Airway microbiota across age and disease spectrum in cystic fibrosis

Edith T. Zemanick, Brandie D. Wagner, Charles E. Robertson, Richard C. Ahrens, James F. Chmiel, John P. Clancy, Ronald L. Gibson, William T. Harris, Geoffrey Kurland, Theresa A. Laguna, Susanna A. McColley, Karen McCoy, George Retsch-Bogart, Kurtis T. Sobush, Pamela L. Zeitlin, Mark J. Stevens, Frank J. Accurso, Scott D. Sagel, J. Kirk Harris

Online Data Supplement

Methods

Molecular Biology: DNA extractions were performed using the Qiagen EZ1 Advanced automated extraction platform (Qiagen Inc., Valencia, CA) with the bacterial card and tissue extraction kit. All sample manipulation was done in the BSL2 hood with appropriate laminar flow. Frozen samples were thawed at 4 °C and vortexed to ensure mixing. An aliquot of 200 µl for extraction was transferred into the tube provided with the EZ1 kit. Remaining sample was placed in a clean 2 ml tube and stored at -70 ºC. Extraction reagent cartridges, elution tubes and tip holders were loaded into the EZ1 sample rack as instructed by the manufacturer. Elution volume of 100µl was selected and EZ1 DNA Tissue Kit program was run. Elution tubes with DNA extract were stored at -20 °C. DNA extraction reagents were confirmed free of bacterial DNA by performing control extractions utilizing buffer or PCR grade water. These extractions were tested by qPCR and demonstrated copy numbers consistent with the background of the TBL assay.

Sequencing: Bacterial profiles were determined by broad-range amplification and sequence analysis of 16S rRNA genes following our previously described methods [1, 2]. Amplicons were generated using primers that target approximately 300 base pairs of the V1/V2 variable region of the 16S rRNA gene. PCR products were normalized using agarose gel densitometry, pooled, gel purified, and concentrated using DNA Clean and Concentrator Kit (Zymo, Irvine, CA). PCR amplification was run in triplicate for each sample along with a barcode specific negative PCR control. Cycling conditions consisted of 94 °C 2 min followed by 30 cycles of 94 °C (20s), 52 °C (20s) and 65 °C (60s). PCR amplicons were assessed using agarose gel electrophoresis to confirm amplification of an appropriately sized product (~350 bp), and lack of amplification in the

negative control. No amplification was observed for the negative control for any sample. PCR reactions (25 µl) consisted of 1x Hotmastermix (5Prime) 150 nM primers (27F, 338R). Pooled 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. PCR cycles were limited to 30 in order to reduce amplification of background based on validation of each tube of PCR mastermix [3].

Analysis of Illumina Paired-end Reads. Illumina MiSeq paired-end reads were aligned to human reference genome Hg19 with bowtie2 and matching sequences discarded [4, 5]. Illumina MiSeq paired-end sequences were sorted by sample via barcodes in the paired reads with a python script [1]. Sorted paired end sequence data were deposited in the NCBI Short Read Archive under accession number SRP044029. The sorted paired reads were assembled using phrap [6, 7]. 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) [8] using the Schloss Silva reference sequences were removed from subsequent analyses[9]. Assembled sequences were aligned and classified with SINA (1.2.11) [10]using the 629,125 bacterial sequences in Silva 115 [11] as reference configured to yield the Silva taxonomy. Operational taxonomic units (OTUs) were produced by combining sequences with identical taxonomic assignments. The software package Explicet (v2.10.5, www.explicet.org) [12] was used for display, analysis of Good's coverage and

calculation of Shannon diversity and evenness. Due to issues with the phylogenetic guide tree in Silva, *Pseudomonas aeruginosa* sequences occur in two places (taxonomy lines are Bacteria/Gammaproteobacteria/Pseudomonadales and Bacteria/Gammaproteobacteria/Pseudomonadales/Pseudomonadaceae/Pseudomonas). These two taxonomy lines were merged during analysis (*Pseudomonas*+). Similarly, this was done for *Achromobacter* (Bacteria/Betaproteobacteria/Burkholderiales/Alcaligenaceae and Bacteria/Betaproteobacteria/Burkholderiales/Alcaligenaceae/Achromobacter) to make *Achromobacter*+.

Results

Indications for bronchoscopy with BALF for CF and DC subjects are shown in **table S1**. Characteristics of DC subjects are available in **table S2**.

Sequencing:

CF samples were more likely to have adequate bacterial load for sequencing with successful amplification and sequencing data obtained from 66% of pediatric CF compared to 27% of DC BALF ($p<0.01$). Total bacterial load was lower in samples that failed to amplify with median (range) of 7.2 (6.3 – 7.4) log10 rDNA copies/ml compared to 7.9 (6.9 – 10.7) for those that successfully amplified ($p<0.01$). Notably, only two BALF samples that failed to amplify had TBL above the LOD, and both samples were also negative by culture.

We compared patient characteristics of CF subjects between those with samples that successfully sequenced (n=98) versus those that did not sequence (n=48) (**table S3**). There was no difference in age, gender, genotype, lung function, pancreatic status, BMI % predicted, antibiotic use, chronic *Pseudomonas* status or bronchoscopic approach. The primary difference between groups was in culture results, with 89% of BALF cultures positive in the sequence positive group compared to 33% positive in the sequence negative group ($p<0.01$). *Staphylococcus aureus* (MSSA or MRSA) was the pathogen detected by culture in 50% of samples that were culture positive yet sequence negative. This finding is consistent with prior studies showing lower sensitivity of detection of *Staphylococcus* by 16S rDNA gene sequencing particularly when present in low quantities[13]. Of those that amplified, sequencing detected a median (range) of 10,745 (497-90,680) sequences per sample, with Good's coverage greater than 97% at the rarefaction point of 497. Diversity, richness and evenness measures are shown in **table S4**.

Airway inflammation: Total cell count, absolute neutrophil count and percent neutrophils were significantly higher in CF compared to DC BALF **(table S6).** CF samples with negative cultures also had significantly elevated absolute and percent neutrophils compared to DC ($p<0.01$) **(table S6 and figure S4)**. Because 6 pediatric DC subjects had an underlying immunodeficiency which could impact the BALF cell counts, we performed a sensitivity analysis of the remaining DC subjects compared to CF and found that differences in cell count, percent and absolute neutrophils remained statistically significant.

BALF Collection Approach

We further examined the impact of collection approach [endotracheal tube (ETT), laryngeal mask airway (LMA) or nasal] on bacterial communities. Only 11 (8%) of CF samples were collected via the nasal route, and these were all from the same site; 7 of these samples had adequate bacterial load for sequencing. From DC patients, 7 (11%) of BALF samples were collected via the nasal route; only one had adequate bacterial load for sequencing. Using a PCoA plot, we found that samples collected by nasal approach clustered together and primarily contained the taxa *Streptococcus*, *Prevotella*, *Haemophilus* and *Neisseria*; however, samples obtained using LMA and ETT, particularly from young children, also contained these taxa. We did not find bacterial community clusters that differed based on whether samples were obtained by LMA or ETT. Although the approach used from ~30% of CF BALF samples was not recorded, many of these were likely collected via ETT or LMA; only one site reported using the nasal route as their primary approach.

Table S1. Indications for bronchoscopy in CF and DC participants: Most common indications and underlying diagnoses for DC participants were recorded by physician preforming bronchoscopy. Participants may have more than one indication and diagnosis recorded.

CF Indications for BAL (N = 146)	DC Indications and underlying diagnoses (N = 45)
Persistent cough: 60 (41%)	Asthma/ wheezing: 17 (38%)
Need for lower airway culture: 40 (27%)	Pneumonia/ Recurrent pneumonia: 16 (36%)
Unresponsive to current therapy: 43 (29%)	Dysphagia/ GERD/ Aspiration 12 (27%)
Concurrent Surgery: 31 (21%)	Cough: 11 (24%)
Increased sputum production: 27 (18%)	Croup: 8 (18%)
Radiology findings: 25 (17%)	Oncologic/ immunodeficiency: 6 (13%)
Lung function decline: 12 (8%)	Interstitial lung disease: 4 (9%)
Airway evaluation: 7 (5%)	Trisomy 21: 4 (9%)
Pneumonia: 2 (1%)	Non-CF bronchiectasis: 2 (4%)
Hemoptysis: 1 (1%)	Hemoptysis: 2 (4%)
Involuntary weight loss: 1 (1%)	

Table S2. Age, diagnoses, culture and molecular results for disease control subjects

Age, years	Primary Dx	Microbiologic culture result	Total Bacterial Load, Log10 copies/ mL	Total cell count, cells/µL	Percent neutrophils	Absolute neutrophil count, cells/µL	Sequencing succesful
9.0	Pneumonia	Streptococcus	7.2	998	24	240	Yes
4.0	Oncologic, pneumonia	Negative	7.4	1424	86	1225	No
19.0	Oncologic, chest CT changes	Negative	7.0	1095	5	55	No
3.0	Interstitial lung disease, pneumonia	Negative	6.4	2210	74	1635	No
2.0	Interstitial lung disease	Negative	7.2	54	8	4	No
17.0	Immunodeficiency, pneumonia	Coryneform bacteria	7.2	508	54	274	No
21.0	Immunodeficiency, pneumonia	Negative	7.0	450	11	50	No
7.0	Immunodeficiency, interstitial lung disease, non-CF bronchiectasis	Negative	8.4	1125	100	1125	Yes
14.0	Immunodeficiency, interstitial lung disease	Negative	6.6	2555	3	77	No
16.0	Hemoptysis	Negative	7.3	97	0	0	No
6.0	GERD, recurrent pneumonia	P. aeruginosa	7.2	178	25	45	No
1.3	Dysphagia, hypoxemia	S. pneumoniae	7.4	52	1	1	Yes
1.4	Dysphagia	Negative	7.3	34	2	1	No
3.0	Croup, recurrent pneumonia	Negative	7.3	176	9	16	No
5.0	Croup, recurrent pneumonia	Negative	6.3	268	1	3	No
4.0	Croup	Negative	6.4	774	5	39	No
1.3	Cough, wheezing	Moroxella catarrhalis, S. pneumoniae	7.4	159	69	110	No
12.0	Cough, pneumonia, hemoptysis	Negative	7.1	55	2	1	No
15.0	Cough, hypoxemia , pulmonary nodules	Negative	7.2	183	25	46	Yes
6.0	Cough, croup	Negative	7.4	32	3	1	No
4.0	Cough, croup	Negative	6.9	148	2	3	No
12.0	Cough	H. influenzae	7.7	2625	80	2100	Yes
5.0	Cough	Negative	7.3	96	0	0	No
4.0	Chronic lung disease, achondroplasia	Mixed Upper Respiratory Flora	7.2	166	0	0	No
9.0	Asthma, recurrent pneumonia	Negative	7.2	103	1	1	No
6.0	Asthma, pneumonia	H. influenzae, Alpha Streptococcus	7.3	339	1	3	No
11.0	Asthma, pneumonia	Negative	7.3	152	4	6	No
10.0	Asthma, croup	Negative	6.4	155	2	3	No
3.0	Asthma, croup	Negative	7.3	289	7	20	No
9.0	Asthma, cough, Stevens Johnson Syndrome	Negative	6.5	177	1	2	No
9.0	Asthma, cough, GERD	Negative	8.3	242	8	19	Yes
11.0	Asthma, cough	Negative	7.1	117	2	2	No
4.0	Asthma, chronic lung disease of prematurity, recurrent pneumonia	Mixed Upper Respiratory Flora	7.7	187	4	7	Yes
4.0	Asthma, chronic lung disease of prematurity, croup,	Bacillus, not cereus/ anthracis	7.2	148	0	0	No
15.0	Asthma	Negative	7.1	108	1	1	Yes
7.0	Asthma	Mixed Upper Respiratory Flora	7.4	220	2	4	No
6.0	Aspiration, Trisomy 21, wheezing	Beta Hemolytic Group A Streptococcus	8.0	290	3	9	Yes
2.0	Aspiration, pneumonia, Trisomy 21	H. influenzae	7.3	97	0	0	No
5.0	Aspiration, non-CF Bronchiectasis	Negative	7.1	1269	86	1091	No
0.8	Aspiration, GERD, wheezing	Negative	7.2	140	3	4	Yes
1.4	Aspiration, dysphagia, pneumonia, Trisomy 21	Streptococcus Pneumoniae	8.7	810	54	437	Yes
4.0	Aspiration, asthma, dysphagia, Trisomy 21	Negative	7.4	23	5	1	No
1.8	Aspiration, asthma, cough, pneumonia	Negative	7.3	51	3	2	Yes
15.0	Aspiration	Negative	7.1	470	2	9	No
4.0	Airway evaluation, tracheostomy	S. aureus, S. maltophilia	7.1	60	2	1	No

Table S3. Comparison of CF patient characteristics between sequence positive and sequence negative groups.

	CF Sequence positive (n = 98)	CF Sequence negative (n = 48)	p-value
Age, median (range)	12 (0.2 – 42)	10 (1.8 – 33)	0.15
Genotype, n (%)			0.78
F508del/F508del	52 (54%)	28 (60%)	
F508del/other	36 (37%)	15 (32%)	
Female, n (%)	46 (47%)	25 (52%)	0.60
FEV_1 % predicted, median (range)	97 (38 – 129)	97 (52 – 120)	0.77
PES, median (range)	6 (0 – 16)	3 (0 – 16)	0.26
Pancreatic insufficient, n (%)	95 (97%)	45 (94%)	0.66
Culture positive, n (%)	87 (89%)	16 (33%)	<0.01
P. aeruginosa	28 (29%)	2 (4%)	<0.01
MSSA	23 (23%)	3 (6%)	0.01
S. maltophilia	19 (19%)	2 (4%)	0.01
MRSA	19 (19%)	5 (10%)	0.24
A. xylosoxidans	4 (4%)	1 (2%)	0.99
B. cepacia complex	2 (2%)	0	0.99
H. influenzae	10 (10%)	0	0.03
Antibiotics in prior 14 days, n (%)	75 (77%)	40 (83%)	0.40
BMI %, median (range)	50 (1 – 96)	50 (3 – 92)	0.59
Bronchoscopic approach, n (%)			0.86
Endotracheal tube	26 (27%)	12 (25%)	
LMA	35 (36%)	20 (42%)	
Nasal/oral	7 (7%)	4 (8%)	
Not recorded	30 (31%)	12 (25%)	
Chronic *P. aeruginosa* infection	29 (30%)	9 (19%)	0.31

Table S4. Shannon Diversity, Richness and Evenness for CF and DC BALF

	Disease Control			Pediatric CF			Adult CF
	All (n=12)	Positive culture (n=6)	Negative culture (n=6)	All (n=90)	Positive culture (n=79)	Negative culture (n=10)	All/ Positive cultures (n=8)
Diversity[A, B] Median (range)	2.1 (0.1-3.6)	1.9 (0.1-3.3)	2.5 (1.2-3.6)	0.4 (0-3.4)	0.3 (0-3.3)	2.3 (0.2-3.3)	0.28 (0.07-0.59)
Richness[C, D] Median (range)	19.2 (4.8-36.9)	18.7 (4.8-29.2)	23.4 (9.0 – 36.9)	5.4 (1-32.3)	5.0 (1-28.9)	18.9 (5.2 – 32.3)	4.2 (3.0-9.1)
Evenness[E] Median (range)	0.5 (0.05-0.7)	0.5 (0.05-0.7)	0.6 (0.4 – 0.7)	0.2 (-1 – 0.9)	0.1 (-1 – 0.9)	0.6 (0.1- 0.7)	0.1 (0.04-0.4)

A. Diversity statistically significantly higher in DC + compared to CF + ($p=0.04$).

B. Diversity statistically significantly higher in CF – compared to CF + ($p=0.0001$).

C. Richness statistically significantly higher in DC + compared to CF + ($p<0.01$).

D. Richness statistically significantly higher in CF - compared to CF + ($p<0.001$).

E. Evenness statistically significantly higher in CF - compared to CF + ($p<0.01$).

Table S5. Prevalence and relative abundance (RA) of most frequently detected taxa and taxa associated with traditional CF pathogens. Median and range include only samples with detectable sequences from the specified taxa. Shaded comparisons were statistically significantly different between DC and CF Pediatric groups with $p<0.05$, adjusted for multiple comparisons.

Taxa	Patient group	BALF Samples (n)	Prevalence (n, %)	RA, Median	RA, Minimum	RA, Maximum
Achromobacter+	DC	45	2 (4%)	0.00	0.00	0.33
	CF PEDS	136	38 (28%)	0.00	0.00	99.89
	CF ADULT	10	1 (10%)	0.00	0.00	2.40
Burkholderia	DC	45	3 (7%)	0.01	0.01	0.05
	CF PEDS	136	16 (12%)	0.02	0.00	100.00
	CF ADULT	10	1 (10%)	0.00	0.00	0.15
Haemophilus	DC	45	12 (27%)	2.63	0.13	78.92
	CF PEDS	136	66 (49%)	0.27	0.01	99.83
	CF ADULT	10	1 (10%)	0.00	0.00	0.47
Mycobacterium	DC	45	1 (2%)	0.01	0.01	0.01
	CF PEDS	136	5 (4%)	0.22	0.01	0.40
	CF ADULT	10	0	--	--	--
Neisseria	DC	45	11 (24%)	2.77	0.01	18.51
	CF PEDS	136	49 (36%)	0.03	0.00	23.82
	CF ADULT	10	2 (20%)	0.00	0.00	0.15
Pseudomonas^{+}	DC	45	11 (24%)	0.10	0.02	0.42
	CF PEDS	136	67 (49%)	0.17	0.01	99.70
	CF ADULT	10	7 (70%)	43.90	0.00	99.19
Porphyromonas	DC	45	11 (24%)	3.86	0.04	9.81
	CF PEDS	136	46 (34%)	0.04	0.00	18.34
	CF ADULT	10	1 (10%)	0.00	0.00	0.40
Prevotella	DC	45	12 (27%)	8.15	0.06	42.71
	CF PEDS	136	73 (54%)	0.08	0.01	67.69
	CF ADULT	10	1 (10%)	0.00	0.00	98.85

Staphylococcus	DC	45	12 (27%)	0.04	0.01	0.15
	CF PEDS	136	75 (55%)	0.32	0.00	99.37
	CF ADULT	10	5 (50%)	0.10	0.00	98.49
Stenotrophomonas	DC	45	8 (18%)	0.06	0.01	0.17
	CF PEDS	136	64 (47%)	0.17	0.00	99.96
	CF ADULT	10	2 (20%)	0.00	0.00	12.25
Streptococcus	DC	45	12 (27%)	29.66	3.62	98.98
	CF PEDS	136	74 (54%)	0.48	0.01	87.05
	CF ADULT	10	3 (30%)	0.00	0.00	0.85
Veillonella	DC	45	12 (27%)	2.92	0.10	17.97
	CF PEDS	136	72 (53%)	0.14	0.00	31.99
	CF ADULT	10	0	--	--	--

Pseudomonas^{+} includes sequences assigned to the taxa *Pseudomonas* and *Pseudomonadaceae.*

Achromobacter^{+} contains sequences assigned to the taxa *Alcaligenaceae and Achromobacter*.

Table S6. BALF cell count, neutrophil count and percent neutrophils from DC and CF.

	Total Cell Count, median (IQR) cells/ µL	Absolute Neutrophil Count, median (IQR) cells/ µL	Percent Neutrophils, median (IQR) %
DC N=43	176 (97-450)	4.3 (1.1-45.8)	3 (1-24)
DC Positive culture N = 15	187 (148 – 508)	7.5 (0.5 – 240)	3 (1 – 54)
DC Negative culture N = 28	154 (97 – 370)	3.7 (1.1 – 29.5)	3 (2 – 9)
CF: Pediatric N=98	988 (386-2,589) P<0.01 (compared to DC)	521.4 (111.3-2,057.3), P<0.01	68 (35-88) P<0.01, n=108*
CF Peds: Positive culture N = 65	1,270 (515 – 3,165)	942 (174-2,227)	77% (55-91%), n = 73*
CF Peds: Negative culture N = 32	561 (351 – 1,130)	206 (56-563)	42% (17-67), n= 34*
CF: Adult N=9	2,002 (908-3,150)	1,602 (599-2,202)	80 (66-92)

*Some centers reported only % neutrophils from BALF rather than total cell count and absolute neutrophils

Table S7. Age, *P. aeruginosa* culture, microbial ecology and airway inflammation by pulmonary exacerbation status. p-values calculated from Kruskall-Wallis or Fisher's exact test for continuous and categorical variables, respectively. Shaded boxes indicate p-values < 0.05.

	Clinically Stable (n = 44)	**Pulmonary Exacerbation (n = 90)**	**Unknown (n = 12)**	**p-value**
Age, yrs, median (IQR)	11 (4 – 14)	13 (7 – 15)	10 (5 – 17)	0.57
Female, n (%)	19 (43%)	48 (53%)	4 (33%)	0.32
Pulmonary exacerbation score (PES), median (IQR)	0 (0 – 0)	8.5 (3 – 11)	3 (0 – 6)	<0.01
FEV_1 % predicted, median (IQR) Age ≥ 5 years	96 (79 – 110)	98 (85 – 107)	94 (76 – 113)	0.98
BAL culture positive, n (%)	29 (66%)	64 (71%)	10 (83%)	0.70
P. aeruginosa culture positive, n (%)	6 (14%)	19 (21%)	5 (42%)	0.10
Chronic *Pseudomonas* (>50% cultures positive in past year)	6 (14%)	27 (30%)	5 (42%)	0.03
Antibiotic use in past 14 days, n (%)	32 (73%)	74 (82%)	9 (75%)	0.39
Total Bacterial Load, median (IQR) Log_{10} copies/ml	7.4 (7.2 – 7.9)	7.6 (7.2 – 8.8)	7.9 (7.5 – 8.7)	0.11
n (%) with sequencing	26 (59%)	63 (70%)	9 (75%)	0.01
Diversity, median (IQR)	0.7 (0.2 – 2.4)	0.3 (0.1 – 1.1)	0.3 (0.2 – 1.0)	0.09
Pseudomonas RA, median (IQR)	0.00 (0.00 - 0.60)	0.12 (0.03 – 0.83)	0.84 (0.03 – 98.29)	0.01
BALF Total Cell Count, cells/ µL median (IQR)	779 (412 – 1770)	1270 (400 – 3360)	623 (223 – 3440)	0.37
Absolute neutrophils, cells/ µL median (IQR)	346 (92 – 1083)	835 (109 – 2792)	356 (85 – 3209)	0.21
Neutrophil, %, median (IQR)	55 (17 – 77)	74 (35 – 91)	90 (66 – 95)	0.03

Table S8. Patient characteristics by age group

	Age < 2 (n = 11)	2-5 (n = 22)	6-10 (n = 26)	11-17 (n = 57)	18-24 (n = 26)	25 (n = 4)
F508del/ F508del, n (%)	6 (55%)	13 (59%)	14 (54%)	31 (55%)	15 (60%)	1 (25%)
Female, n (%)	2 (18%)	11 (50%)	13 (50%)	30 (53%)	12 (46%)	3 (75%)
FEV_1 % predicted, median (range)	n/a	n/a	103 (71 – 117)	99 (38 – 129)	90 (52 – 116)	85 (47 – 94)
PES, median (range)	6 (0 -10)	0 (0 – 11)	8 (0 – 16)	3 (0 – 16)	10 (0 – 16)	6 (5 – 13)
Pancreatic insufficient, n (%)	11 (100%)	21 (95%)	23 (88%)	56 (98%)	26 (100%)	3 (75%)
Sequence Positive, n (%)	10 (91%)	13 (59%)	13 (50%)	36 (63%)	23 (88%)	3 (75%)
Culture positive, n (%)						
P. aeruginosa	9 (82%)	12 (55%)	16 (62%)	39 (68%)	24 (92%)	3 (75%)
MSSA	0	2 (9%)	4 (15%)	11 (19%)	11 (42%)	2 (50%)
S. maltophilia	2 (18%)	3 (14%)	4 (15%)	12 (21%)	5 (19%)	0
MRSA	2 (18%)	4 (18%)	3 (12%)	8 (14%)	3 (12%)	1 (25%)
A. xylosoxidans	1 (9%)	0	6 (23%)	9 (16%)	7 (27%)	1 (25%)
B. cepacia complex	0	0	1 (4%)	3 (5%)	1 (4%)	0
H. influenzae	0	0	0	1 (2%)	1 (4%)	0
Antibiotics in prior 14 days, n (%)	5 (45%)	2 (9%)	0	3 (5%)	0	0
BMI %, median (range)	7 (64%)	11 (50%)	20 (77%)	48 (84%)	26 (96%)	4 (100%)
Bronchoscopic approach, n (%):	n/a	53 (18- 90)	51 (1 – 92)	50 (3 – 96)	37 (3 – 75)	--
Endotracheal tube	0 (0)	3 (14%)	3 (12%)	17 (30%)	5 (19%)	3 (75%)
LMA	6 (55%)	12 (55%)	10 (38%)	18 (32%)	12 (46%)	0
Nasal/oral	2 (18%)	3 (14%)	3 (12%)	3 (5%)	9 (35%)	0
Not recorded	3 (27%)	4 (18%)	10 (38%)	19 (33%)	0	1 (25%)
Chronic *P. aeruginosa* infection	1 (9%)	0	4 (15%)	12 (21%)	17 (65%)	4 (100%)

Supplemental Figure Legends

Figure S1. Culture, Total bacterial load and sequencing flow chart for all BALF samples. Using standard clinical culture procedures, 62% of BALF samples were culture positive. Of the culture positive samples, 69% had TBL ≥ 7.4 $\log_{10}$ rDNA/ml (limit of detection) whereas only 15% of culture negative samples had TBL ≥ 7.4. Sequencing was positive in 91/93 (98%) of samples with TBL ≥ 7.4 compared to 18/97 (19%) of samples with TBL < 7.4 Culture was positive in 82/93 (88%) of samples with TBL ≥ 7.4 compared to 36/97 (37%) of samples with TBL < 7.4. Only two BALF samples that were negative by sequencing had TBL ≥ 7.4, and both samples were also negative by culture.

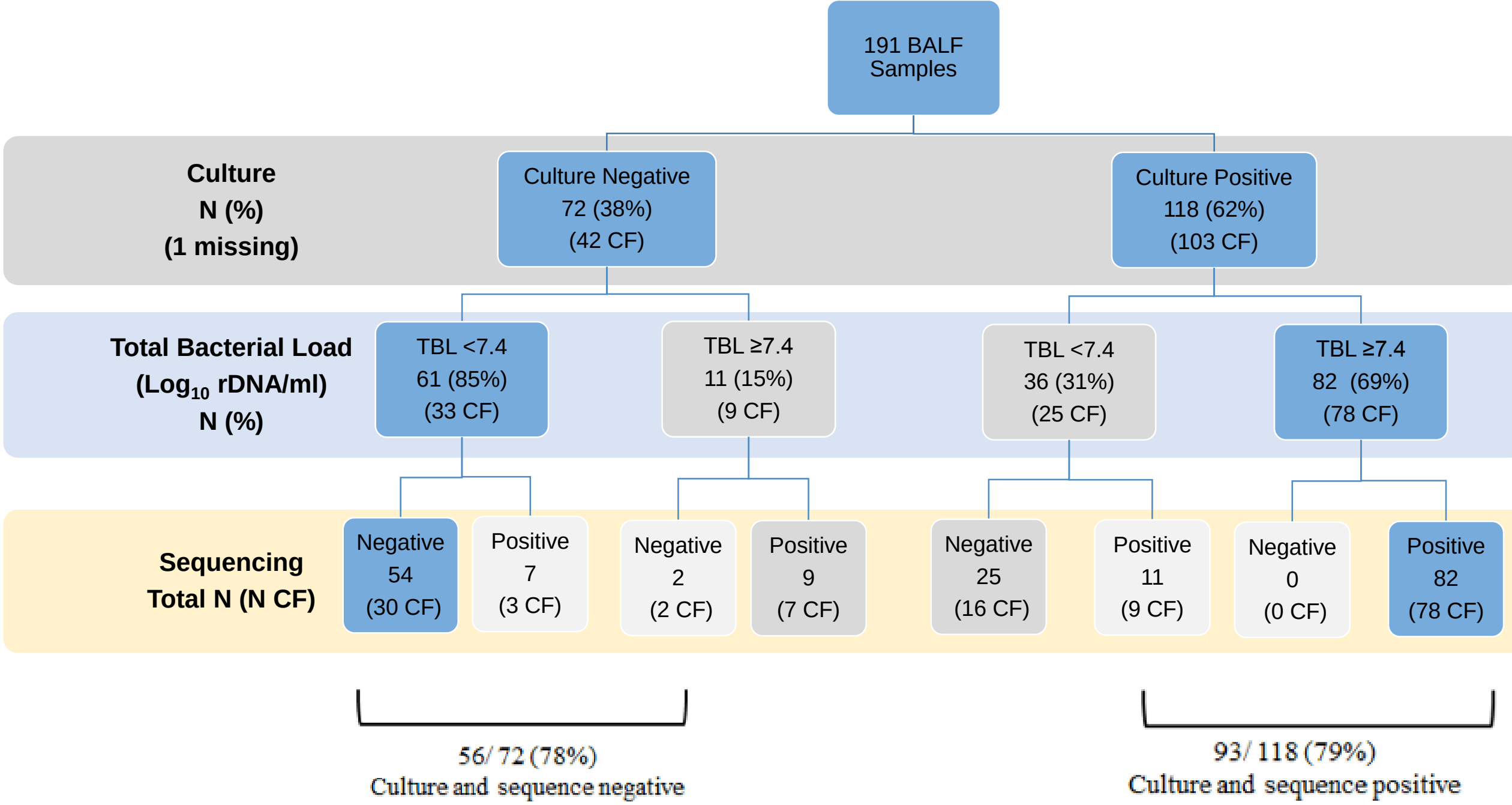

191 BALF Samples
Culture N (%) (1 missing)
Culture Negative 72 (38%) (42 CF)
Culture Positive 118 (62%) (103 CF)
Total Bacterial Load (Log10 rDNA/ml) N (%)
TBL <7.4 61 (85%) (33 CF)
TBL ≥7.4 11 (15%) (9 CF)
TBL <7.4 36 (31%) (25 CF)
TBL ≥7.4 82 (69%) (78 CF)
Sequencing Total N (N CF)
Negative 54 (30 CF)
Positive 7 (3 CF)
Negative 2 (2 CF)
Positive 9 (7 CF)
Negative 25 (16 CF)
Positive 11 (9 CF)
Negative 0 (0 CF)
Positive 82 (78 CF)
56/ 72 (78%)
Culture and sequence negative
93/ 118 (79%)
Culture and sequence positive

Figure S2. Heatmap of sequencing results for all BALF samples (DC, pediatric and adult CF) with sufficient bacterial DNA for amplification (n=109). Taxa detected in ≥0.5% RA are displayed, whereas taxa detected in <0.5% RA are grouped as "other". CF BALF samples are grouped by culture status (positive or negative).

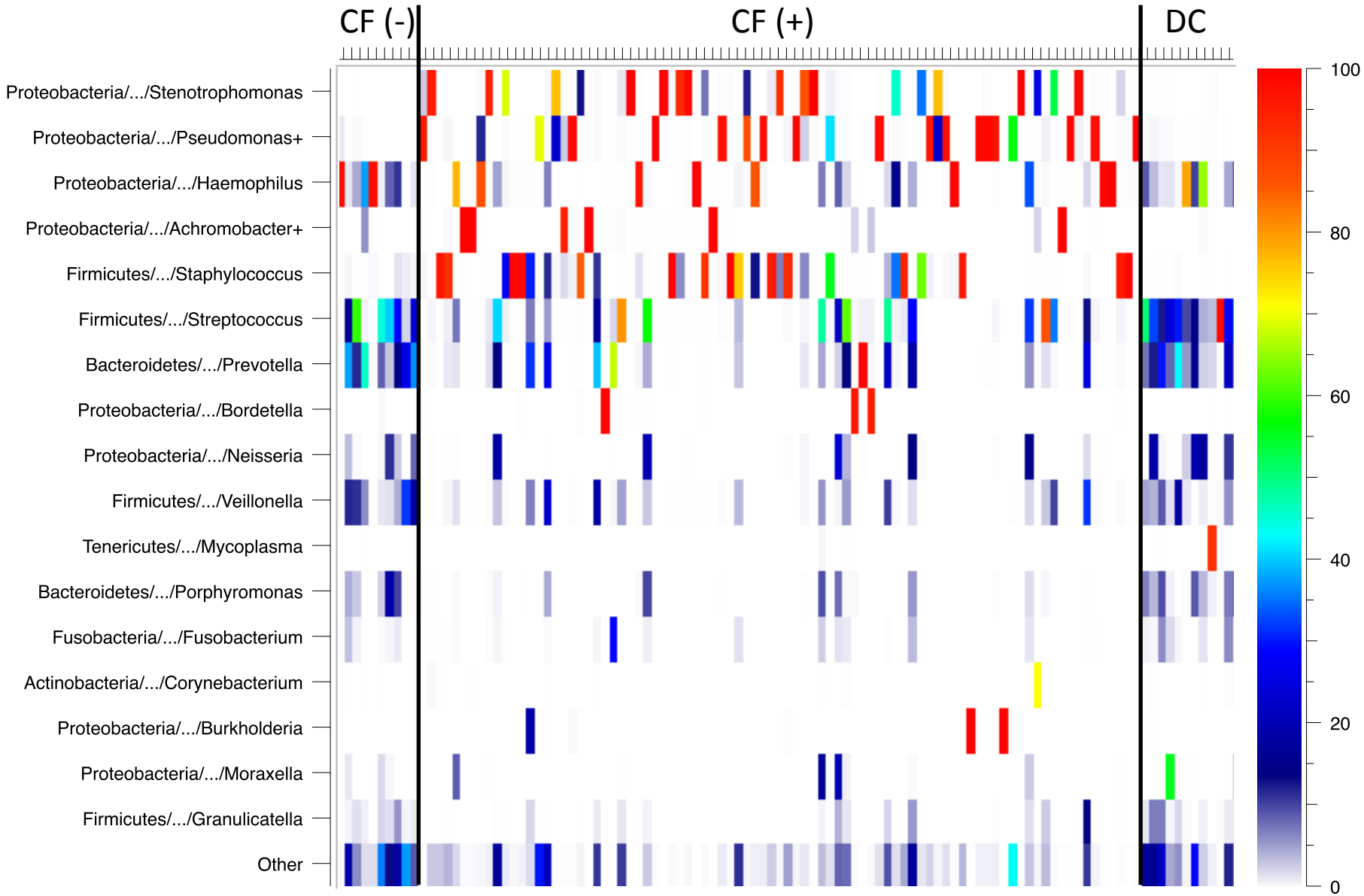
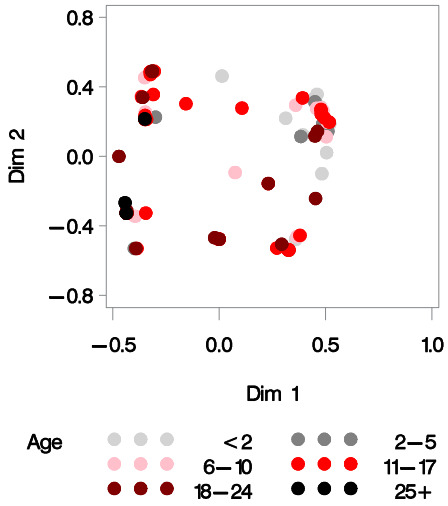
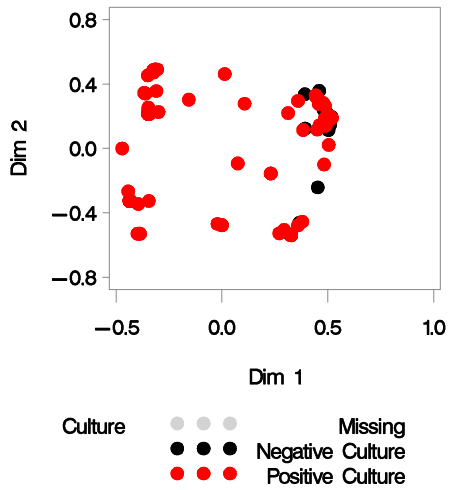

CF (-)
CF (+)
DC
Proteobacteria/.../Stenotrophomonas
Proteobacteria/.../Pseudomonas+
Proteobacteria/.../Haemophilus
Proteobacteria/.../Achromobacter+
Firmicutes/.../Staphylococcus
Firmicutes/.../Streptococcus
Bacteroidetes/.../Prevotella
Proteobacteria/.../Bordetella
Proteobacteria/.../Neisseria
Firmicutes/.../Veillonella
Tenericutes/.../Mycoplasma
Bacteroidetes/.../Porphyromonas
Fusobacteria/.../Fusobacterium
Actinobacteria/.../Corynebacterium
Proteobacteria/.../Burkholderia
Proteobacteria/.../Moraxella
Firmicutes/.../Granulicatella
Other
100
80
60
40
20
0

Figure S3: Differences in overall community structure and distance between each BALF sample shown by Principal Coordinate Analysis (PCoA). PCoA using the Morisita-Horn distance was applied to visualize the relationship between microbiota communities categorized by age, CF center, pulmonary exacerbation status, and disease status (DC or CF). Each circle represents a single BALF sample. Panels are color-coded to indicate variables of interest. Panel A: Colors indicated age of participant, grouped as in Figure 5. Panel B: Colors indicate BALF culture positivity with Red = positive culture, Black = negative culture. Panel C: Colors represent CF Center where BALF was collected; Panel D: Colors indicate approached used for BALF collection. Blue = Laryngeal mask airway (LMA); Green = Endotracheal Tube (ETT); Dark Red = Nasal; Gray = Unknown. Panel E: Colors indicate pulmonary exacerbation status. Blue = clinically stable; Black = disease controls; Dark Red = pulmonary exacerbation (PEx); Green = unknown; Clustering in the upper left corner is seen for culture negative samples. One site collected all samples using the nasal approach, and these samples are grouped together by blue in panel C and dark red in panel D. There was no clustering seen when comparing ETT to LMA approach. Clustering was also not seen by pulmonary exacerbation status.

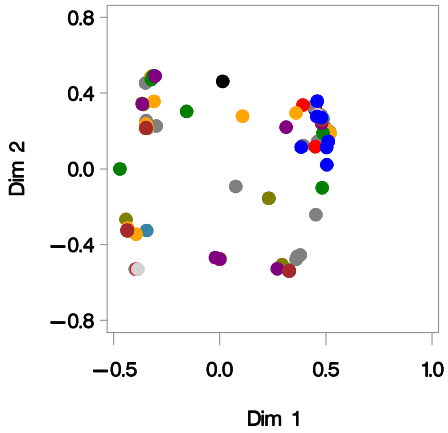
a


Age
0.8
0.4
0.0
−0.4
−0.8
Dim 2
−0.5
0.0
0.5
1.0
Dim 1
Age
<2
2–5
6–10
11–17
18–24
25+

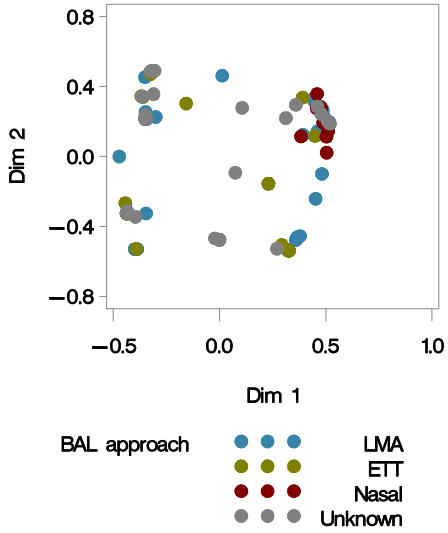
b


Culture Positivity
0.8
0.4
0.0
−0.4
−0.8
Dim 2
−0.5
0.0
0.5
1.0
Dim 1
Culture
Missing
Negative Culture
Positive Culture

c


CF Center
0.8
0.4
0.0
−0.4
−0.8
Dim 2
−0.5
0.0
0.5
1.0
Dim 1

d


BAL Approach
0.8
0.4
0.0
−0.4
−0.8
Dim 2
−0.5
0.0
0.5
1.0
Dim 1
BAL approach
LMA
ETT
Nasal
Unknown

e

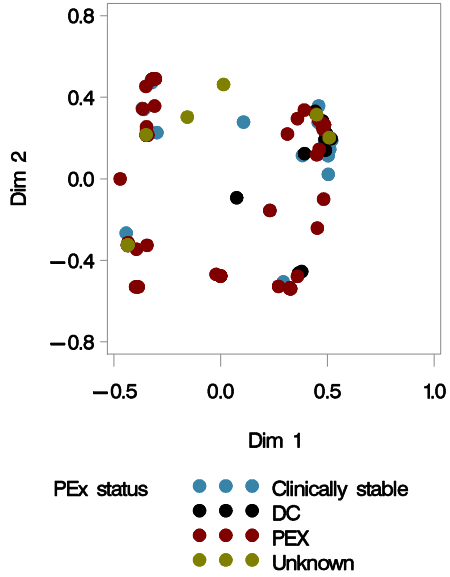

PEx Status
0.8
0.4
0.0
−0.4
−0.8
Dim 2
−0.5
0.0
0.5
1.0
Dim 1
PEx status
Clinically stable
DC
PEX
Unknown

Figure S4. Total cell count, percent neutrophils and absolute neutrophil count detected in BALF samples from DC and CF participants. Panel A: Total cell count for DC and CF BALF. Panel B: total cell count for DC and CF BALF based on positive or negative bacterial culture results. Panel C: Neutrophils (%) in DC and CF BALF. Panel D: Neutrophils (%) in DC and CF BALF based on positive or negative bacterial culture results. Panel E: Absolute neutrophil count in DC and CF BALF. Panel F: Absolute neutrophil count in DC and CF BALF based on positive or negative bacterial culture results. Total cell count, neutrophils (%) and absolute neutrophil count were highest in CF BALF with positive cultures. In negative culture samples, total cell count, neutrophil % and absolute neutrophil count remained higher in CF compared to DC. Median indicated by line. Boxes show 25-75th IQR with whiskers showing 1.5 times the IQR. P-values for statistically significant differences are indicated. CF, cystic fibrosis; DC, disease controls

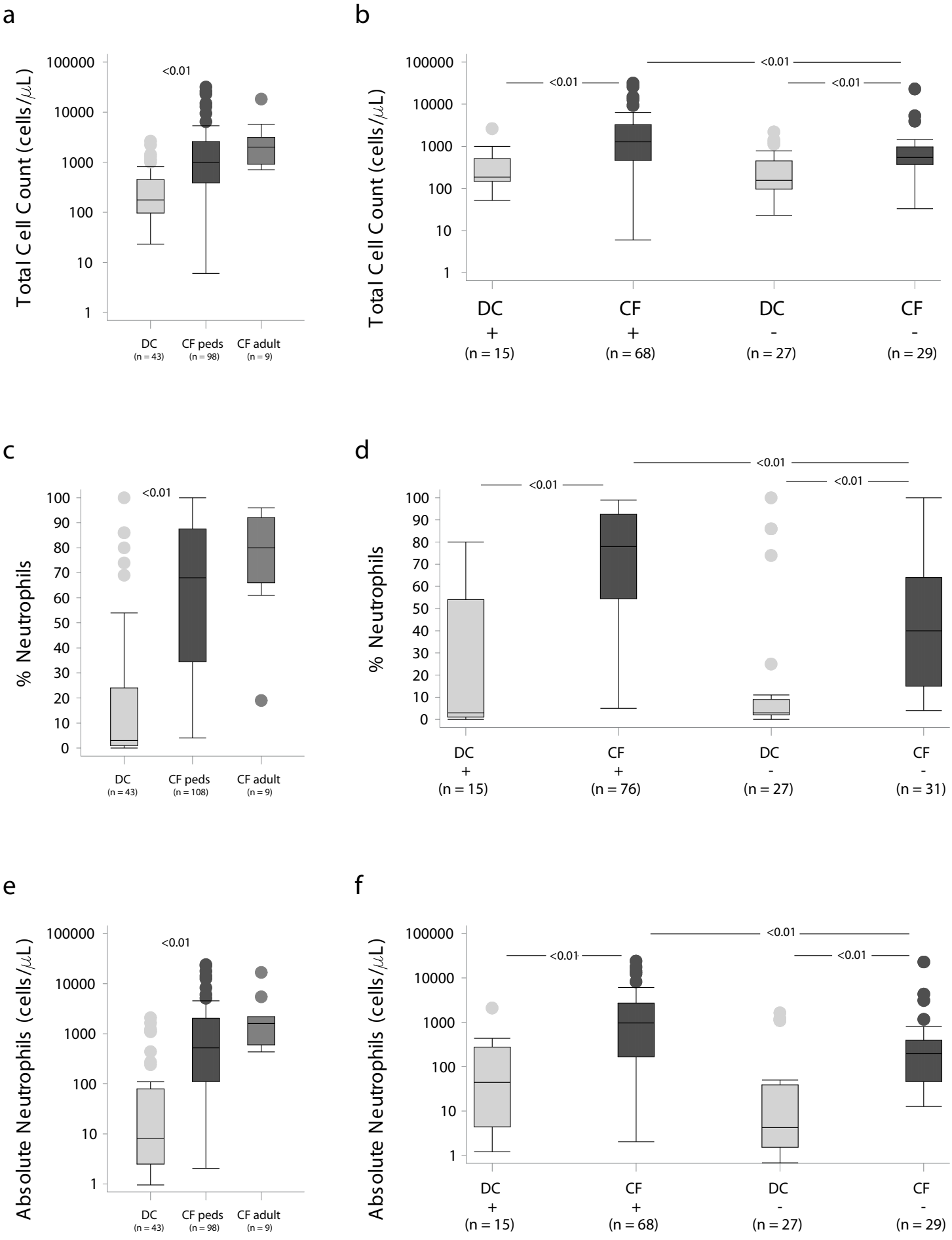
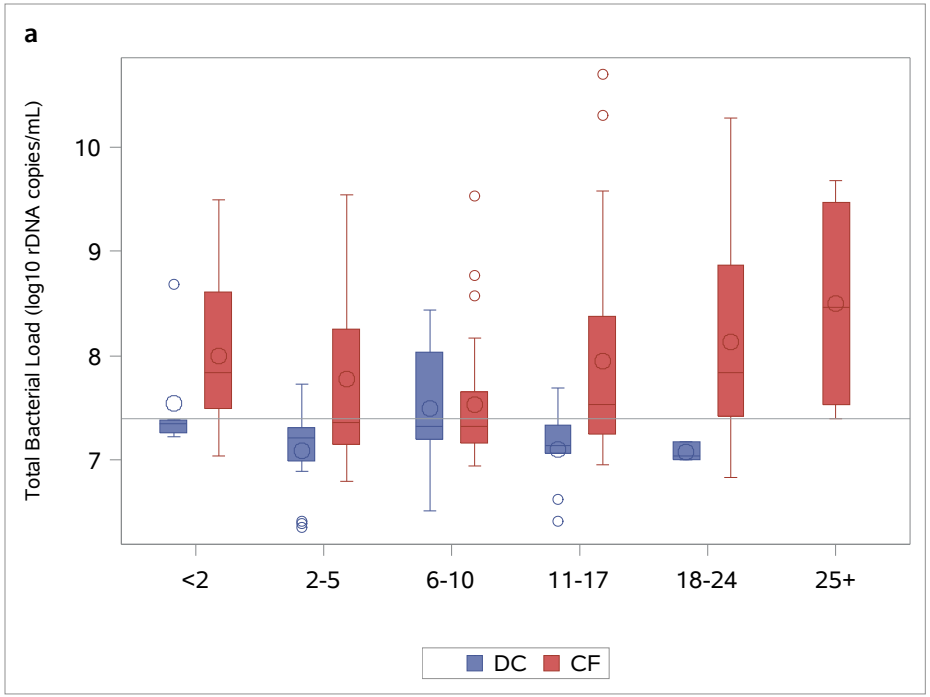

a
b
c
d
e
f
Total Cell Count (cells/μL)
% Neutrophils
Absolute Neutrophils (cells/μL)
<0.01
DC (n = 43)
CF peds (n = 98)
CF adult (n = 9)
DC + (n = 15)
CF + (n = 68)
DC - (n = 27)
CF - (n = 29)
CF peds (n = 108)
CF + (n = 76)
CF - (n = 31)

Figure S5. Cross-sectional results by age group for (A) Total bacterial load and (B) Shannon Diversity index for CF and DC. Median indicated by line, circle indicates mean. Boxes show 25-75th IQR with whiskers showing 1.5 times the IQR. Total bacterial load measured on all BALF samples. Shannon diversity calculated for samples with successful amplification and sequencing.


a
Total Bacterial Load (log10 rDNA copies/mL)
10
9
8
7
<2
2-5
6-10
11-17
18-24
25+
DC
CF

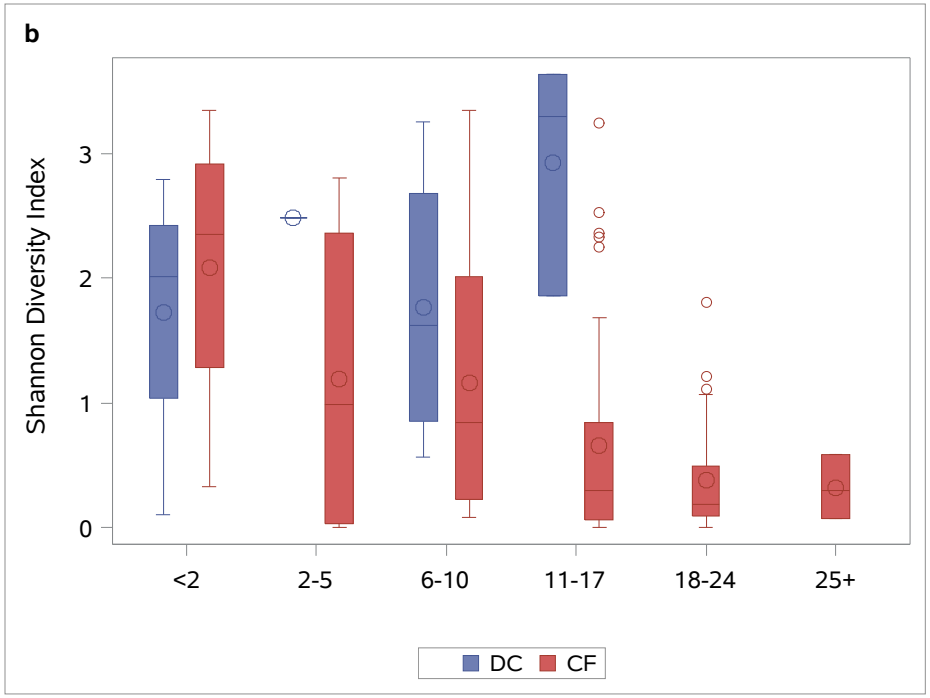
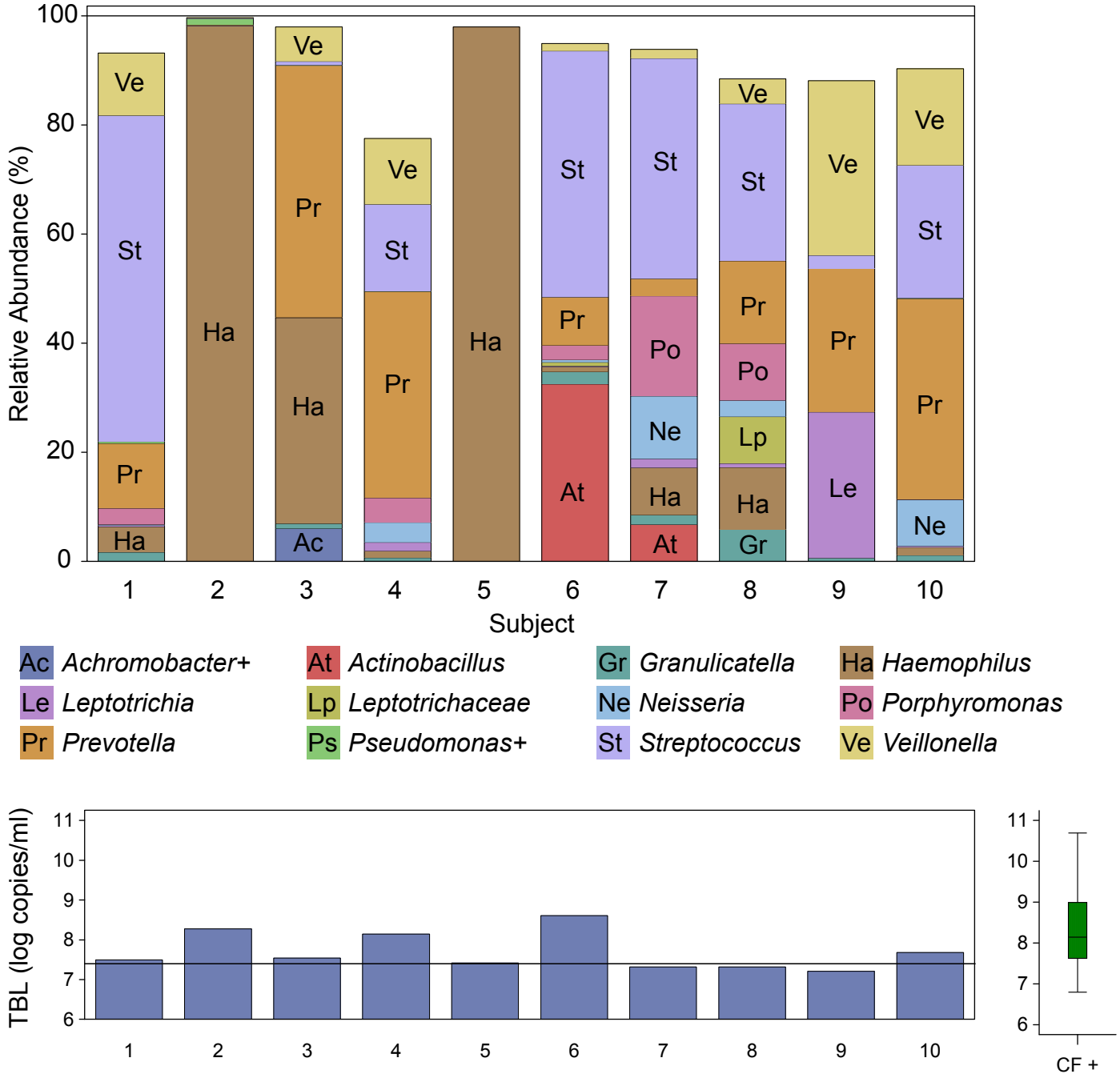

b
Shannon Diversity Index
3
2
1
0
<2
2-5
6-10
11-17
18-24
25+
DC
CF

Figure S6. Total bacterial load and sequencing results from CF culture negative samples. The top graph shows the microbiota community composition for CF BALF samples with negative bacteria cultures and adequate bacterial load for sequencing analysis. Bars show relative abundance (RA) of specific taxa. Two samples contained >98% *Haemophilus* whereas the other 7 samples contained mixed anaerobic taxa. Four samples had *Pseudomonas* detected by sequencing in very low relative abundance (range 0.006-1.13% RA). In specimens where total RA displayed is less than 100%, the proportion not displayed was made up of other taxa present in low RA (<1%). The lower graph shows the total bacterial load detected from each sample. The median bacterial load for CF BALF samples with positive cultures is shown for comparison. TBL limit of detection based on reagent control samples is indicated by gray line. Ac: Achromobacter+; At: *Actinobacillus;* Gr: *Granulicatella*; Ha: *Haemophilus*; Le: *Leptotrichia*; Lp: *Leptotrichiaceae*; Ne: *Neisseria*; Po: *Porphyromonas*; Pr: *Prevotella*; Ps: *Pseudomonas*$^{+}$, St: *Streptococcus*, Ve: *Veillonella*. *Pseudomonas*+ contains sequences assigned to the taxa *Pseudomonadales* and *Pseudomonas*. *Achromobacter*$^{+}$ contains sequences assigned to the taxa *Alcaligenaceae and Achromobacter*.


Relative Abundance (%)
100
80
60
40
20
0
1
2
3
4
5
6
7
8
9
10
Subject
Ac Achromobacter+
At Actinobacillus
Gr Granulicatella
Ha Haemophilus
Le Leptotrichia
Lp Leptotrichaceae
Ne Neisseria
Po Porphyromonas
Pr Prevotella
Ps Pseudomonas+
St Streptococcus
Ve Veillonella
TBL (log copies/ml)
11
10
9
8
7
6
CF +

References

1. Markle JG, Frank DN, Mortin-Toth S, Robertson CE, Feazel LM, Rolle-Kampczyk U, von Bergen M, McCoy KD, Macpherson AJ, Danska JS. Sex differences in the gut microbiome drive hormone-dependent regulation of autoimmunity. *Science* 2013: 339(6123): 1084-1088.
2. 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(8): 3805-3814.
3. Ahn JH, Kim BY, Song J, Weon HY. Effects of PCR cycle number and DNA polymerase type on the 16S rRNA gene pyrosequencing analysis of bacterial communities. *J Microbiol* 2012: 50(6): 1071-1074.
4. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 2012: 9(4): 357-359.
5. Homo Sapiens UCSC Hg19 Human Genome Sequence from iGenome
6. Ewing B, Green P. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res* 1998: 8(3): 186-194.
7. Ewing B, Hillier L, Wendl MC, Green P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* 1998: 8(3): 175-185.
8. Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 2011: 27(16): 2194-2200.
9. 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(10): 3219-3226.
10. Pruesse E, Peplies J, Glockner FO. SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics* 2012: 28(14): 1823-1829.
11. 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 Res* 2013: 41(Database issue): D590-596.
12. 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 metadata-driven management, analysis and visualization of microbiome data. *Bioinformatics* 2013: 29(23): 3100-3101.
13. Johnson EJ, Zemanick ET, Accurso FJ, Wagner BD, Robertson CE, Harris JK. Molecular Identification of *Staphylococcus aureus* in Airway Samples from Children with Cystic Fibrosis. *PLoS One* 2016: 11(1): e0147643.