The cystic fibrosis lower airways microbial metagenome

Patricia Moran Losada[1], Philippe Chouvarine[1], Marie Dorda[1], Silke Hedtfeld[1], Samira Mielke[1], Angela Schulz[1,2], Lutz Wiehlmann[1,2] and Burkhard Tümmler[1,2]

Affiliations: [1]Clinic for Paediatric Pneumology, Allergology and Neonatology, OE 6710, Hannover Medical School, Hannover, Germany. [2]Biomedical Research in Endstage and Obstructive Lung Disease (BREATH), Member of the German Center for Lung Research, Hannover, Germany.

Correspondence: Burkhard Tümmler, Clinic for Paediatric Pneumology, Allergology and Neonatology, OE 6710, Hannover Medical School, Carl-Neuberg-Str. 1, D-30625 Hannover, Germany.
E-mail: tuemmler.burkhard@mh-hannover.de

ABSTRACT   Chronic airway infections determine most morbidity in people with cystic fibrosis (CF). Herein, we present unbiased quantitative data about the frequency and abundance of DNA viruses, archaea, bacteria, moulds and fungi in CF lower airways.

Induced sputa were collected on several occasions from children, adolescents and adults with CF. Deep sputum metagenome sequencing identified, on average, approximately 10 DNA viruses or fungi and several hundred bacterial taxa.

The metagenome of a CF patient was typically found to be made up of an individual signature of multiple, lowly abundant species superimposed by few disease-associated pathogens, such as *Pseudomonas aeruginosa* and *Staphylococcus aureus*, as major components. The host-associated signatures ranged from inconspicuous polymicrobial communities in healthy subjects to low-complexity microbiomes dominated by the typical CF pathogens in patients with advanced lung disease. The DNA virus community in CF lungs mainly consisted of phages and occasionally of human pathogens, such as adeno- and herpesviruses. The *S. aureus* and *P. aeruginosa* populations were composed of one major and numerous minor clone types.

The rare clones constitute a low copy genetic resource that could rapidly expand as a response to habitat alterations, such as antimicrobial chemotherapy or invasion of novel microbes.

@ERSpublications
The CF lung metagenome is composed of few viruses and fungi and hundreds of bacterial species, clones and subclones http://ow.ly/ZiqUE

Lessons for clinicians

- More than 1000 microbial species were identified in the airways of children, adolescents and adults with cystic fibrosis (CF) by high-throughput shotgun metagenome sequencing of induced sputum samples.
- Actinobacteria, Bacteroidetes, Firmicutes, Fusobacteria and Proteobacteria constitute >99% of the bacterial community in CF airways.
- The DNA virome mainly consists of bacteriophages and occasionally of human pathogens, such as adeno- or herpes viruses and the mycobiome mainly consists of *Candida* and *Aspergillus* species.
- A normal microbial airway metagenome lacking any CF-associated pathogens was observed in healthy CF adults with normal spirometry and normal lung clearance index. These index cases have been regularly seen since the age of diagnosis suggesting that continuous surveillance and intervention, if indicated, could prevent or decelerate progressive CF lung disease in some, but not all, subjects.
- *Pseudomonas aeruginosa* was present in all samples from exocrine pancreatic insufficient (PI) CF patients, albeit prevalence was very low in subjects classified as *P. aeruginosa*-negative by standard bacteriology. *P. aeruginosa* is apparently acquired early in life in all PI CF patients. This novel information should be considered in forthcoming best practise guidelines for infection control and *P. aeruginosa* eradication programmes.
- At present, the clonal populations of the common CF pathogens *Staphylococcus aureus* and *P. aeruginosa* in CF airways are more complex than presumed. Besides one or two major clone types that are amenable to standard culture-dependent diagnostics, several rare clone types coexist in CF lungs which could rapidly expand as a response to antimicrobial chemotherapy or invasion of novel microbes.

Introduction

CF is a life-shortening, debilitating, autosomal recessive disease that is caused by mutations in the *cystic fibrosis transmembrane conductance regulator* (*CFTR*) gene [1]. The basic defect of impaired epithelial chloride and bicarbonate secretion predisposes CF patients to chronic airway infections with opportunistic pathogens, which determine most morbidity in people with CF [1]. Epidemiological data drawn from culture-dependent diagnostics of respiratory specimens indicated that the airways of CF patients become colonised with *Haemophilus influenzae* and *Staphylococcus aureus* during early childhood, followed later in life by *Pseudomonas aeruginosa* and sometimes by organisms such as *Burkholderia cepacia* complex or atypical mycobacteria [2]. Culture-independent technologies, however, revealed that the CF respiratory tract is not inhabited by these few pathogens, but rather by complex polymicrobial communities [3–10]. Sequencing of PCR-amplified parts of bacterial 16S ribosomal (r)DNA genes could identify over 100 distinct genera including *Streptococcus* and numerous anaerobes as major players that are routinely not detected during the culture-dependent processing of CF-derived respiratory secretions.

Sequence variations in the ancient and ubiquitous ribosomal RNA genes are considered to reflect the universal molecular clock of life [11]. Correspondingly, the composition of microbial communities is described by the abundance of individual rDNA sequences. However, one has to accept some inherent biases of this approach. First, prokaryotes and eukaryotes need to be analysed separately due to their basic difference of rDNA sequence and, correspondingly, most work has been confined to the bacterial microbiota. Secondly, the sequences of the sample are prepared for analysis by oligonucleotide primer-based amplification steps. Subsequent sequencing of the amplicons has a limited ability to resolve taxonomic identification to the species level, may fail to detect phyla and can skew the estimation of species relative abundance in a community [12, 13].

These limitations can be overcome by whole-genome shotgun sequencing [14], which may provide information about the composition of the community up to the level of clonal complexes. Within the context of the CF lungs, the published studies to date have focused on the sputum metagenomes derived from CF adults with advanced lung disease [15–17]. Each of the 10 so far investigated subjects hosted a unique polymicrobial community.

This study extends the scope to patients of all age groups and grades of disease severity in order to cover the whole range of lower airways microbial metagenomics in CF. Induced sputa collected from exocrine pancreatic sufficient (PS) and PI children, adolescents and adults with CF were investigated by whole-genome shotgun sequencing in order to identify the bacteria, archaea, DNA viruses, moulds and fungi residing in the respiratory secretions. Normalisation provided quantitative data about the relative and absolute abundance of microbial species. Moreover, by focusing on the major CF pathogens *S. aureus* and *P. aeruginosa* the metagenome sequence data sets were examined for the number of coexisting clone types and uncommon, probably *de novo*, mutations in antimicrobial resistance determinants. In future,

such a metagenome-guided deep insight into clone and sequence variations could assist in our management of respiratory tract infections.

Methods

Patients

Subjects with CF were recruited from the CF clinic of Hanover Medical School (Hannover, Germany). All patients had been regularly seen at the CF clinic since the age of diagnosis. The diagnosis of CF had been made by the detection of two disease-causing mutations in the *CFTR* gene [18], and elevated chloride concentrations in the Gibson–Cooke pilocarpine iontophoresis sweat test [19] or a Sermet score in the CF range of nasal transepithelial potential difference measurements [20, 21] and/or chloride secretory responses in the CF range of intestinal current measurements [22]. Exocrine pancreatic status was assessed by the fecal elastase-1 test [23]. Lung function was assessed by spirometry, body plethysmography and, in the healthier subjects, multiple breath nitrogen washout [24, 25]. The 19 exocrine PI patients were homozygous for the most common CF mutation, p.Phe508del [18]. The 11 PS subjects were either compound heterozygous for a PI- and a PS-conferring *CFTR* mutation (n=10) [18] or homozygous for a PS mutation (n=1) [18]. At the date of recruitment patients were either 8–13 years old (children, group A), 18–23 years old (adolescents and young adults, group B) or >28 years old (adults, group C) (table 1). Patients were also classified by disease severity. Healthy CF subjects had normal anthropometry (body mass index >19 kg·m^{-2}) and normal lung function, *i.e.* multiple breath nitrogen washout revealed a normal lung clearance index and spirometry yielded forced expiratory volume in 1 s (FEV_1) values >90% predicted. Mildly affected CF patients (category "mild") exhibited normal anthropometry, an anomalous lung clearance index and FEV_1 values of 70–110% pred on the day of recruitment. Lung function was chronically compromised for ⩾3 years in all moderately or severely affected CF patients (FEV_1 50–70% pred, category "moderate"; FEV_1 30–50% pred in the absence of an acute pulmonary exacerbation, category "severe"; FEV_1 <30% pred in the absence of an acute pulmonary exacerbation, category "end-stage lung disease"). The clinical status of the individual subjects is described in table S1. The study was approved by the Ethics Committee of Hannover Medical School (no. 1510-2012).

Wet lab experimental procedures

The procedures are described in the supplementary material. Briefly, induced sputum was collected by autogenic drainage during cycles of 3-min inhalation of 3% (w/v) hypertonic saline. After sputa had been diluted with buffer and subjected to hypotonic lysis, DNA was purified from the suspension according to the Hard-to-lyse-Bacteria protocol with the NucleoSpin Tissue kit (Machery-Nagel, Düren, Germany). DNA libraries were prepared from sheared DNA according to an in-house protocol. Sequencing was performed on a SOLiD 5500XL system (Life Technologies, Carlsbad, CA, USA) in colour space with 75 bp read length and implemented Exact call chemistry (Life Technologies).

In silico analyses

Approaches and software are described in the supplementary material. In brief, the raw sequence reads were trimmed and then (in this order) low quality reads, human reads, non-human low-complexity reads and non-human reads encoding mobile genetic elements were removed. The remaining microbial reads were normalised by GC content and genome length. This curated data set was then used for the identification of taxa (DNA viruses, bacteria, archaea, moulds and fungi), principal component analysis, search for mutation in antimicrobial resistance genes and analysis of the *S. aureus* and *P. aeruginosa* populations in the respiratory secretions.

Results

Lower airways microbial metagenome of PS and PI individuals with CF

Induced sputa were collected within 1 year in 6- to 20-week intervals from 11 PS (n=1–4, median two samples) and 19 p.Phe508del homozygous PI subjects with CF (n=1–5, median two samples). Deep

TABLE 1 Study participants		
Age group	**PI CF patients**	**PS CF patients**
A: 8–13 years	7	4
B: 18–23 years	5	4
C: >28 years	7	3

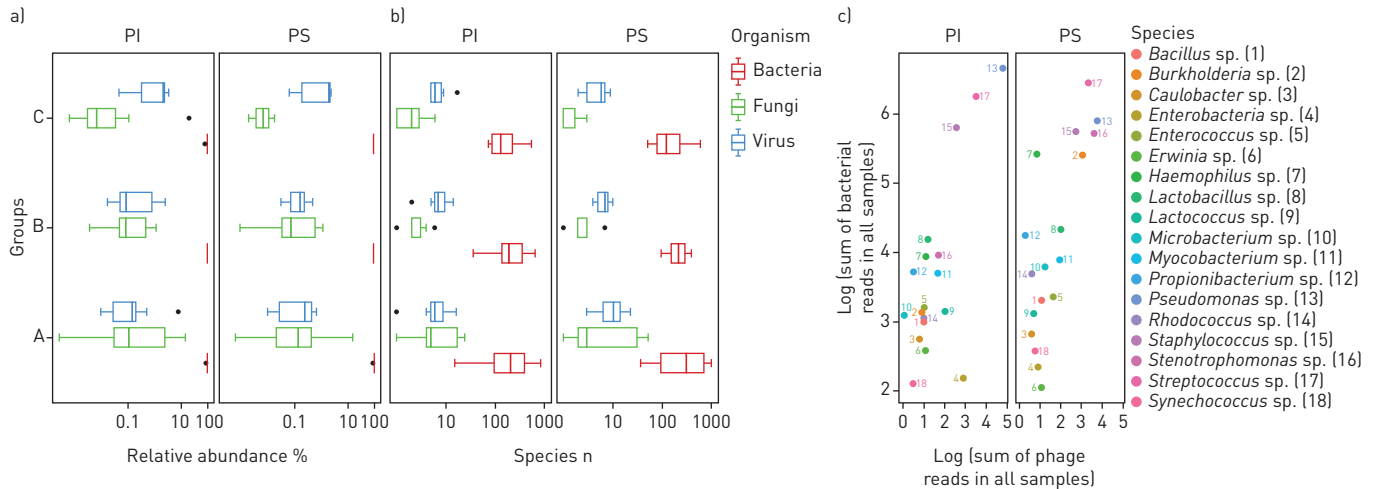PI: exocrine pancreatic insufficient; CF: cystic fibrosis; PS: exocrine pancreatic sufficient.

FIGURE 1 The cystic fibrosis (CF) lower airways metagenome: cumulative presentation of data sets. a, b) Box plot presentation of the absolute frequency and the relative abundance of DNA viruses, bacteria and fungi identified in induced sputa sampled from the lower airways of exocrine pancreas insufficient (PI) and exocrine pancreas sufficient (PS) children (A), adolescents (B) and adults (C) with CF. c) The dots in the double logarithmic plot depict the sum of reads identified for bacteriophages and their corresponding bacterial hosts at the genus level. Genera are differentiated by number and colour.

metagenome sequencing identified, on average, <10 DNA viruses or fungi and several hundred bacterial taxa in samples from children (group A), adolescents (group B) and adults (group C) (figure 1b, figure S1, tables S2 and S3). Bacteria typically made up > 99% of the microbial community (figure 1a) [26]. The median contribution of DNA viruses and fungi was on average <1%, but it varied between 0.002% and 11% in the individual sample. The dominant taxa in the cumulative metagenome of the whole cohort were the bacterial species that are most frequently reported from culture-dependent diagnostics, *i.e.* streptococci, staphylococci, pseudomonads, *Haemophilus* sp., *Burkholderia* sp. and *Stenotrophomonas* sp. (figure 1c, table S3). Thus, quantitative metagenomics fits with culture-based epidemiological data for the most abundant bacterial species in CF lungs.

Figure 2 provides a detailed overview of the frequency of recovery of all detected species and their relative abundance. Taxa and primary data are listed in tables S1 and S3. The viral community consisted primarily of phages (figure 1c), a few human pathogens, primarily herpes virus and adenovirus, and rare cases of viruses infecting non-mammalian eukaryotic hosts. Dominant species in the mycobiome were *Aspergillus* species and Saccharomycetes including *Candida* sp. which had an average abundance of 4.3% and 94.5% in the mycobiome, respectively, consistent with current knowledge of CF mycology [27]. The community of bacteria and archaea turned out to be highly diverse, including numerous very lowly abundant species and phyla that have not yet been reported to inhabit the niche of the human CF lung. In total 1049 and 977 species belonging to 30 phyla were identified in the sputa from PS and PI patients, respectively (table S4). All archaeal phyla and 21 of the 26 bacterial phyla each had an abundance of <0.1%. The remaining five phyla, *i.e.* Actinobacteria, Bacteroidetes, Firmicutes, Fusobacteria and Proteobacteria, made up 99.8% and 99.9% of the bacterial community in PS and PI CF sputa, respectively (table 2). Besides the common CF pathogens *S. aureus*, *P. aeruginosa*, *H. influenzae* and *Burkholderia* spp., species belonging to the genera *Rothia*, *Proteus* and *Streptococcus* were most common among aerobes and facultative anaerobes. The community of anaerobic bacteria consisted, on average, of 98% of *Prevotella melaninogenica* and *Veillonella parvula* (table S4), typical inhabitants of the oral cavity that were also prominent members in the airway metagenome of healthy controls (table S5). Moreover, four further anaerobes were regularly identified as minor members in CF sputum, *i.e.* the oral bacteria *Atopobium parvulum* and *Fusobacterium nucleatum*, the intestinal inhabitant *Megasphaera elsdenii* and the zoonotic airway pathogen *Streptobacillus moniliformis* (table S4).

The lead pathogen *P. aeruginosa* was identified in all specimens from PI patients although six of them had been classified as *P. aeruginosa* negative according to the clinical records. This finding suggests the ubiquitous presence of *P. aeruginosa* in respiratory secretions of PI CF patients. However, in all specimens taken from these misclassified patients *P. aeruginosa* was only present at low abundance on average in 0.02% of all reads. When we processed 24 respiratory specimens from healthy volunteers by the same protocol, no *P. aeruginosa* sequences were detectable in 14 samples and <10 reads could be assigned to *P. aeruginosa* in six samples (table S5). This data strongly indicates that the *P. aeruginosa* sequences detected in the CF specimens were retrieved from the patients' airways and did not originate from any
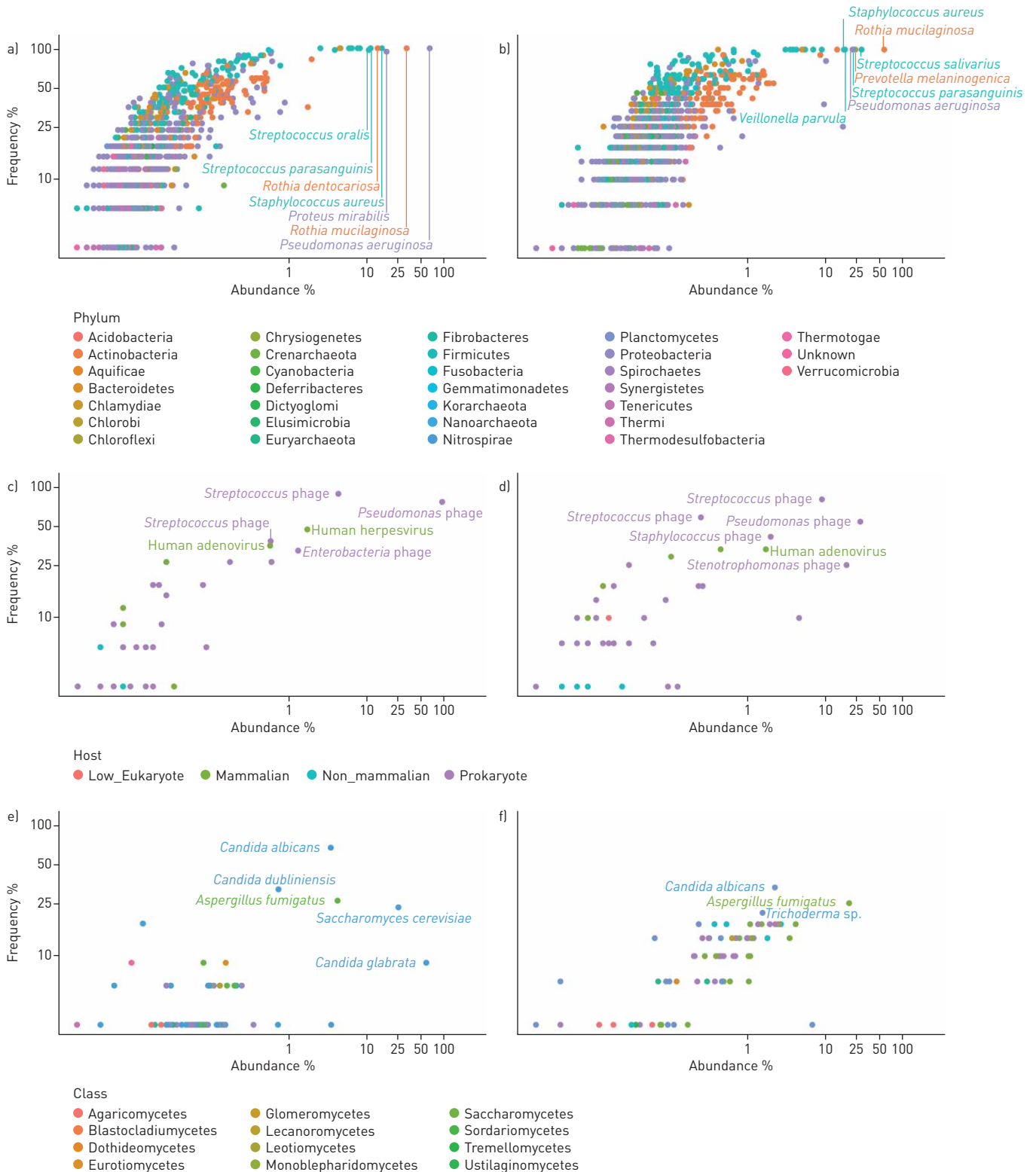
FIGURE 2 Frequency and abundance of microbial species in the cystic fibrosis (CF) lower airways metagenome. Each dot in the double logarithmic graphs depicts the identification frequency of a species and its mean relative abundance in sputa collected from a, c, e) exocrine pancreas insufficient and b, d, f) exocrine pancreas sufficient CF patients. Abundance was separately normalised for a, b) bacteria, c, d) DNA viruses and e, f) moulds and fungi. The name of a taxon is depicted for the most common and most abundant species. Bacterial taxa belonging to the same phylum are indexed by matching colour. Viruses are differentiated by their host. Moulds and fungi are differentiated by class. Please note that the data points depict the relative abundance of individual species found in the whole dataset of all exocrine pancreas insufficient and sufficient samples.

TABLE 2 Common bacterial phyla in sputum from people with PS cystic fibrosis and PI cystic fibrosis[#]

Phylum	Share of total abundance	
	PS sputum	PI sputum
Actinobacteria	36.7%	22.9%
Bacteroidetes	9.2%	1.1%
Firmicutes	35.8%	25.2%
Fusobacteria	0.8%	0.1%
Proteobacteria	17.3%	50.6%

PS: exocrine pancreatic sufficient; PI: exocrine pancreatic insufficient. [#]: note that most bacteria with mechanically labile membranes such as Tenericutes (*Mycoplasma* species) are lost during the processing of the samples. Direct sequencing of unprocessed respiratory specimens yielded a share of ~3% of bacterial reads assigned to Tenericutes. However, species with hard-to-digest cell walls, like *Nocardia* or *Mycobacteria*, were not detected in unprocessed specimens.

contamination during processing of the sample. Our metagenome data point to an acquisition of *P. aeruginosa* early in life. The initially minute reservoir may then expand to clinically relevant numbers during exacerbations and/or at some stage of chronic airway remodelling.

The cladograms in figure 3 focus on the bacteria which made up the top 95% of the cumulative metagenome population of PI (figure 3a) and PS CF patients (figure 3b). The population in PI CF airways was dominated by pseudomonads and staphylococci followed by *Veillonella*, *Streptococci*, *Prevotella*, *Rothia* and other enterobacteriaceae as minor contributors. The spectrum of genera was similar in PS CF patients, but streptococci were more prominent and the population was more diverse and less skewed towards *P. aeruginosa*. Principal component analysis revealed a broadly scattering distribution of data sets retrieved from PS patients and strong clustering of data sets for the samples from PI subjects (figure 4) indicating that bacterial communities are more host-specific in PS CF and more disease-specific in PI CF.

Individual microbial metagenome signatures and CF disease severity

Whole metagenome analysis resolved the microbial signature of the individual patient. The spectrum ranged from normal flora in some healthy subjects with a normal lung clearance index *via* an intermediate stage when the normal community is perturbed by *H. influenzae* or *S. aureus* to a final stage of a low-diversity community dominated by *P. aeruginosa* (table S2, figure 5, figure S2) [28]. This shift from a normal highly diverse metagenome indistinguishable from that of a healthy subject to the CF typical, end-stage of an almost pure culture of *P. aeruginosa* was correlated in our patient cohort with disease severity, but not with age. As shown in figure 5, the diversity of the bacterial communities of the top 90% constituents decreased with increasing lung disease severity.

The metagenome of a CF patient was typically found to be made up of an individual signature of multiple, lowly abundant species superimposed by few disease-associated pathogens, such as *P. aeruginosa* and *S. aureus*, as major components (table S6). This phenotype became more obvious if we normalised the microbial reads to the human DNA in the sample. For example, as is illustrated in figure 6, a PS CF patient appeared to have an unrelated metagenome to that of a PI CF patient with chronic colonisation with *P. aeruginosa* if presented in bar charts as fractions of total microbial reads (figure 6a and b). However, after human DNA normalisation it can be seen that a quantitatively similar pattern of non-pathogenic species ("normal flora") in the two subjects is overshadowed by *P. aeruginosa* in the PI CF patient, whereas no typical CF pathogens are detectable in the healthy PS CF patient (figure 6c and d). This normalisation clarifies some discrepancies in complexity between the microbiomes resolved by comprehensive culture-independent techniques and those based on culture-dependent diagnostics, the latter driven to detect disease-associated microbes and ignore the normal flora.

Clonal diversity of S. aureus and P. aeruginosa populations in CF lungs

The clonal composition of *S. aureus* and *P. aeruginosa* communities in CF sputa was determined from the frequency distribution of single-nucleotide polymorphisms in the metagenomes (figure 7, table S7). The *P. aeruginosa* communities consisted of one (16 sputa) or two major clone types (six sputa) making up 96–100% of the consortium. At least one further rare clone was detectable in 13 sputa. Within-clone variation was detectable for the most prevalent clone, which split up into two (12 specimens), three (six specimens), four (two specimens) or five (one specimen) discriminable clonal variants. The *P. aeruginosa*
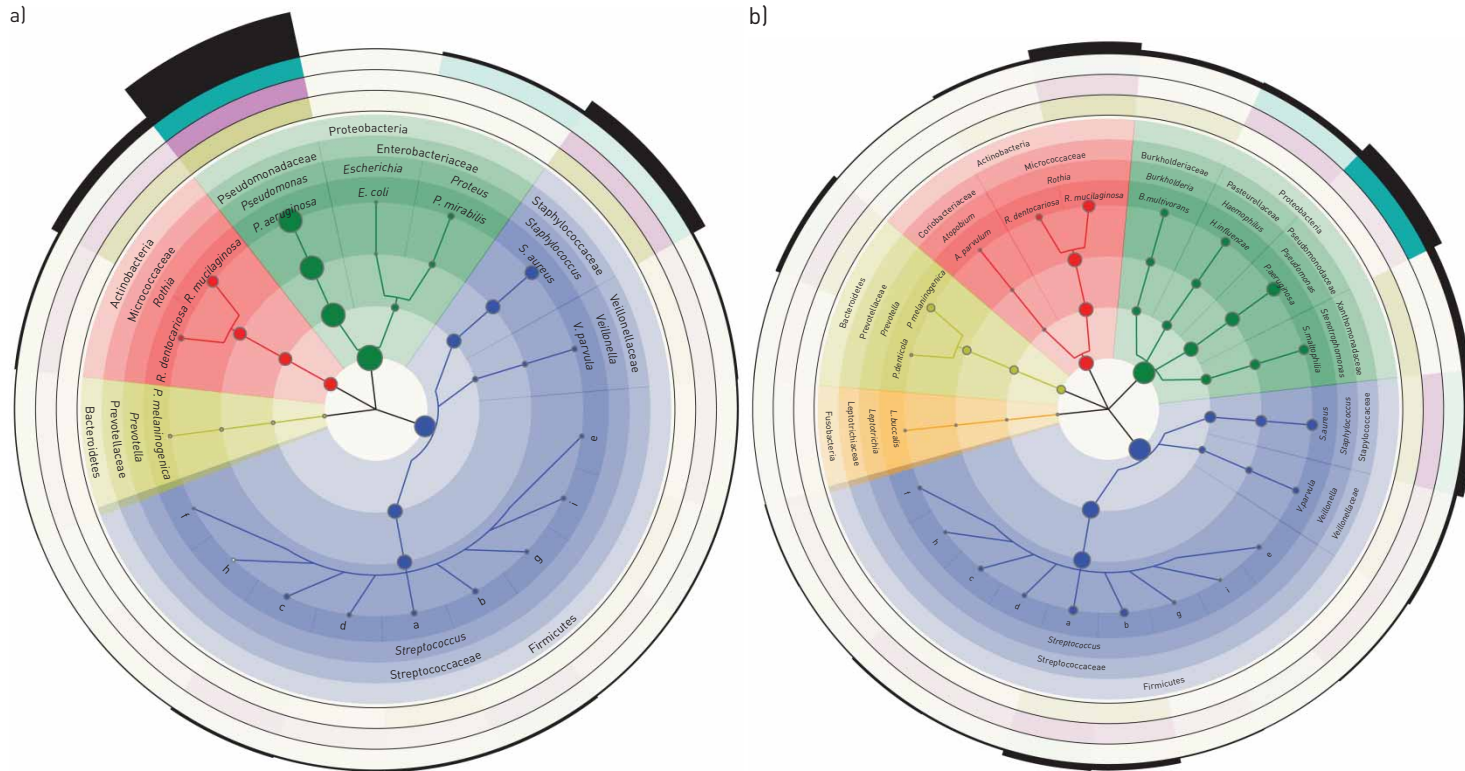
FIGURE 3 Taxonomic cladograms reporting the most abundant species contributing to the top 95% of the bacterial communities of the lower airways of a) exocrine pancreas insufficient and b) exocrine pancreas sufficient cystic fibrosis patients. Circle size is proportional to the log of average abundance of all analysed specimens. The height of the segments of the outermost ring indicates the relative average abundance of the respective genera and phyla in all age groups. The next three outermost circles indicate the relative average abundance of clades by colour intensity for adults (green), adolescents (purple) and children (olive). The analysis is based on 38 and 24 samples from exocrine pancreas insufficient and exocrine pancreas sufficient patients, respectively. Species of the genus *Streptococcus* are: a: *S. parasanguinis*; b: *S. salivarius*; c: *S. oralis*; d: *S. mitis*; e: *S. pneumoniae*; f: *S. thermophilus*; g: *S. sanguinis*; h: *S. pseudopneumoniae*; i: *S. gordonii*. The size of the dots indicates the mean of the average abundance of the three age groups. Within each age group, the mean of the relative abundance of all samples was calculated. Note the different evaluations of the dataset in figures 2 and 3.
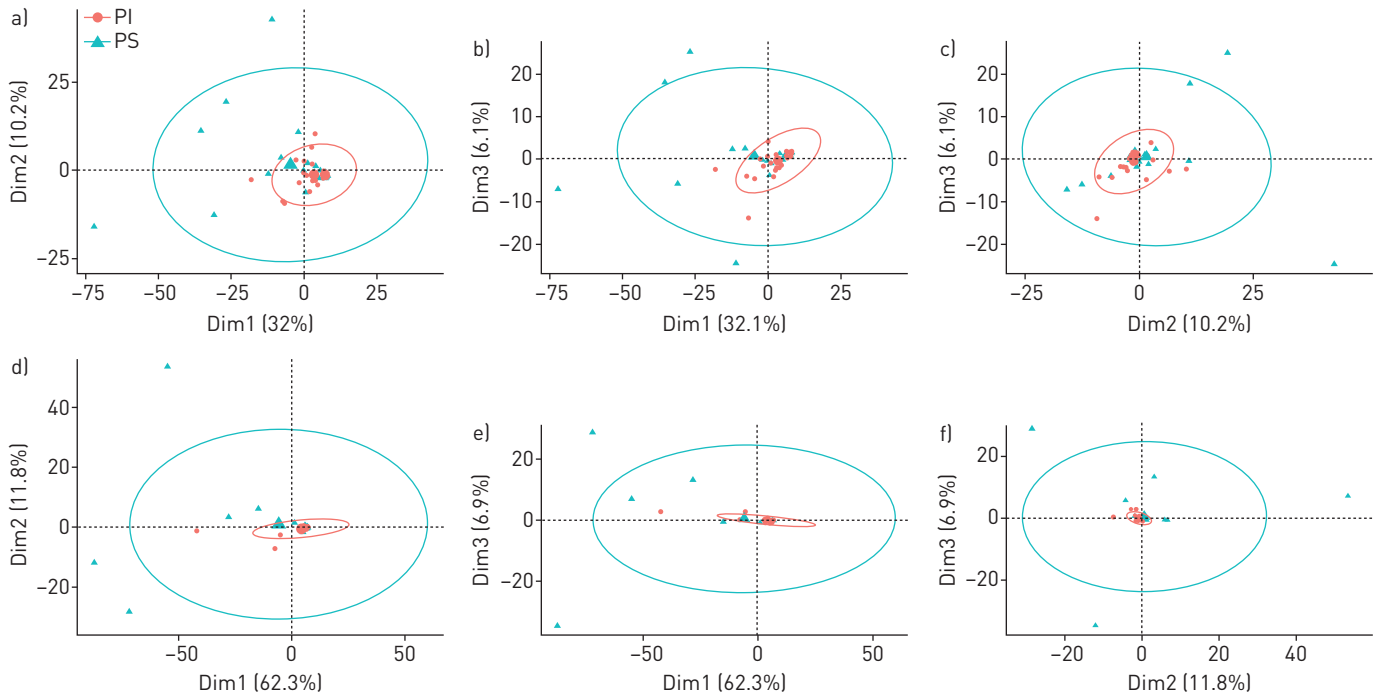
FIGURE 4 Principal component analysis of the sputum metagenome data sets of exocrine pancreas insufficient (PI) and exocrine pancreas sufficient (PS) patients either a–c) not normalised or d–f) normalised to human DNA in the sample. The size of the concentration ellipses in probability was set to 0.95. Dim: dimension.

community remained rather stable in five of the six CF patients who provided serial specimens. A transient shift of the community structure was only observed for subject PIBM5 (figure 7).

The *S. aureus* community consisted in all nine examined cases of one dominant (> 95% of all bacteria) and one minor clone. Two, three or four variants of the dominant clone could be discriminated in four, three and two samples, respectively.

Previous genotyping and subsequent genome sequencing of serial isolates had suggested that the CF lungs are chronically colonised with co-evolving clades of one or, less frequently, two or three unrelated clones [28–31], but our unbiased metagenome data indicate more diverse and more complex compositions of the *S. aureus* and *P. aeruginosa* populations in CF airways. Diversity of the *S. aureus* and *P. aeruginosa* communities is generated by intraclonal diversification [28, 30, 32, 33] and co-colonisation with unrelated clones.

To identify the genotype of the dominant *S. aureus* and *P. aeruginosa* strains within the frame of published typing schemes, the metagenome sequences were searched for matches with a multi-marker array for *P. aeruginosa* [34] and the multilocus sequence typing database for *S. aureus* [35]. Four out of the 10 analysed *P. aeruginosa* strains belonged to ubiquitous clones in the global *P. aeruginosa* population [34] and two pairs of 13 *S. aureus* strains were assigned to the common clone type ST7 and the pandemic methicillin-resistant *S. aureus* lineage ST22 [35], respectively.

Mutations in resistance genes to antimicrobial chemotherapy

The chronic airway infections in CF are treated by chronic or intermittent antimicrobial chemotherapy, at least on the occasion of a pulmonary exacerbation, often accompanied with the emergence of multidrug-resistant bacteria as the unwanted side-effect [1, 2]. We searched the *S. aureus* and *P. aeruginosa* sequences in the metagenomes for non-conservative coding variants in targets of anti-infectives and/or mediators of antimicrobial resistance that have a frequency of <20% in their pangenome (table 3). Mutations in the *P. aeruginosa* genomes affected genes that are known to be prone for mutation during antipseudomonal chemotherapy [36] but, in the case of *S. aureus*, the mutations also emerged in the gyrase-encoding *gyr* loci [37], which are the targets for fluoroquinolones. The patients' clinicians had never prescribed fluoroquinolones as anti-staphyloccocal chemotherapy. Besides the improbable cross-infection with a resistant strain, the treatment of concomitant infections by *P. aeruginosa* with a fluoroquinolone would be the most likely explanation for the collateral mutations in the *S. aureus* *gyr* genes. This example demonstrates the power of non-selective metagenome sequencing to detect genetic variations in traits of interest such as drug resistance, virulence or lifestyle.
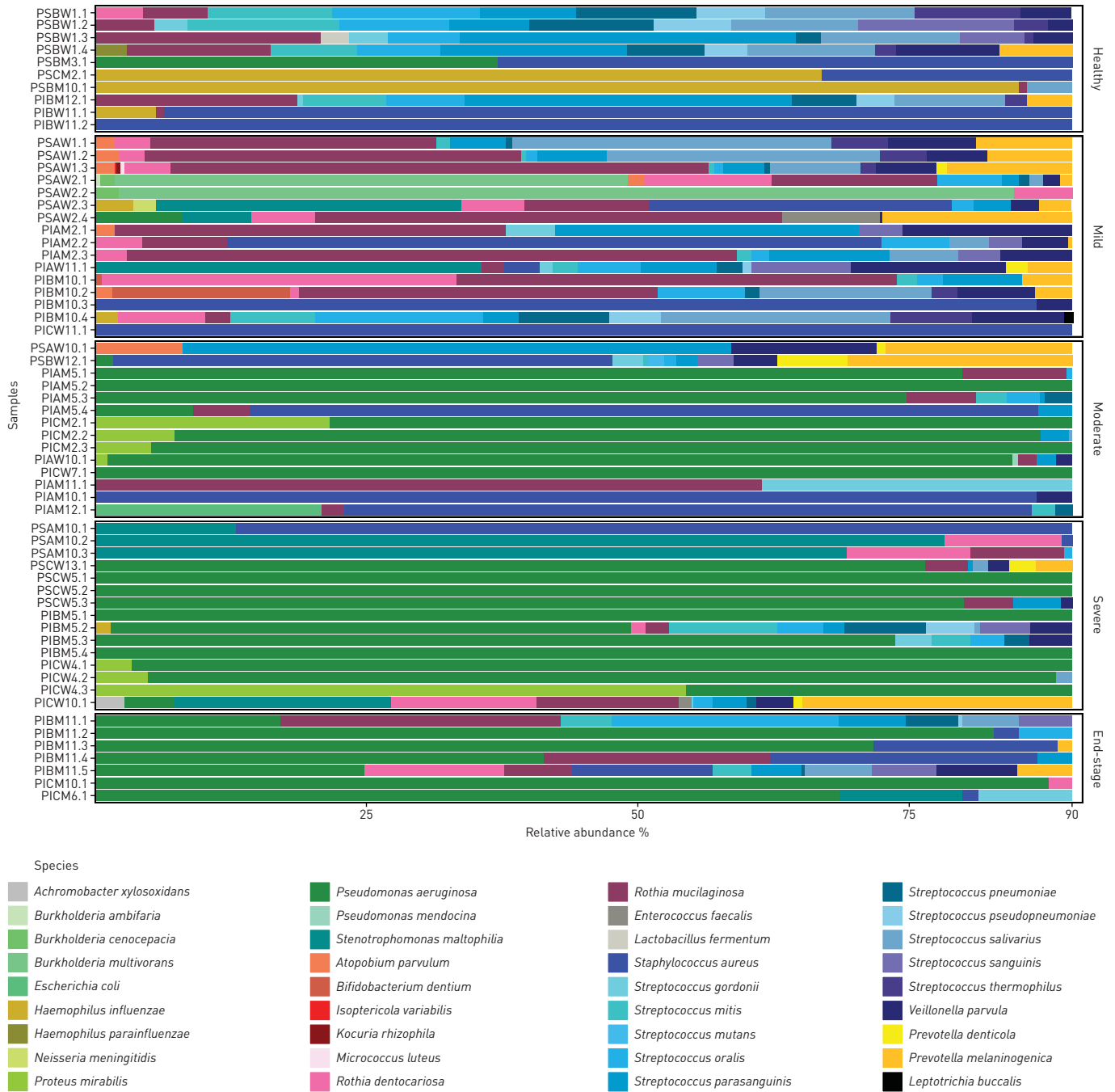
FIGURE 5 Composition of microbial communities in induced sputa of healthy and mild, moderate, severe or very severe (end-stage) lung disease affected individuals with cystic fibrosis. Data are presented in a stacked bar chart of relative abundance as fractions of total reads of the top 90% of species for each patient. The nomenclature of samples is explained in the supplementary material. Classification of disease severity is defined in the Methods section.

Discussion

Deep metagenome sequencing revealed a large repertoire of viruses, moulds, fungi, archaea and bacteria in the CF lung habitat. The lower airways metagenome of a CF patient was typically found to be made up of an individual signature of multiple, lowly abundant species superimposed by few classical CF pathogens, such as *P. aeruginosa* and *S. aureus*, as major components. This phenotype became more obvious if we normalised the microbial reads to the human DNA in the sample. This presentation of data in terms of absolute abundance of microbes, as shown in figure 6, was more similar to the outcome of routine culture-dependent diagnostics that just communicate disease-associated aerobes than the common output
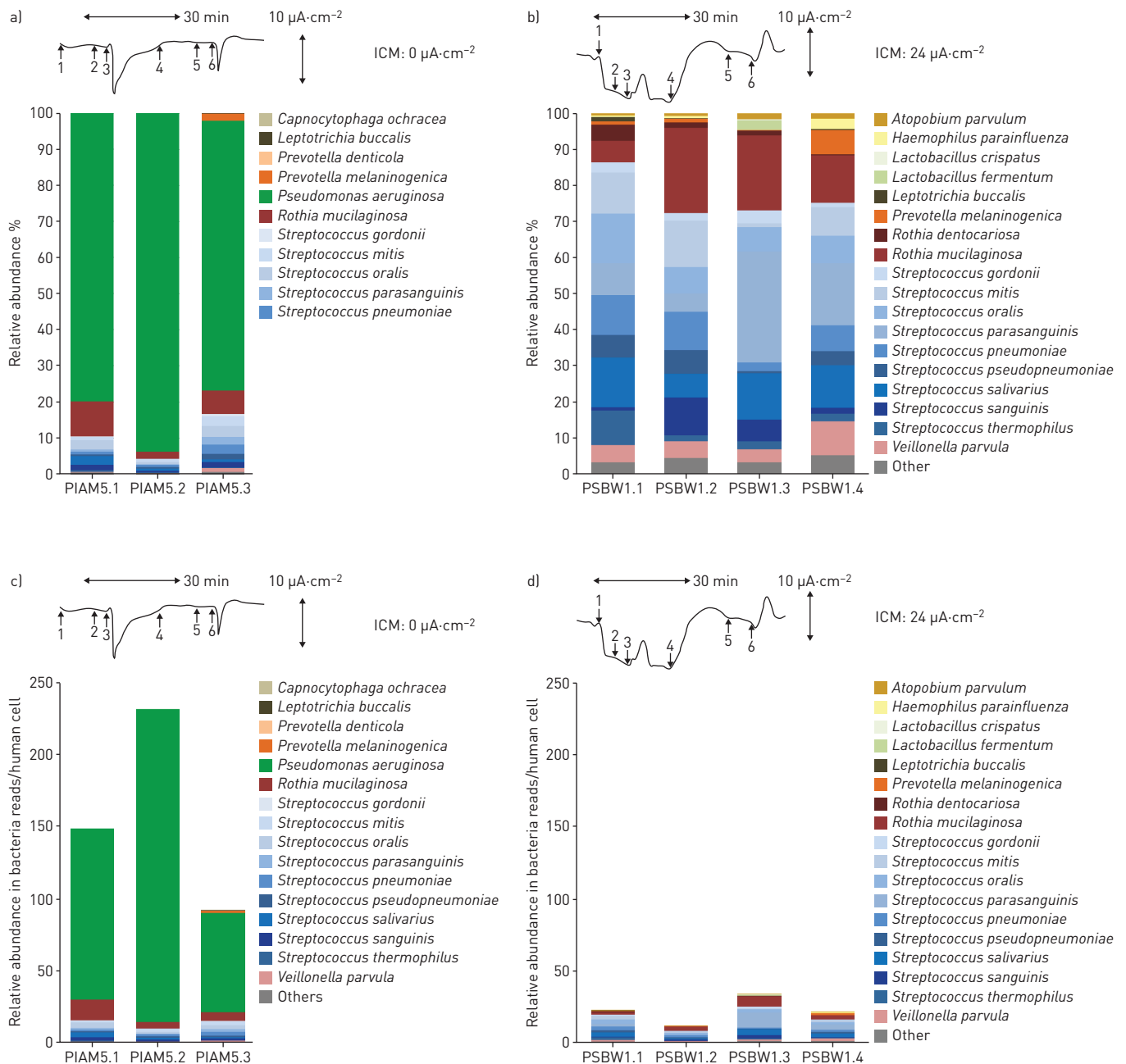
FIGURE 6 Bacterial sputum microbiomes of a, c) an exocrine pancreas insufficient (PI) cystic fibrosis (CF) patient who is homozygous for the most common cystic fibrosis transmembrane conductance regulator (CFTR) mutation p.Phe508del and b, d) an exocrine pancreas sufficient (PS) CF patient who is homozygous for the rare splice mutation in the CFTR gene c.1766+3 A-G. Data are presented as a, b) stacked bar charts of relative abundance as fractions of total bacterial reads or c, d) number of bacterial reads per human cell. Whereas the normalised bacterial microbiomes (a, b) suggest unrelated compositions of the bacterial communities in the respiratory tracts of the two patients, the normalisation per human cell (c, d) unravels similar colonisation modes with anaerobes in the two patients, which are overshadowed by the dominant *Pseudomonas aeruginosa* community in the PI patient. Tracings show original recordings of short circuit currents in rectal suction biopsies [22] to characterise the basic defect and CFTR function in the two individuals with CF. The horizontal arrow corresponds to a recording time of 30 min and the vertical arrow to a current of 10 µA·cm^{-2}. Intestinal current measurements (ICM) were performed per protocol [22] by the addition of amiloride (1), indomethacin (2), carbachol (3), cAMP/forskolin (4), 4,4'-diisothiocyanostilbene-2,2'-disulfonic acid (DIDS) (5) and histamine (6). Carbachol, cAMP/forskolin and histamine induce chloride secretory responses seen in the ICM by signals in the upward direction. The tracings show either no (a, c) or intermediate (b, d) CFTR activity, indicated by the cumulative chloride currents (0 or 24 µA·cm^{-2}) which implies that no functional CFTR is operating in the PI CF patient whereas substantial residual CFTR activity is present in the PS CF patient.
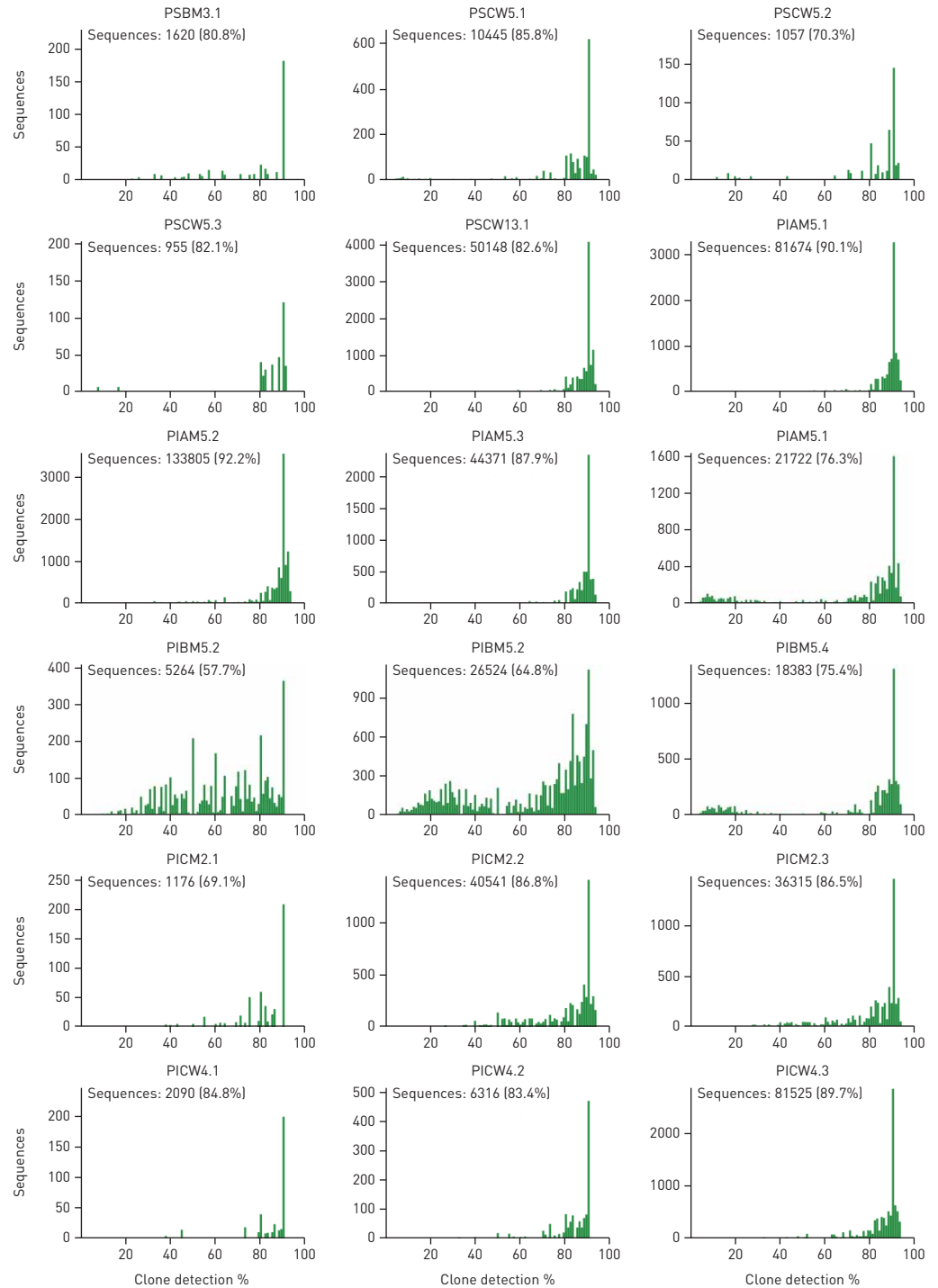
FIGURE 7 Clonal diversity of the *Staphylococcus aureus* and *Pseudomonas aeruginosa* communities in respiratory secretions. Metagenome samples with >40 000 species-specific reads were selected for analysis of single-nucleotide polymorphism (SNP) diversity in *S. aureus* (blue) and *P. aeruginosa* (green) sequences at reference genome positions covered by >10 reads. Data in the corner of each panel indicates the number of "mismatch" sequences, *i.e.* number of reads with at least one nucleotide substitution compared to the *P. aeruginosa* PAO1 and *S. aureus* Newman reference genomes. Reads sharing 100% sequence identity with the reference were designated as "matches". The individual displays show the pooled number of "matches" (*y* axis) plotted in 1% intervals of the truncated ratio of pooled matches to the total number of reads in that interval (*x* axis). Clones are highlighted by a comparatively high number of sequences in neighbouring 1% intervals supported by many SNPs. Conversely, clonal variants are supported by only a few SNPs, thus the number of sequences in neighbouring intervals is comparatively low.
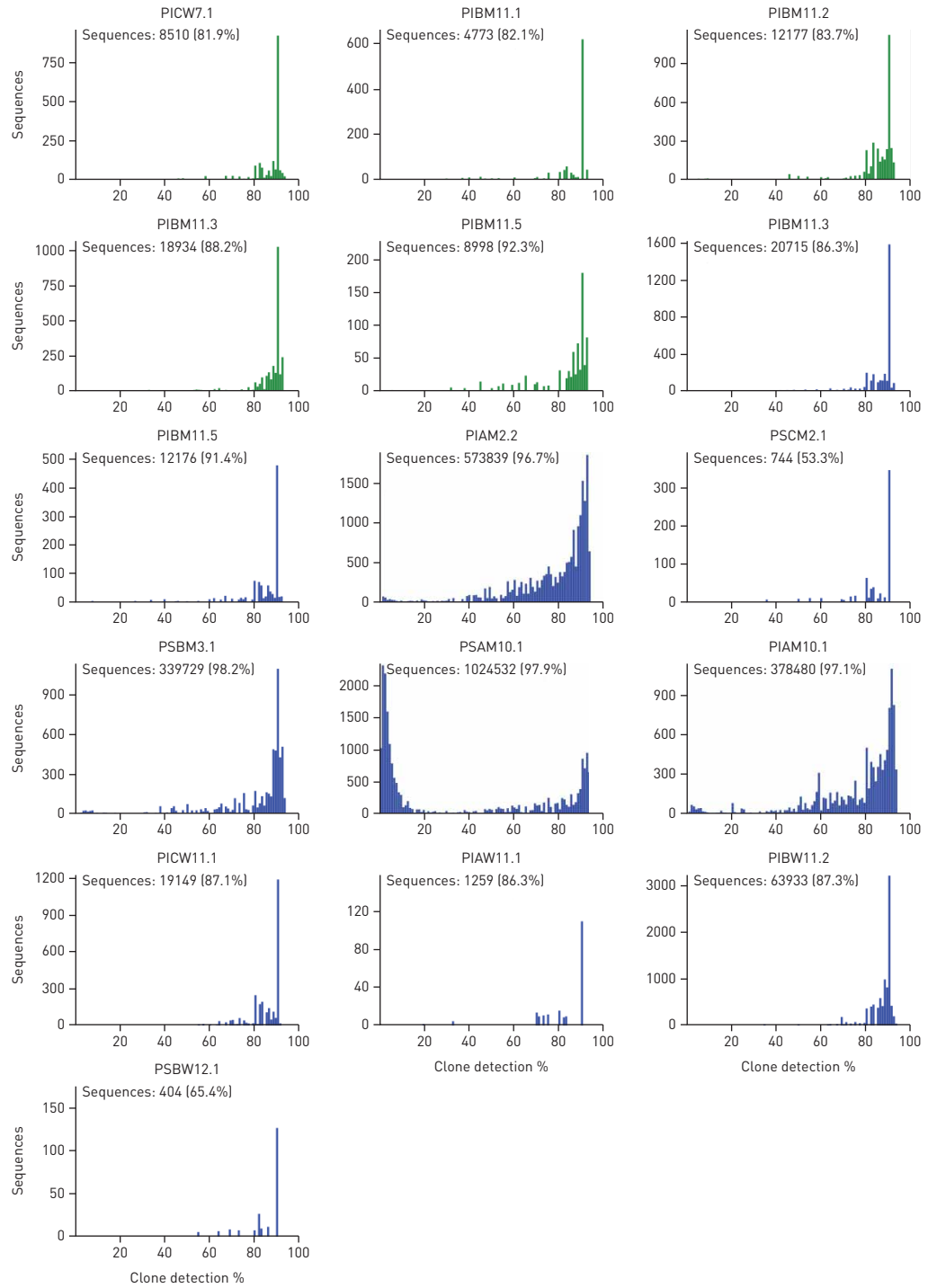
FIGURE 7 Continued.

TABLE 3 Non-synonymous sequence variants of antibiotic resistance genes in *Staphylococcus aureus* and *Pseudomonas aeruginosa* populations of cystic fibrosis sputum metagenomes[#]

Organism	Gene	Amino acid substitution	Patients n (samples n)
S. aureus	*gyrA*	p.Ser90Leu	2 (4 and 2 from each patient)
		p.Leu576Phe	1 (1)
		p.Arg843His	1 (1)
		p.Thr833fs	1 (1)
		p.Asp891_Glu892dup	1 (2)
	rpoB	p.Asp320Asn	1 (1)
	parE	p.Asn139Ser	2 (2 and 1 from each patient)
		p.Gly530Asp	1 (1)
P. aeruginosa	*mexS*	p.Ala235Thr	1 (3)
		p.Arg48His	1 (1)
	mexI	p.Ala782Glu	2 (1 from each patient)

[#]: 353 single-nucleotide polymorphisms were identified in total. The table displays a selection of rare non-conservative coding variants in targets of antimicrobial chemotherapy and antibiotic resistance determinants.

format of microbiomes (figure 5), which normalises all patients' samples to 100% irrespective of the absolute contents of microbes in the respiratory secretions [3–10]. In other words, the outcome of culture-dependent and culture-independent analysis of CF respiratory specimens is more similar than discordant modes of presentation may suggest.

The paediatric CF microbiome has been shown to be more diverse than that of CF adults indicating that there may be a time window for therapeutic intervention, which maintains diversity while reducing total bacterial load [7, 38]. Our study now shows cases of young CF adults who still have a healthy microbial metagenome (samples from subjects PSBW1 and PIBM12 (table S2) compared with those from healthy controls (table S5)). All study participants have been regularly seen since the age of diagnosis by a dedicated team of CF caregivers suggesting that continuous surveillance and intervention, if indicated, could prevent or decelerate progressive CF lung disease in some, but not all subjects.

Consistent with literature reports [39, 40], the lead CF pathogen *P. aeruginosa* became a major member of the microbial community in subjects with compromised lung function. Mucus plugging, airway remodelling, micro-colony and biofilm formation will then drive the regional isolation of the microbial metagenome [28, 41]. Considering this spatial heterogeneity, all sputa were collected by autogenic drainage in order to retrieve metagenomes that are representative for the whole lung.

Within-clone evolution of major clones is thought to trigger the adaptation of *S. aureus* and *P. aeruginosa* to the environment of the CF lungs [17, 28, 30, 31]. Our metagenome study now demonstrates that this concept does not cover the whole scenario. The *S. aureus* and *P. aeruginosa* populations do not only consist of one to three major clones, but also of rare clones. These infrequent clones constitute a low copy genetic resource which could rapidly expand as a response to habitat alterations, such as antimicrobial chemotherapy or invasion of novel microbes. Thanks to the high accuracy of sequencing-by-ligation in the colour space of 99.943% (see Methods section) we could resolve the clonal diversity of *S. aureus* and *P. aeruginosa* in CF airways. The error rates of the more often used sequencing-by-synthesis or single molecule real-time sequencing technologies are too high to reliably detect infrequent sequence variants, which may explain why minor constituents of polymicrobial communities at the rank of strains and clone types have not yet been reported in the literature.

Metagenome sequencing generated quantitative and unbiased data about microbial diversity in CF lungs. Extensive culture-enriched profiling of the CF airway microbiome identified families of bacteria in CF sputa that were not detected by parallel 16S rDNA sequencing [42]. These biases of 16S amplicon sequencing do not apply to metagenome sequencing. Knowing that metagenome sequencing discerned, on average, one order of magnitude more organisms at the species level than 16S rDNA analyses [42], we consider any culture-enriched molecular analysis of CF sputum microbes to be dispensable if a metagenome approach is pursued. However, one should bear in mind that the sensitivity of the detection of rare members depends critically on the total number of microbial read sequences (figure S3). Unless one is interested in specific features, such as the spectrum of sequence variants in loci of interest, about half a million microbial reads are sufficient to provide a comprehensive metagenome analysis of taxa in CF airways.

References

1. Ratjen F, Bell SC, Rowe SM, *et al.* Cystic fibrosis. *Nature Rev Dis Primers* 2015; 1: 15010.
2. Waters V, Smyth A. Cystic fibrosis microbiology: advances in antimicrobial therapy. *J Cyst Fibros* 2015; 14: 551–560.
3. Surette MG. The cystic fibrosis lung microbiome. *Ann Am Thorac Soc* 2014; 11: Suppl. 1, S61–S65.
4. Sibley CD, Parkins MD, Rabin HR, *et al.* A polymicrobial perspective of pulmonary infections exposes an enigmatic pathogen in cystic fibrosis patients. *Proc Natl Acad Sci USA* 2008; 105: 15070–15075.
5. van der Gast CJ, Walker AW, Stressmann FA, *et al.* Partitioning core and satellite taxa from within cystic fibrosis lung bacterial communities. *ISME J* 2011; 5: 780–791.
6. Tunney MM, Klem ER, Fodor AA, *et al.* Use of culture and molecular analysis to determine the effect of antibiotic treatment on microbial community diversity and abundance during exacerbation in patients with cystic fibrosis. *Thorax* 2011; 66: 579–584.
7. Madan JC, Koestler DC, Stanton BA, *et al.* Serial analysis of the gut and respiratory microbiome in cystic fibrosis in infancy: interaction between intestinal and respiratory tracts and impact of nutritional exposures. *MBio* 2012; 3: e00251–12.
8. Zhao J, Schloss PD, Kalikin LM, *et al.* Decade-long bacterial community dynamics in cystic fibrosis airways. *Proc Natl Acad Sci USA* 2012; 109: 5809–5814.
9. Price KE, Hampton TH, Gifford AH, *et al.* Unique microbial communities persist in individual cystic fibrosis patients throughout a clinical exacerbation. *Microbiome* 2013; 1: 27.
10. Rogers GB, Shaw D, Marsh RL, *et al.* Respiratory microbiota: addressing clinical questions, informing clinical practice. *Thorax* 2015; 70: 74–81.
11. Yarza P, Yilmaz P, Pruesse E, *et al.* Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nat Rev Microbiol* 2014; 12: 635–645.
12. Cai L, Ye L, Tong AH, *et al.* Biased diversity metrics revealed by bacterial 16S pyrotags derived from different primer sets. *PLoS One* 2013; 8: e53649.
13. Angly FE, Dennis PG, Skarshewski A, *et al.* CopyRighter: a rapid tool for improving the accuracy of microbial community profiles through lineage-specific gene copy number correction. *Microbiome* 2014; 2: 11.
14. Davenport CF, Tümmler B. Advances in computational analysis of metagenome sequences. *Environ Microbiol* 2013; 15: 1–5.
15. Lim YW, Schmieder R, Haynes M, *et al.* Metagenomics and metatranscriptomics: windows on CF-associated viral and microbial communities. *J Cyst Fibros* 2013; 12: 154–164.
16. Lim YW, Evangelista JSIII, Schmieder R, *et al.* Clinical insights from metagenomics analysis of sputum samples from patients with cystic fibrosis. *J Clin Microbiol* 2014; 52: 425–437.
17. Quinn RA, Lim YW, Maughan H, *et al.* Biogeochemical forces shape the composition and physiology of polymicrobial communities in the cystic fibrosis lung. *MBio* 2014; 5: e00956–13.
18. Sosnay PR, Siklosi KR, Van Goor F, *et al.* Defining the disease liability of variants in the cystic fibrosis transmembrane conductance regulator gene. *Nat Genet* 2013; 45: 1160–1167.
19. Gibson LE, Cooke RE. A test for concentration of electrolytes in sweat in cystic fibrosis of the pancreas utilizing pilocarpine by iontophoresis. *Pediatrics* 1959; 23: 545–549.
20. Rowe SM, Clancy JP, Wilschanski M. Nasal potential difference measurements to assess CFTR ion channel activity. *Methods Mol Biol* 2011; 741: 69–86.
21. Sermet-Gaudelus I, Girodon E, Sands D, *et al.* Clinical phenotype and genotype of children with borderline sweat test and abnormal nasal epithelial chloride transport. *Am J Respir Crit Care Med* 2010; 182: 929–936.
22. Derichs N, Sanz J, Von Kanel T, *et al.* Intestinal current measurement for diagnostic classification of patients with questionable cystic fibrosis: validation and reference data. *Thorax* 2010; 65: 594–599.
23. Leeds JS, Oppong K, Sanders DS. The role of fecal elastase-1 in detecting exocrine pancreatic disease. *Nat Rev Gastroenterol Hepatol* 2011; 8: 405–415.
24. Jensen R, Stanojevic S, Gibney K, *et al.* Multiple breath nitrogen washout: a feasible alternative to mass spectrometry. *PLoS One* 2013; 8: e56868.
25. Jensen R. Standard operating procedure: multiple breath nitrogen washout. Version 1.0. www.mbwtraining.com/ ECFS_MBW_SOP.pdf Date last updated: November 11, 2013. Date last accessed: April 27, 2016.
26. Develioglu ON, Ipek HD, Bahar H, *et al.* Bacteriological evaluation of tonsillar microbial flora according to age and tonsillar size in recurrent tonsillitis. *Eur Arch Otorhinolaryngol* 2014; 271: 1661–1665.
27. Chotirmall SH, McElvaney NG. Fungi in the cystic fibrosis lung: bystanders or pathogens? *Int J Biochem Cell Biol* 2014; 52: 161–173.
28. Jorth P, Staudinger BJ, Wu X, *et al.* Regional isolation drives bacterial diversification within cystic fibrosis lungs. *Cell Host Microbe* 2015; 18: 307–319.
29. Cramer N, Wiehlmann L, Ciofu O, *et al.* Molecular epidemiology of chronic *Pseudomonas aeruginosa* airway infections in cystic fibrosis. *PLoS One* 2012; 7: e50731.
30. Cramer N, Klockgether J, Wrasman K, *et al.* Microevolution of the major common *Pseudomonas aeruginosa* clones C and PA14 in cystic fibrosis lungs. *Environ Microbiol* 2011; 13: 1690–1704.
31. Marvig RL, Johansen HK, Molin S, *et al.* Genome analysis of a transmissible lineage of *Pseudomonas aeruginosa* reveals pathoadaptive mutations and distinct evolutionary paths of hypermutators. *PLoS Genet* 2013; 9: e1003741.
32. Williams D, Evans B, Haldenby S, *et al.* Divergent, coexisting *Pseudomonas aeruginosa* lineages in chronic cystic fibrosis lung infections. *Am J Respir Crit Care Med* 2015; 191: 775–785.
33. Darch SE, McNally A, Harrison F, *et al.* Recombination is a key driver of genomic and phenotypic diversity in a *Pseudomonas aeruginosa* population during cystic fibrosis infection. *Sci Rep* 2015; 5: 7649.

34   Wiehlmann L, Wagner G, Cramer N, *et al.* Population structure of *Pseudomonas aeruginosa*. *Proc Natl Acad Sci USA* 2007; 104: 8101–8106.

35   Enright MC, Day NP, Davies CE, *et al.* Multilocus sequence typing for characterization of methicillin-resistant and methicillin-susceptible clones of *Staphylococcus aureus*. *J Clin Microbiol* 2000; 38: 1008–1015.

36   Poole K. Efflux-mediated multiresistance in Gram-negative bacteria. *Clin Microbiol Infect* 2004; 10: 12–26.

37   Coskun-Ari FF, Bosgelmez-Tinaz G. *grlA* and *gyrA* mutations and antimicrobial susceptibility in clinical isolates of ciprofloxacin- methicillin-resistant *Staphylococcus aureus*. *Eur J Med Res* 2008; 13: 366–370.

38   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.

39   Blainey PC, Milla CE, Cornfield DN, *et al.* Quantitative analysis of the human airway microbial ecology reveals a pervasive signature for cystic fibrosis. *Sci Transl Med* 2012; 4: 153ra130.

40   Stressmann FA, Rogers GB, van der Gast CJ, *et al.* Long-term cultivation-independent microbial diversity analysis demonstrates that bacterial communities infecting the adult cystic fibrosis lung show stability and resilience. *Thorax* 2012; 67: 867–873.

41   Boon M, Verleden SE, Bosch B, *et al.* Morphometric analysis of explant lungs in cystic fibrosis. *Am J Respir Crit Care Med* 2016; 193: 516–526.

42   Sibley CD, Grinwis ME, Field TR, *et al.* Culture enriched molecular profiling of the cystic fibrosis airway microbiome. *PLoS One* 2011; 6: e22702.