16s rRNA Lung Microbiota Study in Mechanically Ventilated Patient: A Pilot Study

Mélanie Fromentin
INSERM UMR 1137 IAME
Antoine Bridier-Nahmias
IAME UMR1137 IAME
Jérôme Legoff
INSERM U976 HIPI
Severine Mercier-Delarue
INSERM U 976 HIPI
Noémie Ranger
INSERM U 976 HIPI
Constance Vuillard
INSERM UMR 1137 IAME
Julien Dovale
INSERM UMR 1137 IAME
Noémie Zucman
Assistance Publique - Hopitaux de Paris
Antonio Alberdi
Université de Paris, IRSL
Jean-Damien Ricard
Assistance Publique - Hopitaux de Paris, hopital Louis Mourier, Intensive care unit
Damien Roux (damien.roux@aphp.fr)
Intensive Care Unit, Louis Mourier Hospital, 178 rue des renouillers 92700 Colombes, France https://orcid.org/0000-0002-6103-6416

Methodology

Keywords: Lung microbiome, 16S rRNA gene, high-throughput sequencing, dysbiosis, metagenomics, ventilator-associated pneumonia, mechanical ventilation

Posted Date: November 4th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-100088/v1

Abstract

Background

Characterization of the respiratory tract bacterial microbiota is in its infancy when compared to the gut microbiota knowledge. As key methodological steps can directly affect the accuracy of the results, it is crucial to determine a robust methodology in order to limit bias.

Two different pairs of primers 515F-806R targeting the V4 hypervariable region of the 16SrRNA gene, “S-V4” for “Stringent V4” primer pair and “R-V4” for “Relaxed V4” primer pair, are respectively used in two ongoing international projects “the Human microbiome project” and “the Earth microbiome project”.

We compared two methods of sample processing using these two different primer pairs for bacterial microbiota analyses of respiratory samples from critically-ill ventilated patients. For the later, Illumina 250 paired-end sequencing was done on a MiSeq platform after DNA extraction using mechanical lysis (bead-beating) and NucliSENS easyMAG. The concordance with conventional microbiology is the criterion of choice to determine the optimal method.

Results

Twenty samples from seven patients and four controls were sequenced. The two primer pair provided highly different results. Only 54% of the samples had a similar microbial composition with both pairs of primers. “S-V4” gave the best agreement with the conventional microbiology for ETA: 89% as compared to 44% for the “R-V4” primer pair. The main difference related to *Enterobacteriaceae*, which were concordant between conventional cultures and microbiota analyses using “S-V4”. *Enterobacteriaceae* detection was poor for “R-V4”. Among patients with VAP, a decrease in alpha diversity in ETA was observed. The mean of pairwise Unifrac distance was higher inside this group of patients at the time of VAP diagnosis as compared to control patients.

Conclusion

Accuracy of the bacterial lung microbiota composition was highly correlated to the pair of primers used for amplification of the 16s rRNA hypervariable sequence. Comparison of microbiota results obtained by sequencing and conventional microbiology allowed us to select the best option for further studies. This work validates our methodology based on 16SrRNA gene amplification with 515F-806R “S-V4” primer pair associated to Illumina® MiSeq 250 paired-end sequencing.

Background

Next-generation sequencing platforms have revolutionized our ability to investigate the microbiota composition and diversity of complex environments, including many human body sites. 16S rRNA gene sequencing is considered one of the two widest used methodologies for phylogenetic studies of bacterial communities. Since 2008, the National Institute of Health launched the Human Microbiome Project (HMP) to sequence and characterize the microbiome of healthy human subjects (http://commonfund.nih.gov/hmp) [1].

Despite remarkable progress in sequencing technologies that produce ever-larger data sets at ever-decreasing cost in metagenomics technology, no definite consensus exists on the reference methods to adopt to study lung microbiota.

Several critical factors, including sampling, DNA extraction [2], primer sequences, amplification batches [3], and sequencing platform [4] can affect the accuracy of the results [5]. Many other challenges relate to bioinformatic analyses and add another level of complexity. Thus, bias can be present at each step of the sampling, sequencing and bioinformatic analyses.

The search for the optimal extraction method has aroused particular interest, and comparison of DNA yield, DNA purity, and most importantly representation of microbial diversity has been largely done [2][6]. One of the most complex questions relates to the hypervariable (V) region(s) of the 16S rRNA gene to be amplified and sequenced, with an objective to identify the entire bacterial population and to accurately reflect the communities. Based on the results of many studies on fecal microbiota, researchers tend to favor the V4 hypervariable region. The International Human Microbiota Standards (IHMS) project aim to homogenize the methods in order to allow a greater comparability between studies [6]. However, the most informative hypervariable 16S rRNA region may differ from an environment to another, especially when high-throughput short-read sequencing technologies (for example, 454 and Illumina) are used [7]. The selection or design of the primer pair is also of critical importance [8].

Because of its diversity, the oral microbiome is a complex community to study, and the selection of which hypervariable region(s) of the 16S rRNA gene to amplify is crucial. Many attempts at identifying the most robust, unbiased and specific V region have been published and researchers mostly used long regions such as V1-V2 [9] V1-V3 [10], V3-V5 [11], or V7-V9 [10]. However, all but one studies were performed on the pioneer but no longer used 454 pyrosequencing [12] and only few studies compared specifics V regions after sequencing with Illumina platforms, which have now took the lead.[7]

As compared to the gut microbiome, knowledge on the lung microbiome is much poorer, especially in the field of intensive care. If lung and oral microbiota differ significantly during chronic pulmonary diseases [8], similarities are observed in healthy subjects [13]. Whether or not they differ in the context of acute pulmonary infection has not been addressed. To date, several unresolved methodological issues exist in the study of lung microbiota. Based on results from the two most recent methodological studies concerning mock DNA community or mock DNA extracted from the mock community cells [14] [5] it appeared that the V4 region could be the best region to amplify when high throughput sequencing is performed with Illumina Miseq plateform. If V4 region seems consensual, no methodological study addressed the question of the optimal pair of primers to choose for consistent and reliable results.

As a standardized approach for sequencing in situations where a comparison across multiple sequencing runs is required, we aimed to validate a robust methodology ahead of conducting longitudinal studies in ICU patients under mechanical ventilation at different time points.

In the present methodological pilot study, we wanted to compare two pairs of primers targeting the V4 region for 16SrRNA gene to assess oral and lung microbiome of infected and non-infected mechanically-ventilated patients.

Methods

Design of the study and ethical aspects

We designed a prospective observational study in the 12-bed ICU of Louis Mourier hospital (Assistance Publique - Hôpitaux de Paris, Colombes, France). All patients hospitalized in our ICU who required invasive mechanical ventilation were screened. Inclusion criteria were age over 18, intubation in our unit or within the six hours before admission, expected duration of mechanical ventilation above 72 hours with at least one of the following reasons for intubation: trauma with or without neurologic lesion, pulmonary sepsis, acute respiratory distress syndrome, and septic shock. Exclusion criteria were a known chronic lung disease (chronic obstructive pulmonary disease, emphysema, bronchiectasis, cystic fibrosis, lung transplant or a restrictive lung disease), an active lung tuberculosis, an invasive ventilation for more than 72 hours in the last six months, an immunosuppression (HIV with less than 200 CD4 lymphocytes/mm^3, neutropenia below 500/mm^3, treatment by immunosuppressive agents), and absence of consent to participate.

The investigator informed the patients or their surrogates about the study both orally and with a written document. In accordance with French law, written informed consent was not required. Patients or their surrogates were informed that they could decline to participate at any time, and their decision was recorded in patient files. The ethics committee of the French

Intensive Care Society approved the study (reference CE SRLF 15–41). Before analyzing all samples, we aimed at determining the best methodology to apply for lung microbiota analysis. We thus designed this preliminary methodological pilot study. Samples of a subset of the first included patients were selected. Control patients (C) were patients who were eventually ventilated no more than 3 days because of death or ventilation weaning. These patients were of no interest for the longitudinal study and were also selected for the methodological study. We also selected only three patients who developed a VAP (P) for this preliminary study, in order to keep the maximum of VAP patients included in the longitudinal study.

VAP definition

We used the modified CPIS (clinical pulmonary infection score) when a VAP was clinically suspected [15]. A score above six with a bacterial growth of a pathogen above 10^4 cfu/mL in a bronchoalveolar lavage (BAL) defined a VAP.

Samples

Two endotracheal aspirates (ETA) for studying lower respiratory tract (LRT) and one oropharyngeal swab (OS) for studying upper respiratory tract (URT) were sampled within the first six hours after intubation and every 72 hours afterward. One ETA and the OS were immediately stored at -80 °C. The second ETA was sent to the microbiology lab for standard culture. When a VAP was suspected, a BAL was performed in addition to the other samples and one vial was stored at -80 °C. The other vials were studied in the clinical microbiology laboratory of Louis Mourier hospital in order to detect the largest set of bacterial species. For conventional microbiology, ETA were diluted to 10^{-3} or to 10^{-5} and plated on four different media: aerobic blood agar, anaerobic blood agar, BY-chocolate agar, Drigalski agar.

For control patients, only one couple of ETA and OS on day 0 (D-0) was selected. For VAP patients, two couples of URT and LRT samples, one on a day between 3 or 6 days before the onset of the pneumonia (D-before) and one on the day of VAP diagnosis (D-VAP) were selected.

Demographic and clinical data

Clinical data including demographic data, underlying conditions, admission diagnosis in ICU, community-acquired pneumonia (CAP) and VAP diagnosis, type and duration of antibiotic administration, length of mechanical ventilation, ICU length of stay (ICU LOS), ICU mortality were extracted from the electronic medical file.

Positive controls strains (PC) and mock community

Five bacterial strains were selected to create a mock community. The strains were selected to represent a great diversity with five different genera typically involved in nosocomial pneumonia (Table 1).The different bacterial strains were grown independently in lysogeny broth (LB medium) with constant shaking at 220 rpm. Bacteria were grown in aerobic atmosphere in an INFORS HT Multitron during 18 h at 37 °C. They were then centrifuged and washed with saline solution. All cultures were individually diluted to obtain an optical density at 600 nm (OD_{600}) of 0.4 and then pooled. This pool constitutes a mock community that was used as a positive control (PC) in each batch of extraction.

Table 1
Bacterial strains cultivated independently and pooled to constitute bacterial positive control

Species	Strain	Gram-stain
Staphylococcus aureus	NCTC 8325	positive
Escherichia coli	P5.21 (B2 phylogenetic group, serogroup O6)	negative
Enterobacter cloacae	ESBL clinical strain	negative
Klebsiella pneumoniae	clinical strain	negative
Pseudomonas aeruginosa	PAO1	negative

All bacterial strains were cultivated independently in Broth medium LB (lysogeny broth) in an Infors aerobic chamber at a temperature of 37 °C. All cultures, after normalization to an OD (optical density) at 1.2, were pooled to constitute bacterial positive control. *K.pneumoniae* and *E. cloacae* are clinical strain which are chosen to have one bacteria of each class of *Enterobacteriaceae*. ESBL: Extended-Spectrum β-Lactamase.

Nucleic acid extraction

Samples were removed from − 80 °C storage and were put in 2 ml lysis matrix tube Y (MP Biomedicals®) for bead beating. For ETA only, 500 microliters of PBS EDTA free with Ca2 + and Mg2+ (Invitrogen®) were added to the lysis matrix tube. Samples were homogenized for 30 seconds two times at 6000 rpm using PlexIDbb instrument (Precellys Bertin technologies®). After centrifugation (one min at 8000 rpm), 400 microliters of supernatant were transferred in polypropylene tubes and incubated with 60 microliters of a proteinase K solution (20 microgram/ml) for 90 min at 56 °C. Bacterial DNA was extracted from each sample using the nucleic extraction platform based on magnetic silica technology NucliSENS® easyMAG® (Biomerieux®) allowing simultaneous extraction of 24 samples. For each nucleic acid extraction, negative controls (NC) made up with 600 µL of PBS EDTA free were used. Positive control (PC) (100 µL of the mock community completed to 600 µL with PBS) were also extracted according to the same protocol. Final DNA amount was quantified by Qubit fluorometric quantification® in 90 µL of eluate.

Specific 16S rRNA gene PCR amplification

Following bacterial DNA extraction, the V4 region of bacterial 16S rRNA gene was amplified using two Illumina MiSeq barcoded primer sets. Noteworthy, the V4 region has been chosen as it is discriminative enough to characterize microbial communities and permitted taxonomic assignment to genus or species-level in many prior studies [14].

These two PCR primer sets both named 515F-806R targeted V4 region of bacterial 16S rRNA and were used in different microbioma project. The first 515F-806R primer pair, called here “S-V4” primer pair for “Stringent-V4” primer pair is used in the gut microbioma project with the following specific sequences: 515F: 5’-GTG**C**CAGCMGCCGCGGTAA-3’; 806R: 5’-GGACTAC**H**VGGGTWTCTAAT- 3’ [16]. The second 515F-806R primer pair, called here “R-V4” primer pair for “Relaxed-V4” primer pair is used in protocol applied for sequencing microbial communities from host-associated and free-living environments as part of the Earth microbioma project with the following sequence: 515F: 5’-GTG**Y**CAGCMGCCGCGGTAA-3’; 806R: 5’-GGACTAC**N**VGGGTWTCTAAT- 3’[17][18]. These sequences differ only on two bases which were replaced by two ambiguous bases (in bold) in the “R-V4” primer pair to make them less stringent. First PCR reactions (PCR1) contained 12,5 µL KaPa HiFi HotStart ReadyMix, 5 µL forward primer (1 nanomol) (Sigma Aldrich, Dublin Ireland), 5 µL reverse primer (1 nanomol) (Sigma Aldrich, Dublin Ireland), 10 µL of template DNA (1 to 80 ng) and DNase RNase free distilled water (UltraPure™ Invitrogen®). They were realized on T100 thermal Cycler (BIORAD→). Variable quantity of extracted DNA was used in order to optimize bacterial amplification, because of the heterogeneity of the sample matrices.

Thermocycler conditions were as follows: heated lid 115 °C, 95 °C x 3 min, followed by 32 cycles of 95 °C x 30 s, 55 °C x 30 s, 72 °C x 30 s), followed by 72 °C x 5 min ad held at 4 °C. Thirty-two cycles was chosen as the best compromise between nonspecific amplification and the greatest number of samples successfully amplified.

PCR products were purified using a 0.9 volume of AMpure magnetic bead-based purification system (Beckman Coulter, Inc, Atlanta, Georgia) and eluted in low Tris-EDTA buffer. DNA yield were quantified in purified samples using Qubit® fluorometric quantification (Qubit dsDNA Hs Assay Invitrogen→). Successful specific amplification and purity were verified using DNA 1000 high sensitivity reagents in 4200 Tape Station Agilent bioanalyseur (Agilent Technologies®).

16S amplicon preparation for next generation sequencing

16SDNA amplicons were normalized for the second PCR as recommended, except in case of small amounts of DNA yielded from the first PCR. Second PCR reaction contained 5 µL of the first PCR products, 1 µL Acutaq polymerase, 5 µL Acutaq 10x buffer, 2 µL dNTP, 5 µL of each Illumina sequencing primer (Nextera XT Index Kit v2 Set D, 96 indexes) and 27 µL DNase RNase free distilled water (UltraPure™ Invitrogen®). Thermocycler conditions on C1000™ Thermal Cycler (BIO-RAD®) were as follows: heated lid 115 °C, 95 °C x 3 min, followed by 12 cycles of 95 °C x 10 s, 55 °C x 30 s, 72 °C x 30 s), followed by 72 °C x 5 min and held at 10 °C. DNA yield were quantified in purified samples using Qubit® fluorometric quantification (Qubit dsDNA Hs Assay Invitrogen→). Successful specific amplification and purity were verified using DNA 1000 high sensitivity reagents in 4200 Tape Station Agilent bioanalyseur (Agilent Technologies®). Samples were prepared for sequencing using standard protocols as recommended by Illumina. First samples were normalized to 2 nM and pooled in an equimolar concentration, then denatured using NaOH 0,2N to obtain a denatured library at 20pM. Libraries at 5pM were obtained combining library at 20pM and HT1 (hybridation buffer) and were mixed with Illumina generated PhiX control libraries (5% of 12pM PhiX solution). Eighty-four samples were loaded at 5pM and sequenced in a single Illumina sequencing run, using the indexing strategy and 250-bp paired-end reads (V2 500 cycle kit) permitting high quality coverage of the 370 bp V4 amplicon. Median read lengths were 251 bp from each end; paired reads were filtered to require minimum overlap of 130 bp.

Bioinformatics analysis

The quality of all the sequences was systematically verified (additional file 1). Briefly, sequences were processed using Mothur pipeline version 1.39.5 [19]. Paired-end sequencing reads of 16S rRNA were first merged and the resulting concatenates were then subsequently filtered and only those that met the following criteria were analyzed further : (a) sequence were at least 170 bp in length; (b) sequence were shorter than 265 bp; (c) had no ambiguous base; (d) had a maximum homopolymer length of 5. Chimeric sequences were filtered out using the chimera.vsearch function of mothur. Sequence alignment to the 16S database: silva.nr_v132 and then de novo operational taxonomic unit (OTU) clustering was performed on the basis of sequence similarity with a threshold of 3% using Mothur's opticlust algorithm.

Taxonomic assignment was also performed with mothur, up to the rank of family genus or species according a threshold identity of ≥ 80%. The resulting OTU table count was then filtered further by removing all OTU's represented by less than three couples of paired-end reads overall. Rarefaction was done by random subsampling with 5000 reads repeated 1000 times. All the script of the analysis is deposited on https://github.com/RespiratoryMicrobiomeMethods. All sequences are deposited in the European nucleotid archive under the study submission number XXXX *(it will be determined soon)*.

Data analysis

Statistical analysis was performed in R v3.6.3 with the help of the tidyverse packages [20][21] For each set of primers and for each samples, number of sequences initially obtained, number of sequences finally conserved for analysis and rarefaction curves were compared. Rarefaction curves revealed for each sample: (a) the $\log_{10}$ of the number of read generated versus the number of OTUs identified (b) the $\log_{10}$ of the number of read generated versus the Shannon's H' (Shannon diversity index). Patients samples were classified in three groups: control patients, VAP patients on day before VAP (D-before) and VAP patients on VAP day (D-VAP). For each patients LRT and URT are represented.

The relative abundance table ranked by order were represented by histograms, at the level of family for all VAP and control patients' samples, and at the level of genus for PC. In order to facilitate understanding and readability of the figures OTUs identified were presented at the family level, not at the genus level, and only families with relative abundances greater than

7% of the total are represented individually, those under 7% being grouped and labeled as ‘other’. This threshold was arbitrarily chosen to keep the figures readable. The full table of all recovered taxa is available as supplementary table (Additional file 2).

For all samples, results obtained with the two primer sets were compared to conventional microbiology. The concordance with conventional microbiology was the main criteria for primer pair validation and comparison and was retained if the pathogen(s) identified in culture was (were) identified by metagenomics. The concordance between the two primer pairs was also assessed and retained if the two most abundant OTUs were similar.

Microbial Alpha diversity was assessed and expressed as the Shannon diversity index (Shannon’s H’) for normalized numbers of sequences for each sample. The Shannon diversity index gives a measure of the species diversity in a given community while taking into account both the abundance and the evenness of species present in a community [22]. Beta diversity was given as weighted (and unweighted) Unifrac distance, and Bray-Curtis distance as these two are the most used indices and represent most features of beta diversity after being combined [23][24].

All qualitative values are expressed as percentages and all quantitative values are expressed as median and interquartile range.

Results

1. **Subject characteristics, samples collection and DNA extraction.**

Eight patients were initially included in this methodological pilot study. They represented a diverse set of underlying diseases and acute indications for mechanical ventilation. Five controls patients (C1 to C5) finally ventilated three days and three patients who developed a VAP were included (P1 to P3). One patient (C1) was finally excluded because of technical failure and the impossibility to perform specific 16S rRNA gene amplification. The patient’s characteristics are presented in the Table 2.

Table 2
Patient's characteristics

patient	age	sex	underlying condition	chronic respiratory disease	CAP	aspiration	VAP	days of VAP	length of MV	ICU LOS
C2	39	F	HCV/cirrhosis/ OH/smoking	COPD	yes	no	no	NA	3	3
C3	58	F	depressiveness	none	no	yes	no	NA	2	4
C4	55	F	HBP	none	no	no	no	NA	2	4
C5	77	M	HBP/PE/smoking	COPD	no	yes	no	NA	3	4
P1	73	M	HBP/pharynx cancer/smoking	COPD	yes	no	yes	10	16	16
P2	76	M	lung cancer	COPD	yes	no	yes	12	16	18
P3	56	M	OH/smoking	none	yes	no	yes	7	10	12

Age, sex, underlying condition, presence chronic respiratory disease, prescription antibiotics before admission, CAP or aspiration diagnosis, VAP diagnosis, length of MV, ICU LOS and vital status are described for each subject. LRTI diagnosed after 48 h of intubation s classified as VAP.

HBP: high blood pressure, HCV: hepatitis C virus, OH: alcohol, COPD: chronic obstructive pulmonary disease, PE: pulmonary embolism, MV: mechanical ventilation, LOS: length of stay, CAP: community-acquired pneumonia, VAP: ventilated-associated pneumonia

All patients but one had CAP or aspiration pneumonia on ICU admission. Median LOS was 4.5 days [4;13] and median duration of invasive mechanical ventilation was 4.5 days [2.7;11.5].

All VAP patients developed a late-onset VAP, after the sixth day of invasive mechanical ventilation (additional file 3). For each control patient, a couple of samples, one from URT and one from LRT within the first six hours of initiation of the mechanical ventilation, were studied. For each VAP patients two couples of samples, URT and LRT, on a day before VAP diagnosis (D-before), and on day of VAP diagnosis (D-VAP) were sequenced (Fig. 1).

Generally, ETA contained high amount of human host DNA, explaining why, despite high Qbit quantification results (between 10 ng/µL and 50 ng/µL), only a small amount of bacterial DNA could be amplified. In contrast, OS contained little amount of global extracted DNA (between 0,1 ng/µL and 10 ng/µL) which were sufficient for specific V4 amplification.

One positive control (PC) named PC1 and three negative control (NC) named NC1 to NC3 were also processed and sequenced. Overall, 24 samples were amplified and sequenced twice using both primer sets 515F-806R "S-V4" primer pair and "R-V4" primer pair.

2. Determining the optimal V4 primer pair between "S-V4" primer pair and the "R-V4" primer pair

Assessment of sequencing quality

For the "S-V4" primer pair and the "R-V4" primer pair the median number of bacterial reads used for taxonomic classification total were 121721 (29166 ; 150342) and 98983 (34324 ; 129888), respectively. Table 3 details the total number of sequences obtained, finally conserved for analysis and length of amplicon obtained per sample for each primer pair. The ratio of the median number of reads analyzed to the median number of total reads was higher for the "S-V4" primer pair 0.74 (0.38;0.85) than for the "R-V4" primer pair 0.64 (0.23;0.76). The median length of the amplicon obtained

with “S-V4” primer pair is closer to the expected length of 440 bp than with “R-V4” primer pair with 435 vs 428 bp respectively. Sequence quality was similar with a mean Phred score of 30 per sequence and respectively 152/168 and 109/168 sequences passing the filter without warning considering median Phred score per sequence and mean Phred score across each base position. Details are presented in additional file 1.

Table 3

Sequencing qualitative results and alpha diversity estimates for the two primer pair

	Initial reads (n)	Analyzed reads (n)	Ratio	Length	OTU (n)	Shannon's H'
« S-V4 »						
primer pair						
C2_URT_D0	162693	123237	0.76	433	124	0.58
C2_LRT_D0	178146	129929	0.73	435	35	0.02
C3_URT_D0	30983	25797	0.83	447	146	3.19
C3_LRT_D0	90694	5201	0.06	447	50	1.65
C4_URT_D0	165825	121959	0.74	436	230	2.75
C4_LRT_D0	127189	54106	0.43	444	276	3.40
C5_URT_D0	172577	146421	0.85	429	213	1.77
C5_LRT_D0	191893	162104	0.84	435	31	0.03
P1_URT_DBefore	205548	164260	0.80	435	165	1.33
P1_LRT_DBefore	187458	169036	0.90	435	194	1.88
P1_URT_DVAP	235152	207423	0.88	435	147	1.65
P1_LRT_DVAP	117827	18498	0.16	468	89	0.37
P2_URT_DBefore	134394	121484	0.90	426	132	2.60
P2_LRT_DBefore	117072	30289	0.26	441	108	3.10
P2_URT_DVAP	126424	92302	0.73	432	162	2.39
P2_LRT_DVAP	123010	71063	0.58	433	140	2.73
P3_URT_DBefore	205913	144602	0.70	446	328	3.79
P3_LRT_DBefore	174215	130433	0.75	433	186	1.64
P3_URT_DVAP	193855	168662	0.87	444	226	2.90
P3_LRT_DVAP	211316	183681	0.87	437	201	2.34
PC1	132833	110070	0.83	410	39	1.15
NC1	304	38	0.13	x	13	1.67
NC2	141	12	0.09	x	6	1.54
NC3	641	92	0.14	x	32	3.09
TOTAL	**148543.5**	**121721.5**	**0.74**	**435,0**	**143**	**1.82**
« R-V4 »						
primer pair						

For “S-V4” primer pair and “R-V4” primer pair qualitative sequencing parameters are represented by number of total reads observed (column 1), number of bacterial reads conserved for analysis (column 2), ratio between these two parameters (column 3) and median length of the amplicon obtained (column 4). The length of amplicons expected theoritecally is 440pb. For control patient (C) day before diagnosis of VAP “D-before” is always D-0 which means day of intubation. Alpha diversity estimates are represented by number of OTU identified and Shannon’s H’.

	Initial reads (n)	Analyzed reads (n)	Ratio	Length	OTU (n)	Shannon's H'
C2_URT_D0	161290	113741	0.71	427	147	0.54
C2_LRT_D0	145641	92276	0.63	429	46	0.02
C3_URT_D0	168197	132362	0.79	427	296	3.18
C3_LRT_D0	156063	9305	0.06	435	149	2.68
C4_URT_D0	183466	137794	0.75	430	268	2.67
C4_LRT_D0	127550	19164	0.15	471	212	2.89
C5_URT_D0	182096	142471	0.78	429	136	1.75
C5_LRT_D0	179734	141710	0.79	429	50	0.04
P1_URT_DBefore	172808	44636	0.26	425	196	1.89
P1_LRT_DBefore	175079	129063	0.74	432	261	1.70
P1_URT_DVAP	180591	60722	0.34	427	155	1.96
P1_LRT_DVAP	103098	4215	0.04	441	183	4.11
P2_URT_DBefore	126703	102250	0.81	428	118	2.53
P2_LRT_DBefore	145811	39377	0.27	425	150	3.03
P2_URT_DVAP	164020	95716	0.58	424	158	2.25
P2_LRT_DVAP	160447	86035	0.54	428	153	2.74
P3_URT_DBefore	175155	112845	0.64	428	406	3,88
P3_LRT_DBefore	149086	103028	0.69	425	259	1.27
P3_URT_DVAP	165471	135036	0.82	429	333	3.03
P3_LRT_DVAP	153986	127217	0.83	424	250	2.52
PC1	175524	17435	0.10	417	31	0.21
NC1	71176	8847	0.12	457	61	2.62
NC2	159018	15279	0.10	459	70	2.79
NC3	213501	145988	0.68	447	223	3.95
TOTAL	**162655**	**98983**	**0.64**	**428.5**	**156.5**	**2.57**

For “S-V4” primer pair and “R-V4” primer pair qualitative sequencing parameters are represented by number of total reads observed (column 1), number of bacterial reads conserved for analysis (column 2), ratio between these two parameters (column 3) and median length of the amplicon obtained (column 4). The length of amplicons expected theoritecally is 440pb. For control patient (C) day before diagnosis of VAP “D-before” is always D-0 which means day of intubation. Alpha diversity estimates are represented by number of OTU identified and Shannon’s H’.

Rarefaction curve for number of observed OTUs and Shannon index showed similar profile for the “S-V4” primer pair (Fig. 2A-B) and the “R-V4” primer pair (Fig. 2 C-D) for all samples. For both primer pairs, the majority of rarefaction curves based on the number of observed OTUs showed similar profiles and began to plateau, indicating sufficient sampling depth for each sample. For both primer pairs, rarefaction curve showing the Shannon’s H’ versus the decimal logarithm of the number of reads indicated precisely the threshold of 10^3 reads to reach the plateau, thus we select the rarefaction threshold of 5.10^3 for further analysis.

With the rarefaction threshold of 5.10^3, number of OTU and Shannon index identified in each sample are compared for the two primer pairs and are noticeably different (additional file 4). For “S-V4” primer pair no OTUs were identified in the NC.

Concordance between 16S sequencing and conventional microbiology

The main criteria was the concordance between metagenomics and conventional microbiology for PC and for lower respiratory tract samples (ETA).

For PC, relative abundance of OTUs greater than 1% of total number of sequences were represented at the taxonomic level of genus on the Fig. 3. With the “R-V4” primer pair, *Staphylococcus* was not identified and relative abundance of *Pseudomonas* reached 98%. With the “S-V4” primer pair all five bacteria were identified but their relative abundance were not strictly equal as *Staphylococcus* and *Pseudomonas* were less represented. For PC, “S-V4” primer pair provides the best agreement with conventional microbiology.

For the patient’s samples, concordance between 16S sequencing and conventional bacterial microbiology was retained if the known bacterial species identified using conventional microbiology were also identified by metagenomics.

Of the ten lower respiratory tract samples, one was not cultured by conventional bacteriology and only one was sterile. The agreement between 16S sequencing and conventional microbiology was excellent with 89% (8/9) concordant samples for “S-V4” primer pair versus 44% (4/9) for “R-V4” primer pair. Major differences between the two primer pairs were due to the lack of identification of *Enterobacteriaceae* with the “R-V4” primer pair. For patient P1 on D-9 and for patient P2 on D-12 *Serratia Marcescens* and *Escherichia coli* were not identified with “R-V4” primer pair (Table 4). For P3, *Escherichia coli* was also not identified with both pair of primers. *Staphylococci* were also underrepresented with the “R-V4” primer pair.

Table 4. Concordance between metagenomics and conventional microbiology for LRT samples

patient	day of MV	culture (cfu/ml)	« S-V4 » primer pair	Concordance culture « S-V4 » primer pair	"R-V4" primer pair	Concordance culture "R-V4" primer pair
C2	0	*Streptococcus pneumoniae (10^6)*	***Streptococcus***	Yes	***Streptococcus***	Yes
C3	0	*Staphylococcus (10^6)*	*Streptococcus,* ***Staphylococcus***	Yes	*Streptococcus*	No
C4	0	normal respiratory flora	*Streptococcus*	Yes	*Streptococcus*	Yes
C5	0	*Haemophilus influenzae (10^6)*	***Haemophilus***	Yes	***Haemophilus***	Yes
P1	6	*Klebsiella pneumoniae (10^6) Streptococcus anginosus (10^6)* normal respiratory flora	*Veillonella,* ***Streptococcus, Enterobacteriacae_unclassified***	Yes	*Veillonella*	No
P1	9	*Serratia marcescens*	***Enterobacteriaceae_unclassified***	Yes	*Bacteroides*	No
P2	6	negative	*Prevotella, Enterobacteriaceae*	x	*Prevotella*	x
P2	12	*Klebsiella pneumoniae (10^4),* normal respiratory flora	*Neisseria,* ***Enterobacteriaceae***	Yes	*Neisseria*	No
P3	0	*E coli (4.10^5),* normal respiratory flora	*Prevotella*	No	*Prevotella*	No
P3	6	*Streptococcus alpha haemolitycus (10^5), E.coli (10^3)*	***Streptococcus***	Yes	***Streptococcus***	Yes

Results of identification by conventional microbiology are presented on the left. Results obtained by metagenomics and the concordance with conventional microbiology are represented for "S-V4" primer pair (column 2-3) and for "R-V4" primer pair (column 4-5). For metagenomics only the most abundant OTUs and the OTU corresponding to the pathogen identified in conventional microbiology are represented. OTUs corresponding to the pathogen identified in conventional microbiology are highlighted using bold text. OTUSs corresponding to genus identified only in metagenomics are highlighted in a different font.

NA: no culture was obtained for the sample or when there was a single species; ESBL: extended spectrum betalactamase; MV: mechanical ventilation

Comparison for taxonomic composition and OTU relative abundance for all patients and respiratory tract site samples

Sequences were classified to the lowest taxonomic designation possible. The relative abundance of bacteria classified at family taxonomic level, is represented for control patients in Fig. 4 and for VAP patients in Fig. 5 for both pairs of primers. Agreement between the two primer pairs for all patients samples was retained if the two most abundant OTUs were similar.

For both primer pairs and considering all patients samples, three major phyla were found: 24% *Proteobacteria*, 43% *Firmicutes* and 24% Bacteroidetes. Metagenomics revealed nonpathogenic bacteria belonging to the genus of *Veillonella*, *Bacteroides*, *Prevotella* and *Neisseria* globally identified as “normal respiratory flora” or “mixed flora” or “oropharyngeal flora” in conventional microbiology (Table 4).

However, the composition of the microbiota was similar in only about half of the cases between the two primer pairs (additional file 5). For control patients, results were identical between the two primer pairs for C2, C4 and C5 but different for C3, especially for the lower respiratory tract due to the identification of *Staphylococcaceae* for C3 with “S-V4” primer pair. Generally, for VAP patient on D-before results with both primer pairs were similar except for P1 because an additional OTU was identified at family level as *Enterobacteriaceae* with “S-V4 primer pair”. For VAP patient on D-VAP, major differences appeared for P1 for the same reason. For P2, the differences were minimal and mainly due to the difference in relative abundance of each OTU, although the latter were identical in both cases.

3. Assessment of 16S rRNA gene profiling for all patient samples with “S-V4” primer pair

Assessing taxonomic composition and OTU relative abundance across respiratory tract site and time

As “S-V4” primer pair provides the best agreement with conventional culture for PC and also with conventional microbiology for patients samples, URT and LRT bacterial composition was compared only with this primer pair. A crucial point was that for each patients URT and LRT profiles were different.

For VAP patients, more OTUs were identified in the URT as compared to the LRT on D-before and on D-VAP. On D-VAP, a single OTU was largely dominant in the LRT: *Enterobacteriaceae* for P1, *Neisseriaceae* for P2, and *Streptococcaceae* for P3. These dominant OTUs represented the pathogen responsible for VAP for P1 and P3 and were also present in the URT. For P2 patient *Enterobacteriaceae* that represented the pathogen responsible for the VAP was identified but represented less than 7% of all OTUs. It was thus represented in the ‘other’ category on Fig. 4A. For P1 patient *Enterobacteriaceae* became predominant in URT and LRT between D-0 and D-VAP whereas for P3 patient *Streptococcaceae* was not present in both samples on D-before and appeared on D-VAP (Fig. 5A-B).

Overall, the pathogen identified by conventional microbiology was not always found to be the most abundant in the microbiome by metagenomics analysis.

Assessing bacterial alpha diversity across respiratory tract site and time

Alpha diversity was higher in URT as compared to LRT with an average Shannon of 2,5 [1,7; 2,7] and 1,8 [0,7; 2,8], respectively. For control patients at D-0, alpha diversity was higher in URT than in LRT except for patient C4. Indeed, all patients but C4 were admitted for community-acquired pneumonia (Fig. 6-A). Changes in diversity over time for VAP patients are represented in Fig. 6-B.

Alpha diversity slightly decreased over time for all VAP patients in LRT with an average Shannon of 2,5 [2;2,8] versus 2 [1,6;2,6]. This was not observed for URT.

Assessing weighted Unifrac distance across respiratory tract site and time

Pairwise Weighted Unifrac distance and Bray-Curtis distance were measured between all samples within each group. For VAP patients on D-VAP weighted Unifrac distance is higher in LRT than URT. This corresponds to the fact that microbial composition of each LRT samples becomes completely different on D-VAP and that the predominant OTU in LRT on D-VAP was different for each VAP patient (Fig. 7).

Considering the comparison between these three groups of samples weighted Unifrac distance was different only for LRT. Weighted Unifrac distance was higher for VAP patients on D-VAP. For the later, there were a raise in weighted Unifrac distance between D-before and D-VAP only in LRT. This means that in LRT a specific OTU for each patient become predominant between D-before and D-VAP. Results were similar using Bray-Curtis distance (additional file 6).

Discussion

While DNA sequencing continues to decrease in cost and more and more sequencing platforms become available, the number of studies investigating the microbiota of diverse environments have considerably increased. However, the difference in protocols and analysis between studies limits largely the comparison between publications. This heterogeneity in the methods is the main obstacle for generalizing conclusions and for eventually providing useful data for clinical practice. In addition, only few methodological studies were conducted on mock communities [5] [14] [2]. Methodological studies are thereby essential to determine the optimal method for each type of analysis, in order to define a consensual methodology.

Numerous factors influence results of 16S rRNA gene amplicon sequencing. Previous studies have independently examined one specific factor such as extraction procedure [25][26], the amplified region of the 16S rRNA gene or the choice of the sequencing platform [27]. Considerable differences occurred based on the sequencing platform: 454-pyrosequencing platforms, Illumina platforms or Ion Torrent platform [28]. Recently, Fouhy et al highlighted that differences occurred between data sets from two studies on a same mock DNA templates, likely due to a combination of factors including DNA extraction procedure, primer choice and sequencing plateform [5].

In contrast to the huge number of microbiome studies in the environment [29] or the digestive tract [6] [30], the pulmonary microbiota especially in the field of respiratory infections remain largely unexplored. Due to its low microbial density, some specificities exist concerning the methods and analysis for this particular microbiota. Indeed, bacterial density seemed to interfere with microbiota analyses at < than 10^6 bacteria per ml or DNA < 1 pg/ml [31].

Our preliminary work has revealed issues from targeting the V3-V4 hypervariable region using a 300-bp paired-end sequencing (V3 600 cycle kit) performed on Illumina Miseq plateform. Since V4 region provides one of the best accuracy in several studies [18] and in the most recent study of Fouhy et al [5], we decided to focus on this hypervariable region. The following step was to determine the primer pair to use. We thus compare two primer pairs based on the Human microbiome project and the Earth microbiome project (515F-806R primer pair).

This work is the first methodological study concerning pulmonary microbiota performed on both the URT and the LRT. Trying to identify the best methodology, we believed it was crucial to use different comparators or positive controls: an internal bacterial positive control (mock population) and conventional microbiology. Because many studies did not do so, it seems hard to interpret the results of many studies revealing bacterial genus or species which are usually not identified in respiratory samples [32][33]. It is thereby impossible to determine which methodology or primer set give the most reliable results. In our work, the results obtained using the “R-V4 primer pair” and the “S-V4 primer pair” were strikingly different, the later providing overall much reliable data. First, the median length of the amplicon obtained with “S-V4 primer pair” (435 bp) was closer to the expected length of 440 bp. Second, the “R-V4 primer pair” was not able to reveal *Staphylococcus* genus

with an important disequilibrium in favor of the *Pseudomonas* genus (almost 98% of the OTU identified while the three *Enterobacteriaceae* represented almost 2%). In contrast, the “S-V4 primer pair” allowed identification of all five bacterial species present in the mock community. These results confirmed that *Enterobacteriaceae* are somehow difficult to identify [14] This is a crucial point for lung microbiome study because they represent one of the main etiology of VAP.

When comparing sequencing results with conventional microbiology, the “S-V4 primers pair” also provided a satisfying agreement between the two methods (89%). Obviously, the sequencing allowed identification of many more bacteria than the conventional microbiology, such as *Prevotella*, *Neisseria* or *Fusobacteria*, which are less cultivable.

Considering negative control, results were also discordant. For “S-V4 primer pair” no OTUs were identified as expected whereas few OTUs corresponding to non-clinical bacteria were identified with the “R-V4 primer pair”. These OTUs could be real contaminants but more probably artificial identification due to errors of amplification as accuracy of the “S-V4 primer pair” is better for PC.

For each patients URT and LRT profiles were not similar whatever the primer pair. We believe this is of high importance as many teams working on pulmonary microbiota based their results only on the URT microbiota and did not explore, for practical reasons, the LRT microbiota. [34]. Our study highlighted that this concept seems false for ventilated patients as for patients with chronic pulmonary diseases.

Thereafter we focused our analyses using the “S-V4 primer pair”. Shannon diversity index I observed in our patients were concordant with previous studies (between 1.5 and 3 depending on the presence of a lung infection or not) [33][32][35]. We observed a higher alpha diversity in URT microbiota than in LRT microbiota for control patient. Interestingly, as previous studies, we found lower alpha diversity in LRT microbiota on D-VAP than in D-before, meaning that alpha diversity tended to decrease in LRT samples over time in case of VAP development [35]. Of course, our data on only three patients are insufficient to describe a typical evolution but, using our methodology, we found results similar to previous larger studies [35][36].

Finally, considering Weighted Unifrac distance, results were similar with both for VAP patient, with increased beta diversity from D-before to D-VAP. It can be explained by a more profound dysbiosis on D-VAP with the emergence of VAP bacterial pathogens with an advantage of growth, different for each patient. Zakharkina et al which found similar results, hypothesized that the bacteria responsible for VAP (*Staphylococcus*, *Acinetobacter* and *Pseudomonas*) would exclude other bacterial communities. [35]

This methodological study has limits. First, the small number of samples limits the conclusion we may draw from this study. As this study has for main objective to validate a global methodology for larger cohort study we decided to limit samples number included. We have insufficient number of patients and samples to apply statistical test and show significant differences between URT and LRT or between the beginning of mechanical ventilation and the day of VAP development. This prevented us from drawing conclusions based on the analysis of alpha and beta diversities. Second, for several OTUs, there is no identification until the species level. Obviously, 16SrRNA gene sequencing is not the most suitable technique for identification at the species level and our results are concordant to previous results. This limit depends of the database used and as SILVA database is updated we can hope to identified more OTUs at the species level in the close future. Third, we focus only on PCR primer choice and our study exclude the potential effects of others aspects of sample processing. As there is no previous methodological study concerning lung microbiota it would have been interesting to evaluate all the potential bias to found the optimal methodology. However, as our global methodology leads us to process all type of samples (BAL, ETA and OS) with concordant results with conventional microbiology, we conclude that it was adapted to conduct larger cohort study. In addition, preliminary data confirmed that we are able to study the virome and the lung mycobiota using the same extracted samples (unpublished data).

Conclusion

This is the first methodological study conducted on the LRT from ventilated patients in order to validate a suitable method for lung microbiota analysis. This study highlights the importance to evaluate the protocol and to confront metagenomics data with different positive controls such as a mock population and conventional microbiology.

Our methodology based on NucliSENS easyMAG total nucleic acid automated extraction, V4 16S rRNA gene amplification with the 515F-806R “S-V4 primer pair” and sequencing on Illumina® MiSeq plateform provided satisfying sequencing depth and identification of pathogens responsible for VAP at family and genus level.

List Of Abbreviations

ETA endotracheal aspirate

OS oropharyngeal swab

VAP ventilator associated pneumonia

HMP Human Microbiome Project

IHMS International Human Microbiota Standards

BAL Bronchoalveolar lavage

LRT lower respiratory tract

URT upper respiratory tract

CAP community acquired pneumonia

ICU LOS intensive care length of stay

PC positive control

NC negative control

Declarations

Ethics approval and consent for participate

All samples provide from patients initially included in a cohort study untitled “ Project Lung microbiota of ventilated ICU patients : towards a new understanding of nosocomial pneumonia”. The ethics committee of the French intensive care society approved the study (reference CE SRLF 15-41). Informed consent was obtained from patients’ next-of-kin and confirmed by the patients whenever possible.

Consent for publication

All patient (or next-of-kin by default) included were orally informed and consent obtained for participation and data publication.

Availability of data and material (data transparency)

All sequences are deposited in the European nucleotid archive under the study accession numbers XXXX *(will be done very soon)*. Analysis script has been uploaded on github.com. and is available at https://github.com/RespiratoryMicrobiomeMethods.

All supplementary experimental data are available in additional files linked to this article.

Conflicts of interest/Competing interests: DR received personal fees from Astellas, JDR received travel support by Fisher and Paykel. Other authors declare that they have no competing interests.

Funding: MF is the recipient of the 2017 MSD/Société de Réanimation de Langue Française Award and of the 2017 Young Investigator award from the Société de Pathologie Infectieuse de Langue Française. JDR is the recipient of the European Society of Intensive Care Medicine 2014 Established Investigator award. DR is the recipient of the European Society of Intensive Care Medicine 2016 Basic Science award.

Authors' contributions

MF, JDR and DR designed the study. MF performed all experiments, reviewed and appraised the literature regarding lung bacterial microbiota and respiratory samples processing. MF drafted the manuscript and approves the final version. ABN performed all bioinformatic analysis. JL designed and coordinated library preparation. SMD, NR and JD performed experiments. CV, NZ and DR designed the clinical sampling and included the patients. AA and DR participated in experiment design and sequencing. ABN, JDR and DR revised the article. All authors approved the submitted version.

Acknowledgments

We thank the Plateforme Technologique IRSL for hosting this work. We thank Pr. Harry Sokol for critical review of this manuscript.

References

[1] The Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. Nature 2012;486:207–14. https://doi.org/10.1038/nature11234.

[2] Yuan S, Cohen DB, Ravel J, Abdo Z, Forney LJ. Evaluation of Methods for the Extraction and Purification of DNA from the Human Microbiome. PLoS ONE 2012;7:e33865. https://doi.org/10.1371/journal.pone.0033865.

[3] Pinto AJ, Raskin L. PCR Biases Distort Bacterial and Archaeal Community Structure in Pyrosequencing Datasets. PLoS ONE 2012;7:e43093. https://doi.org/10.1371/journal.pone.0043093.

[4] Clooney AG, Fouhy F, Sleator RD, O' Driscoll A, Stanton C, Cotter PD, et al. Comparing Apples and Oranges?: Next Generation Sequencing and Its Impact on Microbiome Analysis. PLoS ONE 2016;11:e0148028. https://doi.org/10.1371/journal.pone.0148028.

[5] Fouhy F, Clooney AG, Stanton C, Claesson MJ, Cotter PD. 16S rRNA gene sequencing of mock microbial populations-impact of DNA extraction method, primer choice and sequencing platform. BMC Microbiol 2016;16:123. https://doi.org/10.1186/s12866-016-0738-z.

[6] Santiago A, Panda S, Mengels G, Martinez X, Azpiroz F, Dore J, et al. Processing faecal samples: a step forward for standards in microbial community analysis. BMC Microbiol 2014;14:112. https://doi.org/10.1186/1471-2180-14-112.

[7] Soergel DAW, Dey N, Knight R, Brenner SE. Selection of primers for optimal taxonomic classification of environmental 16S rRNA gene sequences. ISME J 2012;6:1440–4. https://doi.org/10.1038/ismej.2011.208.

[8] Dickson RP, Martinez FJ, Huffnagle GB. The role of the microbiome in exacerbations of chronic lung diseases. The Lancet 2014;384:691–702. https://doi.org/10.1016/S0140-6736(14)61136-3.

[9] Li J, Quinque D, Horz H-P, Li M, Rzhetskaya M, Raff JA, et al. Comparative analysis of the human saliva microbiome from different climate zones: Alaska, Germany, and Africa. BMC Microbiol 2014;14:316. https://doi.org/10.1186/s12866-014-0316-1.

[10] Kumar PS, Brooker MR, Dowd SE, Camerlengo T. Target Region Selection Is a Critical Determinant of Community Fingerprints Generated by 16S Pyrosequencing. PLoS ONE 2011;6:e20956. https://doi.org/10.1371/journal.pone.0020956.

[11] Huse SM, Ye Y, Zhou Y, Fodor AA. A Core Human Microbiome as Viewed through 16S rRNA Sequence Clusters. PLoS ONE 2012;7:e34242. https://doi.org/10.1371/journal.pone.0034242.

[12] Lazarevic V, Whiteson K, Huse S, Hernandez D, Farinelli L, Østerås M, et al. Metagenomic study of the oral microbiota by Illumina high-throughput sequencing. Journal of Microbiological Methods 2009;79:266–71. https://doi.org/10.1016/j.mimet.2009.09.012.

[13] Charlson ES, Bittinger K, Haas AR, Fitzgerald AS, Frank I, Yadav A, et al. Topographical Continuity of Bacterial Populations in the Healthy Human Respiratory Tract. Am J Respir Crit Care Med 2011;184:957–63. https://doi.org/10.1164/rccm.201104-0655OC.

[14] Barb JJ, Oler AJ, Kim H-S, Chalmers N, Wallen GR, Cashion A, et al. Development of an Analysis Pipeline Characterizing Multiple Hypervariable Regions of 16S rRNA Using Mock Samples. PLoS ONE 2016;11:e0148047. https://doi.org/10.1371/journal.pone.0148047.

[15] Luna CM, Blanzaco D, Niederman MS, Matarucco W, Baredes NC, Desmery P, et al. Resolution of ventilator-associated pneumonia: Prospective evaluation of the clinical pulmonary infection score as an early clinical predictor of outcome*: Critical Care Medicine 2003;31:676–82. https://doi.org/10.1097/01.CCM.0000055380.86458.1E.

[16] Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, et al. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proceedings of the National Academy of Sciences 2011;108:4516–22. https://doi.org/10.1073/pnas.1000080107.

[17] Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. Nat Methods 2010;7:335–6. https://doi.org/10.1038/nmeth.f.303.

[18] Walker AW, Martin JC, Scott P, Parkhill J, Flint HJ, Scott KP. 16S rRNA gene-based profiling of the human infant gut microbiota is strongly influenced by sample processing and PCR primer choice. Microbiome 2015;3:26. https://doi.org/10.1186/s40168-015-0087-4.

[19] Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, et al. Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. AEM 2009;75:7537–41. https://doi.org/10.1128/AEM.01541-09.

[20] R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. n.d.

[21] Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, et al. Welcome to the Tidyverse. JOSS 2019;4:1686. https://doi.org/10.21105/joss.01686.

[22] Spellerberg IF. Shannon–Wiener Index. Encyclopedia of Ecology, Elsevier; 2008, p. 3249–52. https://doi.org/10.1016/B978-008045405-4.00132-4.

[23] Lozupone C, Knight R. UniFrac: a New Phylogenetic Method for Comparing Microbial Communities. AEM 2005;71:8228–35. https://doi.org/10.1128/AEM.71.12.8228-8235.2005.

[24] Barwell LJ, Isaac NJB, Kunin WE. Measuring β - diversity with species abundance data. J Anim Ecol 2015;84:1112–22. https://doi.org/10.1111/1365-2656.12362.

[25] McOrist AL, Jackson M, Bird AR. A comparison of five methods for extraction of bacterial DNA from human faecal samples. Journal of Microbiological Methods 2002;50:131–9. https://doi.org/10.1016/S0167-7012(02)00018-0.

[26] Salonen A, Nikkilä J, Jalanka-Tuovinen J, Immonen O, Rajilić-Stojanović M, Kekkonen RA, et al. Comparative analysis of fecal DNA extraction methods with phylogenetic microarray: Effective recovery of bacterial and archaeal DNA using mechanical cell lysis. Journal of Microbiological Methods 2010;81:127–34. https://doi.org/10.1016/j.mimet.2010.02.007.

[27] Hahn A, Sanyal A, Perez GF, Colberg-Poley AM, Campos J, Rose MC, et al. Different next generation sequencing platforms produce different microbial profiles and diversity in cystic fibrosis sputum. Journal of Microbiological Methods 2016;130:95–9. https://doi.org/10.1016/j.mimet.2016.09.002.

[28] Salipante SJ, Kawashima T, Rosenthal C, Hoogestraat DR, Cummings LA, Sengupta DJ, et al. Performance Comparison of Illumina and Ion Torrent Next-Generation Sequencing Platforms for 16S rRNA-Based Bacterial Community Profiling. Appl Environ Microbiol 2014;80:7583–91. https://doi.org/10.1128/AEM.02206-14.

[29] Pershina EV, Ivanova EA, Korvigo IO, Chirak EL, Sergaliev NH, Abakumov EV, et al. Investigation of the core microbiome in main soil types from the East European plain. Science of The Total Environment 2018;631–632:1421–30. https://doi.org/10.1016/j.scitotenv.2018.03.136.

[30] Buelow E, Bello González T d. j., Fuentes S, de Steenhuijsen Piters WAA, Lahti L, Bayjanov JR, et al. Comparative gut microbiota and resistome profiling of intensive care patients receiving selective digestive tract decontamination and healthy subjects. Microbiome 2017;5:88. https://doi.org/10.1186/s40168-017-0309-z.

[31] Biesbroek G, Sanders EAM, Roeselers G, Wang X, Caspers MPM, Trzciński K, et al. Deep Sequencing Analyses of Low Density Microbial Communities: Working at the Boundary of Accurate Microbiota Detection. PLoS ONE 2012;7:e32942. https://doi.org/10.1371/journal.pone.0032942.

[32] Kelly BJ, Imai I, Bittinger K, Laughlin A, Fuchs BD, Bushman FD, et al. Composition and dynamics of the respiratory tract microbiome in intubated patients. Microbiome 2016;4. https://doi.org/10.1186/s40168-016-0151-8.

[33] Smith AD, Zhang Y, Barber RC, Minshall CT, Huebinger RM, Allen MS. Common Lung Microbiome Identified among Mechanically Ventilated Surgical Patients. PLOS ONE 2016;11:e0166313. https://doi.org/10.1371/journal.pone.0166313.

[34] Man WH, de Steenhuijsen Piters WAA, Bogaert D. The microbiota of the respiratory tract: gatekeeper to respiratory health. Nat Rev Microbiol 2017;15:259–70. https://doi.org/10.1038/nrmicro.2017.14.

[35] Zakharkina T, Martin-Loeches I, Matamoros S, Povoa P, Torres A, Kastelijn JB, et al. The dynamics of the pulmonary microbiome during mechanical ventilation in the intensive care unit and the association with occurrence of pneumonia. Thorax 2017;72:803–10. https://doi.org/10.1136/thoraxjnl-2016-209158.

[36] Emonet S, Lazarevic V, Leemann Refondini C, Gaïa N, Leo S, Girard M, et al. Identification of respiratory microbiota markers in ventilator-associated pneumonia. Intensive Care Medicine 2019;45:1082–92. https://doi.org/10.1007/s00134-019-05660-8.

Figures

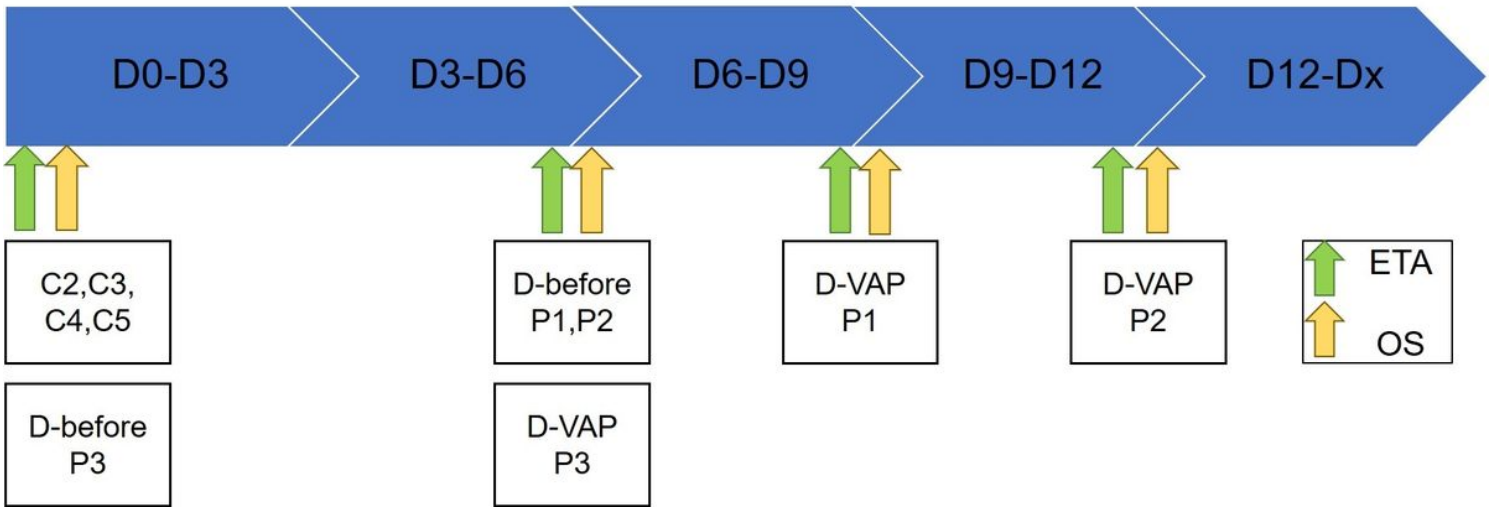

Figure 1

Samples collection D-before for day before VAP diagnosis, D-VAP for day of VAP diagnosis. A couple of sample represents an ETA (endotracheal aspirate) and a OS (oropharyngeal swab) sampled on the same day. The first control patient C1 was excluded because of a technical problem during the specific amplification procedure of the V4 region.

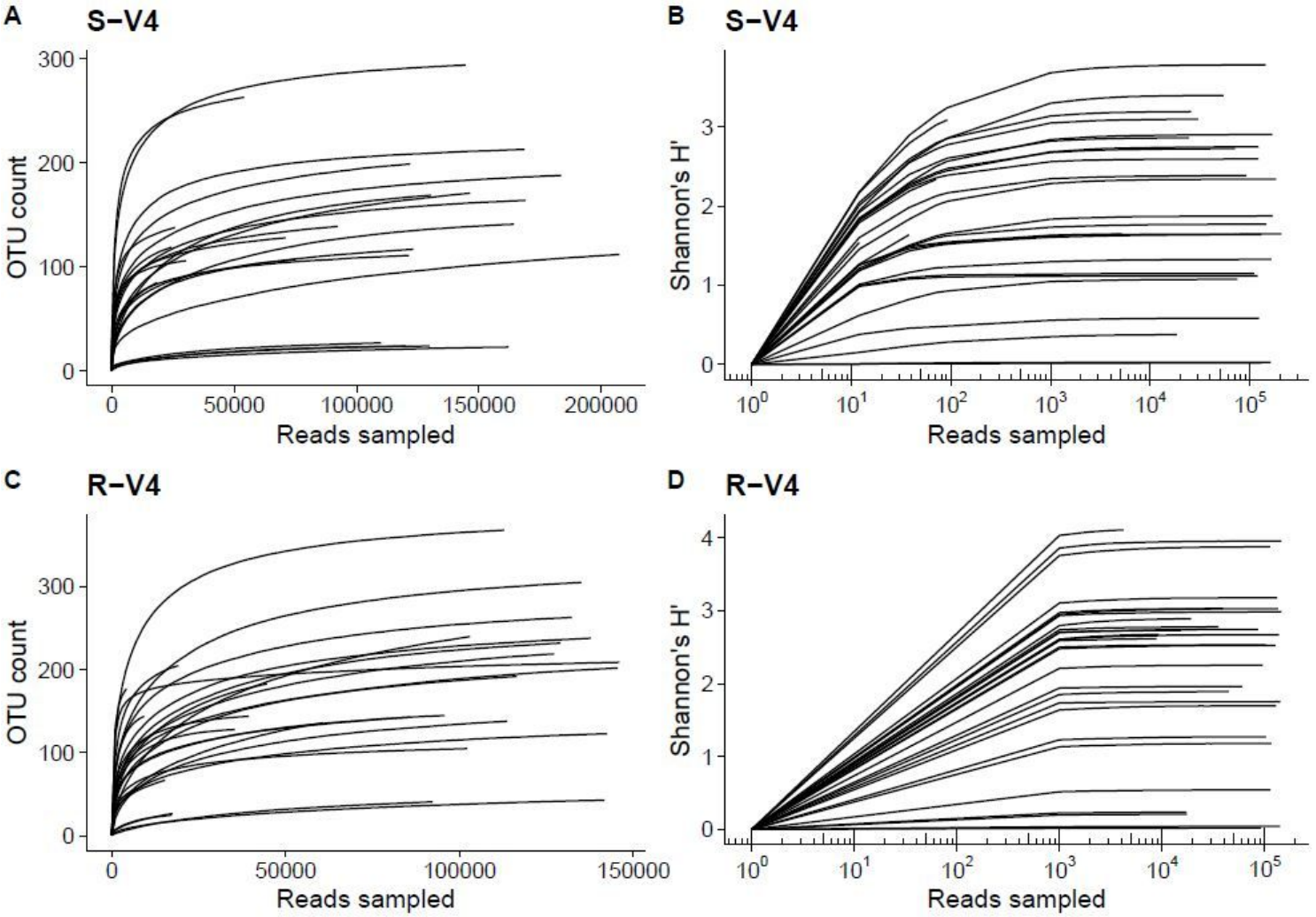

Figure 2

Rarefaction curves for “S-V4” primer pair Rarefaction curves based on the sample identity showing number of generated sequences versus the number of identified OTUs and number of sequences generated versus the Shannon’s H’ index for “S-V4” primer pair are represented Figure 1A and 1B for “S-V4” primer pair and Figure 1C and 1D for “R-V4” primer pair. Curves are approaching or are horizontal with the x axis indicating no need of additional sequencing because the maximal number of OUT are identified.

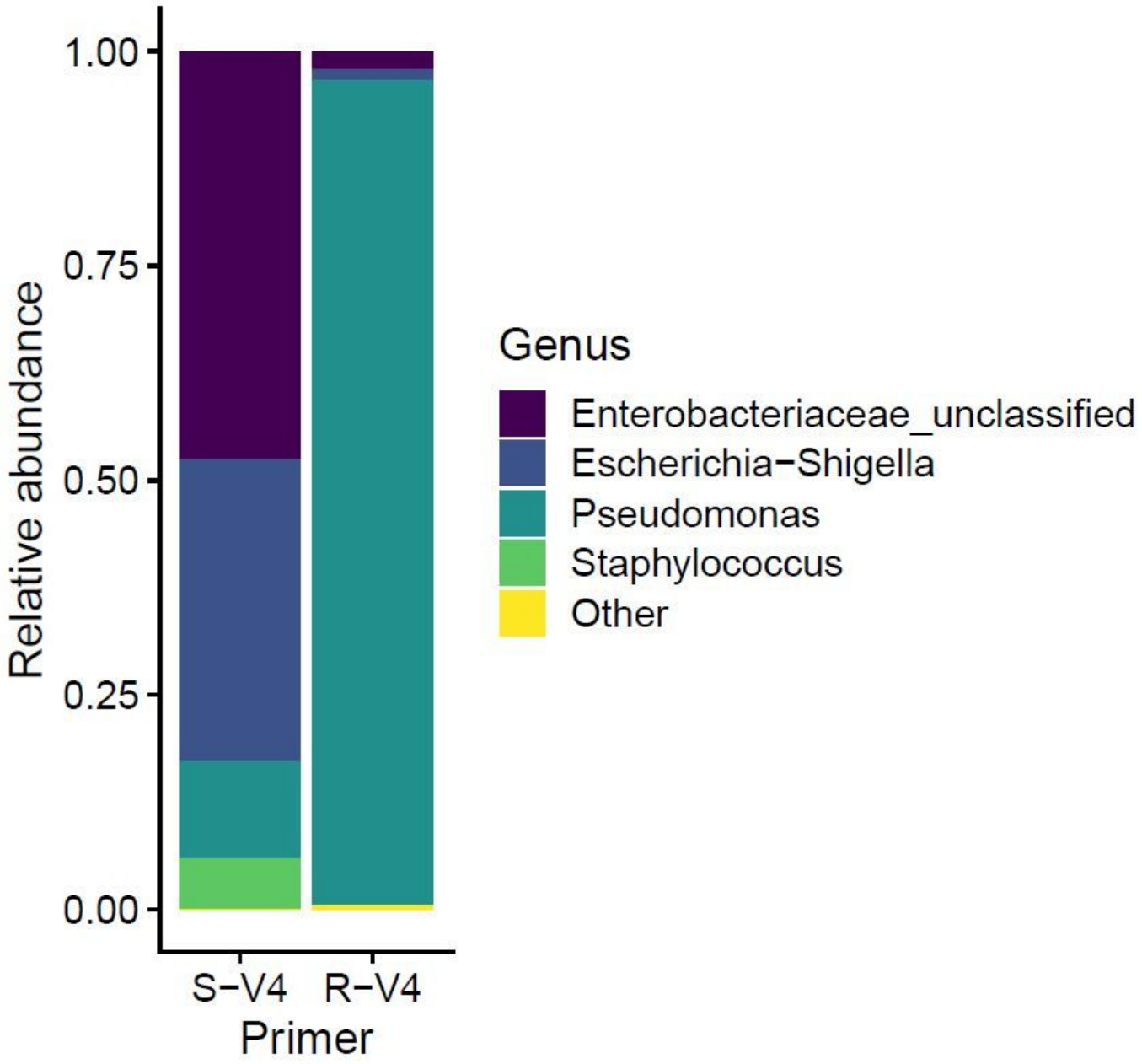

Figure 3

Percentage relative abundance of expected genus (n=5) detected in positive controls according to the V4 primer pair used Percentage of expected genus (n=5) detected in positive controls OTUs greater than 1% of total number of sequences were represented.

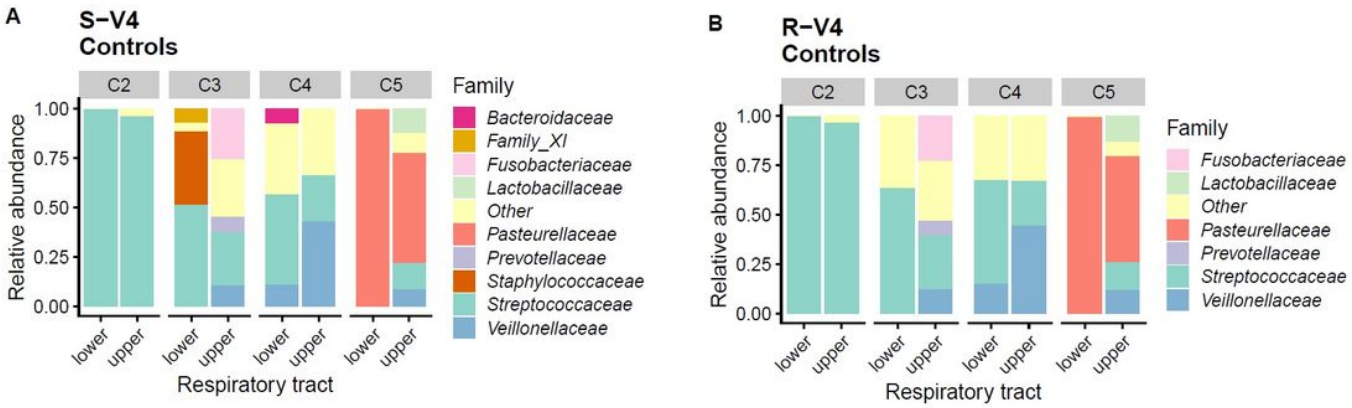

Figure 4

Relative abundance of control patients group samples according to the V4 primer pair used Family level relative abundance of all LRT (low respiratory tract infection) and URT (upper respiratory tract infection) samples for “S-V4” primer pair (figure 4A-B) and for “R-V4” primer pair Figure (figure 4C-D). The number of sequences of bacteria were converted to percentages of the total. Only taxa represented > 7 % of a sample were represented on the graph; any other < 7% is listed as other.

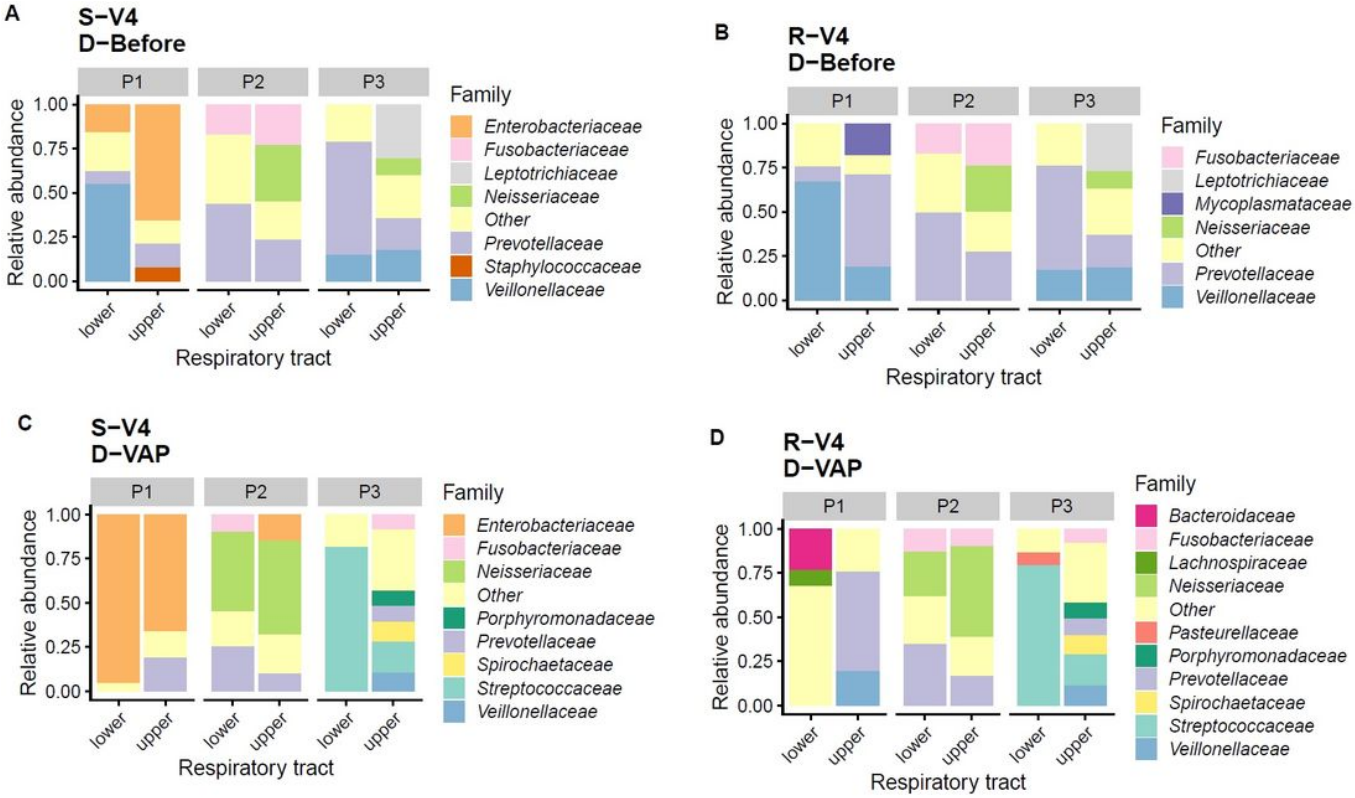

Figure 5

Relative abundance of VAP patients group samples according the primer pair used Family level relative abundance of all LRT (low respiratory tract infection) and URT (upper respiratory tract infection) samples for “S-V4” primer pair (figure 5A-B) and for “R-V4” primer pair Figure (figure 5C-D). Clinical sample are classified in VAP patients D-before and VAP patients D-VAP. The number of sequences of bacteria were converted to percentages of the total. Only taxa represented > 7 % of a sample were represented on the graph; any other < 7% is listed as other.

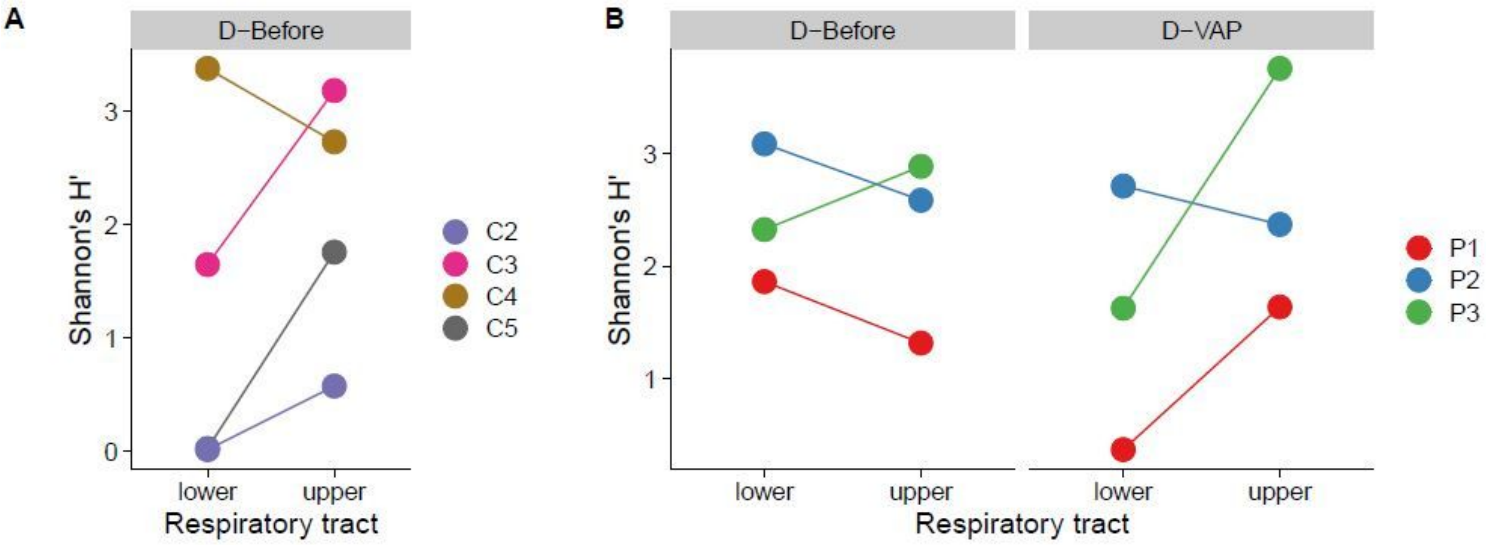

Figure 6

Alpha diversity estimates for upper respiratory tract (URT) and lower respiratory tract (LRT) for “S-V4” primer pair Alpha diversity is measured by the Shannon index (Shannon’s H’) for control patient (figure 6A) and for VAP patients over time (figure 6B). Comparison is made between URT and LRT for all groups and between D-Before and D-VAP for VAP patients.

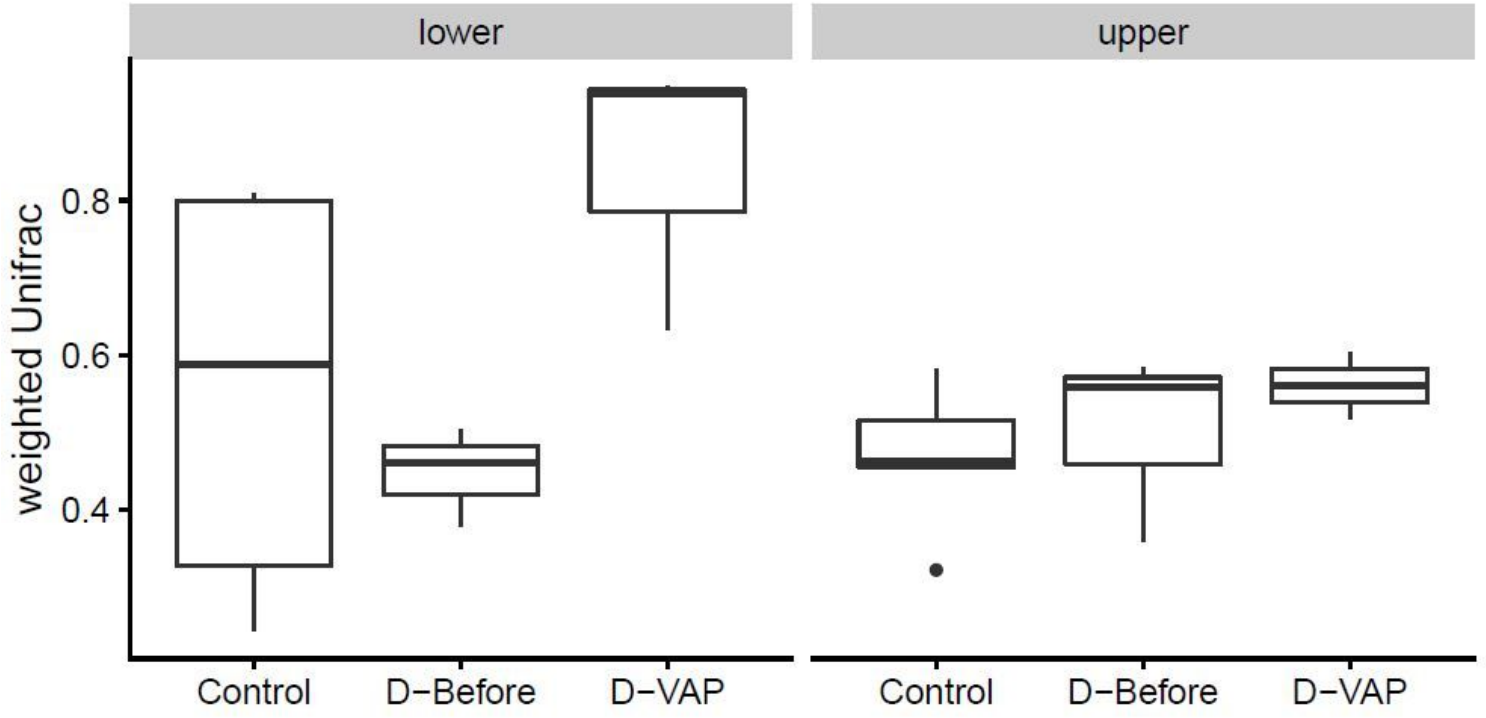

Figure 7

Pairwise Weighted Unifrac distance for “S-V4” primer pair across respiratory tract site and time Weighted UniFrac distance between all URT samples (left) and between LRT samples (right) within each group : control patient at D0, VAP patients at D-Before and VAP patients at D-VAP.

Supplementary Files

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

- Additionalfile1.docx
- Additionalfile2.xlsx
- Additionalfile3.docx
- Additionalfile4.pdf
- Additionalfile5.xlsx
- Additionalfile6.tif