TEMPORAL AIRWAY MICROBIOME CHANGES RELATED TO VENTILATOR ASSOCIATED PNEUMONIA IN CHILDREN

ONLINE DATA SUPPLEMENT

P.M. Mourani[1*], M.K. Sontag[2], K.M. Williamson[3], J.K. Harris[1], R. Reeder[4], C. Locandro[4], T.C. Carpenter[1], A.B. Maddux[1], K. Ziegler[2], E. A. F. Simões[1,2], C.M. Osborne[1], L. Ambroggio[1,2], M.K. Leroue[1], C.E. Robertson[5], C. Langelier[6,7], J.L. DeRisi[7,8] J. Kamm[7], M.W. Hall[9], A.F. Zuppa[10], J. Carcillo[11], K. Meert[12], A. Sapru[13], M.M. Pollack[14], P. McQuillen[15], D.A. Notterman[16], J.M. Dean[4], B.D. Wagner[3], Eunice Kennedy Shriver National Institute of Child Health and Human Development Collaborative Pediatric Critical Care Research Network (CPCCRN);

CONTENTS

Supplemental Tables

MATERIALS and METHODS

VAP Diagnosis

Subjects were screened daily to identify VAP defined by the pediatric 2008 Center for Disease Control (CDC) criteria (1). VAP diagnosis was assigned on the day all criteria were met within the CDC defined window. Chest radiographs obtained for clinical purposes during the study period and the accompanying radiologist interpretations were reviewed by the site investigators to determine whether the radiographic criteria were met. Presence of fever and white blood cell count were obtained from the electronic medical record. Other criteria were obtained by daily surveys of bedside nurses. Physicians (attending or fellow) were also surveyed daily to determine whether they initiated antibiotics for suspected or physician diagnosed VAP. They were asked: “Is the subject receiving antibiotics today for a suspected or diagnosed hospital acquired lower respiratory tract infection?” If the question was answered in the affirmative, they were then asked whether the antibiotics were being administered for “rule out of infection” (suspected VAP) or for a “dedicated treatment course” (physician diagnosed VAP).

Data Collection

Prospectively collected clinical data were recorded in a web-based research database maintained by the Collaborative Pediatric Critical Care Research Network (CPCRRN) data coordinating center at the University of Utah. Clinical data included demographics, primary diagnosis, chronic illnesses/disabilities, baseline functional status score (FSS) (2), Pediatric Risk of Mortality (PRISM) III score (3), antibiotic administration, physician suspected and diagnosed VAP, and data elements utilized to define VAP based on the Center for Disease Control (CDC) pediatric definition (1). Outcome measures included use of extracorporeal membrane oxygenation (ECMO) support, length of MV and other respiratory support, PICU and hospital lengths of stay, PICU discharge FSS, and hospital survival status.

Antibiotics

Antibiotic data were collected for each patient and included the drug name, administration route, and start and stop date. These data were summarized in two ways, 1) the number of drugs given on each day of intubation and 2) total coverage score that was based on whether the antibiotic included coverage against gram-positive bacteria, gram-negative bacteria and/or anaerobic bacteria. Coverage in each area was graded on a 3-point scale, 0 for no activity, 1 for narrow activity, and 2 for broad activity. For patients treated with multiple antibiotics simultaneously, the score for each antibiotic received on a given day was summed. Cumulative scores were summed across days. The calculation of the score for each drug is included as supplementary data (supplementary data file 1). These summary measures resulted in four covariates that were all included the statistical analysis. Two were daily variables (i.e. number of antibiotics administered, and coverage score on each sequential day of MV) and two were cumulative variables (i.e. the cumulative number of antibiotics administered, and the coverage score from intubation to the day of assessment).

Protocol for Specimen Collection

The first tracheal aspirate specimens were collected with routine suctioning of the endotracheal tube (ETT) as soon as possible after intubation, but within 24 hours. Subsequent collections were obtained with the first morning routine suctioning performed by the clinical caregiver (nurse or respiratory therapist). All sites utilized in-line suctioning equipment for routine ETT suctioning. All clinical personnel at each site were trained in the collection procedure utilizing universal training materials and processes. Primary site leads were trained via webinar, and they in turn, trained the remaining local site personnel. To collect the tracheal aspirate, a 40 mL sterile suction trap was attached to the in-line suctioning tubing. Per routine clinical suctioning protocol, the suction catheter was inserted through the ETT to 0.5 cm below the tip of the ETT and

suction applied as catheter is slowly withdrawn. If contents of suction did not enter the suction trap, then 0.5 mL of sterile saline could be instilled to clear the suction catheter of retained secretions into the suction trap. The trap was then aseptically sealed and delivered to the clinical research coordinator for processing and storage. Briefly, under sterile conditions, the tracheal aspirate specimen was pipetted from the suction trap and transferred to a 2 mL sterile microtube labeled with the subject ID and immediately frozen. Specimens were permitted to be placed in a 20°C Freezer for up to 72 hours prior to moving to -80°C freezer, where specimens remained until shipping to the centralized Pediatric Microbiome Laboratory at the Children's Hospital of Colorado.

Microbiome Methods

DNA extraction and Quantitative PCR: DNA extractions were performed using the Qiagen EZ1 Advanced automated extraction platform (Qiagen Inc., Valencia, CA) with the bacterial card and tissue extraction kit per manufacturer's instructions. Total bacterial load was measured using a quantitative real-time PCR assay that has been previously published (4), and evaluated for use in human airway samples (5). DNA extracts were diluted 1:40, and 4 µl of the diluted DNA (dilution factor 10) was used as template in triplicate. A cloned 16S rRNA gene was used to establish copy number based on a standard curve (10^3 to 10^8 copies). Fidelity of the molecular biology preparation was monitored using triplicate blanks in each plate.

16S rRNA sequencing: Bacterial profiles were determined by broad-range amplification and sequence analysis of 16S rRNA genes (5, 6). Amplicons were generated using primers that target approximately 300 base pairs of the V1/V2 (27F/338R) variable region of the 16S rRNA gene. This fragment was selected for numerous reasons. First, the highest level of sequence variation is in the V1-3 region of the 16S rRNA gene. Sequence variation is the most critical issue for taxonomic resolution, which is the main limitation of sequence length constraints from

next generation sequencing platforms. Second, novel diversity was anticipated, and targeting the 5' or 3' ends of the gene would allow primer design to attempt capture of near full-length sequences. Finally, simulation studies performed to target high-throughput sequencing primers during the transition from Sanger sequencing to 454 identified the region ~100 – 350 of the 16S rRNA gene as the best performing fragment compared to full length sequences (7, 8). More recently, ability to assemble the paired-end reads generated by MiSeq was used as a proxy for sequence quality in our workflow when we transitioned from 454 to Illumina. The original 454 validation experiment described in Hamady et al. (9) used samples that had existing V1-5 Sanger data, and the community composition observed by 454 sequencing was very similar. This agreement, presumably driven by the 27F primers sequence common across all sequencing platforms, allowed for comparison between 454 studies and earlier studies based on Sanger sequences. Our internal data confirms that transition from 454 to Illumina platforms did not significantly impact community composition from a range of sample types. PCR was performed using dual indexed primers in triplicate along with a negative PCR control for each sample. Reagent and processing controls were included and evaluated identically to the project samples. The triplicate PCR for each sample was combined, and product was evaluated by agarose gel electrophoresis along with the negative control. Any reaction that exhibited amplification in the negative control was repeated. PCR and DNA extraction controls were performed in parallel using the same PCR design as samples. There was no amplification apparent when assessed by agarose gel electrophoresis. *PCR* products were normalized using agarose gel densitometry by combining approximately equimolar amounts into sequencing pools (PCR and extraction controls were added at maximum volume since there was no band available to assess concentration). The pooled amplicons were gel purified and concentrated using a DNA Clean and Concentrator Kit (Zymo, Irvine, CA). Pooled and concentrated amplicons were quantified using Qubit Fluorometer 2.0 (Invitrogen, Carlsbad, CA). The pool

was diluted to 4nM and denatured with 0.2 N NaOH at room temperature. The denatured DNA was diluted to 20pM and spiked with 10% of the Illumina PhiX control DNA prior to loading the sequencer. Illumina paired-end sequencing was performed on the MiSeq platform using a 500-cycle version 2 reagent kit.

Analysis of Illumina Paired-end Reads. Illumina MiSeq paired-end reads were aligned to human reference genome Hg19 with bowtie2 and matching sequences discarded (10). As previously described, Illumina MiSeq paired-end sequences were sorted by sample via barcodes in the paired reads with a python script (11). Sorted paired end sequence data were deposited in the NCBI Short Read Archive under accession number PRJNA533819 (data generated from methods development [PRJNA436139] and comparison of gastric and tracheal aspirates [PRJNA508231] were also utilized by this study). The sorted paired reads were assembled using phrap (12, 13). Pairs that did not assemble were discarded. Assembled sequence ends were trimmed over a moving window of 5 nucleotides until average quality met or exceeded 20. Trimmed sequences with more than 1 ambiguity or shorter than 200 nt were discarded. Potential chimeras identified with Uchime (usearch6.0.203_i86linux32) (14) using the Schloss (15) Silva reference sequences were removed from subsequent analyses. Assembled sequences were aligned and classified with SINA (1.2.11) (16) using the 418,497 bacterial sequences in Silva 115NR99 (17) as reference configured to yield the Silva taxonomy. Sequences with identical taxonomic assignments were clustered into Operational taxonomic units (OTUs). This process yielded 182,567,651 sequences from 2,202 samples (average size: 82,948 sequences/sample; min: 6,266; max: 494,800). The median Goods coverage score was ≥ 99.4% at the rarefaction point of 6,266 with 100 resamplings. The software package Explicet (v2.9.4, www.explicet.org) (18) was used to calculate rarefied Good's coverage, alpha diversity measures (Shannon diversity and evenness index) and beta diversity (Morisita-Horn index). These indices were chosen a priori based on the relatively equal weighting between evenness

and richness (19). Given the depth of sequencing with current approaches, we chose to use indices that assign less weight to rare taxa.

Species Calls. Rather than applying an arbitrary distance cutoff (e.g., 97% or 99% identity) across all genera, we pre-compute such differentiation criteria for each genus of interest based on the genomic reference sequences provided in Silva and using only the subsequences of these genomic sequences bounded by the primer pair employed in the study. We then apply these individual genus-specific species classification criteria to all sequences falling within the genera of interest. The pre-computation process flags which species can be differentiated within a genus for a given primer pair and which cannot by evaluation of aligned positional differences; the latter sequences remain classified only to the genus level. This is a very conservative process requiring that the MiSeq derived sequences are essentially exact matches to the genomic subsequences in order to achieve a given species level attribution. In some cases the best that can be reported is that, e.g., the species is likely to be in a group of similar species within the genus. Binomial names, when available, or accession numbers are used to identify species groups.

Background assessment. Background was assessed using the 20 most common taxa observed in control data (Figure S9) compared by extraction controls, PCR controls and samples. Likewise, we have provided the top 20 taxa observed in samples (Figure S10) to demonstrate the limited overlap observed between controls and samples.

To limit the impact of background we excluded samples with no evidence of bacterial DNA. Initially, we attempted PCR for all samples irrespective of TBL, but samples with less than approximately 1.5x the TBL background were not successful in amplification as assessed using agarose gel electrophoresis (n=319). The two independent results showing little/no signal to detect were interpreted as quantity not sufficient (QNS) for microbiota determination. Once this

threshold was established, we did not attempt PCR for 505 TA samples with inadequate load to obtain clear amplification.

Statistical Analyses

Children with VAP and without VAP were compared using t-test or Wilcoxon rank sum test as appropriate for continuous variables and chi-square or Fisher's Exact test for categorical variables. To account for differences in sequencing depth, the relative abundance (RA) of each taxon was calculated (number of sequences for specific taxa/ total number of sequences*100).

Change in microbial factors prior to development of VAP: Time to event modeling

The association between changes in microbial factor measures (TBL, Shannon Diversity Index [diversity], and Shannon Evenness [evenness] over time and development of VAP was estimated using a joint longitudinal time to event model (JointModel package in R, R Foundation, Vienna). The longitudinal model included covariates for number of antibiotics administered on each successive day of MV, the cumulative number of antibiotics administered up to the day of assessment in the model, total coverage score on each successive day of MV, and cumulative coverage score up to the day of assessment in the model. Random intercepts and slopes were included for subjects. The survival model included age at intubation and PRISM III score as covariates. Use of a joint model accounts for change and variability in the longitudinal outcome and measurement error. Several approaches to incorporate the longitudinal outcome (TBL, diversity, evenness,) in the time to event model were evaluated and included using the corresponding sequential value, lagged days' values (from 1 or 2 days before the current sequential day), or slope. Models that included splines were also evaluated to assess results that did not include a linear trend constraint.

Matched Case-Control Analysis

In addition to the joint model which used data on all subjects up to extubation or VAP diagnosis, a sub-analysis of subjects with similar characteristics was performed. To make comparisons, we assigned a reference day for each non-VAP subject to correspond with the day of diagnosis for matched VAP subjects. To ensure comparisons at a similar stage of illness, the controls had to remain mechanically ventilated for at least 2 additional days past their assigned reference day. For example, a case who developed VAP on day 6 could only be matched to a control patient who was mechanically ventilated for a minimum of 8 days. Specifically, the assignment was made after sorting the controls, in descending order, by number of days intubated. The distribution of the identified reference day in the non-VAP group was targeted to match the distribution of the day of VAP diagnosis in the VAP group for our comparisons to be as useful as possible. For presentation purposes, results for the analyses evaluating microbial risk factors for VAP are presented in relation to the day of diagnosis (in VAP subjects) or the reference day (in non-VAP subjects), which was termed “Day 0”. We group matched non-VAP subjects to VAP subjects to also have similar distribution of the following characteristics: age at intubation, PRISM III score (3), and infectious admitting diagnosis. Twenty-eight non-VAP subjects were removed because they were ventilated for less than 4 days. An additional 80 non-VAP subjects were removed to make the distribution of the infectious admitting diagnosis similar between cases and controls. Priority for inclusion in the matched analysis was given to controls with longer ventilation times and higher PRISM III scores because these were less frequent occurrences in the cohort. The proportion of cases that were diagnosed for each day was multiplied by the total number of controls to obtain the number of controls that needed to be assigned to match the distribution of case diagnosis days. The resulting group matched cohort consisted of 280 controls and 66 cases.

Mixed effects models were used to evaluate the changes in microbial factors over time. These models were adjusted for antibiotic exposure and included a random subject intercept.

Least squared means for the microbial factors were compared between non-VAP and VAP groups up the three days prior to Day 0. Because sequence data were missing on Day 0 for some VAP patients (n =18 [27%]), a sensitivity analysis was performed that included subjects with at least 3 samples, with the requirement that VAP subjects had one of the samples available on the day of or the day prior to VAP diagnosis. Morisita-Horn (MH, Beta-diversity) for pairwise samples within each subject were calculated and compared across the VAP groups using a mixed model with B-splines to estimate trends over time. Time was included as a continuous variable and modeled using cubic B-splines with internal knots placed at the quintiles and boundary knots placed at the extremes. The model included a random individual intercept and slope, which were assumed to have a multivariate normal distribution and heterogeneous within individual variances that differed by VAP diagnosis. Cholesky decomposition was used to constrain the covariance matrix of the random effects to be positive-definite. Methods for modeling ecological measures over time are described in further detail in Wagner et al (20)

Clustering analysis: Identifying VAP phenotypes

To understand the relationship between clinical and microbial factors at intubation, an unsupervised random forest clustering algorithm was performed. Factors included in the random forest were baseline characteristics (age at intubation, PRISM III score, primary diagnosis, infectious diagnosis, and site of enrollment) and microbial factors (total bacterial load, presence of a pathogen (supplementary data file 2), and dominant taxa in sample). The resulting proximity matrix was transformed into a distance (21) and Ward agglomeration clustering using complete linkage and was applied to generate a hierarchical clustering of subjects.

REFERENCES

1. Pinto AJ, Raskin L. PCR biases distort bacterial and archaeal community structure in pyrosequencing datasets. *PLoS One* 2012; 7: e43093.

2. Pollack MM, Holubkov R, Glass P, Dean JM, Meert KL, Zimmerman J, Anand KJ, Carcillo J, Newth CJ, Harrison R, Willson DF, Nicholson C, Eunice Kennedy Shriver National Institute of Child H, Human Development Collaborative Pediatric Critical Care Research N. Functional Status Scale: new pediatric outcome measure. *Pediatrics* 2009; 124: e18-28.

3. Pollack MM, Patel KM, Ruttimann UE. PRISM III: an updated Pediatric Risk of Mortality score. *Crit Care Med* 1996; 24: 743-752.

4. Nadkarni MA, Martin FE, Jacques NA, Hunter N. Determination of bacterial load by real-time PCR using a broad-range (universal) probe and primers set. *Microbiology (Reading, England)* 2002; 148: 257-266.

5. Zemanick ET, Wagner BD, Robertson CE, Ahrens RC, Chmiel JF, Clancy JP, Gibson RL, Harris WT, Kurland G, Laguna TA, McColley SA, McCoy K, Retsch-Bogart G, Sobush KT, Zeitlin PL, Stevens MJ, Accurso FJ, Sagel SD, Harris JK. Airway microbiota across age and disease spectrum in cystic fibrosis. *Eur Respir J* 2017; 50.

6. Wagner BD, Sontag MK, Harris JK, Miller JI, Morrow L, Robertson CE, Stephens M, Poindexter BB, Abman SH, Mourani PM. Airway Microbial Community Turnover Differs by BPD Severity in Ventilated Preterm Infants. *PLoS One* 2017; 12: e0170120.

7. Liu Z, Lozupone C, Hamady M, Bushman FD, Knight R. Short pyrosequencing reads suffice for accurate microbial community analysis. *Nucleic acids research* 2007; 35: e120.

8. Liu Z, DeSantis TZ, Andersen GL, Knight R. Accurate taxonomy assignments from 16S rRNA sequences produced by highly parallel pyrosequencers. *Nucleic acids research* 2008; 36: e120.

9. Hamady M, Walker JJ, Harris JK, Gold NJ, Knight R. Error-correcting barcoded primers for pyrosequencing hundreds of samples in multiplex. *Nat Methods* 2008; 5: 235-237.

10. Illumina. iGenomes. 2017. Available from: https://support.illumina.com/sequencing/sequencing_software/igenome.html.

11. Hara N, Alkanani AK, Ir D, Robertson CE, Wagner BD, Frank DN, Zipris D. Prevention of virus-induced type 1 diabetes with antibiotic therapy. *J Immunol* 2012; 189: 3805-3814.

12. Ewing B, Hillier L, Wendl MC, Green P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* 1998; 8: 175-185.

13. Ewing B, Green P. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res* 1998; 8: 186-194.

14. Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 2011; 27: 2194-2200.

15. Schloss PD, Westcott SL. Assessing and improving methods used in operational taxonomic unit-based approaches for 16S rRNA gene sequence analysis. *Appl Environ Microbiol* 2011; 77: 3219-3226.

16. Pruesse E, Peplies J, Glockner FO. SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics* 2012; 28: 1823-1829.

17. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glockner FO. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic acids research* 2013; 41: D590-596.

18. Robertson CE, Harris JK, Wagner BD, Granger D, Browne K, Tatem B, Feazel LM, Park K, Pace NR, Frank DN. Explicet: Graphical user interface software for the management, analysis and visualization of microbial ecology data and metadata. *Bioinformatics* 2013.

19. Magurran AE. Measuring Biological Diversity. Malden, MA: Blackwell Science Ltd.; 2004.

20. Wagner BD, Grunwald GK, Zerbe GO, Mikulich-Gilbertson SK, Robertson CE, Zemanick ET, Harris JK. On the Use of Diversity Measures in Longitudinal Sequencing Studies of Microbial Communities. *Front Microbiol* 2018; 9: 1037.

21. Conrad DJ, Bailey BA. Multidimensional clinical phenotyping of an adult cystic fibrosis patient population. *PLoS One* 2015; 10: e0122705.

Figure S1. Subject specific reports. Individual summary reports are generated for each subject in the analysis cohort and are accessible at https://wkayla.shinyapps.io/subject_specific/. A report for a single subject is included here as an example. The clinical data for the subject are printed at the top of the report, followed by graphs displaying the bacterial composition, culture data, total bacterial load, antibiotics administered, and ecological measures over time during intubation for all available samples.

```
[1] "Subject: 1031376, Primary Admitting Diagnosis: Trauma, Age at Intubation
: 4.0246, VAP Case?: No, VAP Day of Diagnosis: 4, PRISM Score: 7 Matched?: NA
"
```

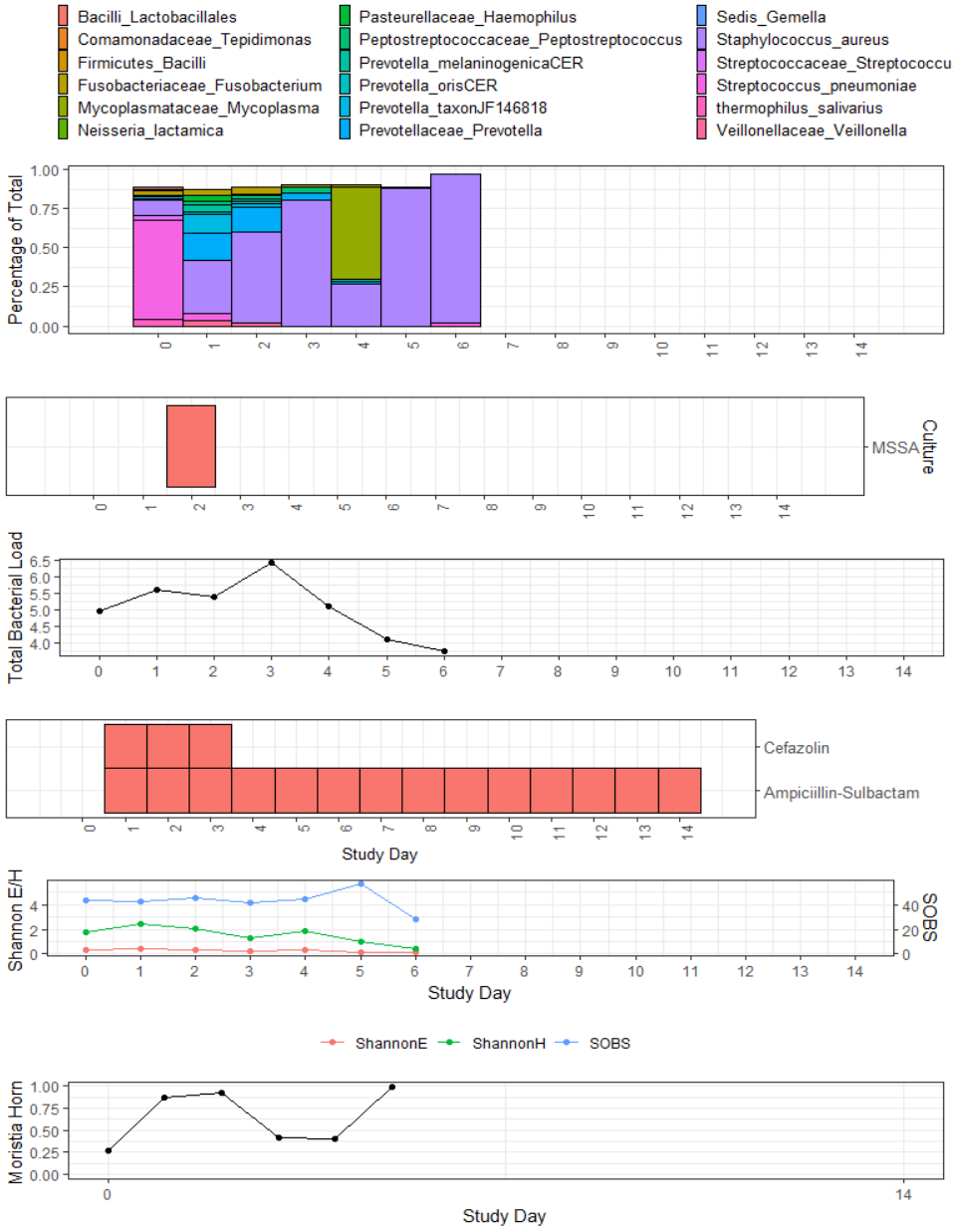


Figure S2. Distribution of VAP diagnosis day. The median day of VAP diagnosis for the 66 subjects diagnosed with VAP was 4.5 days (Range 2 - 14 days).

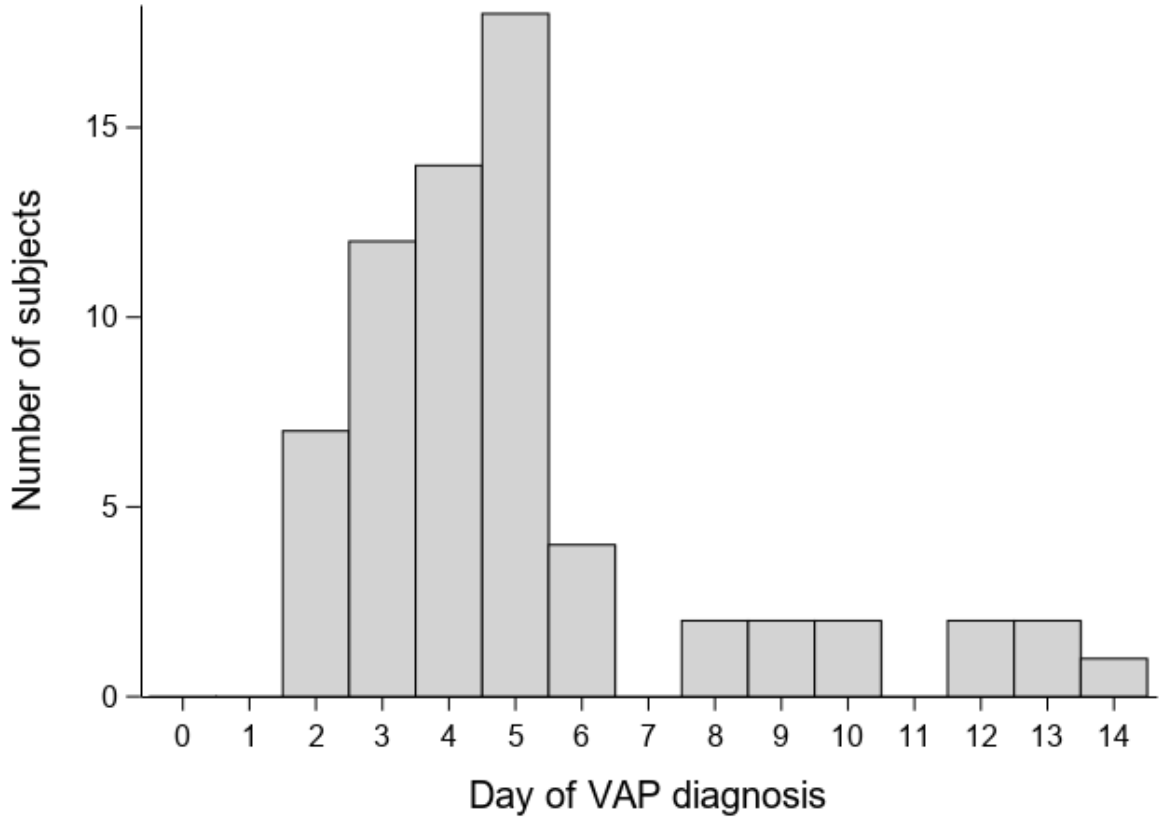
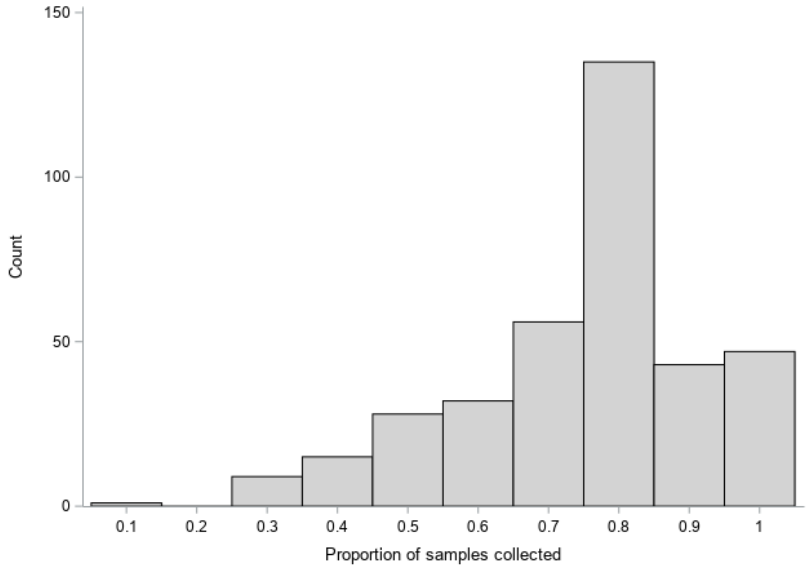


Figure S3. Sample Collection Distribution. Upper panel: Histogram of the proportion of daily samples collected during the first intubation period for each of the 366 subjects in the supervised analytic cohort. 2,330 samples (85%) were collected out of a possible 2,739 ventilator days. The figure shows the distribution of the proportion of daily samples collected during the first intubation period, the y-axis corresponds to the number of subjects and the x-axis represents the proportion of possible samples collected. There were 225 (61%) subjects with at least 75% of their total daily samples collected. Lower panel: The distribution of the proportion of samples with sequencing data, corresponding to 1,693 samples (73% of collected samples). The y-axis corresponds to the number of subjects and the x-axis represents the proportion of possible samples that had sequencing data.



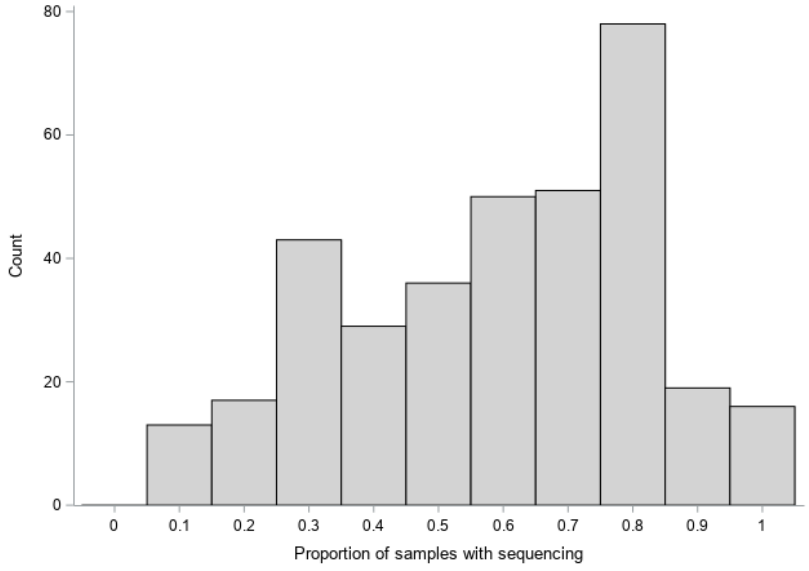


Figure S4: Distribution of total bacterial load for tracheal aspirate samples. The distribution of total bacterial load across samples are displayed in comparison to the values from 348 total negative controls using PCR water as template run in triplicate on each plate (N = 116) of qPCR assay (blue). Samples with sequencing available are shown in red, and samples that failed to amplify sequence are shown in green. Lower panel shows the distribution of TBL does not vary depending on the age of the subject.

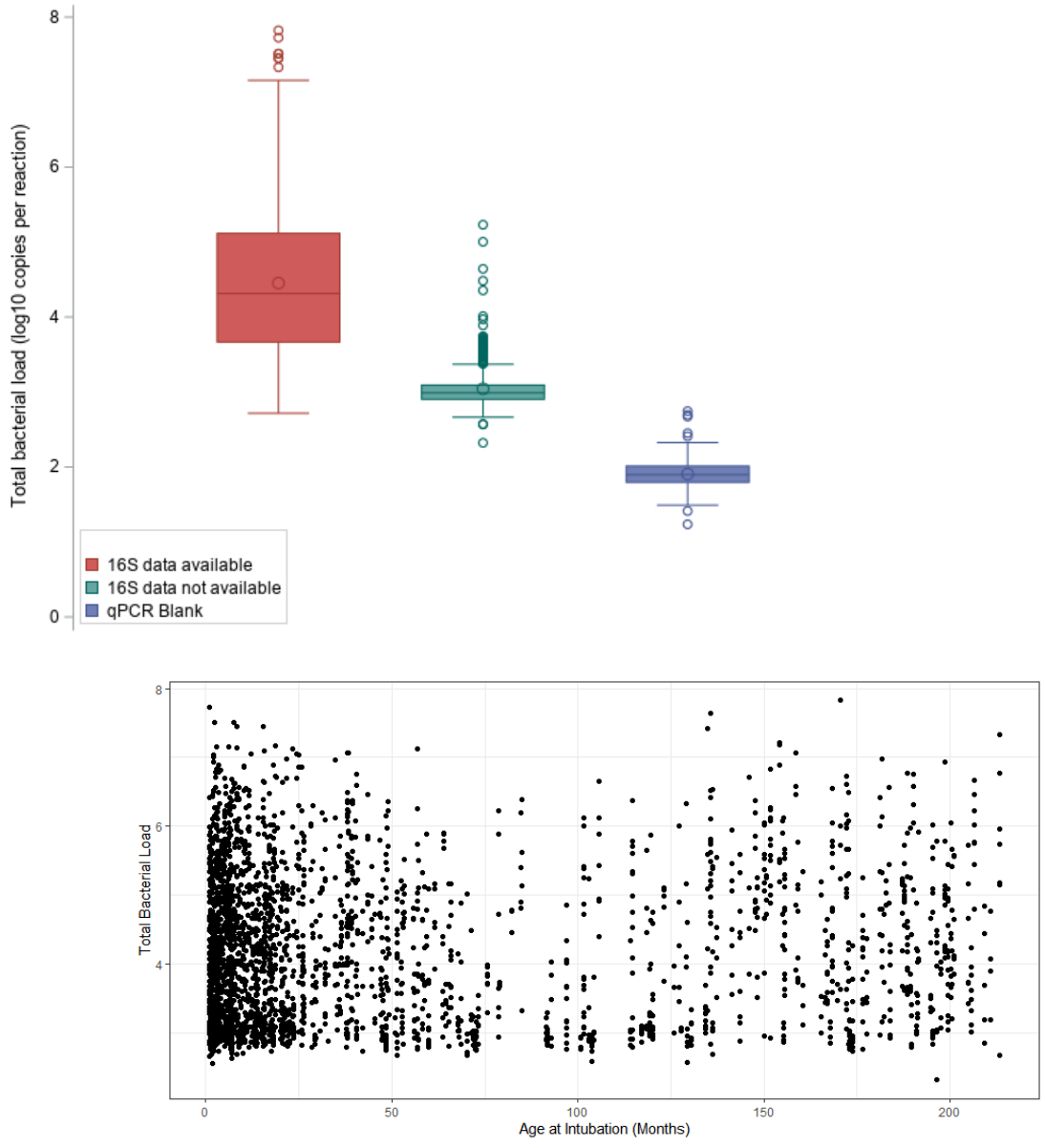


Figure S5: Matched case-control analyses: Comparison of microbiota communities at time of intubation and Day 0. Heatmap comparing microbiota communities at time of intubation (left) and Day 0 (VAP group: day of VAP diagnosis, non-VAP group: corresponding MV day; right) in matched analysis. Samples from non-VAP cases are displayed in columns to the left of the vertical lines (annotated with the blue bar below) and VAP cases are displayed to the right (annotated with the red bar below). Prominent taxa (at least 80% relative abundance (RA) in one sample) for each sample are displayed in rows. The darker the color of the bar, the higher the RA. Microbial taxa appear similar between VAP and non-VAP samples at intubation. Dominant taxa detected at VAP diagnosis are also prominent in non-VAP samples. There does not appear to be a single taxon that contributes to VAP. Often, these taxa are similarly represented in the non-VAP samples as well, suggesting that they may not be pathogenic in some subjects.

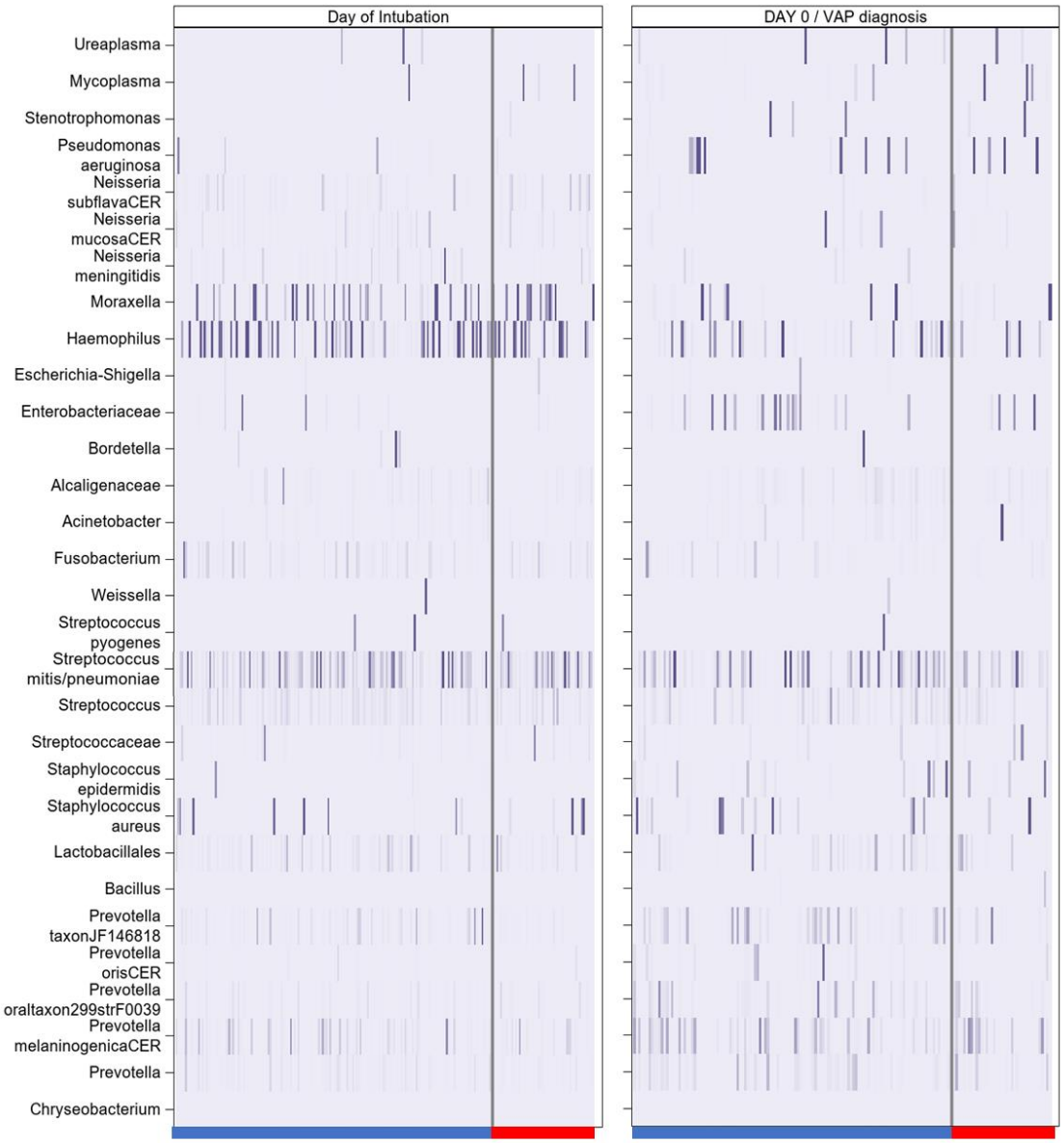
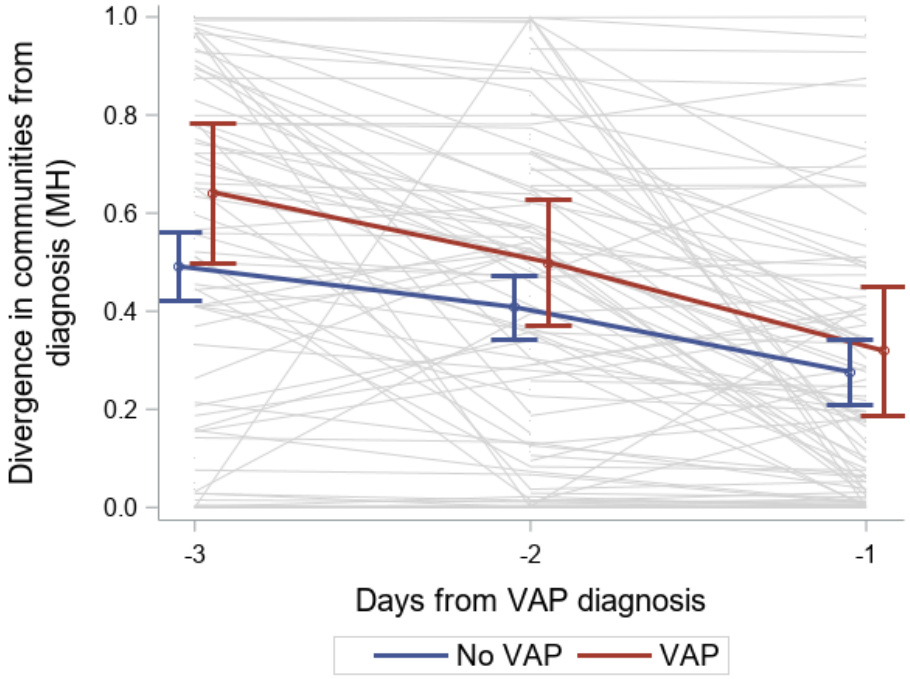
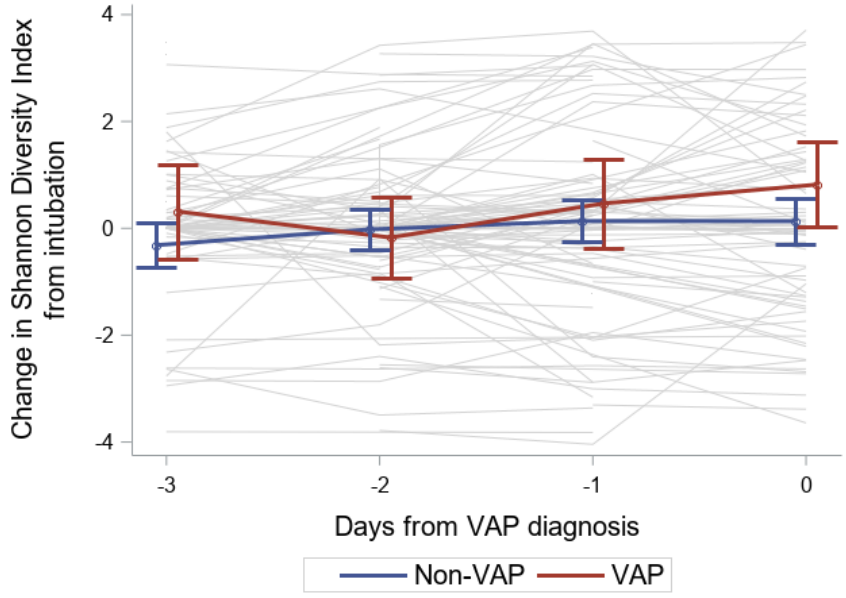


Figure S6: Comparison of beta diversity (MH) between VAP and non-VAP subjects relative to Day 0. Each individual (grey lines) and the average trends for the VAP groups (colored lines) denote the MH between each time point sample (Days -1 to -3) and the Day 0 sample. Day 0 denotes day of diagnosis in VAP cases (n = 66) and the reference day of mechanical ventilation in controls (n=227). Bars represent 95% confidence intervals. The table below presents the statistical comparison between groups at each time point. The degree of divergence relative to Day 0 samples was greatest 3 days prior to Day 0 and was more divergent in the VAP subjects, but was not statistically higher in comparison to non-VAP subjects.



Day relative to Day 0	Estimate (difference between VAP and No VAP)	Standard Error	Pr > |t|	Lower	Upper
-3	-0.1513	0.08043	0.0619	-0.3102	0.007624
-2	-0.09139	0.07320	0.2138	-0.2360	0.05324
-1	-0.04272	0.07401	0.5647	-0.1890	0.1035

Figure S7: Matched case-control analyses: Comparison of microbial factor changes from intubation between subjects who do and do not develop VAP. Mean trajectories of change from intubation in microbial factors do not differ dramatically between subjects who did and did not develop VAP in the matched cohort. Bars represent 95% confidence intervals. Day 0 denotes day of diagnosis in VAP cases (n = 66) and the reference day of mechanical ventilation in controls (n=227). Panels represent comparisons in Shannon Diversity and Total Bacterial Load (Shannon Evenness not shown) for changes from day of intubation for Days -3 to 0.



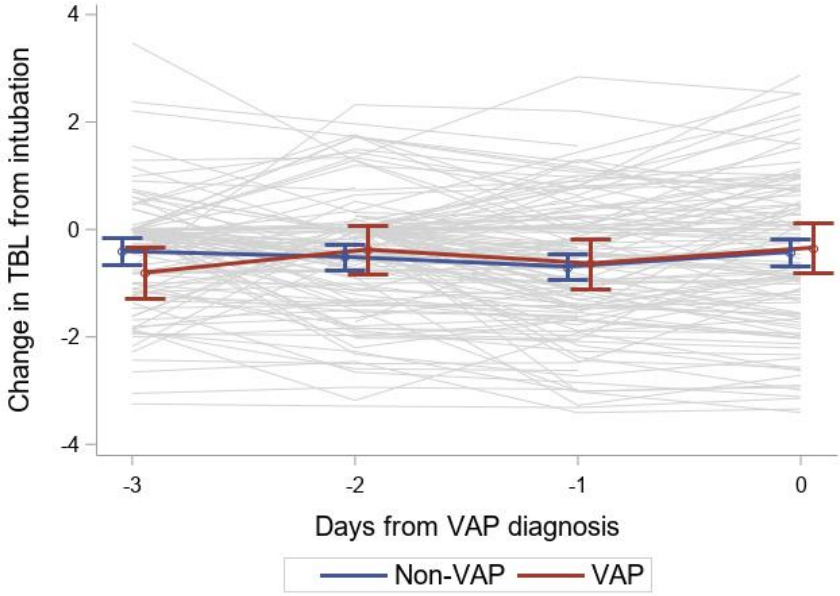


Figure S8: Clustering analyses: Hierarchical clustering of subjects at intubation. We used an unsupervised machine learning algorithm (random forest) to evaluate whether subjects with similar characteristics at the time of intubation cluster together. Subjects with similar characteristics are closest together at the lowest branch of the tree. The features included in the random forest algorithm to define cluster membership are listed on the right side of the dendrogram. Notably, in this heterogeneous cohort, there are no apparent microbial differences across site of enrollment. Abbreviations: Site: site of enrollment; PRISM: Pediatric Risk and Mortality III; TBL: total bacterial load; Infectious Admit: infectious diagnosis at PICU admission; Pathogen: Pathogenic organism present as defined in File 2; Diversity, Shannon diveristy; RA, Relative abundance.

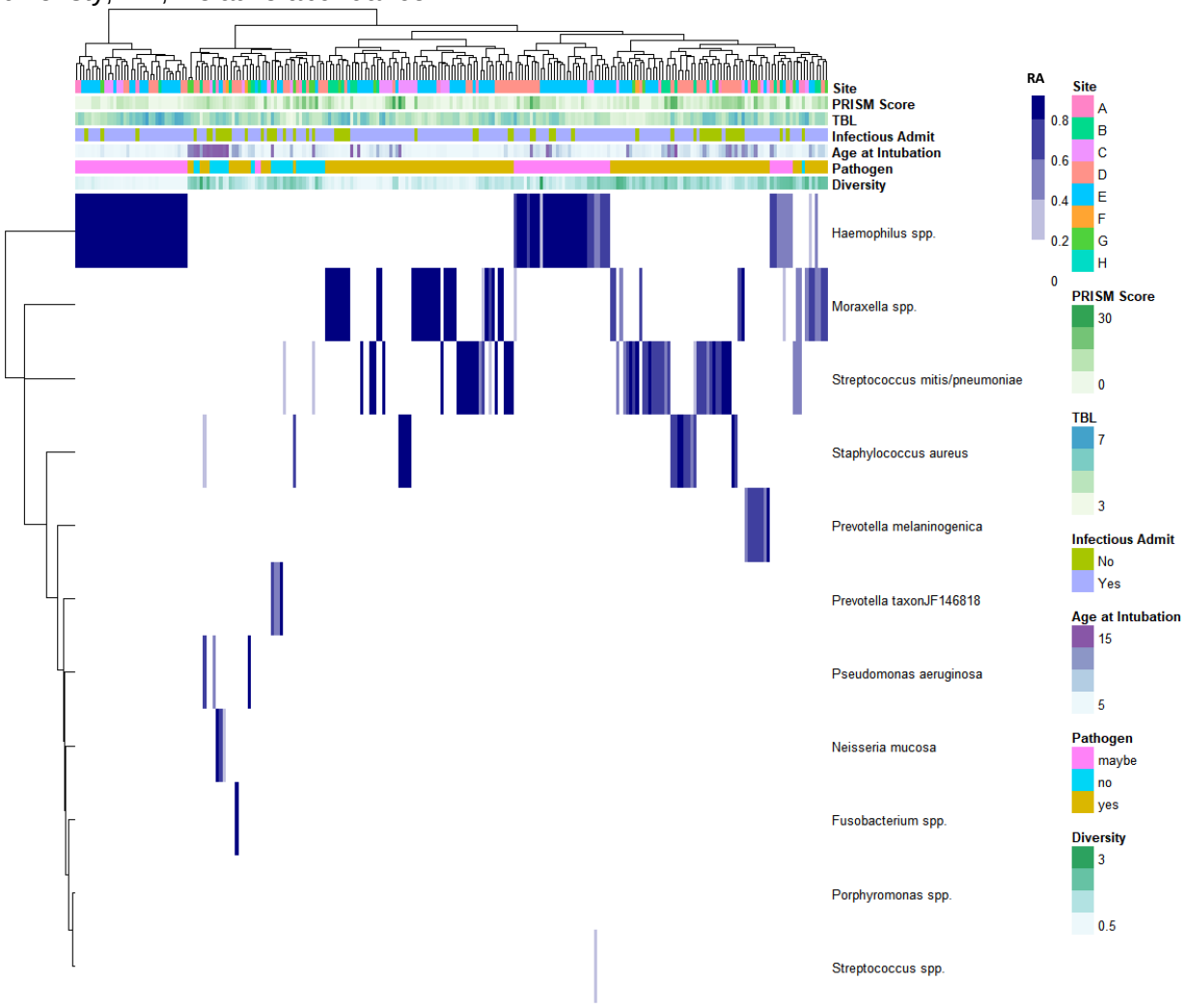


Figure S9: Background comparison of average relative abundance of taxa based on the 20 most abundant taxa observed in controls. Bar plots show PCR and extraction controls for each taxon identified along with the distribution of RA from clinical samples. The majority of taxa observed are consistent with prior reports for background. There was a single predominant taxon associated with the extraction controls (*Alcaligenaceae*). The other taxa identified in the extraction controls appear to originate from the PCR mix, and in two cases (*S. pneumoniae*|*mitis, Haemophilus*) human source (most likely the concurrently tested samples or potentially research staff). There is limited overlap between controls and clinical samples when comparing RA of these taxa. Origin of *Stenotrophomonas* is challenging to interpret. This taxon represents known contaminant sequences and pathogens associated with chronic lung infections. The range observed in samples is higher than any control suggesting both sources contributed to detection of this taxon. *Pseudomonas stutzeri* is represented by accession number (CP007509).

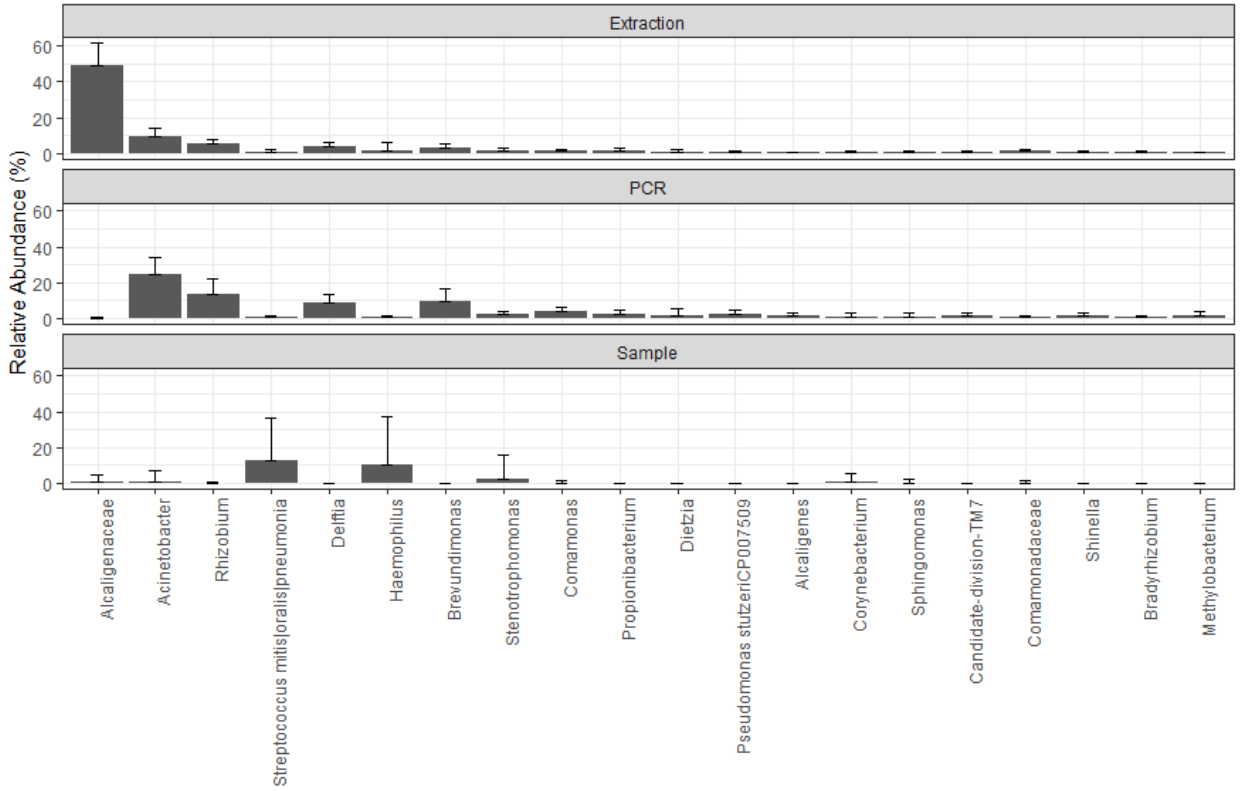


Figure S10: Background comparison of average relative abundance of taxa based on the 20 most abundant taxa in subject samples. Bar plots show PCR and extraction controls for each taxon identified along with the distribution of RA from clinical samples. There is limited overlap between controls and clinical samples when comparing RA of these taxa. The predominant taxa observed in the clinical samples are also present at low levels in controls. None of the controls demonstrated amplification by agarose gel electrophoresis, and overall sequence counts obtained from these samples were much lower than those from TA samples that demonstrated amplification. Exact source tracing is challenging and was most likely due to very low levels of transfer during PCR. However, we can't rule out low background during the extraction step either. Sequences that did not match to any available species sequences were reported at the Genus level (e.g. *Streptococcus*).

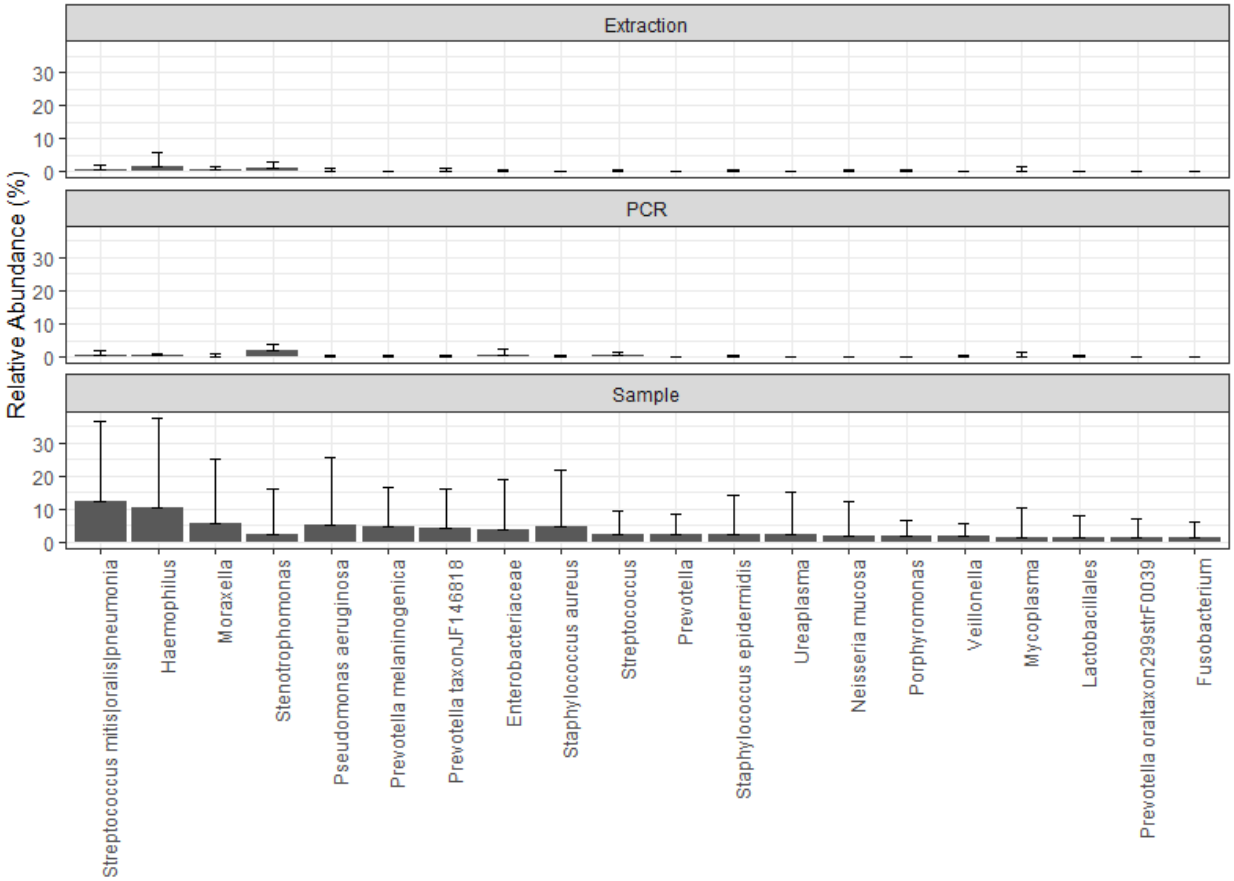


Table S1: Specific Comorbid Diagnoses

Comorbidities	All Subjects (n=454)	Supervised Analytic Cohort (n=366)		Subjects with Physician diagnosed or suspected VAP not meeting CDC criteria (n=88)
		No VAP (n=300)	VAP (n=66)	
Airway/lung disease	60 (27%)	40 (13%)	10 (29%)	10 (11%)
Cardiovascular disease	37 (17%)	19 (6%)	7 (20%)	11 (12%)
Neurological disease	32 (15%)	21 (7%)	3 (9%)	8 (9%)
Congenital anomaly or chromosomal defect	23 (10%)	13 (4%)	9 (26%)	1 (1%)
Cancer	17 (8%)	12 (4%)	3 (9%)	2 (2%)
Immunological disorders	9 (4%)	6 (2%)	0 (0%)	3 (3%)
Liver disease	8 (4%)	7 (2%)	0 (0%)	1 (1%)
Metabolic disease	7 (3%)	3 (1%)	1 (3%)	3 (3%)
Renal disease	3 (1%)	2 (1%)	1 (3%)	0 (0%)
Hematological disease	3 (1%)	2 (1%)	0 (0%)	1 (1%)
Gastrointestinal disease	2 (1%)	2 (1%)	0 (0%)	0 (0%)
Psychiatric disorder	2 (1%)	2 (1%)	0 (0%)	0 (0%)
Transplant	1 (0.5%)	1 (0.3%)	0 (0%)	0 (0%)
Other	15 (7%)	13 (4%)	1 (3%)	1 (1%)

Table S2: Description of sample collection, number of samples sequenced by day of collection

A. Full Cohort (n=454)			
Days after intubation	**Number of subjects intubated**	**Number of subjects with a sample collected**	**Number of subjects with a sequenced sample**
1	454	454 (100%)	406 (89%)
2	454	399 (88%)	273 (68%)
3	425	369 (87%)	260 (70%)
4	347	302 (87%)	216 (72%)
5	280	235 (84%)	169 (72%)
6	214	197 (92%)	146 (74%)
7	151	126 (83%)	95 (75%)
8	113	94 (83%)	66 (70%)
9	87	75 (86%)	51 (68%)
10	72	62 (86%)	43 (69%)
11	61	49 (80%)	36 (73%)
12	57	47 (82%)	37 (79%)
13	51	30 (59%)	23 (77%)
14	32	11 (34%)	9 (82%)
B. Supervised Analytic Cohort (n=366)			
Days after intubation	**Number of subjects intubated**	**Number of subjects with a sample collected**	**Number of subjects with a sequenced sample**
1	366	366 (100%)	330 (90%)
2	366	321 (88%)	218 (68%)
3	339	297 (88%)	206 (69%)
4	270	239 (89%)	169 (71%)
5	212	183 (86%)	133 (73%)
6	156	149 (96%)	108 (72%)
7	103	84 (82%)	60 (71%)
8	75	62 (83%)	40 (65%)
9	59	51 (86%)	32 (63%)
10	49	40 (82%)	28 (70%)
11	43	33 (77%)	24 (73%)
12	40	32 (80%)	25 (78%)
13	36	21 (58%)	16 (76%)
14	22	7 (32%)	5 (71%)

Table S3: Comparison of microbial factors between VAP and non-VAP subjects at intubation. Total Bacterial Load (TBL), Shannon diversity, and Shannon evenness were tested using a two-sample t-test, means and standard deviations are reported. Taxa were tested using a Wilcoxon Rank Sum test, medians, and 75th percentile and 100% percentile (maximum relative abundance) of the highest-ranking taxa are reported. Diversity and evenness at intubation are lower in VAP subjects as compared to non-VAP subjects. No differences in dominant taxa were found between the two groups.

	All Subjects (n=454)	Supervised Analytic Cohort (n=366)			Subjects with Physician diagnosed or suspected VAP not meeting CDC criteria (n=88)
Microbial factor Means (std) or median (75%, 100%)		**No VAP (n=300)**	**VAP (n = 66)**	**P-value**	
Shannon Diversity	1.74 (1.18)	1.83 (1.19)	1.35 (1.16)	0.005	1.7 (1.08)
Shannon Evenness	0.32 (0.2)	0.34 (0.2)	0.25 (0.2)	0.004	0.32 (0.18)
Total Bacterial Load	4.65 (0.99)	4.58 (0.97)	4.47 (0.95)	0.394	5.04 (1.02)
Moraxella	0.0 (0.3, 99.6)	0.0 (0.2, 99.6)	0.0 (2.7, 98.9)	0.754	0.0 (0.2, 98.3)
Haemophilus	0.7 (6.7, 100)	0.7 (5.3, 100)	0.5 (33.7, 100)	0.356	0.9 (15.1, 99.9)
Streptococcus mitis/pneumoniae	3.7 (18.3, 99.9)	3.4 (19.2, 99.9)	3.2 (15.2, 99.7)	0.879	5.6 (17.8, 98.9)
Staphylococcus aureus	0.0 (0.0, 99.7)	0.0 (0.0, 99.6)	0.0 (0.0, 99.7)	0.632	0.0 (0.0, 92.2)
Prevotella melaninogenica	0.2 (2.5, 93.4)	0.2 (3.0, 93.4)	0.1 (0.9, 42.0)	0.17	0.5 (3.1, 74.2)
Ureaplasma	0.0 (0.0, 84.3)	0.0 (0.0, 84.3)	0.0 (0.0, 0.4)	0.238	0.0 (0.0, 0.8)
Mycoplasma	0.0 (0.0, 99.5)	0.0 (0.0, 99.5)	0.0 (0.0, 97.6)	0.759	0.0 (0.0, 18.5)
Pseudomonas aeruginosa	0.0 (0.0, 99.5)	0.0 (0.0, 63.7)	0.0 (0.0, 99.5)	0.928	0.0 (0.0, 0.4)
Neisseria subflava	0.0 (0.6, 97.4)	0.0 (0.6, 97.4)	0.0 (0.1, 30.7)	0.073	0.0 (0.9, 62.0)
Shigella	0.0 (0.0, 86.8)	0.0 (0.0, 9.2)	0.0 (0.0, 20.8)	0.123	0.0 (0.0, 86.8)
Enterobacteriaceae	0.0 (0.0, 0.93)	0.0 (0.0, 93.0)	0.0 (0.0, 4.8)	0.342	0.0 (0.0, 12.5)
Bordetella	0.0 (0.0, 99.8)	0.0 (0.0, 99.8)	0.0 (0.0, 0.0)	0.366	0.0 (0.0, 0.0)
Alcaligenaceae	0.0 (0.3, 60.5)	0.0 (0.4, 60.5)	0.0 (0.3, 5.9)	0.404	0.0 (0.1, 9.2)
Acinetobacter	0.0 (0.1, 4.3)	0.0 (0.2, 4.3)	0.0 (0.2, 1.2)	0.899	0.0 (0.0, 4.3)
Fusobacterium	0.1 (1.5, 82.6)	0.1 (1.6, 82.6)	0.0 (0.5, 20)	0.078	0.2 (1.9, 12.6)
Weissella	0.0 (0.0, 98.4)	0.0 (0.0, 98.4)	0.0 (0.0, 0.2)	0.398	0.0 (0.0, 0.0)
Streptococcus pyogenes	0.0 (0.0, 99.6)	0.0 (0.0, 97.1)	0.0 (0.0, 64.8)	0.637	0.0 (0.0, 99.6)
Streptococcus	0.8 (3.8, 29.7)	1.0 (4.1, 23.7)	0.2 (1.5, 29.7)	0.02	0.9 (4.2, 28.3)
Staphylococcus epidermidis	0.0 (0.0, 61.4)	0.0 (0.0, 57.1)	0.0 (0.0, 0.5)	0.896	0.0 (0.0, 61.4)
Lactobacillales	0.1 (1.4, 92.1)	0.1 (1.6, 92.1)	0.0 (0.6, 64)	0.03	0.2 (0.8, 10.5)
Bacillus	0.0 (0.0, 0.3)	0.0 (0.0, 0.3)	0.0 (0.0, 0.1)	0.343	0.0 (0.0, 0.1)

Prevotella oraltaxon299strF0039	0.0 (0.4, 66.4)	0.0 (0.4, 23.2)	0.0 (0.1, 66.4)	0.052	0.1 (0.7, 13.2)
Prevotella	0.2 (1.4, 38.4)	0.3 (1.4, 33.6)	0.1 (0.4, 4.8)	0.012	0.4 (2.2, 38.4)
Prevotella taxonJF146818	0.1 (0.9, 83.1)	0.1 (1.0, 83.1)	0.0 (0.2, 3.3)	0.019	0.1 (1.2, 12.8)
Prevotella oris	0.0 (0.0, 40.7)	0.0 (0.0, 40.7)	0.0 (0.0, 6.5)	0.184	0.0 (0.0, 37.7)

Table S4: Parameter estimates from the joint model of longitudinal outcome assuming a linear trend over time and time to VAP diagnosis that includes the sequential value of the longitudinal outcome.

	Variable	Total Bacterial Load		Shannon Diversity		Shannon Evenness	
		Est (Interval)	**p-value**	**Est (Interval)**	**p-value**	**Est (Interval)**	**p-value**
	Intercept	4.31 (4.20, 4.43)	**<0.01**	2.25 (2.08, 2.42)	**<0.01**	0.38 (0.31, 0.45)	**<0.01**
Predictors of Longitudinal Outcomes (Microbial Factors)	Time (days)	0.12 (0.04, 0.19)	**<0.01**	-0.07 (-0.15, 0.02)	0.12	0.01 (-0.30, 0.30)	0.99
	Number of antibiotics by day	0.01 (-0.08, 0.11)	0.83	-0.25 (-0.41, -0.08)	**<0.01**	-0.05 (-0.08, -0.02)	**<0.01**
	Cumulative days of antibiotic exposure	-0.04 (-0.10, 0.01)	0.08	0.12 (0.02, 0.20)	**0.01**	0.04 (0.02, 0.06)	**<0.01**
	Cumulative antibiotic coverage score	-0.02 (-0.04, 0.01)	0.10	-0.04 (-0.07, -0.01)	**0.02**	-0.01 (-0.02, -0.01)	**<0.01**
	Total antibiotic coverage score by day	-0.01 (-0.04, -0.03)	0.83	0.10 (0.04, 0.16)	**<0.01**	0.02 (0.01, 0.03)	**<0.01**
Predictors of VAP	Age	-0.03 (-0.10, 0.04)	0.44	-0.07 (-0.14, -0.01)	**0.02**	-0.07 (-0.14, -0.01)	**0.03**
	PRISM III	0.02 (-0.02, 0.06)	0.35	0.01 (-0.03, 0.04)	0.69	0.01 (-0.03, 0.05)	0.62
	Sequential value for longitudinal outcomes	-0.95 (-1.48, -0.44)	**<0.01**	0.02 (-0.26, 0.27)	0.88	-0.58 (-1.38, 0.67)	0.19

Table S5: Time to event analyses: Hazard ratios. Hazard ratios (95% credible intervals) for association between longitudinal outcomes (TBL, Shannon Diversity and Shannon Evenness) and time to VAP diagnosis from the joint models are presented. Different approaches to incorporate the microbial factors into the time to event model were evaluated to determine if they were informative to VAP diagnosis, including lagged values (evaluating whether microbial factors 1 or 2 days prior to the current sequential value day), slope of the linear trend in sequential values, and splines (a curve function to fit non-linear trends). The hazard ratios are displayed for the different approaches to include the longitudinal variable in the time to VAP model. Inclusion of the sequential or lagged TBL values was associated with time to VAP suggesting that lower TBL is associated with development of VAP. Bolded values indicate the 95% credible interval for the hazard ratio excludes 1.

Model	**TBL**	**Shannon Diversity**	**Shannon Evenness**
sequential value-linear trend	**0.39 (0.23, 0.64)**	1.02 (0.77, 1.31)	0.56 (0.25, 1.95)
Lag 1 day – linear trend	**0.55 (0.34, 0.85)**	0.92 (0.68, 1.20)	0.40 (0.14, 1.16)
Lag 2 day – linear trend	**0.65 (0.42, 0.98)**	0.86 (0.65, 1.10)	0.31 (0.06, 1.20)
Slope – linear trend	**0.52 (0.30, 0.88)**	**1.91 (1.31, 3.03)**	**31.1 (6.7, >100)**
sequential value - spline	0.71 (0.39, 1.14)	1.04 (0.80, 1.34)	0.72 (0.24, 7.00)
Time dependent slopes - spline	1.38 (0.67, 2.51)	**1.90 (1.27, 3.19)**	**44.0 (9.85, >100)**

Table S6: Description of group matched cohort. Subjects that did not develop VAP either by the CDC diagnostic criteria or by physician diagnosis or suspicion of VAP were compared to those who developed VAP by CDC criteria based on PRISM III scores, age at intubation, infection status at PICU admission, and time on mechanical ventilation. Non-VAP subjects who did not exhibit similarities with VAP cases based on these characteristics were not included in the analysis (n =73).

N (%) or median (range)	**No VAP (n = 227)**	**VAP (n = 66)**
Age at intubation	1.3 (0.1 – 17.8)	1.2 (0.1 – 16.6)
Noninfectious admitting diagnosis	21 (9.3%)	6 (9.1%)
PRISM III score	5.0 (0 – 31)	5.5 (0 – 28)
Day of VAP diagnosis or matched sample day for no VAP	4.0 (2 – 14)	4.5 (2 – 14)

PRISM – Pediatric Risk of Mortality; VAP – ventilator-associated pneumonia

Table S7. Matched case-control analyses: Results. Comparisons of microbial factors between those that did and did not develop VAP up to 3 days prior to development of VAP

variable	Day from VAP diagnosis	Control (N = 227)			VAP (n = 66)			p-value[1]
		Mean[2]	95% CI		Mean[2]	95% CI		
Shannon H Diversity Index	-3	1.93	1.66	2.19	1.83	1.39	2.27	0.712
Shannon H Diversity Index	-2	2.09	1.85	2.32	1.59	1.17	2.00	0.040
Shannon H Diversity Index	-1	2.09	1.85	2.33	1.70	1.25	2.14	0.124
Shannon H Diversity Index	0	2.14	1.90	2.37	1.93	1.52	2.34	0.377
Shannon Evenness	-3	0.34	0.29	0.38	0.32	0.25	0.39	0.635
Shannon Evenness	-2	0.36	0.32	0.40	0.28	0.21	0.34	0.040
Shannon Evenness	-1	0.36	0.32	0.40	0.29	0.22	0.36	0.099
Shannon Evenness	0	0.38	0.34	0.41	0.35	0.28	0.41	0.430
Total Bacterial Load	-3	4.03	3.87	4.19	3.67	3.40	3.95	0.027
Total Bacterial Load	-2	3.90	3.76	4.04	3.96	3.71	4.21	0.687
Total Bacterial Load	-1	3.82	3.69	3.96	3.72	3.46	3.97	0.461
Total Bacterial Load	0	3.93	3.79	4.07	4.08	3.82	4.34	0.309
Change from intubation								
Change in Shannon H Diversity	-3	-0.32	-0.73	0.10	0.32	-0.57	1.20	0.196
Change in Shannon H Diversity	-2	-0.02	-0.40	0.35	-0.17	-0.93	0.59	0.725
Change in Shannon H Diversity	-1	0.14	-0.26	0.53	0.47	-0.37	1.31	0.484
Change in Shannon H Diversity	0	0.13	-0.28	0.55	0.82	0.02	1.62	0.131
Change Shannon Evenness	-3	-0.05	-0.12	0.02	0.04	-0.10	0.19	0.257
Change Shannon Evenness	-2	-0.00	-0.06	0.06	-0.02	-0.15	0.10	0.793
Change Shannon Evenness	-1	0.02	-0.04	0.09	0.07	-0.06	0.21	0.500
Change Shannon Evenness	0	0.04	-0.03	0.11	0.16	0.03	0.29	0.114
Change in Total Bacterial Load	-3	-0.40	-0.66	-0.15	-0.80	-1.29	-0.32	0.149
Change in Total Bacterial Load	-2	-0.51	-0.74	-0.27	-0.37	-0.82	0.07	0.597
Change in Total Bacterial Load	-1	-0.69	-0.92	-0.46	-0.63	-1.09	-0.17	0.821
Change in Total Bacterial Load	0	-0.42	-0.67	-0.17	-0.34	-0.80	0.13	0.768

[1]P-values are based on standard linear regression with microbial factors included as the outcome, development of VAP as the primary predictor and time and antibiotics variables as covariates. [2]Values correspond to the least square means from the regression model
Abbreviations: VAP: ventilator-associated pneumonia.