



Concordance between upper and lower airway microbiota in infants with cystic fibrosis

Sabine M.P.J. Prevaes ¹, Wouter A.A. de Steenhuijsen Pipers^{1,4}, Karin M. de Winter-de Groot^{1,4}, Hettie M. Janssens², Gerdien A. Tramper-Stranders¹, Mei Ling J.N. Chu¹, Harm A. Tiddens², Mireille van Westreenen³, Cornelis K. van der Ent¹, Elisabeth A.M. Sanders¹ and Debby Bogaert¹

Affiliations: ¹Dept of Paediatrics, Wilhelmina Children's Hospital, University Medical Centre Utrecht, Utrecht, The Netherlands. ²Dept of Paediatric Pulmonology and Allergology, Sophia Children's Hospital, Erasmus University Medical Centre, Rotterdam, The Netherlands. ³Dept of Medical Microbiology and Infectious Diseases, Erasmus University Medical Centre, Rotterdam, The Netherlands. ⁴Both authors contributed equally.

Correspondence: Debby Bogaert, Dept of Paediatric Immunology and Infectious Diseases, Wilhelmina Children's Hospital, University Medical Centre Utrecht, PO Box 85090, 3508 AB Utrecht, The Netherlands. E-mail d.bogaert@umcutrecht.nl



@ERSpublications

Lungs of CF infants have a microbiome that seems seeded by, but is not identical to, the URT microbiome <http://ow.ly/1NIA306DuPv>

Cite this article as: Prevaes SMPJ, de Steenhuijsen Pipers WAA, de Winter-de Groot KM, *et al.* Concordance between upper and lower airway microbiota in infants with cystic fibrosis. *Eur Respir J* 2017; 49: 1602235 [<https://doi.org/10.1183/13993003.02235-2016>].

ABSTRACT Nasopharyngeal and oropharyngeal samples are commonly used to direct therapy for lower respiratory tract infections in non-expectorating infants with cystic fibrosis (CF).

We aimed to investigate the concordance between the bacterial community compositions of 25 sets of nasopharyngeal, oropharyngeal and bronchoalveolar lavage (BAL) samples from 17 infants with CF aged ~5 months (n=13) and ~12 months (n=12) using conventional culturing and 16S-rRNA sequencing.

Clustering analyses demonstrated that BAL microbiota profiles were in general characterised by a mixture of oral and nasopharyngeal bacteria, including commensals like *Streptococcus*, *Neisseria*, *Veillonella* and *Rothia* spp. and potential pathogens like *Staphylococcus aureus*, *Haemophilus influenzae* and *Moraxella* spp. Within each individual, however, the degree of concordance differed between microbiota of both upper respiratory tract niches and the corresponding BAL.

The inconsistent intra-individual concordance between microbiota of the upper and lower respiratory niches suggests that the lungs of infants with CF may have their own microbiome that seems seeded by, but is not identical to, the upper respiratory tract microbiome.

This article has supplementary material available from erj.ersjournals.com

Received: May 17 2016 | Accepted: Nov 23 2016

Support statement: The Wilhelmina Children's Hospital Foundation and the Dutch Cystic Fibrosis Foundation supported this work. Funding sources had no role in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the paper for publication. Funding information for this article has been deposited with the Crossref Funder Registry.

Conflict of interest: Disclosures can be found alongside this article at erj.ersjournals.com

Copyright ©ERS 2017

Introduction

Cystic fibrosis (CF) is a life-limiting hereditary disease that causes chronic lung infections. During early childhood, the lungs of patients with CF become colonised by a complex array of microbes comprising many more bacterial species than those considered as “typical CF pathogens” [1–3]. The core lung microbiome of paediatric patients with CF seems to consist of the genera *Streptococcus*, *Rothia*, *Prevotella*, *Actinomyces*, *Veillonella*, *Gemella*, *Neisseria* and *Haemophilus* [1], which presumably spread from the upper respiratory tract (URT) to the lower respiratory tract (LRT) through dispersal by micro-aspiration or inhalation [4–7].

In theory, direct lung sampling in infants with CF by bronchoalveolar lavage (BAL) [8] would be most representative of the local presence and relative abundance of respiratory pathogens within LRT microbial communities. However, 1) because of the invasive nature of BAL, 2) because infants and young children with CF do not spontaneously expectorate sputum and 3) because of the lack of practical experience with obtaining induced sputum in these patients [8, 9], clinicians generally use alternative specimens including nasal, nasopharyngeal (NP) and oropharyngeal (OP) samples for microbial surveillance [10–15]. Recent cross-sectional studies [16, 17] in older children with CF have shown that the upper and lower airways on average harbour similar bacterial communities. However, intra-individual concordance analyses, which in our opinion are more relevant for the individual patient, were absent until recently and are only available for adults with end-stage CF disease [18]. These analyses have shown that the microbiota of the throat and (explanted) lung tissue are highly discordant.

We investigated the NP, OP and BAL microbiota profiles in infants with CF using both next-generation sequencing techniques and routine culturing at two consecutive time-points in the first year of life. We aimed to 1) assess both the inter- and intra-individual concordance of bacterial communities harbouring various respiratory niches, 2) study the potential additive value of the use of 16S-rRNA-based sequencing over traditional culturing in the management of patients with CF and 3) evaluate the tempo-spatial variation in lung microbiota in infants with CF.

Methods

Study population and design

Infants with CF were recruited through a prospective two-centre observational study, established around a microbial surveillance programme for infants diagnosed with CF by newborn screening in the Netherlands [8]. The programme and study population are described in detail in the supplementary methods [19]. For the current study, we collected successive intra-individual sets of NP, OP and BAL samples from infants with CF during their first year of life. Questionnaires on baseline characteristics and the patient’s respiratory status were obtained at all sampling moments.

Ethical statement

The study (NL10/337) was approved by the paediatric medical ethics committees of the University Medical Centre Utrecht and the Erasmus Medical Centre and undertaken according to Good Clinical Practice guidelines.

Sampling and storage methods

Flexible fibre-optic bronchoscopy was performed under general anaesthesia in accordance with standard safety and monitoring procedures for bronchoscopy. Preceding the BAL procedure and following sedation, the posterior nasopharynx was swabbed transnasally [20] and the oropharynx was swabbed transorally with an ESwab 482CE nylon flocked flexible sterile swab (Copan Diagnostics, Brescia, Italy) and stored in Amies Transport Medium [21]. At the start of the BAL procedure (for the detailed protocol, please refer to the supplementary methods), the tip of the scope was wiped with a sterile gauze to serve as a negative control for bacterial DNA background noise [22]. The bronchoscope was introduced through the laryngeal mask; BAL was performed by instilling 1 mL·kg⁻¹ of sterile 0.9% saline and retrieved using low-pressure suction. To prevent URT contamination, suction was not performed during the introduction of the bronchoscope until the tip passed the carina. BAL fluid was collected during different lavage cycles successively from at least two and a maximum of five locations: the right middle lobe, lingula/left upper lobe, left lower lobe and other lung segments [23]. In one of the two study centres, BAL fluids from different lobes were pooled per individual for logistic reasons. All BAL samples were immediately sealed in a sterile container and, together with the separate NP and OP swabs, transported to the clinical microbiology laboratories of each of the two CF reference centres. BAL fluid and the Amies medium of both URT samples were subsequently vortexed and aliquoted. One aliquot was frozen within 8 h at –80°C until further sequence analysis. Another aliquot was used for plating on Columbia blood, chocolate, MacConkey and mannitol salt agar media and Bur-cep plates (for details, see supplementary methods).

Microbiomic analysis

Bacterial DNA isolation and quantification

DNA was isolated from 200 μL of NP Amies medium [24], 200 μL of OP Amies medium, 300 μL BAL fluid and from our negative control samples (200 μL lysis buffer (NP and OP) and 300 μL phosphate-buffered saline from the tip-of-the-bronchoscope gauze (BAL)). The total bacterial load of the samples was established by quantitative PCR using a primer-probe set targeting the bacterial 16S-rRNA gene [24–26]. To avoid environmental contamination in our 16S-rRNA sequencing analyses, we considered only respiratory samples with a sufficient overall bacterial load, defined as samples with a bacterial density of $\geq 0.3 \text{ pg}\cdot\mu\text{L}^{-1}$ above the background environmental control samples. To confirm that low bacterial density NP and BAL samples were not driven by environmental cross-contamination, we concurrently sequenced a subset of our lowest density NP and BAL samples ($n=22$) with background environmental control samples ($n=27$), including DNA isolation controls and pre-procedural tip-of-the-bronchoscope DNA controls (*i.e.* gauze-controls). For further details, refer to the supplementary methods.

Amplicon library preparation

A PCR amplicon library was generated by amplifying the V5–V7 hypervariable region of the bacterial 16S-rRNA gene as described in the supplementary methods using a 454 GS-FLX-Titanium Sequencer (Life Sciences, Roche, Hongkong, China).

Data processing

The raw sequences obtained were processed using QIIME version 1.8 [27]. Sequences were checked for quality using `split_library.py` using default parameters (supplementary methods). Subsequently, barcodes and primers were trimmed off and chimeric sequences were identified and removed using `chimeraslayer`. Next, reads were aligned and clustered into operational taxonomic units (OTUs) using UCLUST at 97% similarity. Aligned sequences were taxonomically annotated using the Greengenes 16S-rRNA database (version 13.8) and categorised into potential pathogens of CF or general commensals based on literature findings [28–31]. Data are available through the NCBI GenBank database: accession number SRP081002. For each sample, α -diversity indices (Shannon diversity, number of observed species and Chao 1 diversity index) were calculated at a rarefaction depth of 1100 sequences per sample [32, 33].

Statistical analysis

Data analyses were performed in either IBM SPSS version 21 (IBM, Armonk, NY, USA) or R version 3.1 software. We used (paired) t-tests or Mann–Whitney U-tests to calculate the statistical significance of differences between groups for continuous data. Correlation coefficients were calculated using a non-parametric Spearman correlation. The statistical significance of differences in α -diversity indices between niches was evaluated by repeated measures ANOVA with Bonferroni *post hoc* tests. A non-metric multidimensional scaling (nMDS) plot and hierarchical clustering dendrogram (iTol) [34] based on the Bray–Curtis dissimilarity matrix were used to visualise differences in the overall microbial community composition between niches. The statistical significance of these differences was calculated by permutational ANOVA (PERMANOVA, `vegan` package R) [35]. Next, we used Kruskal–Wallis tests with Benjamini–Hochberg correction for multiple testing and subsequent Nemenyi *post hoc* tests to detect significant differences in the relative abundance of the 100 highest-ranking OTUs between niches (NP or OP *versus* BAL). The top 100 OTUs were based on the mean abundance of these OTUs within our sample set of 75 samples ($n=25$, per respiratory niche). Bray–Curtis similarity ($1 - (\text{Bray–Curtis dissimilarity})$), where 1.0 means perfect similarity between both samples, was used to determine overall similarities in microbial composition 1) within individual sample sets of NP, OP and BAL (as reference for BAL samples we used the lavage of the lingula, if available, or alternatively another location or a pooled sample); 2) between different lung lavage locations per BAL; and 3) between early and later BAL samples. To assess the concordance of the most abundant bacterial species between niches, \log_2 -transformed ratios (NP over BAL or OP over BAL) of the relative abundance ($+0.00001$) of these bacteria were calculated.

Results

Characterisation of study population

A total of 25 NP-OP-BAL sets from 17 infants with CF were collected during routine elective bronchoscopy at baseline; comprising 13 sets from infants <9 months of age (“Early”, range 3–8 months) and 12 sets from infants ≥ 9 months of age (“Later”, 9–13 months) [36, 37]. Eight of the infants were sampled at two time points with a median interval between both sampling moments of 8 months (range 5–10). BAL samples were taken from on average three (range 2–5) lung lobes. Baseline characteristics of infants in both age groups are depicted in table 1.

TABLE 1 Patients' characteristics

Parameter	Infants	BAL-early ^f	BAL-later ^f
Total	17	13	12
Sex male	8 (47)	7 (54)	4 (33)
CFTR mutation			
Severe: mostly F508del/F508del [#]	15 (88)	11 (85)	11 (92)
Milder: R334W - c1132 C>T	1 (6)	1 (8)	1 (8)
Mild: F508del/R117H	1 (6)	1 (8)	0 (0)
Elective BAL[¶]		13 (100)	12 (100)
Respiratory symptoms during BAL[*]		2 (15)	2 (17)
(Prophylactic) antibiotic use in previous month[§]		8 (62)	11 (92)
Age months mean (range)		4.6 (3–8)	12.4 (9–13)
Weight kg		6 (5–9)	10 (9–11)
Height cm		63 (57–73)	76 (72–84)

Data are presented as n (%), unless otherwise indicated. BAL: bronchoalveolar lavage; *CFTR*: cystic fibrosis transmembrane conductance regulator. #: F508del/F508del (n=10), F508del/y1092X (n=1), F508del/2183AA>G (n=1), 1717–1G>A/2183AA>G (n=1), DF508/G461R, 1513>A (n=1), F508del/N1303K (n=1); ¶: according to a Dutch clinical surveillance protocol; *: patient's respiratory status (presence of cough, rhinorrhoea or shortness of breath) was defined as asymptomatic or symptomatic; §: by protocol, patients on antibiotics prophylaxis (mostly trimethoprim-sulfamethoxazole or amoxicillin) were asked to stop antibiotic prophylaxis for 3 days preceding the BAL procedure; ^f: number of BAL procedures.

Density data and sequencing results

On average, we observed the highest bacterial density in OP samples, followed by NP and BAL samples. The density of samples and environmental control samples (gauze-controls and DNA isolation controls) are depicted in supplementary figure S1. Concurrent sequencing of our lowest density samples and environmental controls confirmed that there was negligible cross-contamination between the environment and our lowest density samples (supplementary results).

Using 16S-based sequencing we obtained a total of 590227 sequences (mean number of sequences per sample \pm SD: NP 5754 \pm 2084, OP 4490 \pm 1717 and BAL 4705 \pm 1610), grouped into 259 OTUs (excluding singletons) and representing 15 taxonomic phyla with more than 10 attributed sequences assigned.

Nasopharyngeal and oropharyngeal communities compared to bronchoalveolar lavage communities at population level

Variation in ecological diversity between niches

The Shannon diversity index, accounting for both species richness and evenness, differed significantly between sample sites (repeated-measures ANOVA, $p < 0.001$): the lowest diversity was observed for NP microbiota (mean diversity index (DI) 1.9, 95% CI 1.4–2.5, $p = 0.001$ for NP versus BAL) followed by BAL (DI 3.1, 95% CI 2.6–3.5) and OP microbiota (DI 3.3, 95% CI 2.9–3.8, $p = 0.352$ for OP versus BAL). A similar pattern was observed for other diversity measures (figure 1).

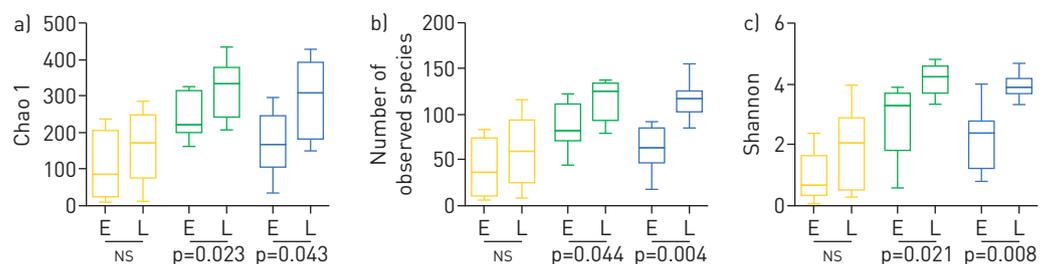


FIGURE 1 Boxplot of α -diversity analyses for nasopharyngeal (NP, yellow), oropharyngeal (OP, green) and bronchoalveolar lavage (BAL, blue) samples over time. a) Chao 1 diversity index, b) number of observed species, c) Shannon diversity index. Eight infants with cystic fibrosis were sampled twice, *i.e.* when 3–8 months of age (E) and when ≥ 9 months (L). Statistical significance in relative abundance between early and later obtained samples was assessed by paired t-tests (p values depicted) and between the niches by repeated ANOVA with Bonferroni *post hoc* tests (data in main text). NS: not significant.

Similarity of bacterial community structure between niches

Phylum-level comparison of NP, OP and BAL bacterial communities is shown in supplementary figure S2. Because NP and OP samples were used as proxy for BAL samples [10–14], we evaluated the general concordance between microbiota profiles of these two niches and BAL samples. A nMDS plot based on the Bray–Curtis dissimilarity matrix was used to visualise differences in the overall microbiota composition between the three niches (PERMANOVA unadjusted $R^2=0.11$, $p<0.001$); this showed that BAL microbiota were generally positioned between the groups of NP and OP microbiota (figure 2), suggesting that BAL microbiota represent a mixture of bacterial communities originating from both URT niches. On average, the NP microbiota differed more than the OP microbiota from the BAL microbiota (NP versus BAL PERMANOVA unadjusted $R^2=0.07$, $p=0.002$; OP versus BAL PERMANOVA unadjusted $R^2=0.04$, $p=0.07$). In conjunction with these results, hierarchical clustering of samples overlapped between all three niches, with BAL microbiota profiles mixing in with both NP and OP profiles (figure 3).

Genus-level analysis showed that the OP and BAL samples contained comparable abundances of the commensals *Neisseria*, *Streptococcus*, *Rothia*, *Veillonella*, *Gemella* and *Prevotella* spp., whereas the NP and BAL samples contained comparable abundances of potential pathogens like *S. aureus*, *H. influenzae* and *Moraxella* (Kruskal–Wallis (with Benjamini–Hochberg correction) and Nemenyi *post hoc* tests) (table 2 and supplementary figure S3). *Corynebacterium* and *Dolosigranulum* spp. were almost exclusively detected in the NP samples. *Pseudomonas* spp. abundance was low for all three niches; therefore, no clinically relevant differences could be detected.

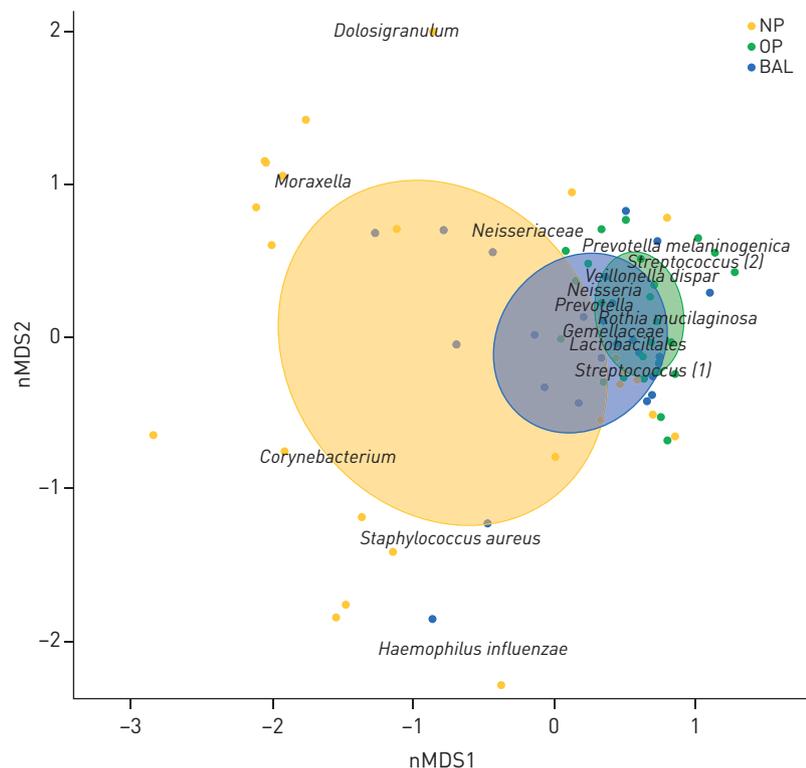


FIGURE 2 Two-dimensional non-metric multidimensional scaling (nMDS) plot of the microbial community composition in the nasopharyngeal (NP), oropharyngeal (OP) and bronchoalveolar lavage (BAL) samples. A nMDS plot based on the Bray–Curtis dissimilarity matrix was used to simultaneously visualise individual samples (dots) originating from NP (yellow), OP (green) and BAL (blue); the co-clustering of the 15 most abundant bacterial species among the three niches (based on the $n=75$ cohort); and the standard deviations around the geometric mean of samples stratified by niche (ellipses). The beta-diversity (*i.e.* inter-individual variability of microbiota) was highest in the NP followed by OP and BAL samples. The overall bacterial community composition of BAL samples shows overlap with community compositions of both the NP and OP niches. Furthermore, *S. aureus*, *Corynebacterium*, *Dolosigranulum* and *Moraxella* are more strongly related to NP samples, whereas streptococci, *Rothia*, *Prevotella* and *Neisseria* are enriched in OP and BAL samples. The numbers behind the operational taxonomic unit (OTU) names represent the hierarchical number of that OTU for its genus based on average relative abundance [*e.g.* *Streptococcus* (2) is the second dominant *Streptococcus* observed on group level].

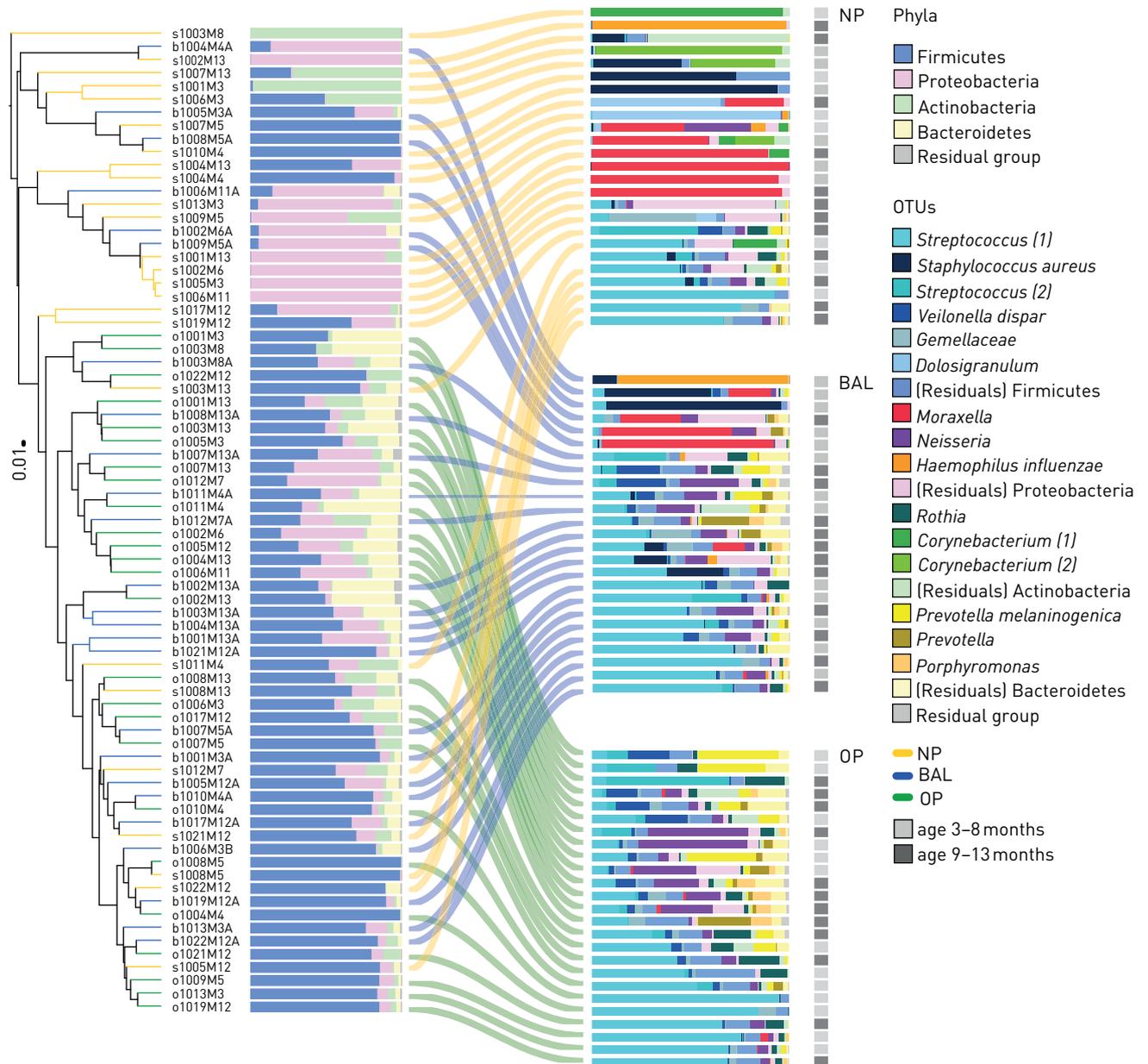


FIGURE 3 Niche-based hierarchical clustering of microbiota profiles. Branches and edges are coloured according to the niche of origin. The length of the branches corresponds with the (average) Bray-Curtis dissimilarity between (clusters of) samples (scale bar equal to a value of 0.01 shown). The first column of stacked bar charts shows the relative abundance of microbiota on a phylum level; the second column represents the 15 overall highest ranked operational taxonomic units (OTUs) per niche (nasopharyngeal (NP), oropharyngeal (OP) and bronchoalveolar lavage (BAL) ($n=25$ each)) of the originating dataset. Early- and later-obtained samples are depicted in light and dark grey, respectively. The numbers behind the operational taxonomic unit (OTU) names represent the hierarchical number of that OTU for its genus based on average relative abundance (e.g. *Corynebacterium* [2] is the second dominant *Corynebacterium* observed on group level).

Nasopharynx or oropharynx microbiota compared to bronchoalveolar lavage at the individual level

For potential clinical applications, we additionally calculated the within-individual concordance between NP, OP and BAL microbiota by Bray-Curtis similarity measures. Owing to the high inter-lobe concordance within subjects, as described in the following section, we used the lingula, sampled in 24 out of 25 of the lavage samples, as representative of the lung microbiome in these analyses. We observed variable intra-individual concordance between NP and BAL microbiota as well as between OP and BAL microbiota for the majority of patients, with most samples failing to exceed an overall microbiota similarity index of 0.6 (figure 4). We also observed a modest increase in overall similarity indices between NP and BAL samples and OP and BAL samples from early to later sampling moments.

At the species level, we observed large differences in concordance between niches: e.g. we observed a high intra-individual concordance between OP and BAL samples for the abundance of *Streptococcus* (*S. mitis*),

TABLE 2 Differences in microbial community composition between niches

OTUs	q value		p value		Relative abundance %		
	All three niches	NP-BAL	OP-BAL	NP	OP	BAL	
<i>Streptococcus</i> (1)	0.006	0.012	0.842	17.5	34.2	30.2	
<i>Moraxella</i>	0.393	0.708	0.862	20.4	0.4	8.9	
<i>Neisseria</i>	0.011	0.007	0.956	2.5	10.6	7.6	
<i>Staphylococcus aureus</i>	0.004	0.941	0.007	9.6	0.1	8.5	
<i>Streptococcus</i> (2)	0.020	0.051	0.820	2.5	5.8	3.5	
<i>Veillonella dispar</i>	0.001	0.002	0.698	0.8	5.1	4.0	
<i>Prevotella melaninogenica</i>	0.008	0.018	0.878	0.6	6.9	2.3	
<i>Rothia mucilaginosa</i>	<0.001	0.003	0.203	1.1	5.8	2.5	
Gemellaceae	0.005	<0.001	0.614	2.7	2.6	4.1	
<i>Haemophilus influenzae</i>	0.242	0.398	0.415	4.4	0.0	3.9	
<i>Dolosigranulum</i>	0.136	0.438	0.887	6.9	0.0	0.0	
<i>Corynebacterium</i> (1)	0.003	0.040	0.984	6.3	0.0	0.0	
<i>Corynebacterium</i> (2)	0.010	0.085	0.962	5.7	0.0	0.0	
<i>Prevotella</i>	0.028	0.022	0.808	0.2	2.4	2.3	
Lactobacillales	0.005	0.006	0.942	0.8	1.5	1.4	
<i>Porphyromonas</i> (1)	0.011	0.010	0.995	0.2	2.0	1.2	
<i>Streptococcus</i> (3)	0.002	0.016	0.314	0.3	1.9	0.8	
Actinomyces	<0.001	0.002	0.500	0.4	1.7	0.8	
<i>Corynebacterium</i> (3)	0.620	0.967	0.969	2.3	0.0	0.0	
<i>Porphyromonas</i> (2)	0.004	0.015	0.901	0.0	0.9	1.4	

This table depicts the differences with regard to the 20 most abundant OTUs between NP, OP and BAL samples (n=25 each) calculated by Kruskal–Wallis, with Benjamini–Hochberg correction for multiple testing (q value) and Nemenyi *post hoc* tests (only considering NP–BAL and OP–BAL comparisons). $p < 0.05$ are shown in bold. OTU: operational taxonomic units; NP–BAL: comparison between nasopharyngeal and bronchoalveolar lavage samples; OP–BAL, comparison between oropharyngeal and bronchoalveolar lavage samples; NP: nasopharyngeal; OP: oropharyngeal; BAL: bronchoalveolar lavage.

Neisseria and *Rothia*, whereas the concordance between the abundance of *Moraxella* and *S. aureus* was more similar between NP and BAL samples. The intra-individual concordance of *H. influenzae* was relatively low for NP and BAL as well as for OP and BAL samples (supplementary figure S4). Microbiota profiles of all NP, OP and BAL samples of all participants are provided in the supplementary results (supplementary figure S5).

Comparison of microbial composition between different lobes

Based on the European Respiratory Society recommendation to execute a BAL routinely in two or more lobes [8], we hypothesised that different lung locations might have distinct microbial compositions. We studied this

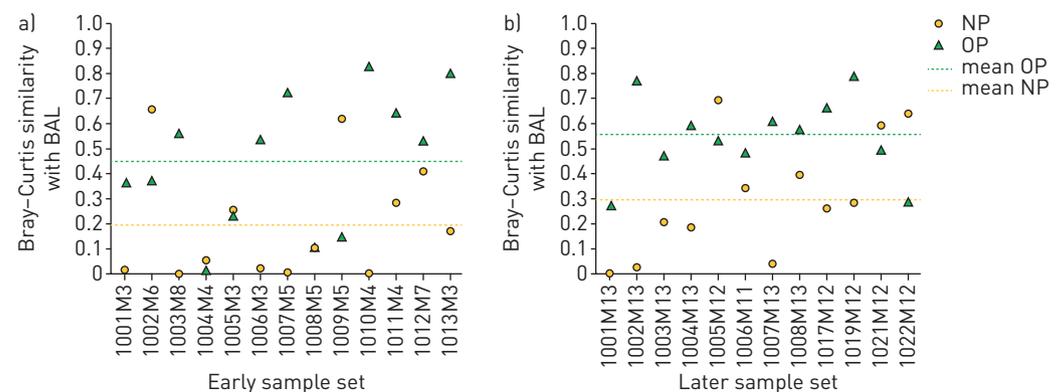


FIGURE 4 Intra-individual concordance between nasopharyngeal (NP) and oropharyngeal (OP) microbiota profiles and their paired bronchoalveolar lavage (BAL) microbiota based on Bray–Curtis similarity. a) “Early” sets, b) “Later” sets. For each individual, we calculated the intra-individual Bray–Curtis similarity between NP and BAL samples (reference: lingula when available) (yellow) and OP and BAL samples (reference: lingula when available) (green), respectively. The similarity index ranges from 0 (no similarity between both samples) to 0.85 (where 1.0 would mean perfect similarity between the samples). We defined a similarity score of >0.6 as high. The mean \pm SD similarity index, represented by a horizontal dashed line, of early OP–BAL (green) sets was 0.45 ± 0.27 and of later sets was 0.55 ± 0.16 , and of NP–BAL (yellow) sets was 0.20 ± 0.23 and of later sets was 0.31 ± 0.24 .

by comparing the different lung lobes within and between individuals. In total 71 BAL locations from 25 BAL procedures (two to five lung regions per BAL) were available for analysis: concordance between microbiota profiles of different lobes from the same BAL procedure within individuals was high (figure 5), and significantly higher than the similarity of lung microbiota between subjects (mean similarity: within-subject 0.73, between-subject 0.28, $p < 0.001$) (supplementary figure S6).

Temporal variation of BAL microbial composition

We studied temporal variation in BAL microbiota over an average of 8 months in eight infants during their first 13 months of life. On average, we observed an increase in microbiota diversity over time (Shannon index: Early 2.2, Later 3.9; figure 1). Comparable increases in diversity were observed over time for the NP and OP niches (figure 1). Additionally, we calculated the stability of microbiota profiles over time by calculating the Bray–Curtis similarity index between the early and the later BAL sample of each individual, and compared this to the similarity of the samples from each child with those from the other infants. We observed low intra-individual concordance between early and later BAL samples (supplementary figure S7), indicating significant changes in lung microbiota over time (supplementary figure S6).

Comparability between conventional culturing and 16S-based sequencing results

To allow comparison with the current standard of practice, *i.e.* conventional culturing, we compared our sequencing with culturing results (supplementary methods) comparing the most abundant *Staphylococcus*, *Moraxella* and *Haemophilus* OTUs with *S. aureus*, *M. catarrhalis* and *H. influenzae* culture results, confirming that culture results for *S. aureus*, *M. catarrhalis* and *H. influenzae* correlated strongly with the respective *S. aureus* ($r_s = 0.451$), *M. catarrhalis* ($r_s = 0.517$) and *H. influenzae* OTUs ($r_s = 0.276$, all $p < 0.001$; supplementary figure S8). Moreover, the sensitivity of sequencing in detecting potential pathogens was higher compared to conventional culture, as shown by supplementary figure S8.

In line with testing intra-individual concordance between microbiota of the three respiratory niches, we repeated these analyses for the conventional culture results. In general the positive predictive value (PPV) and negative predictive value (NPV) was moderate to high for all cultured bacteria (table 3). However, conventional cultures often failed to detect the complete set of pathogens identified in the lung in the corresponding NP and OP samples (28% and 50% of sample sets, respectively; supplementary figure S9).

Discussion

The most interesting finding of our study was that the degree of concordance between the microbial communities of either the NP or OP with their corresponding BAL microbiota was variable within the individual patient when studied by either conventional culture or sequencing technology. This suggests a limited diagnostic value of both types of URT samples to determine microbial colonisation profiles in infants with CF. The work of ZEMANICK *et al.* [38] is the only other study determining airway microbiota in

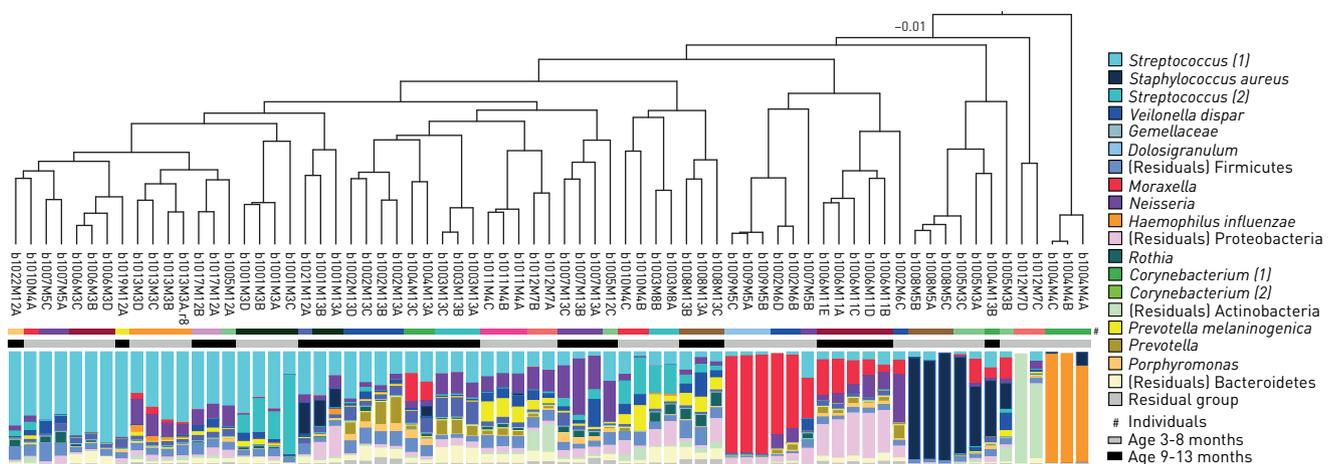


FIGURE 5 Dendrogram visualising the hierarchical clustering of multiple lavage locations (total $n = 71$) of the 25 bronchoalveolar lavage samples. Clustering was based on Bray–Curtis dissimilarity and average linkage. The length of the branches corresponds with the [average] Bray–Curtis dissimilarity between (clusters of) samples (scale bar equal to a value of 0.01 shown). #Different coloured rectangular boxes depict individuals. Eight individuals were sampled twice during their first year of life, represented in grey (Early) and black (Later). Stacked bar charts show the relative abundance of the top 15 operational taxonomic units (OTUs) of the overall dataset. The numbers behind the OTU names represent the hierarchical number of that OTU for its genus based on average relative abundance [e.g. *Corynebacterium* (2) is the second dominant *Corynebacterium* observed on group level].

TABLE 3 Positive and negative predictive values of upper respiratory tract samples for presence of bacteria in the bronchoalveolar lavage

Potential pathogen	Nasopharynx		Oropharynx	
	PPV	NPV	PPV	NPV
<i>S. aureus</i>	100 (52–100)	53 (29–75)	100 (63–100)	61 (32–85)
<i>P. aeruginosa</i> [#]	0	92 (72–99)	100 (5–100)	100 (81–100)
<i>H. influenzae</i>	67 (13–98)	72 (50–88)	100 (40–100)	78 (52–93)
<i>M. catarrhalis</i>	75 (22–99)	100 (81–100)	0	86 (64–96)
Gneg-o [¶]	100 (5–100)	83 (61–95)	75 (22–99)	83 (58–96)

Data are presented as the PPV and NPV (95% CI) of the presence of pathogens and Gneg-o in nasopharyngeal (n=25) and oropharyngeal (n=25) samples for the presence of these bacteria in bronchoalveolar lavage. PPV: positive predictive value; NPV: negative predictive value; Gneg-o: bacteria of the *Enterobacteriaceae* family and non-fermenting bacteria. #: only detected once in an oropharyngeal sample and twice in bronchoalveolar lavage; ¶: *Escherichia coli* was most frequently cultured.

saliva and OP samples, induced sputum and expectorated sputum from children with CF; they concluded that induced sputum may be a better proxy for lung colonisation than OP swabs. Our study confirms their opinion on the limited role of OP swabs, and also shows that NP samples, which according to the World Health Organization (WHO) are the gold standard for detecting respiratory pathogens in the general population, are also not an appropriate substitute.

A second interesting finding was that the concordance between URT and BAL increased over time: this might have been a result of antibiotic use providing similar selective pressure in URT and lung niches. Antibiotics might also explain why with progressing disease microbiota diversity decreased and the microbiota composition of the niches started to look more similar. Alternatively, micro-aspiration of URT secretions and gastro-oesophageal reflux might induce community mixing, although clinically only a quarter of our patients used anti-reflux medication for this problem.

At the population level, however, our study showed that OP microbiota profiles resembled lung microbiota more closely in composition and diversity than did NP microbiota, although we observed a distinction between the plausible origin of commensals (OP microbiota) and those of respiratory pathogens (NP microbiota). The generally higher concordance between OP and lung is consistent with previous studies in young [16] and school-age children [39] and adolescents [17] with early-stage CF. Although we cannot rule out that the resemblance between OP and BAL niches was a result of contamination during BAL procedures, the low intra-individual concordance observed between both niches seems to refute this possibility.

Furthermore, we concluded that the lung microbiota of children changed significantly over an 8-month interval, and was accompanied by an increase in diversity, suggesting there is still a dynamic development of lung microbiota in the second half of the first year of life. From previous research we know that, in CF patients, diversity peaks during childhood and then falls during the second decade of life, when it is more frequently dominated by *P. aeruginosa* [1, 40–42]. These data therefore support the conclusion that regular samples should be obtained to create an accurate picture of the dynamics of bacteria populating the LRT in infants with CF.

We additionally compared our sequencing result with conventional culture results. For decades, a large body of literature has tried to estimate the PPVs and NPVs of oropharynx cultures for lung colonisation. In 1991 RAMSEY *et al.* [43] stated that the presence of *S. aureus*, *P. aeruginosa* and *H. influenzae* in oropharyngeal cultures of very young children with CF is highly predictive for carriage in the lung (high PPV). They also stated that the absence of these organisms in the URT does not rule out their presence in the lower airways (low NPV). However, later studies have generated conflicting information on this subject [13, 14]. In our cohort, we observed high PPVs and NPVs for culture results of both URT niches predicting lung colonisation when measured per bacterial species. However, and more importantly, the complete set of pathogens in the lung was only rarely observed in any of the corresponding URT niches, underlining the low predictive value of URT cultures for lung colonisation.

Altogether, our data suggest that the lungs of infants with CF have their own independent microbiome that seems seeded by, but is not identical to, the URT microbiome. In this study, we determined the additional value of multi-lobar sampling [8] this early in life, to decipher whether there are different ecosystems present within the lungs. We observed high concordance between the microbiota profiles of different lobes within individual infants, which is consistent with other studies in healthy individuals [44, 45] and adult patients with stable CF [46]. This finding might have been a result of the still fairly healthy lungs in this

infant population: it is therefore still plausible that among CF patients with an unstable phenotype and/or chronic infection and inflammation, individual lobes might develop their own microbial community and pathogens over time. To be able to capture this phenomenon, BAL procedures might be advantageous over induced sputum samples. In infants, however, induced sputum samples may be preferred because of their low invasive nature and repeatability.

The strength of this study is that we used 16S-rRNA-based sequencing to answer clinically relevant questions. Using this technique, we were able to characterise the complete microbial communities residing in both the URT and LRT of infants with CF, giving a more detailed picture of the composition and relative contribution of different bacteria than would be possible using conventional culturing. Furthermore, we focused our analyses on the intra-individual rather than the inter-individual concordance of sequencing and culturing results, which strongly corresponds to clinical decision making in individual patients. To appreciate the results of our study, several limitations of our study should be taken into account. It was not possible to assess the potential impact of antibiotic use in our study population owing to the young age and a relatively short follow-up time. For this, prolonged follow-up is necessary. Second, sampling of lung secretions in living subjects requires a bronchoscope to pass through the oropharynx; although we took procedural precautions during BAL and laboratory procedures to avoid confounding by oropharyngeal contamination, this cannot be ruled out completely. Furthermore, because environmental contamination could have had an impact on our findings, we additionally sequenced and subsequently clustered “negative” controls together with our lowest density samples, showing negligible cross-contamination between the environment and samples. In addition, because we only targeted a small fragment of the 16S-rRNA gene, species-level annotations were based on a level of uncertainty; targeting more or longer fragments could solve this limitation. Last, our patients were asymptomatic at the time of the BAL procedure; therefore, we cannot extrapolate our findings to situations of active infection.

In conclusion, we have demonstrated inconsistent intra-individual concordance between microbiota of the upper and lower respiratory niches. Furthermore, we have shown that microbiota are highly concordant between different lobes of the lungs. Finally, we observed a dynamic development of lung microbiota in the second half of the first year of life. More work is needed to design alternative and less invasive methods to monitor lung colonisation, e.g. through the use of induced sputum in non-expectorating infants. Whether these data can be reliably extrapolated to infants with symptomatic CF needs to be evaluated. Moreover, we recommend investigating the additive effect of microbiota profiling above conventional culture methods in order to guide treatment of (infectious) lung disease in infants with CF.

Acknowledgements

The authors are indebted to all the participating children and their families. We thank all the members of the research team, especially Annelotte Visser and Eveline Nieuwhof; the laboratory staff, with special thanks to Raiza Hasrat, Jody van Engelsdorp Gastelaars and Cindy Kok; and the cooperating institutes for their dedication to the project.

Author contributions: C.K. van der Ent, D. Bogaert, G.A. Tramper-Stranders and E.A.M. Sanders designed the study. G.A. Tramper-Stranders, C.K. van der Ent and D. Bogaert wrote the study protocols. S.M.P.J. Prevaes, C.K. van der Ent, K.M. de Winter-de Groot, H.M. Janssens and D. Bogaert were responsible for recruitment of participants and collection of the samples. M.L.J.N. Chu and D. Bogaert were responsible for qPCR and sequencing. S.M.P.J. Prevaes, D. Bogaert and W.A.A. de Steenhuijsen Piters were responsible for the post-processing of sequences and data analysis. All authors were involved in data interpretation and drafting of the manuscript, approved the final manuscript as submitted, and agree to be accountable for all aspects of the work.

References

- 1 Coburn B, Wang PW, Diaz Caballero J, *et al.* Lung microbiota across age and disease stage in cystic fibrosis. *Sci Rep* 2015; 5: 10241.
- 2 Parkins MD, Floto RA. Emerging bacterial pathogens and changing concepts of bacterial pathogenesis in cystic fibrosis. *J Cyst Fibros* 2015; 14: 293–304.
- 3 Tracy M, Cogen J, Hoffman LR. The pediatric microbiome and the lung. *Curr Opin Pediatr* 2015; 7: 348–355.
- 4 Whiteson KL, Bailey B, Bergkessel M, *et al.* The upper respiratory tract as a microbial source for pulmonary infections in cystic fibrosis. Parallels from island biogeography. *Am J Respir Crit Care Med* 2014; 189: 1309–1315.
- 5 Marik PE. Aspiration pneumonitis and aspiration pneumonia. *N Engl J Med* 2001; 344: 665–671.
- 6 Morris A, Beck JM, Schloss PD, *et al.*, Microbiome Project. Comparison of the respiratory microbiome in healthy nonsmokers and smokers. *Am J Respir Crit Care Med* 2013; 187: 1067–1075.
- 7 van der Gast CJ, Cuthbertson L, Rogers GB, *et al.* Three clinically distinct chronic pediatric airway infections share a common core microbiota. *Ann Am Thorac Soc* 2014; 11: 1039–1048.
- 8 Brennan S, Gangell C, Wainwright C, *et al.* Disease surveillance using bronchoalveolar lavage. *Paediatr Respir Rev* 2008; 9: 151–159.
- 9 Zampoli M, Pillay K, Carrara H, *et al.* Microbiological yield from induced sputum compared to oropharyngeal swab in young children with cystic fibrosis. *J Cyst Fibros* 2016; 15: 605–610.
- 10 Forton J. Induced sputum in young healthy children with cystic fibrosis. *Paediatr Respir Rev* 2015; 16: Suppl. 1, 6–8.
- 11 Mainz JG, Naehrlich L, Schien M, *et al.* Concordant genotype of upper and lower airways *P. aeruginosa* and *S. aureus* isolates in cystic fibrosis. *Thorax* 2009; 64: 535–540.

- 12 Taylor L, Corey M, Matlow A, *et al.* Comparison of throat swabs and nasopharyngeal suction specimens in non-sputum-producing patients with cystic fibrosis. *Pediatr Pulmonol* 2006; 41: 839–843.
- 13 Rosenfeld M, Emerson J, Accurso F, *et al.* Diagnostic accuracy of oropharyngeal cultures in infants and young children with cystic fibrosis. *Pediatr Pulmonol* 1999; 28: 321–328.
- 14 Doumit M, Belessis Y, Stelzer-Braid S, *et al.* Diagnostic accuracy and distress associated with oropharyngeal suction in cystic fibrosis. *J Cyst Fibros* 2016; 15: 473–478.
- 15 Armstrong DS, Grimwood K, Carlin JB, *et al.* Bronchoalveolar lavage or oropharyngeal cultures to identify lower respiratory pathogens in infants with cystic fibrosis. *Pediatr Pulmonol* 1996; 21: 267–275.
- 16 Renwick J, McNally P, John B, *et al.* The microbial community of the cystic fibrosis airway is disrupted in early life. *PLoS One* 2014; 9: e109798.
- 17 Boutin S, Graeber SY, Weitnauer M, *et al.* Comparison of microbiomes from different niches of upper and lower airways in children and adolescents with cystic fibrosis. *PLoS One* 2015; 10: e0116029.
- 18 Goddard AF, Staudinger BJ, Dowd SE, *et al.* Direct sampling of cystic fibrosis lungs indicates that DNA-based analyses of upper-airway specimens can misrepresent lung microbiota. *Proc Natl Acad Sci USA* 2012; 109: 13769–13774.
- 19 Prevaes SMPJ, de Winter-de Groot KM, Janssens HM, *et al.* Development of the nasopharyngeal microbiota in infants with cystic fibrosis. *Am J Respir Crit Care Med* 2016; 193: 504–515.
- 20 Satzke C, Turner P, Virolainen-Julkunen A, *et al.* Standard method for detecting upper respiratory carriage of *Streptococcus pneumoniae*: updated recommendations from the World Health Organization Pneumococcal Carriage Working Group. *Vaccine* 2013; 32: 165–179.
- 21 Matsen JM, Ederer GM. Specimen collection and transport. *Hum Pathol* 1976; 7: 297–307.
- 22 Linnane B, Clarke D, Murray P, *et al.* The benefit of taking a control sample when performing bronchoalveolar lavage. *Thorax* 2015; 70: 990.
- 23 de Blic J, Midulla F, Barbato A, *et al.* Bronchoalveolar lavage in children. ERS Task Force on bronchoalveolar lavage in children. European Respiratory Society. *Eur Respir J* 2000; 15: 217–231.
- 24 Wyllie AL, Chu MLJN, Schellens MHB, *et al.* *Streptococcus pneumoniae* in saliva of Dutch primary school children. *PLoS One* 2014; 9: e102045.
- 25 Bogaert D, Keijsers B, Huse S, *et al.* Variability and diversity of nasopharyngeal microbiota in children: a metagenomic analysis. *PLoS One* 2011; 6: e17035.
- 26 Biesbroek G, Sanders EAM, Roeselers G, *et al.* Deep sequencing analyses of low density microbial communities: working at the boundary of accurate microbiota detection. *PLoS One* 2012; 7: e32942.
- 27 Caporaso JG, Kuczynski J, Stombaugh J, *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nat Meth* 2010; 7: 335–336.
- 28 Hornef M. Pathogens, commensal symbionts, and pathobionts: discovery and functional effects on the host. *ILAR J* 2015; 56: 159–162.
- 29 de Steenhuijsen Pitsers WAA, Sanders EAM, Bogaert D. The role of the local microbial ecosystem in respiratory health and disease. *Philos Trans R Soc Lond B Biol Sci* 2015; 370: pii: 20140294.
- 30 Bosch AATM, Biesbroek G, Trzciński K, *et al.* Viral and bacterial interactions in the upper respiratory tract. *PLoS Pathog* 2013; 9: e1003057.
- 31 Biesbroek G, Bosch AATM, Wang X, *et al.* The impact of breastfeeding on nasopharyngeal microbial communities in infants. *Am J Respir Crit Care Med* 2014; 190: 298–308.
- 32 Magurran AE. Measuring Biological Diversity. Oxford, Blackwell Science, 2004.
- 33 Rogers GB, Shaw D, Marsh RL, *et al.* Respiratory microbiota: addressing clinical questions, informing clinical practice. *Thorax* 2015; 70: 74–81.
- 34 Letunic I, Bork P. Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. *Nucleic Acids Res* 2011; 39: W475–W478.
- 35 Oksanen J, Blanchet FG, Kindt R, *et al.* The R-project for statistical computing. Community ecology package “vegan”. 2015; version 2.3.0. <https://cran.r-project.org/src/contrib/Archive/vegan/>
- 36 Sly PD, Brennan S, Gangell C, *et al.*, on behalf of the Australian Respiratory Early Surveillance Team for Cystic Fibrosis (AREST-CF). Lung disease at diagnosis in infants with cystic fibrosis detected by newborn screening. *Am J Respir Crit Care Med* 2009; 180: 146–152.
- 37 Sly PD, Gangell CL, Chen L, *et al.* Risk factors for bronchiectasis in children with cystic fibrosis. *N Engl J Med* 2013; 368: 1963–1970.
- 38 Zemanick ET, Wagner BD, Robertson CE, *et al.* Assessment of airway microbiota and inflammation in cystic fibrosis using multiple sampling methods. *Ann Am Thorac Soc* 2015; 12: 221–229.
- 39 Klepac-Ceraj V, Lemon KP, Martin TR, *et al.* Relationship between cystic fibrosis respiratory tract bacterial communities and age, genotype, antibiotics and *Pseudomonas aeruginosa*. *Environ Microbiol* 2010; 12: 1293–1303.
- 40 Cox MJ, Allgaier M, Taylor B, *et al.* Airway microbiota and pathogen abundance in age-stratified cystic fibrosis patients. *PLoS One* 2010; 5: e11044.
- 41 VanDevanter DR, LiPuma JJ. Microbial diversity in the cystic fibrosis airways: where is thy sting? *Future Microbiol* 2012; 7: 801–803.
- 42 Hampton TH, Green DM, Cutting GR, *et al.* The microbiome in pediatric cystic fibrosis patients: the role of shared environment suggests a window of intervention. *Microbiome* 2014; 2: 14.
- 43 Ramsey BW, Wentz KR, Smith AL, *et al.* Predictive value of oropharyngeal cultures for identifying lower airway bacteria in cystic fibrosis patients. *Am Rev Respir Dis* 1991; 144: 331–337.
- 44 Dickson RP, Erb-Downward JR, Freeman CM, *et al.* Spatial variation in the healthy human lung microbiome and the adapted island model of lung biogeography. *Ann Am Thorac Soc* 2015; 12: 821–830.
- 45 Charlson ES, Bittinger K, Haas AR, *et al.* Topographical continuity of bacterial populations in the healthy human respiratory tract. *Am J Respir Crit Care Med* 2011; 184: 957–963.
- 46 Hogan DA, Willger SD, Dolben EL, *et al.* Analysis of lung microbiota in bronchoalveolar lavage, protected brush and sputum samples from subjects with mild-to-moderate cystic fibrosis lung disease. *PLoS One* 2016; 11: e0149998.