected infection. Our quantitative mNGS process showed good diagnostic capabilities (specificity: 96.8 %; sensitivity: 100 %). 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* [40]. However, some “false positive” mNGS results might also reflect the presence of VBNC and antibiotic persister cells, and their potential for infection recurrences [43]. 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 solve several limitations. One of them is a relatively long sample processing time that includes steps to control and quantify the extracted DNA and sequencing libraries. Our SPC allowed to skip these steps, reducing the hands-on time around 4–5 h. The use of a SPC is also compatible with the development of integrated platforms which can perform all steps, from sample preparation to the reporting, without any additional intervention.

Abbreviations

BAL: bronchoalveolar lavage; CFU: Colony forming unit; DT: Detection threshold; GEq: Genome equivalent; HAP: Hospital acquired pneumoniae; mNGS: Metagenomics next-generation sequencing; MT: Metagenomics threshold; RPMB: Reads per million of reads classified to bacteria; SOI: Species of interest; SPC: Sample processing control; VAP: Ventilator associated pneumonia; VBNC: Viable but non-culturable; WGS: Whole genome sequencing.

Declarations

- Ethics approval and consent to participate

(mini-)BAL samples were collected at the Geneva University Hospitals (Switzerland) for diagnostic purpose. The sample leftovers were then used in this research study. Because of the difficulty to obtain the consent in the short time frame required for the right execution of the DNA extraction (no sample freezing possible) and because there were 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 Swiss Ordinance on research on humans with the exception of clinical trials (ORH).

The DNA of the sample were extracted promptly after reception in the Genomic Research Laboratory of the Geneva University Hospitals. Remaining material were fully discarded.

- Consent for publication

Not applicable

- Availability of data and material

The datasets supporting the conclusions of this article are available in the NCBI Sequence read archive (SRA) repository under study number PRJNA771365 (<https://dataview.ncbi.nlm.nih.gov/object/PRJNA771365?reviewer=6noeoicp2ggcl2a2u312alrl2d>). The sequence reads were deposited after filtering out the read pairs that matched the human genome.

- Competing interests

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.

- Funding

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.

- Authors' contributions

MT, ER, SS and GGu developed the bioinformatics. MT, VLa, SH, ESA, VLaz, MG, CM, GGe, and JS conceived and designed the study, and performed the analysis. MT, SH, ESA and SS wrote the article and all the authors reviewed it.

- Acknowledgements

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

References

1. Rea-Neto A, Youssef NCM, Tuche F, Brunkhorst F, Ranieri VM, Reinhart K, et al. Diagnosis of ventilator-associated pneumonia: a systematic review of the literature. *Critical Care*. 2008;12:R56. doi: <https://dx.doi.org/10.1186/cc6877>.
2. IDSA/ATS. Guidelines for the management of adults with hospital-acquired, ventilator-associated, and healthcare-associated pneumonia. *Am J Respir Crit Care Med*. 2005;171:388-416. doi: <https://doi.org/10.1164/rccm.200405-644ST>.
3. Ruppe E, Baud D, Schicklin S, Guigon G, Schrenzel J. Clinical metagenomics for the management of hospital- and healthcare-acquired pneumonia. *Future Microbiol*. 2016. doi: <http://dx.doi.org/10.2217/fmb.15.144>.
4. Lung M, Codina G. Molecular diagnosis in HAP/VAP. *Current Opinion in Critical Care*. 2012;18:487-94. doi: <https://dx.doi.org/10.1097/MCC.0b013e3283577d37>.
5. Gadsby NJ, Russell CD, McHugh MP, Mark H, Conway Morris A, Laurenson IF, et al. Comprehensive Molecular Testing for Respiratory Pathogens in Community-Acquired Pneumonia. *Clin Infect Dis*.

- 2016;62:817-23. doi: <http://dx.doi.org/10.1093/cid/civ1214>.
6. Pezzulo AA, Kelly PH, Nassar BS, Rutland CJ, Gansemer ND, Dohrn CL, et al. Abundant DNase I-sensitive bacterial DNA in healthy porcine lungs and its implications for the lung microbiome. *Appl Environ Microbiol*. 2013;79:5936-41. doi: <https://dx.doi.org/10.1128/AEM.01752-13>.
 7. Potgieter M, Bester J, Kell DB, Pretorius E. The dormant blood microbiome in chronic, inflammatory diseases. *FEMS Microbiol Rev*. 2015;39:567-91. doi: <https://dx.doi.org/10.1093/femsre/fuv013>.
 8. Ruppe E, Greub G, Schrenzel J. Messages from the first International Conference on Clinical Metagenomics (ICCMg). *Microbes and infection*. 2017;19:223-8. doi: <https://dx.doi.org/10.1016/j.micinf.2017.01.005>.
 9. Chiu CY, Miller SA. Clinical metagenomics. *Nature Reviews Genetics*. 2019. doi: <https://doi.org/10.1038/s41576-019-0113-7>.
 10. Rehm HL, Bale SJ, Bayrak-Toydemir P, Berg JS, Brown KK, Deignan JL, et al. ACMG clinical laboratory standards for next-generation sequencing. *Genet Med*. 2013;15:733-47. doi: <https://doi.org/10.1038/gim.2013.92>.
 11. Gargis AS, Kalman L, Berry MW, Bick DP, Dimmock DP, Hambuch T, et al. Assuring the quality of next-generation sequencing in clinical laboratory practice. *Nat Biotechnol*. 2012;30:1033-6. doi: <https://doi.org/10.1038/nbt.2403>.
 12. FDA. Draft Guidance for Infectious Disease Next Generation Sequencing Based Diagnostic Devices: Microbial Identification and Detection of Antimicrobial Resistance and Virulence Markers. 2016.
 13. Turlousse DM, Yoshiike S, Ohashi A, Matsukura S, Noda N, Sekiguchi Y. Synthetic spike-in standards for high-throughput 16S rRNA gene amplicon sequencing. *Nucleic Acids Res*. 2017;45:e23. doi: <https://dx.doi.org/10.1093/nar/gkw984>.
 14. Jones MB, Highlander SK, Anderson EL, Li W, Dayrit M, Klitgord N, et al. Library preparation methodology can influence genomic and functional predictions in human microbiome research. *Proc Natl Acad Sci U S A*. 2015;112:14024-9. doi: <https://doi.org/10.1073/pnas.1519288112>.
 15. Kozyreva VK, Truong C-L, Greninger AL, Crandall J, Mukhopadhyay R, Chaturvedi V. Validation and Implementation of Clinical Laboratory Improvements Act-Compliant Whole-Genome Sequencing in the Public Health Microbiology Laboratory. *Journal of Clinical Microbiology*. 2017;55:2502-20. doi: <https://doi.org/10.1128/JCM.00361-17>.
 16. Schlaberg R, Chiu CY, Miller S, Procop GW, Weinstock G, Professional Practice C, et al. Validation of Metagenomic Next-Generation Sequencing Tests for Universal Pathogen Detection. *Arch Pathol Lab Med*. 2017;141:776-86. doi: <https://dx.doi.org/10.5858/arpa.2016-0539-RA>.
 17. Leo S, Gaïa N, Ruppé E, Emonet S, Girard M, Lazarevic V, et al. Detection of Bacterial Pathogens from Broncho-Alveolar Lavage by Next-Generation Sequencing. *Int J Mol Sci*. 2017;18:2011. doi: <https://doi.org/10.3390/ijms18092011>.
 18. Lazarevic V, Gaïa N, Emonet S, Girard M, Renzi G, Despres L, et al. Challenges in the culture-independent analysis of oral and respiratory samples from intubated patients. *Front Cell Infect Microbiol*. 2014;4. doi: <https://doi.org/10.3389/fcimb.2014.00065>.

19. Nelson MT, Pope CE, Marsh RL, Wolter DJ, Weiss EJ, Hager KR, et al. Human and Extracellular DNA Depletion for Metagenomic Analysis of Complex Clinical Infection Samples Yields Optimized Viable Microbiome Profiles. *Cell Reports*. 2019;26:2227-40.e5. doi: <https://doi.org/10.1016/j.celrep.2019.01.091>.
20. Chastre J, Fagon JY, Bornet-Lecso M, Calvat S, Dombret MC, al Khani R, et al. Evaluation of bronchoscopic techniques for the diagnosis of nosocomial pneumonia. *Am J Respir Crit Care Med*. 1995;152:231-40. doi: <https://doi.org/10.1164/ajrccm.152.1.7599829>.
21. Rotstein C, Evans G, Born A, Grossman R, Light RB, Magder S, et al. Clinical practice guidelines for hospital-acquired pneumonia and ventilator-associated pneumonia in adults. *Can J Infect Dis Med Microbiol*. 2008;19:19-53. doi: <https://doi.org/10.1155/2008/593289>.
22. Chastre J, Fagon JY. Ventilator-associated pneumonia. *Am J Respir Crit Care Med*. 2002;165:867-903. doi: <https://doi.org/10.1164/ajrccm.165.7.2105078>.
23. Kollef MH, Bock KR, Richards RD, Hearn ML. The Safety and Diagnostic Accuracy of Minibronchoalveolar Lavage in Patients with Suspected Ventilator-Associated Pneumonia. *Annals of Internal Medicine*. 1995;122:743-8. doi: <https://doi.org/10.7326/0003-4819-122-10-199505150-00002>.
24. Vandeputte D, Kathagen G, D'Hoe K, Vieira-Silva S, Valles-Colomer M, Sabino J, et al. Quantitative microbiome profiling links gut community variation to microbial load. *Nature*. 2017;551:507-11. doi: <https://doi.org/10.1038/nature24460>.
25. Stammler F, Glasner J, Hiergeist A, Holler E, Weber D, Oefner PJ, et al. Adjusting microbiome profiles for differences in microbial load by spike-in bacteria. *Microbiome*. 2016;4:28. doi: <https://doi.org/10.1186/s40168-016-0175-0>.
26. Venkataraman A, Parlov M, Hu P, Schnell D, Wei X, Tiesman JP. Spike-in genomic DNA for validating performance of metagenomics workflows. *Biotechniques*. 2018;65:315-21. doi: <https://doi.org/10.2144/btn-2018-0089>.
27. Parkinson NJ, Maslau S, Ferneyhough B, Zhang G, Gregory L, Buck D, et al. Preparation of high-quality next-generation sequencing libraries from picogram quantities of target DNA. *Genome Res*. 2012;22:125-33. doi: <https://doi.org/10.1101/gr.124016.111>.
28. Adey A, Morrison HG, Asan, Xun X, Kitzman JO, Turner EH, et al. Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition. *Genome Biol*. 2010;11:R119. doi: <https://doi.org/10.1186/gb-2010-11-12-r119>.
29. Tournoud M, Ruppé E, Perrin G, Schicklin S, Guigon G, Mahé P, et al. Clinical metagenomics bioinformatics pipeline for the identification of hospital-acquired pneumonia pathogens antibiotic resistance genes from bronchoalveolar lavage samples. *bioRxiv*. 2020:2020.02.26.966309. doi: <https://doi.org/10.1101/2020.02.26.966309>.
30. Jaillard M, Lima L, Tournoud M, Mahé P, van Belkum A, Lacroix V, et al. A fast and agnostic method for bacterial genome-wide association studies: Bridging the gap between k-mers and genetic events. *PLoS genetics*. 2018;14:e1007758. doi: <https://doi.org/10.1371/journal.pgen.1007758>.

31. Wood DE, Salzberg SL. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol.* 2014;15:R46. doi: <https://doi.org/10.1186/gb-2014-15-3-r46>.
32. Davis JJ, Wattam AR, Aziz RK, Brettin T, Butler R, Butler RM, et al. The PATRIC Bioinformatics Resource Center: expanding data and analysis capabilities. *Nucleic acids research.* 2020;48:D606-D12. doi: <https://doi.org/10.1093/nar/gkz943>.
33. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, et al. NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res.* 2016;44:6614-24. doi: <https://doi.org/10.1093/nar/gkw569>.
34. Sichtig H, Minogue T, Yan Y, Stefan C, Hall A, Tallon L, et al. FDA-ARGOS is a database with public quality-controlled reference genomes for diagnostic use and regulatory science. *Nature Communications.* 2019;10:3313. doi: <https://doi.org/10.1038/s41467-019-11306-6>.
35. Peng Y, Leung HC, Yiu SM, Chin FY. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics.* 2012;28:1420-8. doi: <https://dx.doi.org/10.1093/bioinformatics/bts174>.
36. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* 1990;215:403-10. doi: [https://doi.org/10.1016/s0022-2836\(05\)80360-2](https://doi.org/10.1016/s0022-2836(05)80360-2).
37. Truong DT, Franzosa EA, Tickle TL, Scholz M, Weingart G, Pasolli E, et al. MetaPhlan2 for enhanced metagenomic taxonomic profiling. *Nature Methods.* 2015;12:902. doi: <https://dx.doi.org/10.1038/nmeth.3589>.
38. Renzoni A, Francois P, Li D, Kelley WL, Lew DP, Vaudaux P, et al. Modulation of fibronectin adhesins and other virulence factors in a teicoplanin-resistant derivative of methicillin-resistant *Staphylococcus aureus*. *Antimicrob Agents Chemother.* 2004;48:2958-65. doi: <https://doi.org/10.1128/aac.48.8.2958-2965.2004>.
39. Cummings LA, Kurosawa K, Hoogestraat DR, SenGupta DJ, Candra F, Doyle M, et al. Clinical Next Generation Sequencing Outperforms Standard Microbiological Culture for Characterizing Polymicrobial Samples. *Clin Chem.* 2016;62:1465-73. doi: <https://doi.org/10.1373/clinchem.2016.258806>.
40. Jahn K, Kuisma M, Maki M, Grendelmeier P, Hirsch HH, Tamm M, et al. Molecular diagnostics for bacterial infections in bronchoalveolar lavage—a case-control, pilot study. *Swiss Med Wkly.* 2015;145:w14193. doi: <https://doi.org/10.4414/smw.2015.14193>.
41. Kralik P, Ricchi M. A Basic Guide to Real Time PCR in Microbial Diagnostics: Definitions, Parameters, and Everything. *Frontiers in microbiology.* 2017;8:108-. doi: <https://doi.org/10.3389/fmicb.2017.00108>.
42. Li L, Mendis N, Trigui H, Oliver JD, Faucher SP. The importance of the viable but non-culturable state in human bacterial pathogens. *Frontiers in microbiology.* 2014;5. doi: <https://doi.org/10.3389/fmicb.2014.00258>.
43. Ayrapetyan M, Williams T, Oliver JD. Relationship between the Viable but Nonculturable State and Antibiotic Persister Cells. *Journal of Bacteriology.* 2018;200:e00249-18. doi:

<https://doi.org/10.1128/jb.00249-18>.

44. Albrich WC, Madhi SA, Adrian PV, van Niekerk N, Mareletsi T, Cutland C, et al. Use of a rapid test of pneumococcal colonization density to diagnose pneumococcal pneumonia. *Clin Infect Dis*. 2012;54:601-9. doi: <https://doi.org/10.1093/cid/cir859>.
45. Gadsby NJ, McHugh MP, Russell CD, Mark H, Conway Morris A, Laurenson IF, et al. Development of two real-time multiplex PCR assays for the detection and quantification of eight key bacterial pathogens in lower respiratory tract infections. *Clinical Microbiology and Infection*. 2015;21:788.e1-e13. doi: <https://doi.org/10.1016/j.cmi.2015.05.004>.
46. Johansson N, Kalin M, Tiveljung-Lindell A, Giske CG, Hedlund J. Etiology of community-acquired pneumonia: increased microbiological yield with new diagnostic methods. *Clin Infect Dis*. 2010;50:202-9. doi: <https://doi.org/10.1086/648678>.
47. Uprety P, Mitchell SL, Krock BL, Graf EH. The Current State of Metagenomics in Infectious Disease. *Advances in Molecular Pathology*. 2018;1:81-9. doi: <https://doi.org/10.1016/j.yamp.2018.07.001>.
48. Willis AL, Calton JB, Carr TF, Chiu AG, Chang EH. Dead or Alive: Deoxyribonuclease I Sensitive Bacteria and Implications for the Sinus Microbiome. *American Journal of Rhinology & Allergy*. 2016;30:94-8. doi: <https://doi.org/10.2500/ajra.2016.30.4278>.

Figures

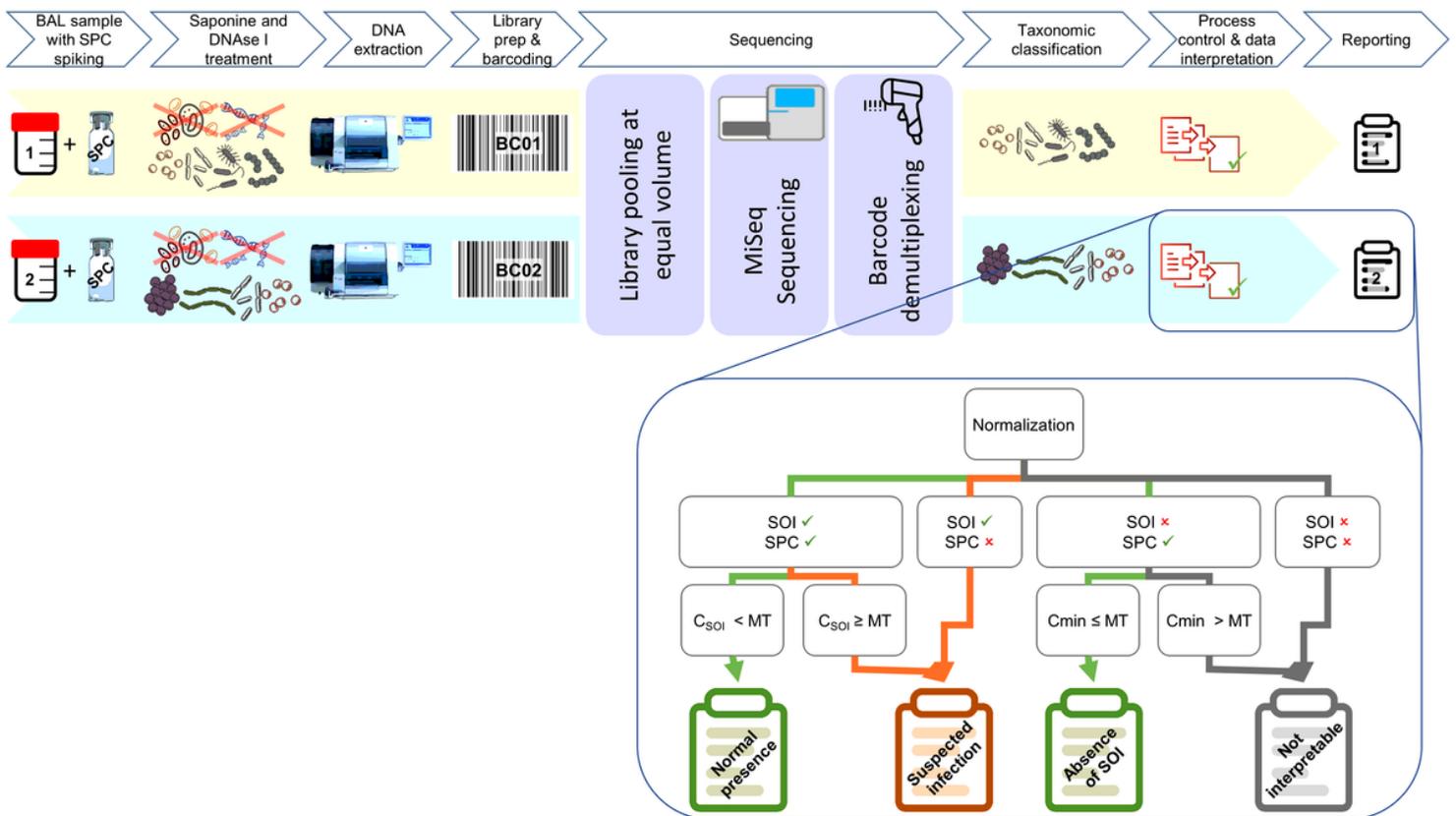


Figure 1

Complete workflow for quantitative detection of pathogen causing VAP by metagenomics.

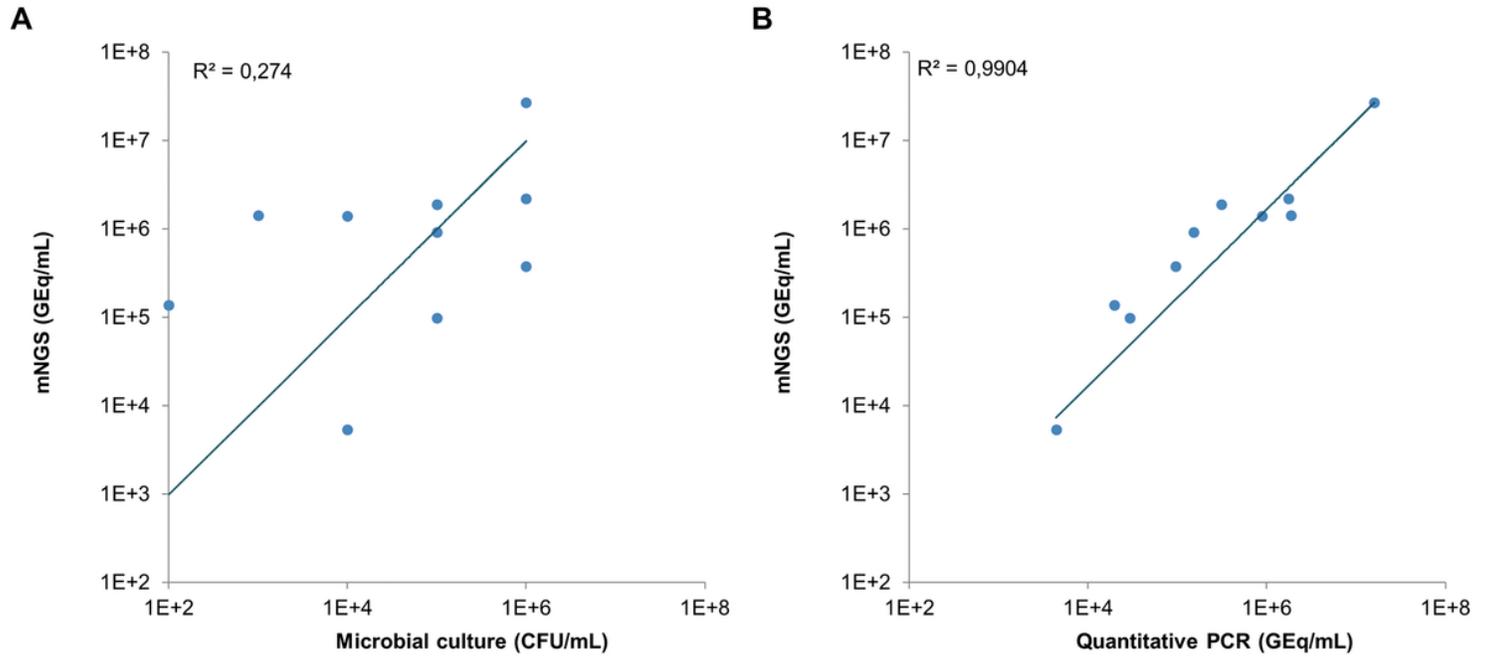


Figure 2

Comparison of *S. aureus* quantification. (A) Comparison of quantification in genome equivalent (GEq/mL) by mNGS to colony forming unit (CFU/mL) by microbial culture. (B) Comparison of quantification by mNGS to quantitative PCR, both in GEq/mL.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [AddInfomNGSquantdetectPathogensVAPV1.docx](#)