Supplemental Methods:

Clinical data

Lung function (best of three measures) was entered as litres and transformed into percent of predicted value for age, gender, height and race based on Global Lung Initiative (GLI) reference equations [1]. Level of lung function was categorised as >80%, 40–80%, and <40% predicted. Medication use was obtained from clinical records and/or per patient report.

2.3. Bacterial isolation and identification

Patients were enrolled continuously at each site based on ability to expectorate a sufficient volume (300 mg). Typically the patients who provided sputum samples, were those who routinely expectorate sputum. In case of samples with salivary contamination, only the obvious mucus plugs were used for culture or, if insufficient material, the sample was not cultured. Standardised culture protocols were used across the three sites with training of staff from UNC and Dublin undertaken by Belfast staff at study initiation. Samples were placed in anaerobic pouches and transferred to the laboratories for processing in anaerobic cabinets. Culture was performed at each site under aerobic, microaerophilic and anaerobic conditions as follows.

Sputum and BALF samples were treated for 15 minutes with Sputalysin® (Calbiochem, La Jolla, CA) in accordance with the manufacturer's instructions, and serial 10-fold dilutions prepared in quarter-strength Ringers lactate (Oxoid, Basingstoke, UK) supplemented with 0.05% (wt/vol) L-cysteine (Sigma-Aldrich,Dorset,UK). One hundred–microliter aliquots were spread plated onto the agars and incubated as indicated in the table. After incubation, the total viable counts of each distinct colony type were determined. For each condition (medium and aerobic and anaerobic growth), isolates of each distinct colony type were subcultured under the appropriate conditions and were identified by full-length 16s rRNA sequencing.

Table S1: Incubation conditions for each medium and oxygen status

Target organism	**Media**	**Incubation**		
		Temp (°C)	**Atmosphere**	**Time**
Aerobes	ABA	35-37	air	2-3 days
Microaerophiles e.g. *Streptococci (milleri group)*	ABA, McKay agar	35-37	5% CO_2	2-3 days
Haemophilus influenzae	BCA	35-37	5% CO_2	2-3 days
Anaerobes	ABA, KVLB	35-37	anaerobic	5-7 days

ABA, Anaerobe Basal Agar; BCA, Chocolate blood agar containing bacitracin; KVLB, Kanamycin-vancomycin laked blood agar

Strict anaerobes were defined as those known not to survive under atmospheric oxygen tension. Facultative anaerobes i.e. those able to grow under aerobic or anaerobic conditions were included as aerobic bacteria, as were any typical aerobic genera growing on anaerobic culture conditions e.g. *Pseudomonas*.
All bacteria detected were quantified (colony forming units/gram sputum; CFU /g or CFU/mL in BAL) by total viable count (TVC) and identified by full length16S rRNA sequencing at a central laboratory to the species level. For statistical analyses genus level results are reported to allow sufficient numbers per group.

2.4. Statistical analysis

Multivariable logistic regression was used to test for characteristics that may predict prevalence of certain bacterial genera. Predictors were identified via stepwise model selection (with significance level for entry or staying in the model of p=0.10 and p=0.05, respectively). Potential predictors included center, gender, continuous age, FEV_1, BMI, number of F508 del alleles, pancreatic status, and prescription of nine chronic medications. To include as many participants as possible, stepwise selection was also performed excluding characteristics with >5 participants with missing values (FEV_1, BMI, and inhaled corticosteroids) when these factors were not selected by the initial analyses. Final models were then fit including only selected predictors.
A sensitivity analysis was also conducted with significance for entry and staying in the model at p=0.1 to evaluate for potential important predictors with a strong odds ratio that nearly missed the $p < 0.05$ criteria. (**Table S6)**.

Analyses that included bacterial quantities were log transformed to account for non-normal distribution using $\log_{10}$(CFU/g +1), to incorporate samples with undetectable bacteria. Diversity measures included richness (number of counted taxa) and evenness and diversity (Shannon-Wiener index). Analyses of bacterial communities (network analyses and factor analysis) were conducted on sputum bacterial quantity for the sixteen most common genres (≥5 % of samples). By applying a filtering step prior to the generation of any potential co-occurrence networks, we reduced both the overall complexity of the data due to rare taxa, as well as the effect of false-positive correlations arising from spurious associations caused by poorly represented genera within the dataset. Network, co-occurrence analysis was generated by calculating all possible Spearman´s rank correlation coefficients (ρ) between the pairs of retained genera. The resulting network inference was generated by calculating all possible Spearman´s rank correlation coefficients (ρ) between the pairs of retained genera. Furthermore, to reduce the likelihood of potential false-positive and/or spurious associations between corresponding taxon pairs, correction for multiple testing was performed by Benjamini-Hochberg-Yekutieli false discovery rate (FDR) correction on the original p-values [3]. Valid co-occurrence, or mutual exclusion (negative association), between two different taxa if the

Spearman's correlation coefficient (ρ) were defined as both >0.2 (positive association), or <-0.2 (negative association), and an adjusted p<0.05. For the final reconstructed co-occurrence network; all nodes represent taxa classified as belonging to the same genera, with the edges (i.e., connections) corresponding to a significant correlation between nodes (i.e., taxa; based on ρ and significance according to the adjusted p-value).

Factor analysis was conducted on quantity of the sixteen most prevalent bacteria. Factor analysis is an unsupervised, exploratory method that can be used to group variables into a smaller number of latent constructs. In this case, genera quantities were tested to obtain co-occurring genera groups, where related bacteria have high loading/weighting on one or more factors. The varimax rotation aligns the factor structure to be orthogonal, so that, as much as possible, each variable occurs in only one factor, and the factors are independent. Genera were assigned to a factor based on the highest magnitude of loading across factors. Factor analyses was conducted using JMP. Factor groups were compared for markers of disease severity using chi-square test for categorical and ANOVA for continuous variables to test of bacterial networks associated with clinical characteristics.

Analyses were conducted using SAS 9.2, factor analysis was performed in JMP® Pro 12.0.1 (SAS). Graphs were prepared in GraphPad Prism 7.02. Co-occurrence network analyses were performed in the R environment (http://www.R-project.org) using vegan (version 2.4-1) and igraph (version 1.1.2) [4]. Post analysis and visualisation of the resulting co-occurrence network was performed within the Gephi package (release 0.9.1) [5]. A Spearman rank correlation coefficient (ρ) was calculated to measure the strength of association between different taxa, as implemented in the Hmisc (version 3.12-2) in R [6].

RESULTS: Supplemental Tables and Figures

Table S2A: Patient Demographics by site for sputum samples

	Overall N=200 N (%)	*Clinical Site B N=75 N (%)*	*C N=41 N (%)*	*D N=84 N (%)*	*p values*[1]
Age(years)					
Mean	26.1	29.1	21.5	25.7	
Std Dev	10.5	12.2	8.6	8.7	
Median	23.9	26.2	19.9	23.5	
Minimum	8.3	8.3	9.1	12.4	
Maximum	68.2	68.2	50.0	61.2	
Age(years)	200				0.04
0–<6	0 (0%)	0 (0%)	0 (0%)	0 (0%)	
6–<13	12 (6%)	4 (5%)	6 (15%)	2 (2%)	
13–<18	22 (11%)	6 (8%)	6 (15%)	10 (12%)	
18–<25	75 (38%)	22 (29%)	18 (44%)	35 (42%)	
25–<30	43 (22%)	19 (25%)	5 (12%)	19 (23%)	
30+	48 (24%)	24 (32%)	6 (15%)	18 (21%)	
Gender	200	75	41	84	0.30
Female	86 (43%)	29 (39%)	22 (54%)	35 (42%)	
Male	114 (57%)	46 (61%)	19 (46%)	49 (58%)	
BMI	193	74	41	78	0.008
undernourished	24 (12%)	3 (4%)	7 (17%)	14 (18%)	
acceptable	96 (50%)	34 (46%)	24 (58%)	38 (49%)	
well nourished	73 (38%)	37 (50%)	10 (24%)	26 (33%)	
F508 del mutation	198	74	40	84	0.001
Homozygote	98 (49%)	29 (39%)	30 (75%)	39 (46%)	
Heterozygote	84 (42%)	34 (46%)	10 (25%)	40 (48%)	
None	16 (8%)	11 (15%)	0 (0%)	5 (6%)	
Pancreatic status	200	75	41	115	0.93
PI	172 (86%)	64 (85%)	35 (85%)	73 (87%)	
PS	28 (14%)	11 (15%)	6 (15%)	11 (13%)	
					0.02
FEV_1 % predicted GLI	192	74	38	80	
<41%	42 (22%)	10 (14%)	5 (13%)	27 (34%)	
41–80%	108 (56%)	48 (65%)	24 (63%)	36 (45%)	
>80%	42 (22%)	16 (22%)	9 (24%)	17 (21%)	
Chronic antibiotics	154 (77%)	58 (77%)	35 (85%)	61 (73%)	0.30
Flucloxacillin	11 (6%)	8 (11%)	0 (0%)	3 (4%)	0.03
Azithromycin	120 (60%)	44 (59%)	32 (78%)	44 (52%)	0.02
Inhaled antibiotics[2]	118 (60%)	42 (56%)	23 (56%)	53 (65%)	0.41
Any mucolytic	174 (87%)	59 (79%)	39 (95%)	76 (90%)	0.02
DNAse	151 (76%)	54 (72%)	33 (41%)	64 (76%)	0.59
Hypertonic saline	82 (41%)	15 (20%)	26 (63%)	41 (49%)	<0.0001
Inhaled corticosteroids	92 (49%)	31 (41%)	36 (55%)	38 (33%)	0.01
Antacid[3]	105 (53%)	32 (43%)	26 (70%)	35 (47%)	0.01
Insulin	31 (16%)	7 (9%)	6 (15%)	18 (21%)	0.12

B=Belfast, Northern Ireland; C=Chapel Hill, USA; D=Dublin, Ireland. PS=pancreatic sufficiency. PI=pancreatic insufficiency. BMI=body mass index. FEV_1-forced expiratory volume in 1 second expressed as % predicted based on GLI=global lung initiative reference values[1]. AZM=azithromycin.
1: Comparisons across sites using chi- square, Mantel-Haenszel mean score chi-square, or Fisher's exact test.
2: tobramycin, colistin and aztreonam. 3: antacids, H2-blockers and proton pump inhibitors.

Table S2B: Patient demographics by site for bronchoalveolar lavage (BAL) samples

	Overall N=55 N (%)	*C N=24 N (%)*	*D N=31 N (%)*	*p values*[1]
Age(years)				
Mean	6.6	6.7	6.5	
Std Dev	6.9	4.7	8.2	
Median	4.0	5.0	3.9	
Minimum	1.0	1.0	1.0	
Maximum	33.8	15.0	33.8	
Age(years)	55	24	31	0.03
0–<6	39 (71%)	14 (58%)	25 (81%)	
6–<13	9 (16%)	7 (29%)	2 (6%)	
13–<18	3 (5%)	3 (13%)	0 (0%)	
18–<25	2 (4%)	0 (0%)	2 (6%)	
25–<30	1 (2%)	0 (0%)	1 (3%)	
30+	1 (2%)	0 (0%)	1 (3%)	
Gender	55	24	31	0.59
Female	29 (53%)	14 (58%)	15 (48%)	
Male	26 (47%)	10 (42%)	16 (52%)	
BMI	42	24	18	0.52
undernourished	2 (5%)	2 (8%)	0 (0%)	
acceptable	22 (52%)	13 (54%)	9 (50%)	
well nourished	18 (43%)	9 (38%)	9 (50%)	
F508 del mutation	53	23	30	0.54
Homozygote	33 (62%)	15 (65%)	18 (60%)	
Heterozygote	15 (28%)	5 (22%)	10 (33%)	
None	5 (9%)	3 (13%)	2 (7%)	
Pancreatic status	55	24	31	0.12
PI	51 (93%)	(100%)	27 (87%)	
PS	4 (7%)	0 (0%)	4 (13%)	
FEV_1 % predicted GLI (n)	20	13	7	
<41%	0 (0%)	0 (0%)	0 (0%)	0.66
41–80%	7 (35%)	4 (31%)	3 (43%)	
>80%	13 (65%)	9 (69%)	4 (57%)	
Chronic antibiotics	19 (35%)	8 (33%)	11 (35%)	1.00
Flucloxacillin	7 (12%)	1 (4%)	6 (19%)	0.12
Azithromycin	7 (13%)	5 (21%)	2 (6%)	0.22
Inhaled antibiotics[2]	12 (22%)	7 (30%)	5 (16%)	0.32
Any mucolytic	31 (56%)	15 (63%)	16 (52%)	0.59
DNAse	13 (24%)	10 (42%)	3 (10%)	0.01
Hypertonic saline	27 (49%)	11 (46%)	16 (52%)	0.79
Inhaled corticosteroids	13 (27%)	10 (50%)	3 (10%)	0.003
Antacid[3]	24 (44%)	17 (71%)	7 (23%)	.0008
Insulin	3 (5%)	2 (8%)	1 (3%)	0.57

B=Belfast, Northern Ireland; C=Chapel Hill, USA; D=Dublin, Ireland. PS=pancreatic sufficiency. PI=pancreatic insufficiency. BMI=body mass index. FEV_1-forced expiratory volume in 1 second expressed as % predicted based on GLI=global lung initiative reference values [1]. AZM=azithromycin.
1: Comparisons across sites using chi- square, Mantel-Haenszel mean score chi-square, or Fisher's exact test.
2: tobramycin, colistin and aztreonam. 3: antacids, H2-blockers and proton pump inhibitors.

Table S3 Prevalence of bacterial genera across study sites

Organism	Total N=255 N (%)	B N=75 N (%)	C N=65 N (%)	D N=115 N (%)	p-value[1]
Any anaerobe	151 (59%)	63 (84%)	40 (62%)	48 (42%)	<.001
Any aerobe	248 (97%)	75 (100%)	62 (95%)	111 (97%)	0.180
Streptococcus	209 (82%)	75 (100%)	45 (69%)	89 (77%)	<.001
Prevotella [A]	129 (51%)	54 (72%)	37 (57%)	38 (33%)	<.001
Pseudomonas	120 (47%)	34 (45%)	36 (55%)	50 (43%)	0.300
Staphylococcus	117 (46%)	41 (55%)	31 (48%)	45 (39%)	0.100
Rothia	106 (42%)	62 (83%)	14 (22%)	30 (26%)	<.001
Actinomyces	70 (27%)	42 (56%)	8 (12%)	20 (17%)	<.001
Haemophilus	64 (25%)	29 (39%)	15 (23%)	20 (17%)	0.004
Gemella	58 (23%)	42 (56%)	7 (11%)	9 (8%)	<.001
Veillonella [A]	57 (22%)	32 (43%)	15 (23%)	10 (9%)	<.001
Neisseria	45 (18%)	6 (8%)	17 (26%)	22 (19%)	0.010
Stenotrophomonas	23 (9%)	5 (7%)	8 (12%)	10 (9%)	0.500
Lactobacillus	22 (9%)	13 (17%)	2 (3%)	7 (6%)	0.007
Fusobacterium [A]	19 (7%)	5 (7%)	8 (12%)	6 (5%)	0.230
Granulicatella	14 (5%)	9 (12%)	0 (0%)	5 (4%)	0.006
Burkholderia	13 (5%)	9 (12%)	3 (5%)	1 (<1%)	0.002
Porphyromonas[A]	13 (5%)	4 (5%)	8 (12%)	1 (<1%)	0.002
Capnocytophaga	9 (4%)	5 (7%)	4 (6%)	0 (0%)	0.006
Escherichia	9 (4%)	1 (1%)	3 (5%)	5 (4%)	0.470
Enterococcus	8 (3%)	4 (5%)	1 (2%)	3 (3%)	0.470
Atopobium [A]	6 (2%)	5 (7%)	0 (0%)	1 (<1%)	0.020
Bacillus	5 (2%)	2 (3%)	2 (3%)	1 (<1%)	0.530
Leptotrichia [A]	5 (2%)	2 (3%)	1 (2%)	2 (2%)	0.860
Achromobacter	4 (2%)	1 (1%)	2 (3%)	1 (<1%)	0.480
Corynebacterium	4 (2%)	3 (4%)	0 (0%)	1 (<1%)	0.200
Eubacterium [A]	4 (2%)	2 (3%)	2 (3%)	0 (0%)	0.120
Parvimonas [A]	4 (2%)	1 (1%)	1 (2%)	2 (2%)	1.000
Peptostreptococcus [A]	4 (2%)	1 (1%)	2 (3%)	1 (<1%)	0.480
Propionibacterium [A]	4 (2%)	3 (4%)	0 (0%)	1 (<1%)	0.200
Aggregatibacter	3 (1%)	3 (4%)	0 (0%)	0 (0%)	0.040
Moraxella	3 (1%)	2 (3%)	0 (0%)	1 (<1%)	0.460
Abiotrophia	2 (<1%)	0 (0%)	1 (2%)	1 (<1%)	0.730
Actinobacillus	2 (<1%)	2 (3%)	0 (0%)	0 (0%)	0.150
Bordetella	2 (<1%)	1 (1%)	0 (0%)	1 (<1%)	1.000
Brevibacterium	2 (<1%)	1 (1%)	0 (0%)	1 (<1%)	1.000
Campylobacter	2 (<1%)	1 (1%)	1 (2%)	0 (0%)	0.300
Cardiobacterium	2 (<1%)	2 (3%)	0 (0%)	0 (0%)	0.150
Enterobacter	2 (<1%)	0 (0%)	2 (3%)	0 (0%)	0.060
Kocuria	2 (<1%)	2 (3%)	0 (0%)	0 (0%)	0.150
Megasphaera [A]	2 (<1%)	0 (0%)	1 (2%)	1 (<1%)	0.730
Micrococcus	2 (<1%)	1 (1%)	0 (0%)	1 (<1%)	1.000
Mogibacterium [A]	2 (<1%)	0 (0%)	2 (3%)	0 (0%)	0.060
Ralstonia	2 (<1%)	0 (0%)	1 (2%)	1 (<1%)	0.730
Acinetobacter	1 (<1%)	0 (0%)	1 (2%)	0 (0%)	0.250
Aerococcus	1 (<1%)	0 (0%)	0 (0%)	1 (<1%)	1.000

Organism	Total N=255 N (%)	B N=75 N (%)	C N=65 N (%)	D N=115 N (%)	p-value[1]
Alloscardovia	1 (<1%)	1 (1%)	0 (0%)	0 (0%)	0.550
Cronobacter	1 (<1%)	1 (1%)	0 (0%)	0 (0%)	0.550
Cryptobacterium [A]	1 (<1%)	1 (1%)	0 (0%)	0 (0%)	0.550
Dermacoccus	1 (<1%)	1 (1%)	0 (0%)	0 (0%)	0.550
Eikenella	1 (<1%)	0 (0%)	1 (2%)	0 (0%)	0.250
Mycobacterium	1 (<1%)	0 (0%)	1 (2%)	0 (0%)	0.250
Olsenella [A]	1 (<1%)	1 (1%)	0 (0%)	0 (0%)	0.550
Peptoniphilus [A]	1 (<1%)	1 (1%)	0 (0%)	0 (0%)	0.550
Proteus	1 (<1%)	1 (1%)	0 (0%)	0 (0%)	0.550
Selenomonas [A]	1 (<1%)	1 (1%)	0 (0%)	0 (0%)	0.550
Shuttleworthi [A]	1 (<1%)	1 (1%)	0 (0%)	0 (0%)	0.550
Solobacterium [A]	1 (<1%)	1 (1%)	0 (0%)	0 (0%)	0.550

B=Belfast, Northern Ireland; C=Chapel Hill, USA; D=Dublin, Ireland. Genera are sorted by overall prevalence. [A] denotes obligate anaerobe.

1: Comparison between sites was by Fisher's exact test.

Table S4: Species level prevalence across all sites

Species	Total n=255 N (%)	Species	Total n=255 N (%)
Any aerobe	248 (97%)		
Any anaerobe	151 (59%)		
Abiotrophia defectiva	2 (<1%)	*Leptotrichia buccalis*	1 (<1%)
Achromobacter spanius	1 (<1%)	*Leptotrichia goodfellowii*	1 (<1%)
Achromobacter xylosoxidans	3 (1%)	*Leptotrichia spp.*	1 (<1%)
Acinetobacter baumannii	1 (<1%)	*Leptotrichia wadei*	3 (1%)
Actinobacillus minor	2 (<1%)	*Megasphaera micronuciformis*	2 (<1%)
Actinomyces graevenitzii	8 (3%)	*Micrococcus luteus*	2 (<1%)
Actinomyces linginae	1 (<1%)	*Mogibacterium neglectum*	2 (<1%)
Actinomyces massiliensis	1 (<1%)	*Moraxella catarrhalis*	3 (1%)
Actinomyces meyeri	3 (1%)	*Mycobacterium abscessus*	1 (<1%)
Actinomyces naeslundii	21 (8%)	*Neisseria lactamica*	2 (<1%)
Actinomyces odontolyticus	40 (16%)	*Neisseria polysaccharea*	5 (2%)
Actinomyces oris	2 (<1%)	*Neisseria subflava*	42 (16%)
Actinomyces viscosus	14 (5%)	*Olsenella profusa*	1 (<1%)
Aerococcus urinaeequi	1 (<1%)	*Parvimonas micra*	4 (2%)
Aggregatibacter aphrophilus	3 (1%)	*Peptoniphilus lacrimalis*	1 (<1%)
Alloscardovia omnicolens	1 (<1%)	*Peptostreptococcus stomatis*	4 (2%)
Atopobium parvulum	3 (1%)	*Porphyromonas catoniae*	10 (4%)
Atopobium rimae	3 (1%)	*Porphyromonas uenonis*	3 (1%)
Bacillus cereus	1 (<1%)	*Prevotella bivia*	1 (<1%)
Bacillus infantis	1 (<1%)	*Prevotella buccae*	1 (<1%)
Bacillus simplex	1 (<1%)	*Prevotella denticola*	10 (4%)
Bacillus spp.	1 (<1%)	*Prevotella disiens*	1 (<1%)
Bacillus subtilis	1 (<1%)	*Prevotella histicola*	59 (23%)
Bordetella hinzii	1 (<1%)	*Prevotella loescheii*	3 (1%)
Bordetella petrii	1 (<1%)	*Prevotella melaninogenica*	67 (26%)
Brevibacterium frigoritolerans	1 (<1%)	*Prevotella nanceiensis*	12 (5%)
Brevibacterium sanguinis	1 (<1%)	*Prevotella nigrescens*	19 (7%)
Burkholderia cenocepacia	2 (<1%)	*Prevotella oralis*	1 (<1%)
Burkholderia cepacia	2 (<1%)	*Prevotella oris*	4 (2%)
Burkholderia diffusa	1 (<1%)	*Prevotella pallens*	9 (4%)
Burkholderia gladioli	1 (<1%)	*Prevotella salivae*	20 (8%)
Burkholderia lata	1 (<1%)	*Prevotella tannerae*	1 (<1%)
Burkholderia latens	1 (<1%)	*Prevotella veroralis*	4 (2%)
Burkholderia multivorans	6 (2%)	*Propionibacterium acidifaciens*	3 (1%)
Burkholderia plantarii	1 (<1%)	*Propionibacterium acnes*	1 (<1%)
Burkholderia vietnamiensis	1 (<1%)	*Proteus mirabilis*	1 (<1%)
Campylobacter showae	2 (<1%)	*Pseudomonas aeruginosa*	117 (46%)
Capnocytophaga gingivalis	3 (1%)	*Pseudomonas beteli*	1 (<1%)

Capnocytophaga ochracea	3 (1%)	*Pseudomonas geniculata*	2 (<1%)
Capnocytophaga spp.	1 (<1%)	*Pseudomonas hibiscicola*	1 (<1%)
Capnocytophaga sputigena	2 (<1%)	*Ralstonia mannitolilytica*	2 (<1%)
Cardiobacterium hominis	2 (<1%)	*Rothia aeria*	15 (6%)
Corynebacterium coyleae	1 (<1%)	*Rothia dentocariosa*	71 (28%)
Corynebacterium diphtheriae	1 (<1%)	*Rothia mucilaginosa*	58 (23%)
Corynebacterium minutissimum	1 (<1%)	*Selenomonas artemidis*	1 (<1%)
Corynebacterium propinquum	1 (<1%)	*Shuttleworthia satelles*	1 (<1%)
Cronobacter sakazakii	1 (<1%)	*Solobacterium moorei*	1 (<1%)
Cryptobacterium curtum	1 (<1%)	*Staphylococcus aureus*	97 (38%)
Dermacoccus nishinomiyaensis	1 (<1%)	*Staphylococcus capitis*	2 (<1%)
Eikenella corrodens	1 (<1%)	*Staphylococcus caprae*	1 (<1%)
Enterobacter hormaechei	2 (<1%)	*Staphylococcus epidermidis*	15 (6%)
Enterococcus faecalis	7 (3%)	*Staphylococcus haemolyticus*	2 (<1%)
Enterococcus faecium	1 (<1%)	*Staphylococcus hominis*	3 (1%)
Escherichia coli	1 (<1%)	*Staphylococcus lugdunensis*	1 (<1%)
Escherichia flexneri	8 (3%)	*Staphylococcus pasteuri*	1 (<1%)
Eubacterium brachy	2 (<1%)	*Stenotrophomonas maltophilia*	23 (9%)
Eubacterium saburreum	2 (<1%)	*Streptococcus agalactiae*	4 (2%)
Fusobacterium gonidiaformans	1 (<1%)	*Streptococcus anginosus*	55 (22%)
Fusobacterium nucleatum	10 (4%)	*Streptococcus australis*	14 (5%)
Fusobacterium periodonticum	8 (3%)	*Streptococcus constellatus*	6 (2%)
Gemella elegans	1 (<1%)	*Streptococcus cristatus*	4 (2%)
Gemella haemolysans	33 (13%)	*Streptococcus dysgalactiae*	1 (<1%)
Gemella morbillorum	2 (<1%)	*Streptococcus gordonii*	15 (6%)
Gemella sanguinis	28 (11%)	*Streptococcus infantis*	18 (7%)
Gemella spp.	1 (<1%)	*Streptococcus intermedius*	13 (5%)
Granulicatella adiacens	11 (4%)	*Streptococcus mitis*	116 (45%)
Granulicatella elegans	3 (1%)	*Streptococcus mutans*	7 (3%)
Haemophilus aegyptius	9 (4%)	*Streptococcus oralis*	14 (5%)
Haemophilus influenzae	23 (9%)	*Streptococcus parasanguinis*	50 (20%)
Haemophilus parahaemolyticus	1 (<1%)	*Streptococcus peroris*	1 (<1%)
Haemophilus parainfluenzae	39 (15%)	*Streptococcus pneumoniae*	3 (1%)
Haemophilus paraphrohaemolyticus	2 (<1%)	*Streptococcus pseudopneumoniae*	23 (9%)
Haemophilus pittmaniae	4 (2%)	*Streptococcus pyogenes*	2 (<1%)
Haemophilus segnis	1 (<1%)	*Streptococcus salivarius*	131 (51%)
Kocuria rhizophila	2 (<1%)	*Streptococcus sanguinis*	45 (18%)
Lactobacillus casei	1 (<1%)	*Streptococcus spp.*	2 (<1%)
Lactobacillus crispatus	1 (<1%)	*Streptococcus vestibularis*	5 (2%)
Lactobacillus gasseri	4 (2%)	*Streptococcus warneri*	1 (<1%)
Lactobacillus gastricus	1 (<1%)	*Veillonella atypica*	18 (7%)
Lactobacillus paracasei	5 (2%)	*Veillonella dispar*	27 (11%)
Lactobacillus pentosus	1 (<1%)	*Veillonella parvula*	9 (4%)
Lactobacillus rhamnosus	6 (2%)	*Veillonella rogosae*	4 (2%)

Lactobacillus salivarius	3 (1%)	*Veillonella spp.*	6 (2%)
Lactobacillus zeae	1 (<1%)		

Species are sorted alphabetically by genus name.

Table S5: **Prevalence by age of genera present in ≥5% of BAL and sputum samples**

Genus	**Total**	**0-<6**	**6-<13**	**13-<18**	**18-<25**	**25-<30**	**30+**	**p-value**
	N=255	**N=39**	**N=21**	**N=25**	**N=77**	**N=44**	**N=49**	**Over all***
	N (%)	**N (%)**	**N (%)**	**N (%)**	**N (%)**	**N (%)**	**N (%)**	
Any aerobe	248 (97%)	35 (90%)	21 (100%)	25 (100%)	76 (99%)	43 (98%)	48 (98%)	0.07
Any anaerobe	151 (59%)	12 (31%)	13 (62%)	16 (64%)	42 (55%)	31 (70%)	37 (76%)	<.001
Streptococcus	209 (82%)	22 (56%)	15 (71%)	19 (76%)	64 (83%)	42 (95%)	47 (96%)	**<.001**
Prevotella [A]	129 (51%)	8 (21%)	12 (57%)	13 (52%)	41 (53%)	28 (64%)	27 (55%)	**0.001**
Pseudomonas	120 (47%)	11 (28%)	4 (19%)	13 (52%)	45 (58%)	23 (52%)	24 (49%)	**0.005**
Staphylococcus	117 (46%)	10 (26%)	12 (57%)	16 (64%)	36 (47%)	23 (52%)	20 (41%)	0.35
Rothia	106 (42%)	8 (21%)	6 (29%)	12 (48%)	30 (39%)	25 (57%)	25 (51%)	**<.001**
Actinomyces	70 (27%)	7 (18%)	4 (19%)	5 (20%)	23 (30%)	14 (32%)	17 (35%)	**0.03**
Haemophilus	64 (25%)	13 (33%)	9 (43%)	3 (12%)	16 (21%)	13 (30%)	10 (20%)	0.15
Gemella	58 (23%)	2 (5%)	3 (14%)	5 (20%)	21 (27%)	10 (23%)	17 (35%)	**<.001**
Veillonella [A]	57 (22%)	4 (10%)	4 (19%)	5 (20%)	15 (19%)	9 (20%)	20 (41%)	**0.002**
Neisseria	45 (18%)	14 (36%)	6 (29%)	5 (20%)	11 (14%)	5 (11%)	4 (8%)	**<.001**
Stenotrophomonas	23 (9%)	1 (3%)	4 (19%)	3 (12%)	5 (6%)	3 (7%)	7 (14%)	0.37
Lactobacillus	22 (9%)	0 (0%)	0 (0%)	3 (12%)	7 (9%)	6 (14%)	6 (12%)	**0.01**
Fusobacterium [A]	19 (7%)	2 (5%)	1 (5%)	3 (12%)	7 (9%)	3 (7%)	3 (6%)	0.87
Granulicatella	14 (5%)	1 (3%)	0 (0%)	1 (4%)	8 (10%)	2 (5%)	2 (4%)	0.47
Burkholderia	13 (5%)	0 (0%)	1 (5%)	0 (0%)	4 (5%)	3 (7%)	5 (10%)	**0.02**
Porphyromonas [A]	13 (5%)	2 (5%)	2 (10%)	2 (8%)	3 (4%)	2 (5%)	2 (4%)	0.51

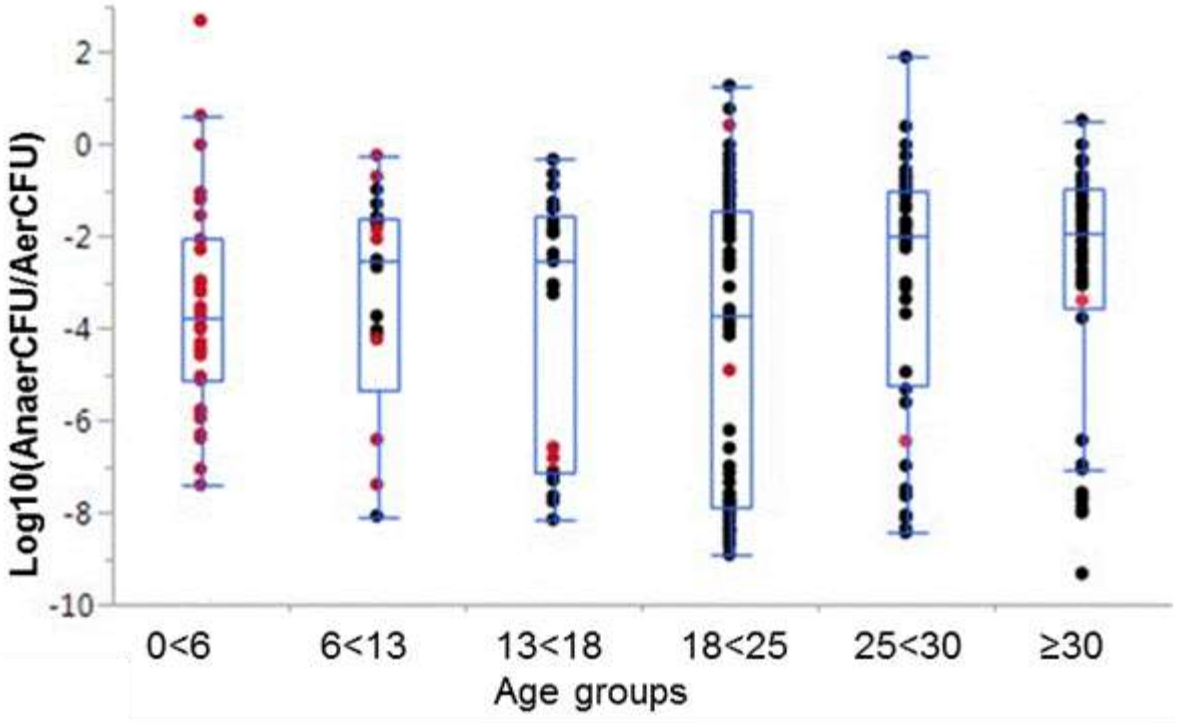
BAL=bronchoalveolar lavage. * p-values indicating differences across age groups were computed using a Mantel-Haenszel mean score Chi-Square test. [A] denotes obligate anaerobes.

Table S6: Odds ratios from logistic regression models of bacterial prevalence in sputum – less stringent criteria

	Anaerobes N=197	*Haemophilus* N=197	*Staphylococcus* N=200	*Pseudomonas* N=187
F508 del (1 vs 0)	--	--	--	**8.41** (1.01, 70.26)
F508 del (2 vs 0)	--	--	--	**22.67** (2.70,190.55)
Pancreatic insuf.	**0.16** (0.03, 0.72)	--	--	--
FEV1 (1 vs 3)	--	--	--	0.37 (0.11, 1.23)
FEV1 (2 vs 3)	--	--	--	1.03 (0.40, 2.64)
Azithromycin	0.49 (0.23, 1.03)	0.44 (0.19, 1.01)	**0.34** (0.19, 0.62)	**2.16** (1.00, 4.69)
Hypertonic saline		**0.40** (0.16, 0.99)		
Inhaled antibiotics	--	**0.25** (0.11, 0.56)	--	**5.25** (2.33, 11.84)
Insulin	0.43 (0.18, 1.04)	--	--	--
Center C vs. B	0.66 (0.25, 1.75)	0.58 (0.20, 1.64)	--	--
Center D vs. B	**0.16** (0.07, 0.37)	**0.20** (0.08, 0.51)	--	--

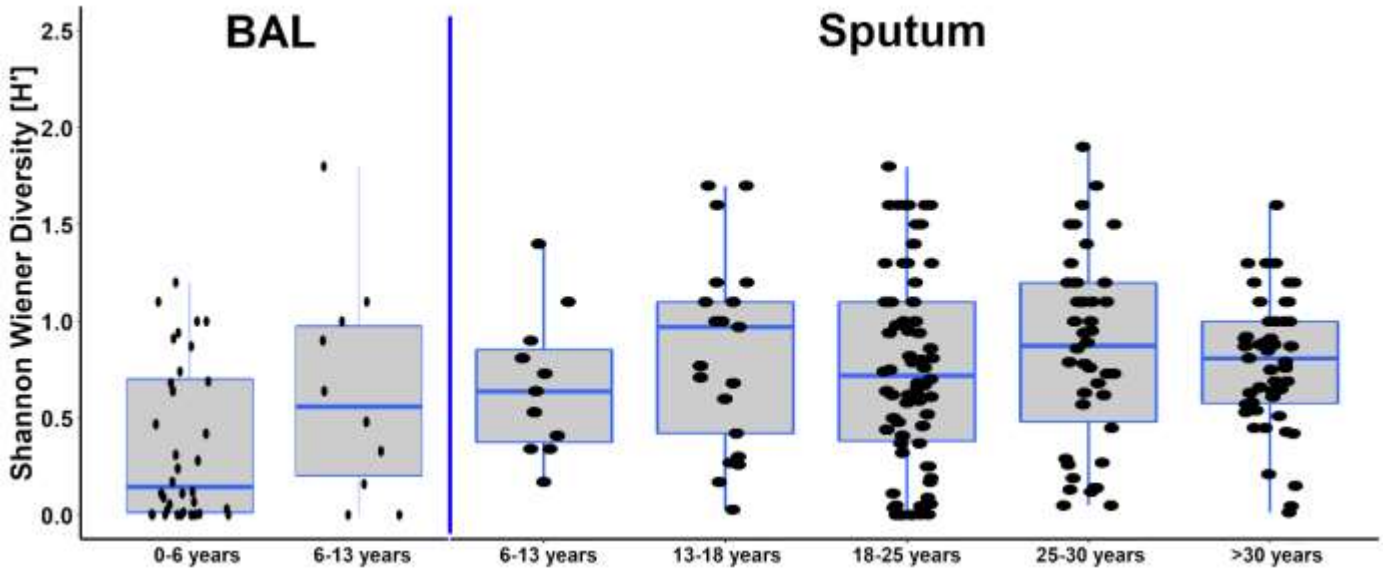
Variables included in the full model were: gender, age, site, FEV_1 and BMI as categories (as in Figure 2), F508 del, pancreatic status, chronic use of flucloxacillin, azithromycin, inhaled antibiotics, DNase, hypertonic saline, corticosteroids and antacids. " – " indicates these variables were not selected for the final model because they were not significant predictors as determined by stepwise model selection. Odds Ratios that are significant are clarified by bold font.

Figure S1A: Ratio of anaerobic CFU to aerobic CFU in BAL and SPU across age groups.



Ratio of total viable count (CFU) (anaerobe+1)/ (aerobe+1) was calculated for each sample. Comparisons by ANOVA and pairwise comparisons controlling for multiple comparisons (Tukey-Kramer) did not show differences between age groups. Red markers indicate BAL samples, black markers sputum.

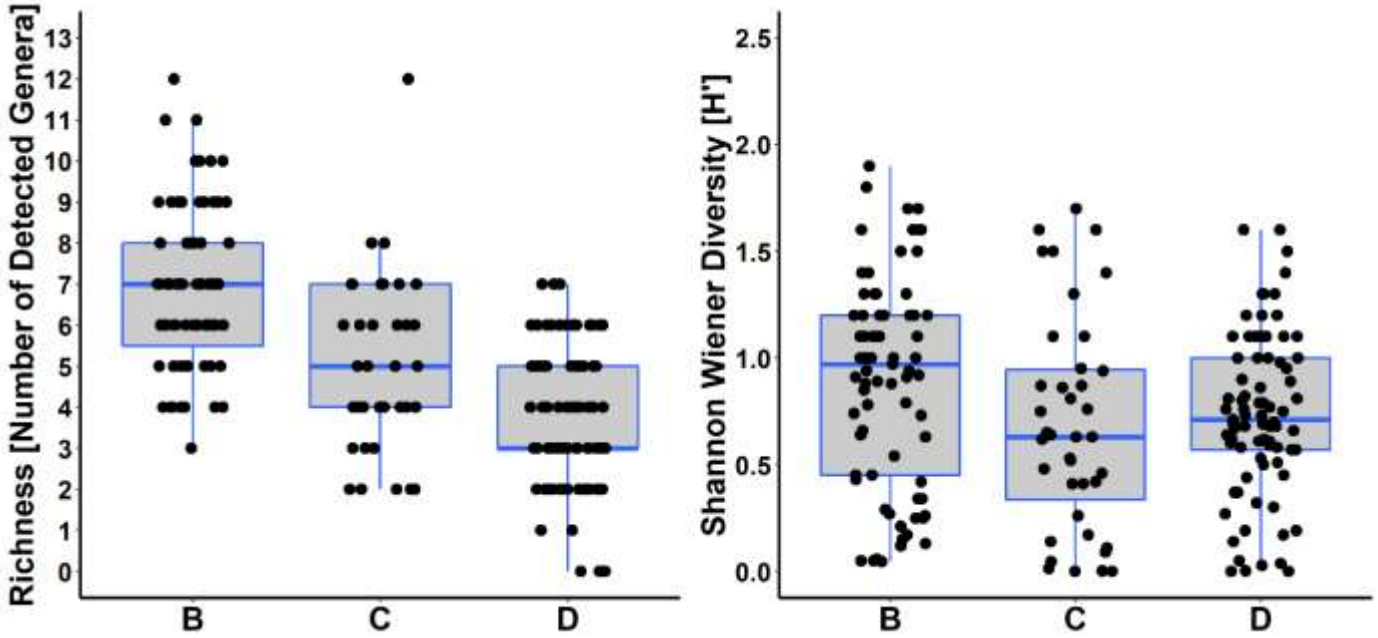
Figure S1B: **Bacterial diversity in bronchoalveolar lavage (BAL) and sputum samples by age sextiles that reflect clinical progression**



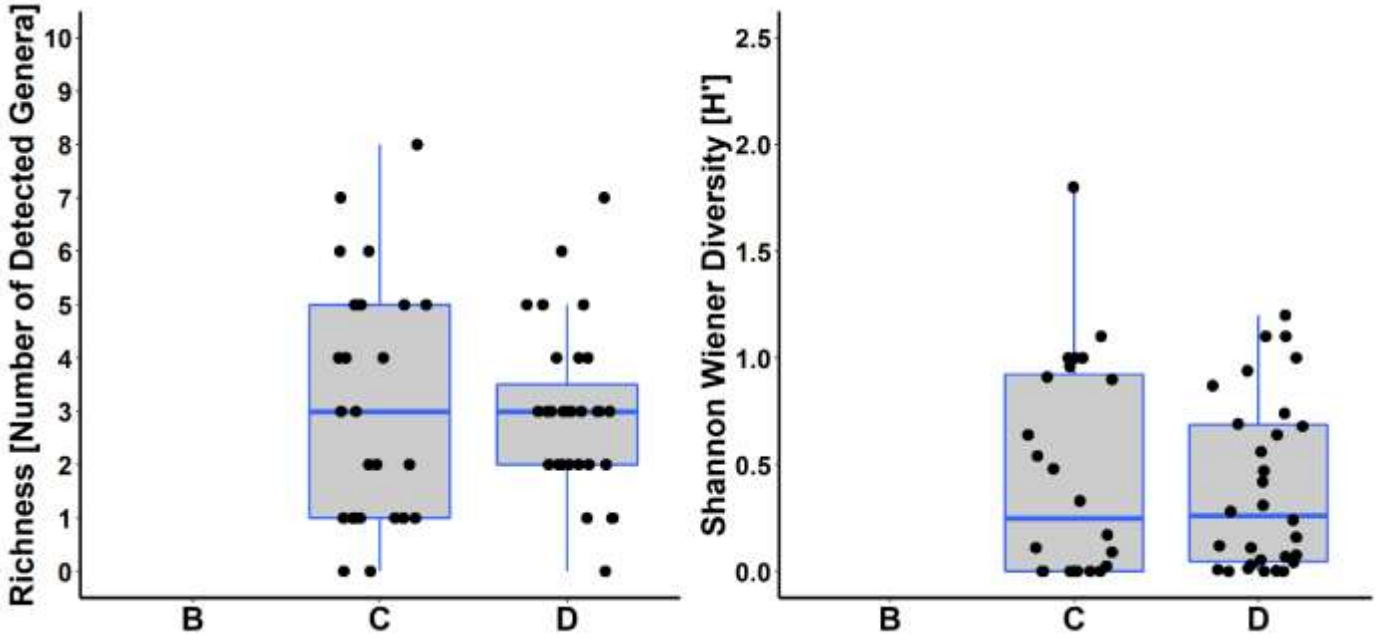
Each sample is represented by a circle. The vertical line indicates bronchoalveolar lavage (BAL) to the left and sputum samples to the right. Box indicates 25–75 quartiles with horizontal lines indicating the median. Whiskers are 1.5 times the interquartile range.

Figure S2: Bacterial richness and diversity across study sites

a)



b)



Each sample is represented by a circle. Box indicates 25-75 quartiles with horizontal lines indicating the median. Whiskers are 1.5 times the interquartile range. Richness: Number of all anaerobic and aerobic genera detected at any colony forming unit per sample.

a) For sputum samples (n=200), richness differed significantly between all sites ($p<0.001$). Shannon-Wiener diversity trended higher at Belfast than UNC ($p=0.05$). **b)** For bronchoalveolar lavage samples (n=55), neither richness nor diversity differed by site ($p=8.7$ and 0.76, respectively). B=Belfast, C=UNC, D-Dublin.

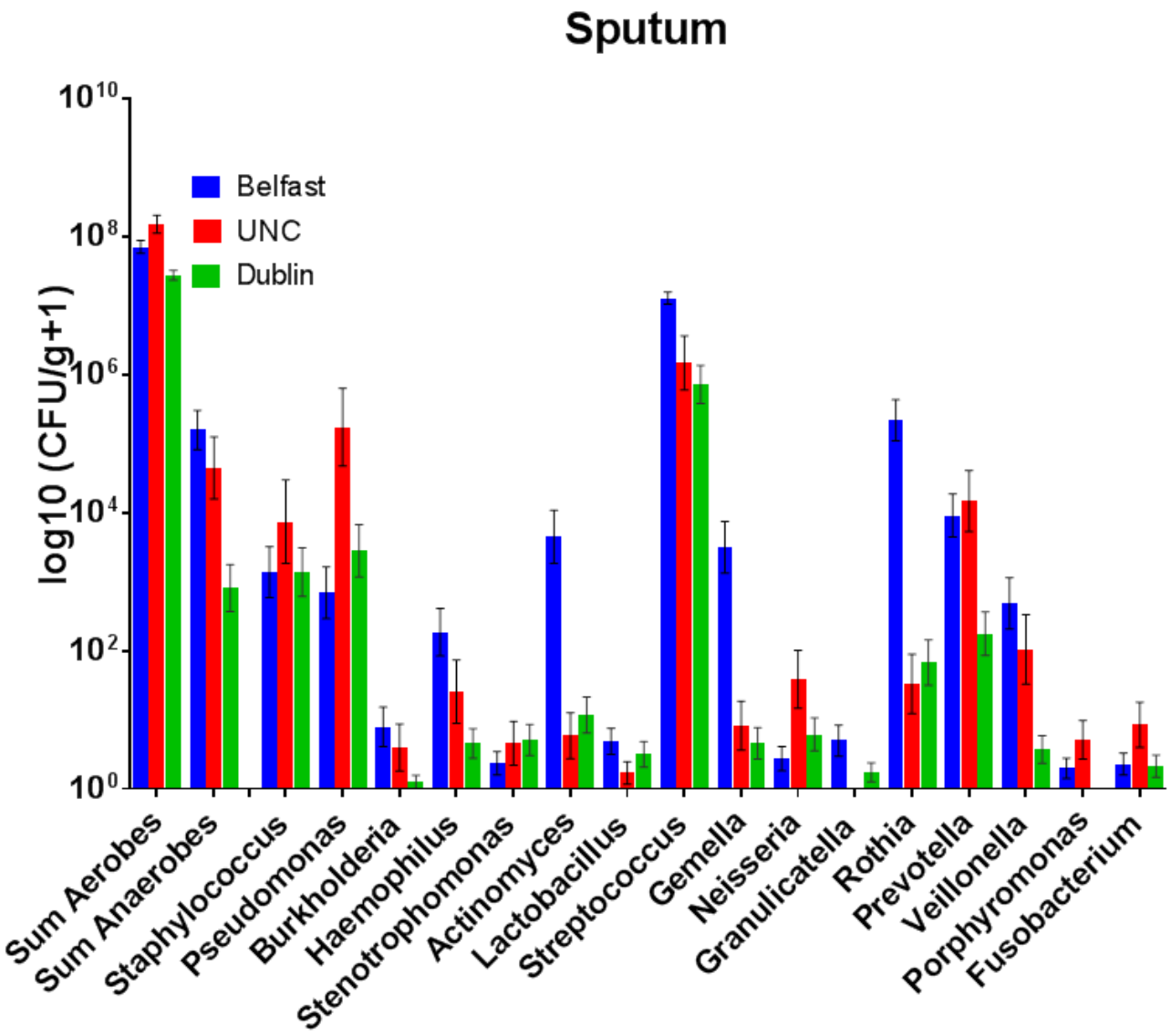
Table S6 – The 16 most prevalent genera by sample type and site:

Organism	Total BAL N=55 N (%)	C BAL N=24 N (%)	D BAL N=31 N (%)	P-value BAL [1]	Total Sputum N=200 N (%)	B Sputum N=75 N (%)	C Sputum N=41 N (%)	D Sputum N=84 N (%)	P-value Sputum [1]
Any aerobe	51 (93%)	21 (88%)	30 (97%)	0.31	197 (99%)	75 (100%)	41 (100%)	81 (96%)	0.23
Any anaerobe	17 (31%)	9 (38%)	8 (26%)	0.39	134 (67%)	63 (84%)	31 (76%)	40 (48%)	<.001
Streptococcus	28 (51%)	9 (38%)	19 (61%)	0.11	181 (91%)	75 (100%)	36 (88%)	70 (83%)	<.001
Prevotella[A]	13 (24%)	8 (33%)	5 (16%)	0.2	116 (58%)	54 (72%)	29 (71%)	33 (39%)	<.001
Pseudomonas	15 (27%)	7 (29%)	8 (26%)	1	105 (53%)	34 (45%)	29 (71%)	42 (50%)	0.03
Staphylococcus	18 (33%)	10 (42%)	8 (26%)	0.26	99 (50%)	41 (55%)	21 (51%)	37 (44%)	0.4
Rothia	11 (20%)	4 (17%)	7 (23%)	0.74	95 (48%)	62 (83%)	10 (24%)	23 (27%)	<.001
Actinomyces	8 (15%)	3 (13%)	5 (16%)	1	62 (31%)	42 (56%)	5 (12%)	15 (18%)	<.001
Haemophilus	18 (33%)	7 (29%)	11 (35%)	0.77	46 (23%)	29 (39%)	8 (20%)	9 (11%)	<.001
Gemella	2 (4%)	1 (4%)	1 (3%)	1	56 (28%)	42 (56%)	6 (15%)	8 (10%)	<.001
Veillonella[A]	5 (9%)	3 (13%)	2 (6%)	0.64	52 (26%)	32 (43%)	12 (29%)	8 (10%)	<.001
Neisseria	18 (33%)	6 (25%)	12 (39%)	0.39	27 (14%)	6 (8%)	11 (27%)	10 (12%)	0.02
Stenotrophomonas	5 (9%)	4 (17%)	1 (3%)	0.16	18 (9%)	5 (7%)	4 (10%)	9 (11%)	0.69
Lactobacillus	0 (0%)	0 (0%)	0 (0%)		22 (11%)	13 (<1%)	2 (5%)	7 (8%)	0.08
Fusobacterium[A]	3 (5%)	1 (4%)	2 (6%)	1	16 (8%)	5 (7%)	7 (17%)	4 (5%)	0.07
Granulicatella	2 (4%)	0 (0%)	2 (6%)	0.5	12 (6%)	9 (12%)	0 (0%)	3 (4%)	0.02
Burkholderia	0 (0%)	0 (0%)	0 (0%)		13 (5%)	9 (12%)	3 (<1%)	1 (1%)	0.01
Porphyromonas[A]	3 (5%)	2 (8%)	1 (3%)	0.57	10 (5%)	4 (5%)	6 (15%)	0 (0%)	0.001

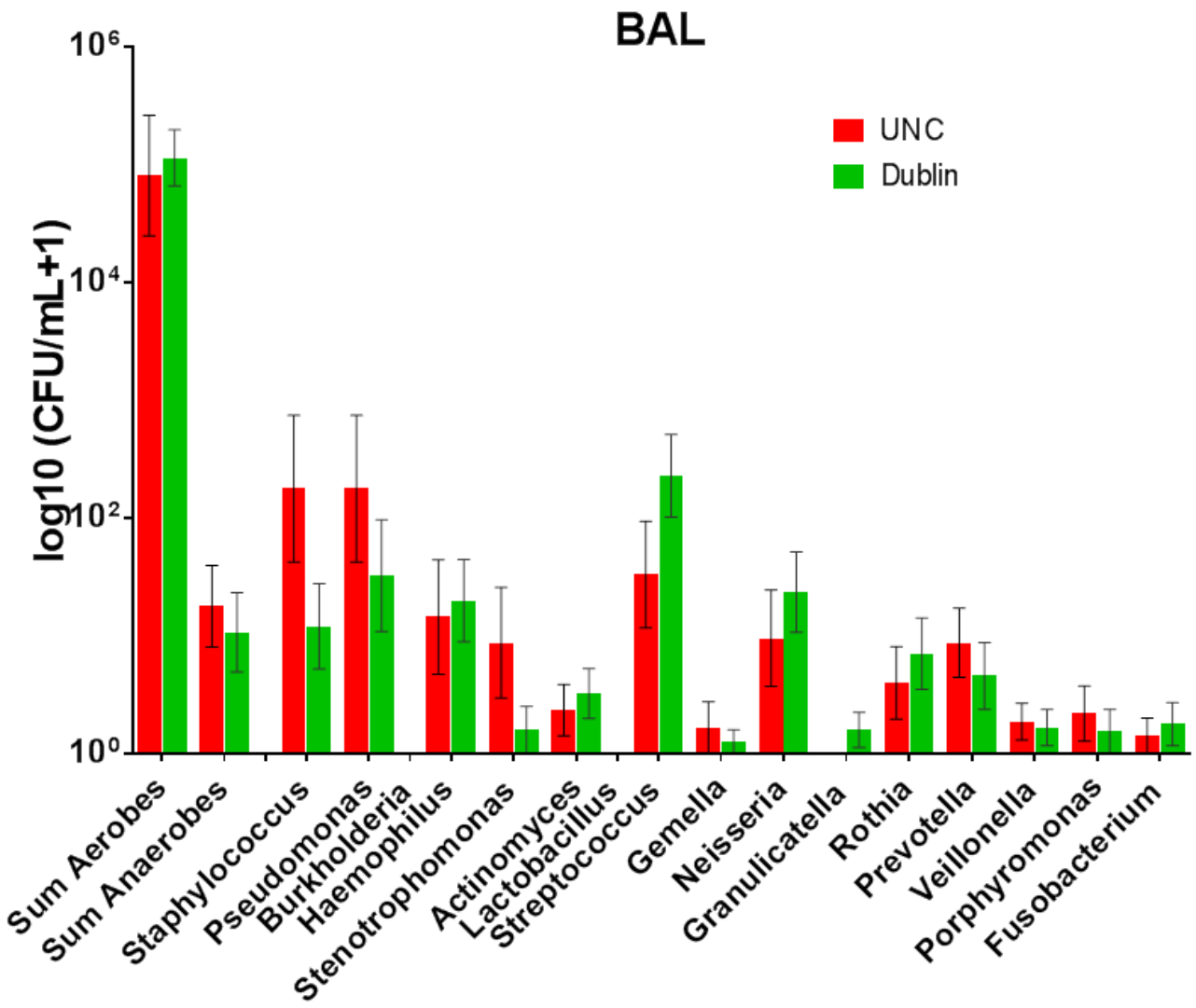
Sixteen genera had a mean prevalence of ≥5% across study sites in sputum. Genera are sorted by overall prevalence in sputum. B=Belfast, Northern Ireland; C=Chapel Hill, USA; D=Dublin, Ireland. BAL= Bronchoalveolar Lavage. [A] indicates obligate anaerobe genera. 1 Comparison between sites was by Fisher's exact test.

Figure S3: Bacterial quantity for 16 most prevalent genera as total viable counts in a) sputum and b) bronchoalveolar lavage (BAL) for each site.

a)

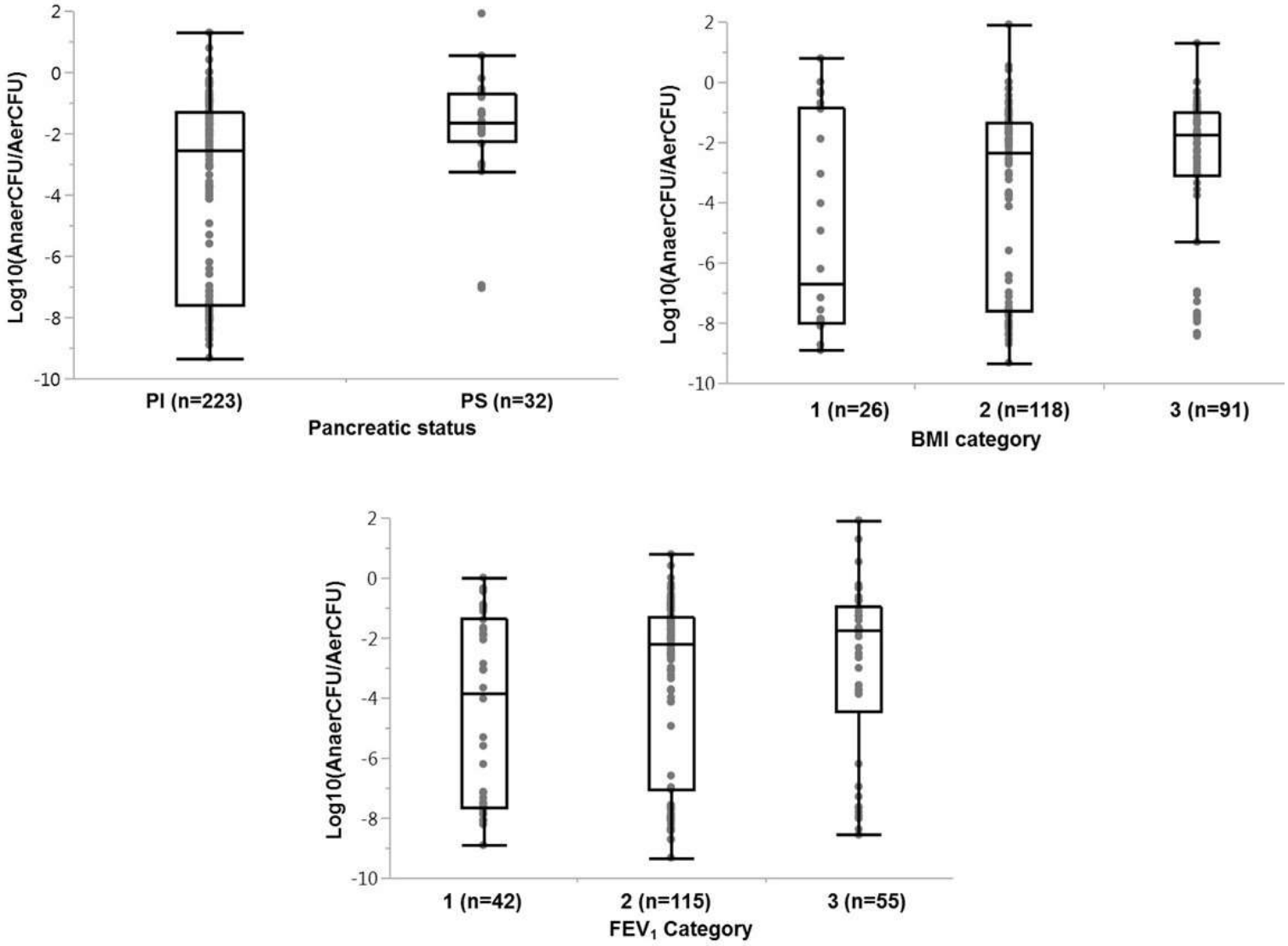


b)



Legend Figure S3: Bars indicate the mean and error bars SEM. N=200 for sputum and 55 for bronchoalveolar lavage (BAL).

Figure S4: **Ratio of bacterial density of anaerobe to aerobe colony forming units**



Ratio of total viable count (anaerobe+1)/total viable count (aerobe+1) was calculated for each sample. Comparison by ANOVA after log(10) transformation. Significant differences were seen by pancreatic and nutritional status but not lung function (FEV_1 categories are 1: <40% predicted; 2: 40–80%; 3: >80%). PI=pancreatic insufficiency. PS=pancreatic sufficiency. Body mass index (BMI) categories are 1 - poor nutritional status; 2 - adequate; 3 - well-nourished.

REFERENCES:

1. Quanjer PH, Stanojevic S, Cole TJ et al. ERS Global Lung Function Initiative. Multi-ethnic reference values for spirometry for the 3-95-yr age range: the global lung function 2012 equations. *Eur Respir J* 2012: 40(6): 1324-1343.
2. Tunney MM, Field TR, Moriarty TF, et al. Detection of anaerobic bacteria in high numbers in sputum from patients with cystic fibrosis. *Am J Respir Criti Care Med* 2008: 177(9): 995-1001.
3. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Roy Statist Soc Ser B Statist Methodol* 1995: 57: 289-300.
4. Csardi G, Nepusz T. The igraph software package for complex network research. *InterJournal* 2006: Complex Systems 1695.
5. Bastian M, Heymann S, Jacomy M. Gephi: an open source software for exploring and manipulating networks. *International AAAI Conference on Weblogs and Social Media* 2009: 1.
6. Harrell Jr F. Harrell miscellaneous. R package version 3.12-2. *Hmisc* 2013: http://cran.r-project.org/web/packages/Hmisc/index.html.