Implementation of microbiota analysis in clinical trials for cystic fibrosis lung infection: experience from the OligoG Phase 2b clinical trials

Authors: Rebecca Weiser[1*], Philip D. Rye[2], and Eshwar Mahenthiralingam[1*]

Affiliations:

[1] Microbiomes, Microbes and Informatics Group, Organisms and Environment Division, School of Biosciences, Cardiff University, The Sir Martin Evans Building, Museum Avenue, Cardiff, Wales, CF10 3AX, UK.

[2]AlgiPharma AS, Industriveien 33, N-1337 Sandvika, Norway

*** Corresponding authors:**

Rebecca Weiser, School of Biosciences, Cardiff University, The Sir Martin Evans Building, Museum Avenue, Cardiff, Wales, CF10 3AX, UK.

Co-correspondence: Eshwar Mahenthiralingam, School of Biosciences, Cardiff University, The Sir Martin Evans Building, Museum Avenue, Cardiff, Wales, CF10 3AX, UK.

Tel: +44 (0)29 2087 5875, Fax: +44 (0)29 2087 4305

E-mail addresses: WeiserR@cardiff.ac.uk (R. Weiser), phil.rye@algipharma.com (P.D. Rye) MahenthiralingamE@cardiff.ac.uk (E. Mahenthiralingam)

ORCID ID:

0000-0003-3983-3272 (R. Weiser)

0000-0001-9014-3790 (E. Mahenthiralingam).

0000-0001-7762-3300 (P.D. Rye)

Highlights

- Microbiota analysis was applied in clinical trials of a novel CF therapeutic
- Paired sputum samples (<2 hours apart) had highly concordant microbiota profiles
- Patients had heterogeneous lung infection communities at recruitment
- Microbiota profiles were patient-specific and stable over time
- The infecting *Burkholderia* species influenced the sputum bacterial diversity

Abstract

Culture-independent microbiota analysis is widely used in research and being increasingly used in translational studies. However, methods for standardisation and application of these analyses in clinical trials are limited. Here we report the microbiota analysis that accompanied two Phase 2b clinical trials of the novel, non-antibiotic therapy OligoG CF-5/20 for cystic fibrosis (CF) lung infection. Standardised protocols (DNA extraction, PCR, qPCR and 16S rRNA gene sequencing analysis) were developed for application to the *Pseudomonas aeruginosa* (NCT02157922) and *Burkholderia cepacia* complex (NCT02453789) clinical trials involving 45 and 13 adult trial participants, respectively. Microbiota analysis identified that paired sputum samples from an individual participant, taken within 2 hours of each other, had reproducible bacterial diversity profiles. Although culture microbiology had identified patients as either colonised by *P. aeruginosa* or *B. cepacia* complex species at recruitment, microbiota analysis revealed patient lung infection communities were not always dominated by these key CF pathogens. Microbiota profiles were patient-specific and remained stable over the course of both clinical trials (6 sampling points over the course of 140 days). Within the *Burkholderia* trial, participants were infected with *B. cenocepacia* (n=4), *B. multivoran*s (n=6), or an undetermined species (n=3). Colonisation with either *B. cenocepacia* or *B. multivorans* influenced the overall bacterial community structure in sputum. Overall, we have shown that sputum microbiota in adults with CF is stable over a 2-hour time-frame, suggesting collection

of a single sample on a collection day is sufficient to capture the microbiota diversity. Despite the uniform pathogen culture-positivity status at recruitment, trial participants were highly heterogeneous in their lung microbiota. Understanding the microbiota profiles of individuals with CF ahead of future clinical trials would be beneficial in the context of patient stratification and trial design.

Keywords: microbiota analysis, cystic fibrosis microbiology, clinical trials, PCR, qPCR and 16S rRNA gene sequencing

Introduction

Cystic fibrosis (CF) is the most common genetically inherited disease in Caucasian populations, affecting approximately 1 in 2500 new-borns (Davies et al., 2007). CF is caused by mutations in the cystic fibrosis transmembrane conductance regulator (CFTR) protein which impair the normal exchange of ions and fluid across epithelial surfaces. In the respiratory tract, this results in a reduced volume of airway surface liquid, inadequate mucociliary clearance and increased susceptibility to infection (Davies et al., 2007). Chronic respiratory infections and accompanying progressive lung damage is the primary cause of morbidity and mortality in individuals with CF (Surette, 2014).

Traditional culture-based techniques have identified a number of typical pathogens associated with the CF lower airway including *Pseudomonas aeruginosa*, *Staphylococcus aureus*, *Haemophilus influenzae* and *Burkholderia cepacia* complex (Bcc) species (Lipuma, 2010), in addition to emergent pathogens such as *Stenotrophomonas maltophilia*, *Mycobacterium abscessus*, *Achromobacter* species, *Streptococcus milleri/anginosus* group and *Aspergillus fumigatus* (Surette, 2014). Culture-independent studies have revealed that lung infections are polymicrobial and can comprise ‘non-typical’ genera such as anaerobes more commonly associated with the upper airways, although their role in disease is still unclear (Lipuma, 2010, Jorth et al., 2019). Microbiota analysis has also shown that the decreased diversity linked with pathogen-dominated infections, correlates to a reduction in lung function and the presence of severe disease (Blainey et al., 2012). Multiple studies have corroborated that this pathogen dominated state is associated with severe lung disease in CF. Straightforward molecular methods such as Ribosomal Intergenic Spacer Analysis (RISA) PCR have proven useful for the rapid identification of pathogen-dominated microbiota and difficult to identify multi-resistant infections (Flight et al., 2015). Integrating the extra information afforded by culture-independent methods over culture-based approaches is of great value to translational studies (Shankar, 2017). There is still work to be done to address how these microbiota methods are applied clinically, and their limitations in terms of clinical decision making (Shankar, 2017).

The median life expectancy of individuals with CF has increased from less than a decade to over 50 years (MacKenzie et al., 2014). A major reason for this increase in survival is the implementation of effective treatments for respiratory infections such as inhaled antibiotics (Fajac and De Boeck, 2017). However, the extensive and long-term use of antibiotics eventually leads to the development of resistant infections. *P. aeruginosa* remains the most prevalent pathogen in CF respiratory disease and once a chronic infection is established it is almost impossible to eradicate (Döring et al., 2012). Infections with bacteria from the Bcc are also highly problematic because they represent a difficult to identify and treat multi-species group (Lipuma, 2010), may spread from one CF individual to another and are associated with poor clinical outcome (Zlosnik et al., 2015). The development of new therapies is therefore of great interest, particularly those that can disrupt the biofilm mode of microbial growth in the CF lung (Bjarnsholt et al., 2013).

The low molecular weight alginate oligosaccharide OligoG CF-5/20 is a novel, non-antibiotic therapy that has antibiotic potentiation (Khan et al., 2012, Pritchard et al., 2017) and anti-biofilm (Pritchard et al., 2017, Powell et al., 2018) properties, and can alter the visco-elastic properties of sputum (Pritchard et al., 2016). Two phase 2b clinical trials have documented the safety and efficacy of OligoG CF-5/20 in CF patients infected with *P. aeruginosa* (clinicaltrials.gov identifier NCT02157922)(van Koningsbruggen-Rietschel et al., 2020) or Bcc species (clinicaltrials.gov identifier NCT02453789) as the primary CF pathogen in the lung. Lung function, specifically the Forced Expiratory Volume in 1 second (FEV1), is widely used as a pivotal outcome measure in the development of drugs to treat CF (Stanojevic and Ratjen, 2016). However, for novel anti-infective agents such as OligoG CF-5/20 which are multimodal in their activity (Pritchard et al., 2017, Powell et al., 2018), understanding which outcome measures are relevant in terms of their efficacy for the treatment of lung infections is challenging. The safety and potential for clinical efficacy of OligoG CF-5/20 within the *P. aeruginosa* trial has been recently described (van Koningsbruggen-Rietschel et al., 2020). Here we report the microbiota analysis accompanying both the *P. aeruginosa* and Bcc OligoG

CF-5/20 clinical trials, and how this was developed and implemented to understand the value of culture-independent microbiology methods in clinical trials for new CF lung infection therapeutics.

Methods

Study design and Patient cohort

The efficacy and safety of inhaled alginate oligosaccharide (OligoG CF-5/20; referred to as OligoG) was evaluated in adults with CF using two randomised, double-blind, placebo-controlled cross-over, multicentre Phase 2b studies. Additional details for the *P. aeruginosa* clinical trial have been described elsewhere (van Koningsbruggen-Rietschel et al., 2020). Recruitment was on the basis of culture positivity for the pathogens of interest, either *P. aeruginosa* or Bcc species. For the *P. aeruginosa* trial (NCT02157922), a culture positive finding of the pathogen was required from expectorated sputum or cough swab within 12 months prior to screening, along with a culture negative result for Bcc bacteria 12 months prior to screening(van Koningsbruggen-Rietschel et al., 2020). For the Bcc trial (NCT02453789), chronic colonisation with Bcc species. was determined by at least two positive microbiological cultures in expectorated sputum 12 months prior to screening. Participant demographics and baseline clinical characteristics for each clinical trial are shown in Table 1. Diagnostic bacteriology for *P. aeruginosa* and Bcc bacteria was performed by each clinical trial centre in the context of clinical therapy, and a diagnostic microbiology service during the trial (Synlab AG, Munich, Germany).

148

149 **Table 1. Patient demographics**

Characteristic	Bcc trial	*P. aeruginosa* trial
No of patients	13	45
Age at start of trial*	33 (23-47)	35 (19-57)
Sex (% male)	31	51
Diabetes (% with diabetes)	54	29
Baseline FEV1% predicted*	49 (25-77)	61 (42-92)

150 Footnotes: *data are given as mean and range; FEV1, forced expiratory volume in 1 second

151

152

153 Expectorated sputum samples were collected at six time points across the trials for culture-
154 independent microbiological investigations: V1 (Screening, day -28 to -7), V2 (Treatment 1
155 start, day 0), V4 (Treatment 1 end, day 28), V5 (Treatment 2 start, day 56), V7 (Treatment 2
156 end, day 84) and V8 (End, day 112). Patients were randomly allocated Treatment 1 as OligoG,
157 followed by Treatment 2 as Placebo, or Treatment 1 as Placebo, followed by Treatment 2 as
158 OligoG. At each time point two sputum samples were taken within 2 hours from each patient
159 where possible, and are referred to as paired samples throughout the study.

160

161 **DNA extraction from CF sputum**

162 Sputum samples were stored frozen at -80°C and processed for DNA extraction within 4
163 weeks of provision. After thawing, samples were weighed and diluted with 4M UltraPure™
164 Guanidine Isothiocyanate Solution (ThermoFisher Scientific) in a 1:1 weight to volume ratio.
165 Samples were centrifuged (1409 *g* for 2 mins) and vortex mixed (1 min), before 1 ml was
166 removed and added to 2 ml tubes with caps and seals (Benchmark Scientific) containing 1 g
167 of 100 µm triple-pure high impact zirconium beads (Benchmark Scientific). Bead-beating was
168 performed using the Beadbug instrument (Benchmark Scientific) for 2 minutes at 2800 rpm.
169 After pulse centrifugation to settle the beads, 400 µl of the mix was added to the Maxwell 16®

tissue kit cartridge (Promega) and DNA extraction achieved using the Maxwell 16® instrument (Promega) according to the manufacturer's instructions. Approximately 200 µl DNA was obtained per sample which was stored frozen at -20°C. Full details are given in Supplementary method S1. Extraction blank controls were also run for Maxwell 16® tissue kit cartridges and evaluated for bacterial DNA contamination by PCR amplification of the 16S rRNA gene (27F and 1492R primers)(Lane, 1991); no background amplification of DNA was observed in these kit blanks.

Identification of *Burkholderia* species

To determine the identity of the Bcc species in V1 samples from patients in the Bcc trial, *recA* and *gyrB* gene sequences were PCR amplified from sputum DNA extracts using previously described primers (Spilker et al., 2009)(Table 2). PCR products were purified and sent to Eurofins Genomics for Sanger sequencing of forward and reverse DNA strands. BioEdit (Hall, 1999) was used to create consensus sequences for *recA* and *gyrB* genes and the *Burkholderia* species identity determined using the BLASTN tool of the *Burkholderia* Genome Database (Winsor et al., 2008). Full details are given in Supplementary method S2.

Quantitative PCR (qPCR)

Quantification of *P. aeruginosa* was performed by targeting the *gyrB* gene using a previously designed TaqMan assay (Anuj et al., 2009). Quantification of *Burkholderia* was achieved using primers targeting the *rpoD* gene (Sass et al., 2013) and a TaqMan probe designed in this study (Table 2). Reaction volumes were 10 µl and comprised 1X Platinum qPCR Supermix-UDG with ROX (Life Technologies), 1.8 µM forward and reverse primers supplied by Eurofins Genomics and 225 nM TaqMan probe (Thermofisher Scientific), 1 µl template DNA and nuclease free water (Severn Biotech Ltd). Reactions were performed in triplicate with negative controls, alongside a standard dilution series of 10^2-10^8 *gyrB (*amplified from *P. aeruginosa* PAO1) or *rpoD* (amplified from *B. cenocepacia* J2315) gene copies per µl (Table 2). The Bio-Rad Chromo4 system was used with the following qPCR cycling conditions: UNG treatment

at 50°C for 3 minutes, Taq activation at 95°C for 10 minutes, followed by 40 cycles of denaturation at 95°C for 30 seconds, annealing and extension at 67°C (*rpoD*) or 60°C (*gyrB*) for 30 seconds and a plate read. Quality control of qPCR results was performed as described previously (Zemanick et al., 2010). Three biological replicates were obtained per sample. Gene copy number was taken as equivalent to the number of cells as *rpoD* and *gyrB* are single copy within the genome. Full details are given in Supplementary method S3.

Table 2. PCR and qPCR primers

Gene	Primers/Probe	Sequence 5'>3'	Annealing temp (°C)	Product size (bp)	Reference
PCR primers for amplification of gene standards					
rpoD	F	GATCTTGCACATCGTCGTC	59	1011	This study
	R	GTTCGTAACGGAGACGCTG			
gyrB	F	GAGTCGATCACTGTCCGC	58	1186	This study
	R	GCATCTTGTCGAAGCGCG			
qPCR primers and TaqMan probes					
rpoD	F	GAGATGAGCACCGATCACAC	67	143	(Sass *et al.* 2013)
	R	CCTTCGAGGAACGACTTCAG			
	PROBE	5'FAM-CTGCGCAAGCTGCGTCACC-3'MGBNFQ			This study
gyrB	F	CCTGACCATCCGTCGCCACAAC	60	220	(Anuj *et al.* 2009)
	R	CGCAGCAGGATGCCGACGCC			
	PROBE	5'FAM-GGTCTGGGAACAGGTCTACCACCACGG-3'MGBNFQ			

16S rRNA gene sequencing and bacterial diversity analysis

The suitability of the DNA extraction protocol for bacterial diversity analysis was initially evaluated using RISA as previously described (Flight et al., 2015) (Supplementary method S4) and applied to the enrolment and a limited number of paired samples. Subsequently, amplification and sequencing of the 16S rRNA gene V1-V2 region was performed by RTL genomics (Lubbock, Texas) using the Illumina MiSeq platform. The 16S rRNA sequencing reads were analysed using Mothur version 1.33 (Schloss et al., 2009) following the Illumina MiSeq SOP pipeline. Each data set was subsampled in Mothur to the lowest number of sequence reads (*P. aeruginosa* trial = 1018, Bcc trial = 5577); samples falling below the minimum threshold of 1000 reads were excluded. Sequences were grouped into operational taxonomic units (OTUs) based on a 0.03 distance limit, which equated to 97% sequence similarity. The RDP MultiClassifier script was used to determine OTU taxonomies to the genus level, with manual searching of the RDP-II sequence database (Wang et al., 2007) (http://rdp.cme.msu.edu/) to corroborate assignments. Following taxonomic assignment, an OTU minimum abundance threshold of 10 was applied which removed OTUs represented fewer than 10 times across the dataset. OTUs were consolidated to the genus level for all downstream analyses. Full details are given in Supplementary method S5. Raw sequence data have been submitted the European Nucleotide Archive under the study number PRJEB38277.

Statistical analysis

Data handling and statistical analysis were carried out in Microsoft Excel and R statistical software (R-Core-Team, 2013). Full details of R scripts are given in Supplementary method S6. Concordance between the proportions of genera in paired samples was determined using the Pearson product-moment correlation coefficient (PPMCC) using the cor.test function with Pearson correlation in R, as previously used to assess bacterial community concordance between samples (Muhlebach et al., 2018). Alpha and beta diversity indices were calculated in R using the vegan package. Differences in alpha diversity (as measured by the Shannon

index), total abundance and relative abundance of the key pathogen (*P.* aeruginosa or Bcc) were determined between three start and end points during the trial: Screening (V1) and End (V8), OligoG start (V2 or V5) and OligoG end (V4 or V7), and Placebo start (V2 or V5) and Placebo end (V4 or V7). For Shannon diversity and total abundance, Wilcoxon signed-rank tests (non-parametric) were performed to determine significant differences between paired time points. Differences in relative abundance were analysed using Generalized Additive Models for Location, Scale and Shape (GAMLSS package in R) with a beta inflated family (BEINF; Zero-One inflated beta model), (mu) logit links and other default options as previously recommended for proportional microbiome data (Ho et al., 2019, Peng et al., 2016, Ho et al., 2018) where the variable of interest (relative abundance) may be zero or one (Ospina and Ferrari, 2012). In the models, the response variable was relative abundance, the fixed effect was trial stage (with start as the reference class to which end was compared) and patient ID was specified as the random effect. The estimates (regression coefficients) of the model are the difference in log odds of genus relative abundances between paired start and end groups (Ho et al., 2018).

To examine differences between bacterial community composition between sample types, non-metric multidimensional scaling (vegan package in R) and hierarchical clustering with Ward's method based on Bray-Curtis dissimilarity were used (vegan and NMF packages in R). Permutational ANOVA (PERMANOVA, adonis function in vegan package) was used to determine the significance of differences between community structures. The betadisper function (vegan package) was used to test homogeneity of group dispersions (variances) as a condition of PERMANOVA. As adonis cannot accommodate random effects, the PERMANOVA model had the following structure to best estimate patient influence on community composition: 'dissimilarity_distance_matrix ~ A + B', first the effects of A (Patient) are evaluated, then the effects of B are evaluated (Trial stage: Screening, OligoG start, OligoG end, Placebo start, Placebo end and End) after removing the effects of A. The relationship between Shannon diversity and lung function (FEV1% predicted) was investigated using linear

regression and linear mixed models with patient as the random effect (lme function in R package nlme). In the Bcc trial lung function measurements were only available for V1, V2, V4, V5 and V7 time points, whereas in the *P. aeruginosa* trial measurements for all 6 time points were recorded.

Results

Patient summary, standardised methods and application of bacterial diversity analysis

The demographics of patients recruited into each trial are given in Table 1, with a total of 666 sputum samples collected from 45 individuals (511 samples) within the *P. aeruginosa* trial, and 13 patients (155 samples) within the Bcc trial. Paired sputum samples (designated S1 and S2) were obtained for nearly all the 6 time points across both trials (Supplementary Table S1; Supplementary Table S2). Standardised protocols were developed and adhered to for multiple sample processing and analysis (Supplementary methods S1-6). Within the Bcc study, one of the Screening (V1_S1) samples was subjected to *recA* and *gyrB* gene sequencing (26 samples) to determine the identity of the infecting *B. cepacia* complex species. All 155 Bcc trial samples were submitted for 16S rRNA gene sequencing bacterial diversity analysis, and one sample (S1) from each time point (78 samples) was used for *rpoD* qPCR to determine Bcc bacterial load in sputum. Within the *P. aeruginosa* trial, 16S rRNA gene sequencing bacterial diversity analysis was performed for all 511 samples and *P. aeruginosa* load was determined by *gyrB* qPCR for one sample (S1) at each time point (270 samples). Two samples were excluded from the bacterial diversity analysis due to the number of reads falling below the threshold level of 1000 (Patient 27610-011, sample V7_S1 from the Bcc and 82601-003, sample V1_S1 from the *P. aeruginosa* trial). Accurate determination of bacterial classification to the genus level was obtained by microbiota analysis, and therefore all

subsequent results will be referred to in context of *Pseudomonas, Burkholderia* or other genera as appropriate.

Microbiota profiles in paired CF sputum samples taken within 2 hours are concordant

Culture-based bacteriology can be highly variable from CF sputum unless rigorous protocols and appropriate standards are employed (Burns and Rolain, 2014). However, the variability of sputum microbiota profiles collected over short time frames such as sampling visits for clinical trials is not known. Therefore, where possible, paired sputum samples were collected within 2 hours of each other at each time point during the trials to understand potential sampling variation. In total, 76 and 241 paired samples were collected for the Bcc and *P. aeruginosa* trials, respectively. Microbiota analysis demonstrated that the paired sputum samples collected within 2 hours of each other had concordant bacterial diversity profiles. Representative examples of microbiota profiles indicating levels of similarity between paired samples are shown in Figure 1. Pearson product-moment correlation coefficients (PPMCC) for the *P. aeruginosa* (mean = 0.92, range = 0.07-1.00) and Bcc (mean = 0.98, range = 0.68-1.00) paired samples were also high, corroborating the similarity and stability of the bacterial microbiota signature over the short time frame of 2 hours (Supplementary Figure S1).

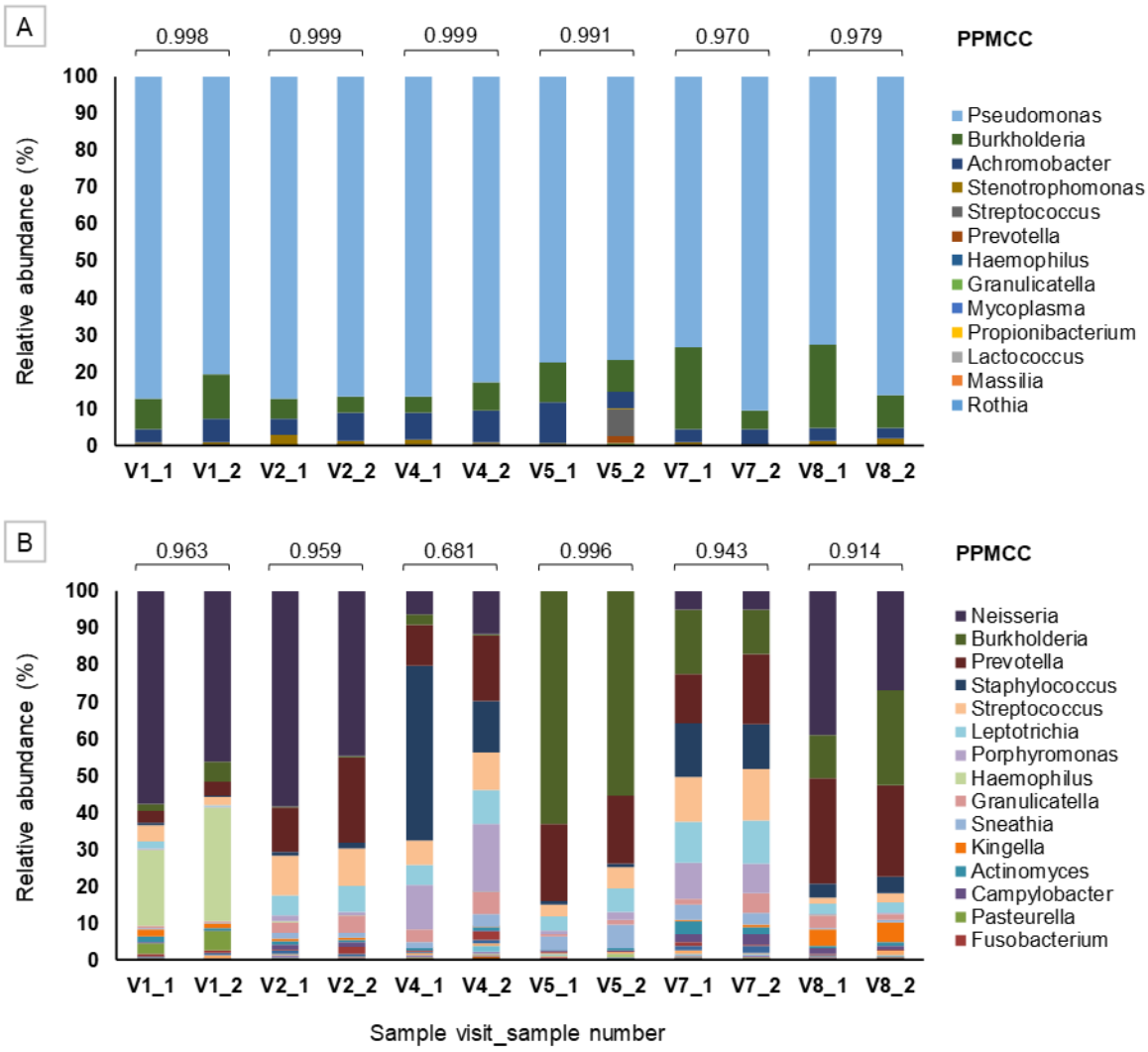


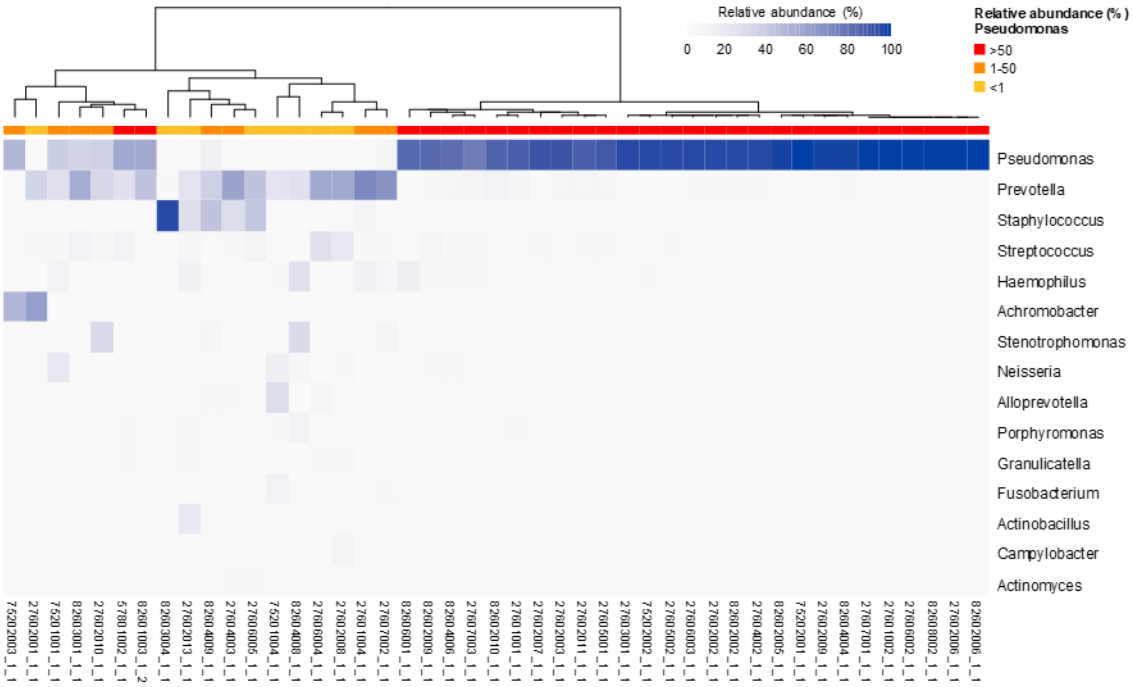
Figure 1. Microbiota stability between paired sputum samples. Relative abundances of genera in paired sputum samples (all time points) are shown for: **(A)** a pathogen dominated patient (Bcc trial patient 27610-007; all genera present shown in key); **(B)** a microbiota diverse patient (Bcc trial patient 27611-006). The top 15 genera for each individual are shown in the key and the Pearson product-moment correlation coefficient (PPMCC) values are displayed above the stacked bar charts to demonstrate the level of concordance between paired samples.

Paired samples were also highly similar in terms of the shared prevalent genera. Within the Bcc trial, 86% of paired samples shared the 5 most prevalent genera at ≥1% relative abundance, which rose to 97% of paired samples at ≥5% relative abundance. Within the *P. aeruginosa* trial, 62% of paired samples shared the same top 5 genera at ≥1% relative

abundance, increasing to 83% of paired samples at ≥5% relative abundance. The demonstration of microbiota stability between paired sputum samples led to the decision to use only 1 sample from each pair for all further analyses. Sample 1 (S1) for each individual and time point was therefore used throughout for qPCR analysis. Sample S1 was also used for all 16S rRNA gene bacterial diversity analyses, except for two samples where only S2 was available (Patient 27610-011, sample V7_S2 from the Bcc trial and 82601-003, sample V1_S2 from the *P. aeruginosa* trial).

The bacterial diversity of CF infection varied in the individuals at screening

Patients were recruited for the clinical trials on the basis that they were colonised with *P. aeruginosa* and not Bcc species for the for *P. aeruginosa* trial, or any Bcc species for the Bcc trial. However, culture-independent molecular diagnostics and bacterial diversity analysis of sputum samples at screening (sample V1) revealed that not all patient samples had substantial evidence of infection with these pathogens. The heterogeneity in bacterial diversity between patients was clearly seen by hierarchical cluster analysis of the 16S rRNA sequencing diversity data. For the *P. aeruginosa* trial, the microbiota profiles split into 2 major groups of individuals, those dominated by the target pathogen in the trial, and those with more diverse microbiota (Figure 2). The same phenomenon was observed for the Bcc trial (Figure 5A), and additional microbiota metrics demonstrating the overall variability of the V1 samples for both trials are given in Table 3.



345

346 **Figure 2. Participants recruited into the *P. aeruginosa* trial show heterogeneous baseline**
347 **microbiota profiles.** Hierarchical cluster analysis of Bray-Curtis dissimilarity values using the ward
348 algorithm for the 15 most abundant genera in 45 V1 samples (all samples were S1 except 82601-
349 003_V1 S2 where 82601-003_V1 S1 was not available) are shown. The column annotation colours
350 indicate groupings of samples with different relative abundances of *Pseudomonas;* <1%, yellow; 1-50%,
351 orange, and; >50%, red.

352

353 Of the 45 patients within the *P. aeruginosa* trial, 6 were not colonised with *P. aeruginosa* when
354 evaluated on microbiota analysis <1% relative abundance for the pathogen and a negative
355 *gyrB* qPCR result (27602-001, 27606-005, 27602-008, 27602-013, 75201-004, 82603-004).
356 In these patients the prevalent bacterial genera (>25% relative abundance) were *Prevotella*,
357 *Achromobacter*, *Staphylococcus*, and *Alloprevotella*. Within the *Burkholderia* trial, 3 out of the
358 13 patients also had very low levels of Bcc in their sputum (27610-003, 27610-011, 27611-
359 005; <1% relative abundance, negative for *recA* and *gyrB* PCR but low levels detected by
360 *rpoD* qPCR). In these Bcc patients the prevalent genera (>25% relative abundance) were
361 *Achromobacter*, *Prevotella*, *Granulicatella* and *Pseudomonas*.

Furthermore, the *P. aeruginosa* clinical trial had a subset of patients (n=10; 27601-004, 27602-010, 27604-003, 27606-004, 27607-002, 75201-001, 75202-003, 82603-001, 82604-008 and 82604-009) where *Pseudomonas* was not the dominant genus (defined as <50% relative abundance in the sample). In these non-dominant *P. aeruginosa* cases, *Prevotella*, *Staphylococcus*, *Achromobacter* and *Stenotrophomonas* were the prevalent genera (>25% relative abundance). The Bcc clinical trial also had a subset of patients where *Burkholderia* was not dominant (n=3; 27610-001, 27610-007 and 27611-006), and *Pseudomonas*, *Streptococcus* and *Neisseria* were instead prevalent. The remaining patients in both trials had either *Pseudomonas* (n=29; 64% of the trial cohort) or *Burkholderia* (n=7; 53% of the trial cohort) as the dominant genus. Overall, the microbiota analysis demonstrated that at screening, despite all participants being unified by a single bacteriology measure of culture positivity for either Bcc or *P. aeruginosa* for the different trials, lung infection in over one third of the trial participants was highly variable.

Lung microbiota was patient-specific and remained stable during the trial

To investigate overarching differences at the community level, samples from all six time-points within the trials were analysed as either patient groups or in the context of trial stage (Screening, OligoG start, OligoG end, Placebo start, Placebo end and End) groups. Using NMDS ordination based on Bray-Curtis dissimilarity distances, bacterial community composition was investigated. In the Bcc trial bacterial community composition was not found to alter due to treatment (Figure 3A; PERMANOVA, R^2=0.03, p=0.98), but was significantly different between individuals (Figure 3B; PERMANOVA, R^2=0.03, p=0.043). In the *P. aeruginosa* trial, the same outcome was observed, with bacterial diversity being driven more by each CF individual (Supplementary Figure S2B; PERMANOVA, R^2=0.01 p=0.034) than the trial treatment stages (Supplementary Figure S2A; PERMANOVA, R^2=0.006, p=0.998). Overall, consideration of multiple microbiota parameters including Shannon diversity, pathogen abundance and the most prevalent genera, were highly similar with respect to each

individual when either Screening only samples (V1), or all trial sample points (V1, V2, V4, V5, V7 and V8) were considered (Table 3).

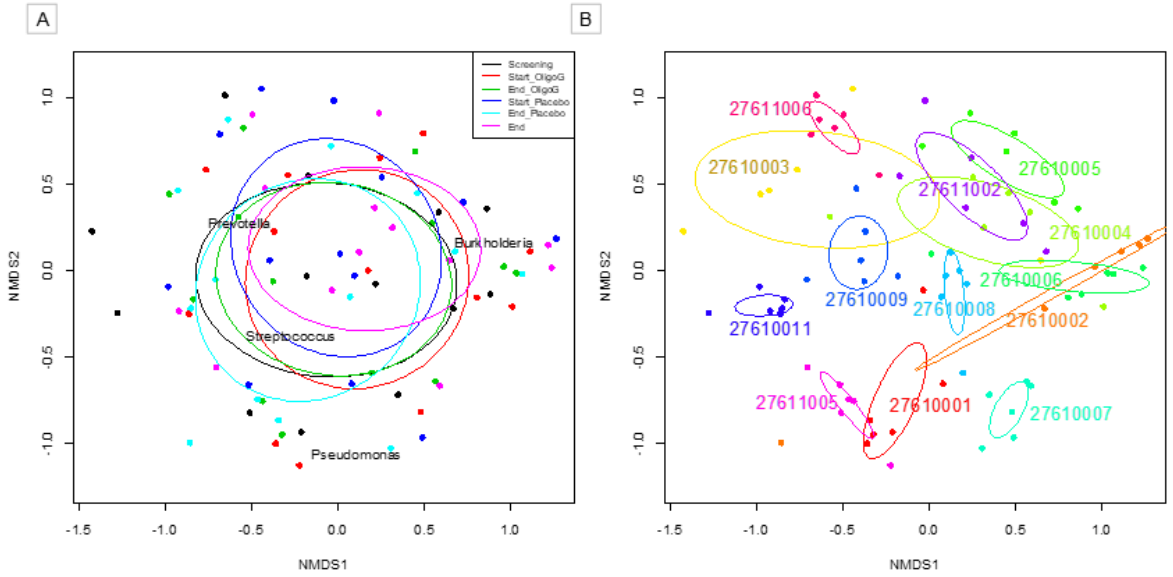


Figure 3. The lung microbiota within the *Bcc* Oligo G trial participants was linked to the individual rather than treatment. NMDS analysis of Bray-Curtis dissimilarity values for S1 samples from all 6 time points for the 13 patients in the Bcc trial (except 27610-011 that only had S2 for V7). Analyses grouped by: **(A)** treatment, and **(B)** patient are shown. Points represent individual samples, ellipses are standard deviations of points scores for each grouping. The top 7% genera based on abundance across the dataset are shown in **(A)**. A significant difference was observed between patient groups (PERMANOVA, R^2=0.03 p=0.043) but not between treatment groups (PERMANOVA, R^2=0.03, p=0.98). For both sample groups (treatment, patients) the group variances were homogeneous, satisfying the conditions of the PERMANOVA model.

Table 3. Variability of key characteristics in S1 samples at the V1 sample point and and across all sample points (V1, V2, V4, V5, V7 and V8) in the trial

Characteristic	Bcc trial	*P. aeruginosa* trial
Screening sample (V1)		

Relative abundance of key pathogen (%)*	46.70 (0-99.62)	62.19 (0.10-100)
Total abundance of key pathogen (Log_{10} cells/g sputum)*	6.54 (0.76-8.55)	6.42 (0-8.53)
Top genera in samples[a]	*Burkholderia, Pseudomonas, Streptococcus, Prevotella, Granulicatella*	*Pseudomonas, Prevotella, Streptococcus, Haemophilus, Porphyromonas*
Shannon diversity index*	0.63 (0.03-1.49)	0.71 (0-2.10)
Lung function (FEV1% predicted) *	48.88 (25-77.08)	61.00 (41.80-91.57)
All samples (V1, V2, V4, V5, V7,V8)		
Relative abundance of key pathogen (%)*	41.65 (0-100)	56.00 (0-100)
Total abundance of key pathogen (Log_{10} cells/g sputum)*	6.43 (0-8.81)	6.34 (0-8.75)
Top genera in samples[a]	*Burkholderia, Pseudomonas, Prevotella, Streptococcus, Granulicatella, Actinomcyes, Leptotrichia*	*Pseudomonas, Prevotella, Streptococcus, Haemophilus, Granulicatella*
Shannon diversity index*	0.72 (0-2.40)	0.78 (0-2.10)
Lung function (FEV1% predicted) *[b]	50.26 (22.09-77.08)	60.00 (34.75-103.61)

Footnotes: *data are given as mean and range of S1 samples (except Bcc 27610-011 V7 and Pa 82601-003 V1 where only S2 samples were available); [a]Top genera were defined as those identified in ≥50% of V1 samples by 16S rRNA gene sequencing analysis, and had a maximum relative abundance of ≥10%; [b] FEV1% predicted values were only available for V1-V7 time points for the Bcc trial

Differences in the Shannon diversity (Supplementary Table S3), total (Supplementary Table S4) and relative abundance (Supplementary Table S5) of Bcc and *P. aeruginosa* were also investigated at the three paired start-end points for the trials: Screening and End, OligoG start and OligoG end, and Placebo start and Placebo end. These paired start-end points measured potential short term (28 days; OligoG start to end or Placebo start to end) and longer term

changes (112 ± 28 days; Screening to End). In the Bcc trial, no differences were observed between any of the paired time-points with the exception of total *Burkholderia* abundance as measured by qPCR, which significantly decreased (p=0.02148) from Placebo start to Placebo end (Supplementary Figure S3). In the *P. aeruginosa* trial, the only difference was the relative abundance of *Pseudomonas* which significantly decreased (p=0.00226) from Screening to End (Figure 4). Overall, the microbiota profiles for each chronically infected individual remained stable over the trial period.

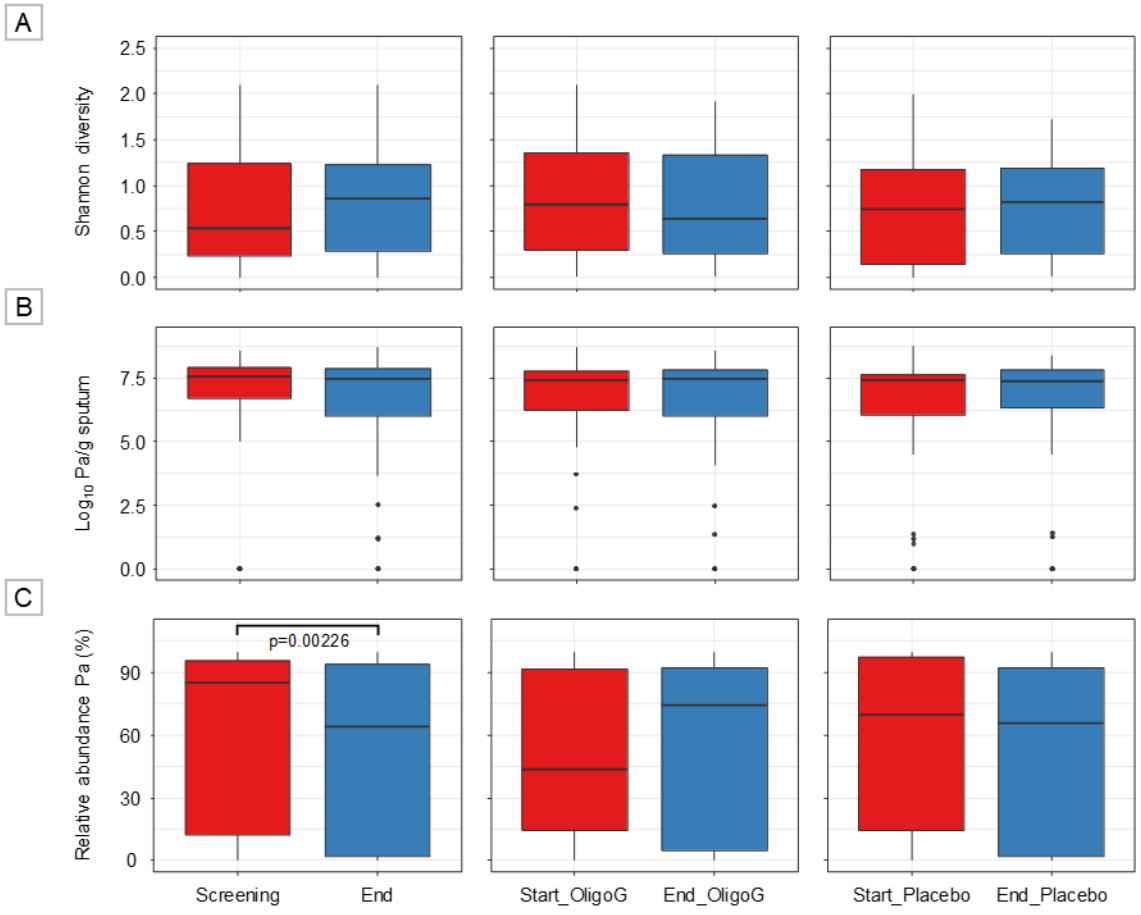


Figure 4. Analysis of microbiota present between paired start and end time-points sputum samples from the *P. aeruginosa* trial. Boxplots show the spread of data for Screening versus End samples, Start OligoG versus End OligoG samples, and Start Placebo versus End Placebo (S1 samples only, n=270; except 82601-003 that only had S2 for V1). **(A)** shows microbiota diversity measured using the Shannon index; **(B)** shows the total abundance of *P. aeruginosa* per gram of sputum measured using qPCR, and; **(C)** shows the relative abundance of *Pseudomonas* from 16S rRNA gene sequencing

analysis. For Shannon diversity and total *Pseudomonas* abundance, Wilcoxon signed-rank tests were used to assess the differences between paired time-points. Differences in relative abundance were determined using GAMLESS-BEINF models with Patient as the random effect, reporting changes in log(odds ratio) between paired time-points. Statistical significance is shown as a bracket above boxplots with the p-value under the bracket.

Microbiota profile trends in the clinical trials: signatures for each Bcc species

The identity of the infecting Bcc species was determined at Screening by *recA* and *gyrB* gene sequence analysis (Supplementary Table S6), and was found to be either *B. cenocepacia* (n=4 patients) or *B. multivorans* (n=6 patients). In three cases, no identification could be made due to negative results for the *recA* and *gyrB* PCRs, which also corresponded to low total and relative abundance levels of *Burkholderia* (Supplementary Table S6). Analysis of V1 samples showed that *B. cenocepacia* was always found as the dominant *Burkholderia* (>50% relative abundance in samples, n=4). In contrast, *B. multivorans* showed mixed infection status in relation to *Burkholderia* abundance, with dominance (>50% relative abundance, n=3) or at low relative abundance (<10%, n=3) seen equally across the 6 positive individuals (Supplementary Table S6).

These differences in relative abundance of *Burkholderia* in relation to *B. cenocepacia* or *B. multivorans* infection impacted the overall bacterial diversity (Figure 5A and 5B). *B. cenocepacia* infected trial participants had bacterial communities with lower diversity, but *B. multivorans* was present in both low and higher diversity communities (visualised by hierarchical cluster analysis; Figure 5A). Whilst sample numbers were small, NMDS ordination of Bray-Curtis dissimilarity values highlighted overlap between *B. cenocepacia* and *B. multivorans* dominated communities (Figure 5B). The higher diversity communities containing *B. multivorans* also overlapped with samples containing <0.1% *Burkholderia* relative abundance, for which no *Burkholderia* species identification was obtained (Figure 5, panel B).

Total abundance was also higher for *B. cenocepacia* than *B. multivorans*, although this was not statistically significant (Figure 5, panel C).

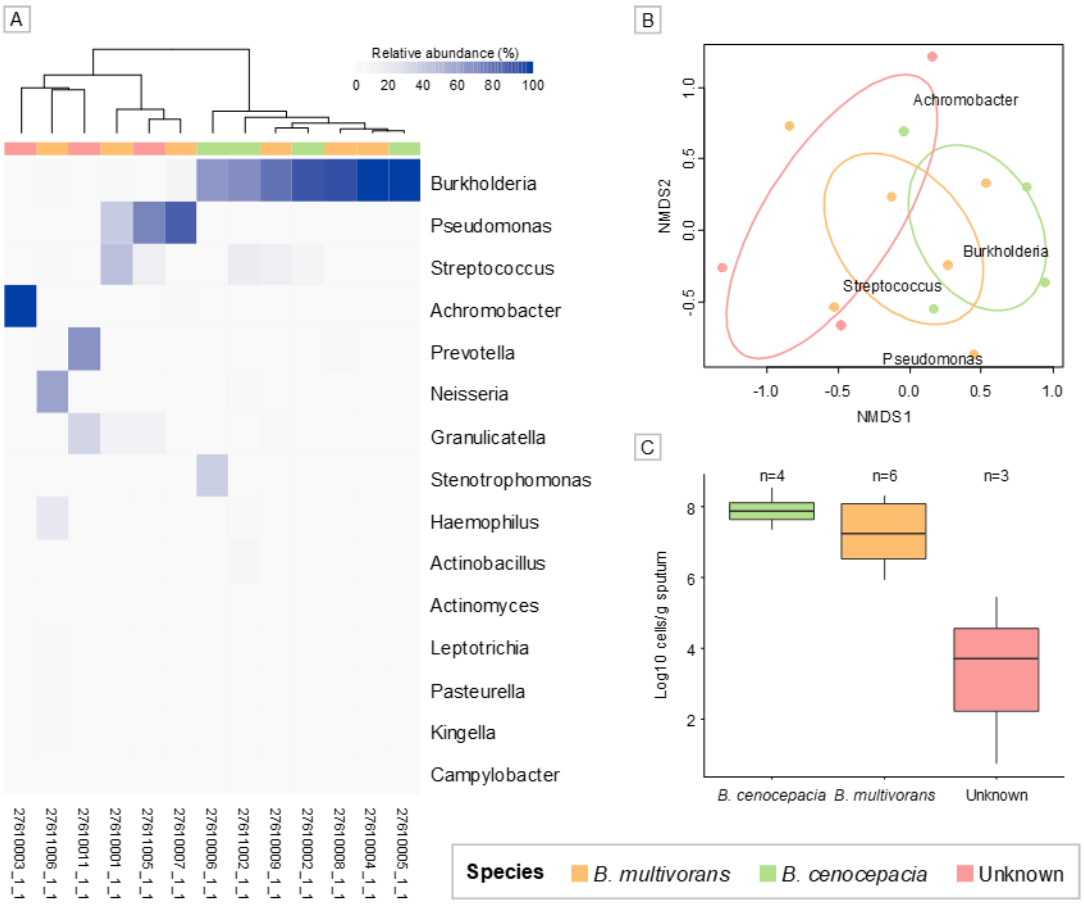


Figure 5. The infecting *Burkholderia* species drives microbiota diversity. Trends identified in microbiota screening (V1_S1) samples from 13 patients in the Bcc trial (n=13) were evaluated using multiple analyses as follows. **(A)** Shows hierarchical clustering of samples with top 15 genera. **(B)** shows an NMDS ordination plot of Bray-Curtis dissimilarity distances. Points represent individual samples, ellipses are standard deviations of points scores for each grouping. The top 10% genera based on total abundance across the dataset are shown. **(C)** shows boxplots of total Bcc abundance as determined by qPCR. Samples were grouped by Bcc species as identified by *recA* gene PCR and sequencing (*B. cenocepacia* or *B. multivorans*). For 3 samples, no Bcc species could be identified by *recA* analysis, but a value was obtained for total abundance by qPCR for Bcc. Consistent colour coding is used to identify *B. multivorans* (orange), *B. cenocepacia* (green) and unknown species (pink).

The correlation between Shannon diversity and lung function (FEV1% predicted) was also explored using all available samples from the Bcc (13 patients; V1, V2, V4, V5 and V7 samples) and *P. aeruginosa* trials (45 patients; V1, V2, V3, V4, V5, V7 and V8 samples). Whilst linear regression identified a trend for decreasing lung function with decreasing Shannon diversity (Supplementary Figure 4), once repeated measures were taken into account using linear mixed models with 'Patient' as the random effect, there were no significant correlations. The small number of patients in each trial did not give the sufficient power needed for identification of statistically significant trends. However, overall microbiota diversity loss and its linkage to reduced lung function was observed and correlated to other studies on adults with CF (Blainey et al., 2012, Flight et al., 2015) (Supplementary Figure 4).

Discussion

Culture-independent molecular diagnostic methods such as qPCR and microbiota analysis have a number of advantages over culture for CF infection microbiology (Mahboubi et al., 2016). Their application in clinical trials of new anti-infective drugs for CF lung disease has been limited. Here we describe the comprehensive protocols used to incorporate microbiota analysis into the OligoG CF-5/20 Phase 2b clinical trials for *P. aeruginosa* and Bcc. Since OligoG is not a conventional antibiotic and works in a multifactorial manner (Khan et al., 2012, Pritchard et al., 2017, Powell et al., 2018), understanding its effect on both the target pathogen and wider CF lung microbiota was important. OligoG was shown to be safe and have promising clinical efficacy in the *P. aeruginosa* clinical trial (van Koningsbruggen-Rietschel et al., 2020) and the Bcc clinical trial (unpublished data). Our microbiota analysis demonstrated that OligoG CF-5/20 did not cause major changes in the bacterial diversity within the CF lung after short term exposure (28 days). A promising overall signature from the *P. aeruginosa* trial was a significant decrease in the relative abundance of *Pseudomonas* from screening to end (Figure 4), although it was not linked to the OligoG administration periods. The lack of major changes within the CF lung microbiota for both the *P. aeruginosa* and Bcc OligoG trials

505 correlates to the lack of bacterial diversity changes seen with the potent inhaled antibiotic,
506 aztreonam (Heirali et al., 2017). Heirali *et al.* (Heirali et al., 2017) observed no significant
507 changes in microbiota diversity in relation to 28 days of aztreonam, with microbiota clustering
508 most associated with the 24 individual patients examined, as also saw with our trial
509 participants (Figure 3B; Supplementary Figure 2). The lack of radical changes in the CF
510 microbiota during the OligoG CF-5/20 trials further supports its overall safety (van
511 Koningsbruggen-Rietschel et al., 2020) and potential for optimisation as long-term therapeutic
512 for chronic infections in CF.
513
514 Application of microbiota methods to OligoG CF-5/20 trials revealed multiple advantages and
515 information useful for standardisation in the context of diagnostic microbiology for CF. For
516 example, sampling method is important to consider as it can influence which microorganisms
517 are recovered from a polymicrobial infection community (Burns and Rolain, 2014). In adult CF
518 patients, the primary microbiological specimen is spontaneously expectorated sputum. The
519 use of expectorated sputum has been criticised as being poorly representative of the lower
520 airway due to contamination with oral bacteria, and because a single sample may not capture
521 the spatial heterogeneity of microbial communities in the lung (Willner et al., 2011, de Dios
522 Caballero et al., 2017). However, sputum is an easy and non-invasive sample type, and
523 typical CF pathogens not found in the upper airway often dominate sputum samples both in
524 culture and microbiota analyses (Goddard et al., 2012, Caverly et al., 2015). We have also
525 demonstrated short term stability for sputum samples, with high similarity in microbiota profiles
526 obtained from paired samples taken within 2 hours of each other (Figure 1). No other studies
527 have looked at sputum samples taken within such as short time frame, but daily sputum
528 microbiota dynamics have been found to be both relatively stable (Goddard et al., 2012,
529 Carmody et al., 2015) and show major fluctuations (Goddard et al., 2012) in CF adults. In
530 these studies, prevalent genera were consistently identified, just in different proportions.
531 Overall, the excellent short-term stability and reproducibility of our sputum microbiota analysis
532 in the identifying key genera present was more consistent than sputum culture (Table 3). It

533 also supported that obtaining only one sputum sample at a given time point would be sufficient
534 for the design of future clinical trials.
535
536 Multiple CF patients are recruited onto clinical trials based on previous sputum culture results.
537 Whilst some studies have shown good concordance in the typical CF pathogens identified by
538 culture and culture-independent techniques (Goddard et al., 2012) others have shown that
539 results can vary by genus (Mahboubi et al., 2016). In our study, patients were selected for the
540 two separate OligoG clinical trials based on whether they were *P. aeruginosa* (but not Bcc)
541 positive or Bcc positive by culture. Good overlap between culture and microbiota identification
542 has been observed for these genera when looking just at presence or absence, but variability
543 in relative abundance has not been considered (Mahboubi et al., 2016). We identified a highly
544 heterogeneous baseline patient population where some patients carried the pathogen of
545 interest as the dominant genus (>50% relative abundance) and others had very low or no
546 carriage (Figure 3 and Figure 5A). These results pose the question, should patients be
547 stratified based on their microbiota status prior to CF clinical trials? Stratification based on
548 predominant microbiota taxa has already been investigated for non-CF bronchiectasis and
549 was more closely correlated to disease, inflammation and disease outcomes than stratification
550 based on culture (Rogers et al., 2014). Our microbiota data is supportive of such testing to
551 ensure a targeted pathogen is both dominant and present within CF patients at the start of an
552 anti-infective clinical trial.
553
554 Microbiota analysis permitted the identification of a number of trends during the OligoG Phase
555 2b clinical trials. The collection of repeated samples from each patient (6 samples over 112 ±
556 28 days) showed that the microbiota was stable over time and patient-specific. Other studies
557 have also largely shown that the adult CF microbiota is resilient to change in the form of
558 antibiotic treatment (Heirali et al., 2017). This is also true during pulmonary exacerbations,
559 where although some studies have found transient changes, the microbiota either remains
560 stable, or returns to its previous state (Bevivino et al., 2019). Therapy with the CF modulator

Ivacaftor has both had limited effects on CF lung infection composition (Bernarde et al., 2015) and been linked to striking microbiota changes (Hisert et al., 2017). While there is clearly variation in microbiota dynamics, the lack of microbiota change in the short (28 day) OligoG or Placebo treatment window during our study could perhaps have been expected. It should also be noted that although samples taken in this study were longitudinal, they were analysed as cross-sectional groups. In this context, it is possible to identify overarching microbiota changes, but variation at the individual level could be missed. Other studies have tried to overcome this issue by subgrouping patients, for example into drug responders and drug non-responders (Heirali et al., 2017), demonstrating another approach to patient stratification.

One clear microbiota pattern is the association of decreased bacterial diversity with reduced lung function (Bevivino et al., 2019), which due to the small patient group sizes in our study was observed but could not be statistically confirmed (Supplementary Figure 4). Our analysis also highlighted another new trend warranting further investigation: that the *B. cepacia* complex species present in *Burkholderia* CF lung infections may influence the overall microbiota community structure. *B. cenocepacia* and *B. multivorans* are the most prevalent *B. cepacia* complex (Bcc) species found in CF patients colonised with *Burkholderia* (Kenna et al., 2017), and both species were identified in the Bcc clinical trial (Figure 5). *B. cenocepacia* has been more closely associated with severe clinical disease and mortality than other Bcc species including *B. multivorans* (Zlosnik et al., 2015). To support this increased pathogenicity and reveal a potential mechanism behind these clinical observations, we observed that when *B. cenocepacia* was present, it was found only as the dominant pathogen by relative abundance and had high total abundance (Figure 5). In contrast, *B. multivorans* could be either the dominant pathogen, or found at lower abundance in a higher diversity community often in association with *Pseudomonas* and *Streptococcus* (Figure 5). No other studies have previously uncovered these microbiota differences in relation to Bcc infection, which could potentially explain variation in clinical outcomes (Zlosnik et al., 2015).

There are many benefits to employing culture-independent analyses for translational studies and in clinical trials. It is also important to mention, however, that microbiota analysis has limitations. Variation in methodologies such as sample collection and handling, DNA extraction, DNA sequencing and microbiota analysis can greatly influence results (Bevivino et al., 2019). The impact of different DNA extraction techniques (Terranova et al., 2018, Oriano et al., 2019) and pre-treatment of sputum to remove 'dead' DNA (Nelson et al., 2019) have recently been explored. Whilst we developed standardised protocols to ensure consistency within our study, others have used varying methods and it is difficult to standardise microbiota analysis as it is a dynamic and developing area. Already studies are starting to expand on 16S rRNA gene sequencing for bacterial diversity by using metagenomic analysis to understand community interactions and functions (Bacci et al., 2019, Bevivino et al., 2019). It is also important to consider microorganisms other than bacteria; culture-independent analysis has identified a broad range of fungi and viruses, including bacteriophages, in the CF lung (Bevivino et al., 2019). Since our study was performed, the importance of controls for DNA extraction reagents that eliminate bias in rare microbiota reads due to DNA within the "kitome" has become clear (Salter et al., 2014). Although we found our DNA extraction negative controls to be free of bacterial DNA amplification by 16S rRNA gene PCR, they were not included in our bacterial diversity sequencing analysis at the time and this should be performed for all future microbiota analysis studies.

Overall, to comprehensively evaluate the impact of therapeutics on the CF lung microbiota there are also a variety of other factors that need to be considered. As previously mentioned the choice of study design (longitudinal/cross-sectional) and potential patient stratification might facilitate a greater understanding of different treatment responses in such a heterogeneous group of patients. Furthermore, although not taken into account in this study, antibiotic or therapeutic treatment regimes, patient factors such as age, CF genotype and clinical status need also to be incorporated into analyses to unravel the complexity of microbiota associations (Bevivino et al., 2019). Our study has shown that microbiota analysis

617 can be applied with standardised protocols and excellent reproducibility to CF adults, showing
618 they expectorate sputum samples with limited variation at the point of collection. We also
619 demonstrate the potential of using microbiota analysis as a means to stratify participants
620 during enrolment for anti-infective clinical trials in CF. Understanding whether participants are
621 dominated with key pathogens or have a diverse microbiota will shed light on the efficacy of
622 new treatments to target specific pathogens.

References

Anuj, S. N., Whiley, D. M., Kidd, T. J., Bell, S. C., Wainwright, C. E., Nissen, M. D. & Sloots, T. P. 2009. Identification of *Pseudomonas aeruginosa* by a duplex real-time polymerase chain reaction assay targeting the ecfX and the gyrB genes. *Diagnostic Microbiology and Infectious Disease,* 63**,** 127-131.

Bacci, G., Taccetti, G., Dolce, D., Armanini, F., Segata, N., Di Cesare, F., Lucidi, V., Fiscarelli, E., Morelli, P., Casciaro, R., Negroni, A., Mengoni, A. & Bevivino, A. 2019. The personalized temporal dynamics of microbiome in the airways of cystic fibrosis patients. *bioRxiv***,** 609057.

Bernarde, C., Keravec, M., Mounier, J., Gouriou, S., Rault, G., Ferec, C., Barbier, G. & Hery-Arnaud, G. 2015. Impact of the CFTR-potentiator ivacaftor on airway microbiota in cystic fibrosis patients carrying a G551D mutation. *PLoS One,* 10**,** e0124124.

Bevivino, A., Bacci, G., Drevinek, P., Nelson, M. T., Hoffman, L. & Mengoni, A. 2019. Deciphering the Ecology of Cystic Fibrosis Bacterial Communities: Towards Systems-Level Integration. *Trends in Molecular Medicine,* 25**,** 1110-1122.

Bjarnsholt, T., Ciofu, O., Molin, S., Givskov, M. & Høiby, N. 2013. Applying insights from biofilm biology to drug development — can a new approach be developed? *Nature Reviews Drug Discovery,* 12**,** 791-808.

Blainey, P. C., Milla, C. E., Cornfield, D. N. & Quake, S. R. 2012. Quantitative Analysis of the Human Airway Microbial Ecology Reveals a Pervasive Signature for Cystic Fibrosis. *Science Translational Medicine,* 4**,** 153ra130.

Burns, J. L. & Rolain, J. M. 2014. Culture-based diagnostic microbiology in cystic fibrosis: can we simplify the complexity? *J Cyst Fibros,* 13**,** 1-9.

Carmody, L. A., Zhao, J., Kalikin, L. M., Lebar, W., Simon, R. H., Venkataraman, A., Schmidt, T. M., Abdo, Z., Schloss, P. D. & Lipuma, J. J. 2015. The daily dynamics of cystic fibrosis airway microbiota during clinical stability and at exacerbation. *Microbiome,* 3**,** 12-12.

Caverly, L. J., Zhao, J. & Lipuma, J. J. 2015. Cystic fibrosis lung microbiome: Opportunities to reconsider management of airway infection. *Pediatric Pulmonology,* 50**,** S31-S38.

Davies, J. C., Alton, E. W. F. W. & Bush, A. 2007. Cystic fibrosis. *BMJ (Clinical research ed.),* 335**,** 1255-1259.

De Dios Caballero, J., Vida, R., Cobo, M., Máiz, L., Suárez, L., Galeano, J., Baquero, F., Cantón, R. & Del Campo, R. 2017. Individual Patterns of Complexity in Cystic Fibrosis Lung Microbiota, Including Predator Bacteria, over a 1-Year Period. *mBio,* 8**,** e00959-17.

Döring, G., Flume, P., Heijerman, H. & Elborn, J. S. 2012. Treatment of lung infection in patients with cystic fibrosis: Current and future strategies. *Journal of Cystic Fibrosis,* 11**,** 461-479.

Fajac, I. & De Boeck, K. 2017. New horizons for cystic fibrosis treatment. *Pharmacology & Therapeutics,* 170**,** 205-211.

Flight, W. G., Smith, A., Paisey, C., Marchesi, J. R., Bull, M. J., Norville, P. J., Mutton, K. J., Webb, A. K., Bright-Thomas, R. J., Jones, A. M. & Mahenthiralingam, E. 2015. Rapid Detection of Emerging Pathogens and Loss of Microbial Diversity Associated with Severe Lung Disease in Cystic Fibrosis. *J Clin Microbiol,* 53**,** 2022-9.

Goddard, A. F., Staudinger, B. J., Dowd, S. E., Joshi-Datar, A., Wolcott, R. D., Aitken, M. L., Fligner, C. L. & Singh, P. K. 2012. Direct sampling of cystic fibrosis lungs indicates that DNA-based analyses of upper-airway specimens can misrepresent lung microbiota. *Proceedings of the National Academy of Sciences,* 109**,** 13769.

Hall, T. A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series,* 41**,** 95-98.

Heirali, A. A., Workentine, M. L., Acosta, N., Poonja, A., Storey, D. G., Somayaji, R., Rabin, H. R., Whelan, F. J., Surette, M. G. & Parkins, M. D. 2017. The effects of inhaled aztreonam on the cystic fibrosis lung microbiome. *Microbiome,* 5**,** 51.

Hisert, K. B., Heltshe, S. L., Pope, C., Jorth, P., Wu, X., Edwards, R. M., Radey, M., Accurso, F. J., Wolter, D. J., Cooke, G., Adam, R. J., Carter, S., Grogan, B., Launspach, J. L., Donnelly, S. C., Gallagher, C. G., Bruce, J. E., Stoltz, D. A., Welsh, M. J., Hoffman, L. R., Mckone, E. F. & Singh, P. K. 2017. Restoring Cystic Fibrosis Transmembrane Conductance Regulator Function Reduces Airway Bacteria and Inflammation in People with Cystic Fibrosis and Chronic Lung Infections. *American journal of respiratory and critical care medicine,* 195**,** 1617-1628.

Ho, N. T., Li, F., Lee-Sarwar, K. A., Tun, H. M., Brown, B. P., Pannaraj, P. S., Bender, J. M., Azad, M. B., Thompson, A. L., Weiss, S. T., Azcarate-Peril, M. A., Litonjua, A. A., Kozyrskyj, A. L., Jaspan, H. B., Aldrovandi, G. M. & Kuhn, L. 2018. Meta-analysis of effects of exclusive breastfeeding on infant gut microbiota across populations. *Nature Communications,* 9**,** 4169.

Ho, N. T., Li, F., Wang, S. & Kuhn, L. 2019. metamicrobiomeR: an R package for analysis of microbiome relative abundance data using zero-inflated beta GAMLSS and meta-analysis across studies using random effects models. *BMC Bioinformatics,* 20**,** 188.

Jorth, P., Ehsan, Z., Rezayat, A., Caldwell, E., Pope, C., Brewington, J. J., Goss, C. H., Benscoter, D., Clancy, J. P. & Singh, P. K. 2019. Direct Lung Sampling Indicates That Established Pathogens Dominate Early Infections in Children with Cystic Fibrosis. *Cell Reports,* 27**,** 1190-1204.e3.

Kenna, D. T. D., Lilley, D., Coward, A., Martin, K., Perry, C., Pike, R., Hill, R. & Turton, J. F. 2017. Prevalence of Burkholderia species, including members of Burkholderia cepacia complex, among UK cystic and non-cystic fibrosis patients. *J Med Microbiol,* 66**,** 490-501.

Khan, S., Tondervik, A., Sletta, H., Klinkenberg, G., Emanuel, C., Onsoyen, E., Myrvold, R., Howe, R. A., Walsh, T. R., Hill, K. E. & Thomas, D. W. 2012. Overcoming drug resistance with alginate oligosaccharides able to potentiate the action of selected antibiotics. *Antimicrob Agents Chemother,* 56**,** 5134-41.

Lane, D. J. 1991. 16S/23S rRNA Sequencing. *In:* STACKEBRANDT, E. & GOODFELLOW, M. (eds.) *Nucleic Acid Techniques in Bacterial Systematics.* New York: John Wiley and Sons.

Lipuma, J. J. 2010. The changing microbial epidemiology in cystic fibrosis. *Clin Microbiol Rev,* 23**,** 299-323.

Mackenzie, T., Gifford, A. H., Sabadosa, K. A., Quinton, H. B., Knapp, E. A., Goss, C. H. & Marshall, B. C. 2014. Longevity of patients with cystic fibrosis in 2000 to 2010 and beyond: survival analysis of the Cystic Fibrosis Foundation patient registry. *Annals of internal medicine,* 161**,** 233-241.

Mahboubi, M. A., Carmody, L. A., Foster, B. K., Kalikin, L. M., Vandevanter, D. R. & Lipuma, J. J. 2016. Culture-Based and Culture-Independent Bacteriologic Analysis of Cystic Fibrosis Respiratory Specimens. *Journal of Clinical Microbiology,* 54**,** 613.

Muhlebach, M. S., Zorn, B. T., Esther, C. R., Hatch, J. E., Murray, C. P., Turkovic, L., Ranganathan, S. C., Boucher, R. C., Stick, S. M. & Wolfgang, M. C. 2018. Initial acquisition and succession of the cystic fibrosis lung microbiome is associated with disease progression in infants and preschool children. *PLOS Pathogens,* 14**,** e1006798.

Nelson, M. T., Pope, C. E., Marsh, R. L., Wolter, D. J., Weiss, E. J., Hager, K. R., Vo, A. T., Brittnacher, M. J., Radey, M. C., Hayden, H. S., Eng, A., Miller, S. I., Borenstein, E. & Hoffman, L. R. 2019. Human and Extracellular DNA Depletion for Metagenomic Analysis of Complex Clinical Infection Samples Yields Optimized Viable Microbiome Profiles. *Cell Reports,* 26**,** 2227-2240.e5.

Oriano, M., Terranova, L., Teri, A., Sottotetti, S., Ruggiero, L., Tafuro, C., Marchisio, P., Gramegna, A., Amati, F., Nava, F., Franceschi, E., Cariani, L., Blasi, F. & Aliberti, S. 2019. Comparison of different conditions for DNA extraction in sputum - a pilot study. *Multidisciplinary Respiratory Medicine,* 14**,** 6.

Ospina, R. & Ferrari, S. L. 2012. A general class of zero-or-one inflated beta regression models. *Computational Statistics & Data Analysis,* 56**,** 1609-1623.

Peng, X., Li, G. & Liu, Z. 2016. Zero-Inflated Beta Regression for Differential Abundance Analysis with Metagenomics Data. *Journal of computational biology : a journal of computational molecular cell biology,* 23**,** 102-110.

Powell, L. C., Pritchard, M. F., Ferguson, E. L., Powell, K. A., Patel, S. U., Rye, P. D., Sakellakou, S.-M., Buurma, N. J., Brilliant, C. D., Copping, J. M., Menzies, G. E., Lewis, P. D., Hill, K. E. & Thomas, D. W. 2018. Targeted disruption of the extracellular polymeric network of Pseudomonas aeruginosa biofilms by alginate oligosaccharides. *npj Biofilms and Microbiomes,* 4**,** 13.

Pritchard, M. F., Powell, L. C., Jack, A. A., Powell, K., Beck, K., Florance, H., Forton, J., Rye, P. D., Dessen, A., Hill, K. E. & Thomas, D. W. 2017. A Low-Molecular-Weight Alginate Oligosaccharide Disrupts Pseudomonal Microcolony Formation and Enhances Antibiotic Effectiveness. *Antimicrob Agents Chemother,* 61.

Pritchard, M. F., Powell, L. C., Menzies, G. E., Lewis, P. D., Hawkins, K., Wright, C., Doull, I., Walsh, T. R., Onsøyen, E., Dessen, A., Myrvold, R., Rye, P. D., Myrset, A. H., Stevens, H. N. E., Hodges, L. A., Macgregor, G., Neilly, J. B., Hill, K. E. & Thomas, D. W. 2016. A New Class of Safe Oligosaccharide Polymer Therapy To Modify the Mucus Barrier of Chronic Respiratory Disease. *Molecular Pharmaceutics,* 13**,** 863-872.

R-Core-Team 2013. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.

Rogers, G. B., Zain, N. M. M., Bruce, K. D., Burr, L. D., Chen, A. C., Rivett, D. W., Mcguckin, M. A. & Serisier, D. J. 2014. A novel microbiota stratification system predicts future exacerbations in bronchiectasis. *Annals of the American Thoracic Society,* 11**,** 496-503.

Salter, S. J., Cox, M. J., Turek, E. M., Calus, S. T., Cookson, W. O., Moffatt, M. F., Turner, P., Parkhill, J., Loman, N. J. & Walker, A. W. 2014. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biology,* 12**,** 87.

Sass, A. M., Schmerk, C., Agnoli, K., Norville, P. J., Eberl, L., Valvano, M. A. & Mahenthiralingam, E. 2013. The unexpected discovery of a novel low-oxygen-activated locus for the anoxic persistence of Burkholderia cenocepacia. *Isme j,* 7**,** 1568-81.

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., Lesniewski, R. A., Oakley, B. B., Parks, D. H., Robinson, C. J., Sahl, J. W., Stres, B., Thallinger, G. G., Van Horn, D. J. & Weber, C. F. 2009. Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. *Applied and Environmental Microbiology,* 75**,** 7537-7541.

Shankar, J. 2017. Insights into study design and statistical analyses in translational microbiome studies. *Annals of Translational Medicine,* 5**,** 2.

Spilker, T., Baldwin, A., Bumford, A., Dowson, C. G., Mahenthiralingam, E. & Lipuma, J. J. 2009. Expanded multilocus sequence typing for *Burkholderia* species. *J Clin Microbiol,* 47**,** 2607-10.

Stanojevic, S. & Ratjen, F. 2016. Physiologic endpoints for clinical studies for cystic fibrosis. *J Cyst Fibros,* 15**,** 416-23.

Surette, M. G. 2014. The Cystic Fibrosis Lung Microbiome. *Annals of the American Thoracic Society,* 11**,** S61-S65.

Terranova, L., Oriano, M., Teri, A., Ruggiero, L., Tafuro, C., Marchisio, P., Gramegna, A., Contarini, M., Franceschi, E., Sottotetti, S., Cariani, L., Bevivino, A., Chalmers, J. D., Aliberti, S. & Blasi, F. 2018. How to Process Sputum Samples and Extract Bacterial DNA for Microbiota Analysis. *International journal of molecular sciences,* 19**,** 3256.

Van Koningsbruggen-Rietschel, S., Davies, J. C., Pressler, T., Fischer, R., Macgregor, G., Donaldson, S. H., Smerud, K., Meland, N., Mortensen, J., Fosbøl, M. Ø., Downey, D. G., Myrset, A. H., Flaten, H. & Rye, P. D. 2020. Inhaled dry powder alginate oligosaccharide in cystic fibrosis: a randomised, double-blind, placebo-controlled, crossover phase 2b study. *ERJ Open Research,* 6**,** 00132-2020.

Wang, Q., Garrity, G. M., Tiedje, J. M. & Cole, J. R. 2007. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and Environmental Microbiology,* 73**,** 5261-7.

Willner, D., Haynes, M. R., Furlan, M., Schmieder, R., Lim, Y. W., Rainey, P. B., Rohwer, F. & Conrad, D. 2011. Spatial distribution of microbial communities in the cystic fibrosis lung. *The Isme Journal,* 6**,** 471.

Winsor, G. L., Khaira, B., Van Rossum, T., Lo, R., Whiteside, M. D. & Brinkman, F. S. L. 2008. The *Burkholderia* Genome Database: facilitating flexible queries and comparative analyses. *Bioinformatics,* 24**,** 2803-2804.

Zemanick, E. T., Wagner, B. D., Sagel, S. D., Stevens, M. J., Accurso, F. J. & Harris, J. K. 2010. Reliability of quantitative real-time PCR for bacterial detection in cystic fibrosis airway specimens. *PloS one,* 5**,** e15101-e15101.

Zlosnik, J. E., Zhou, G., Brant, R., Henry, D. A., Hird, T. J., Mahenthiralingam, E., Chilvers, M. A., Wilcox, P. & Speert, D. P. 2015. Burkholderia species infections in patients with cystic fibrosis in British Columbia, Canada. 30 years' experience. *Ann Am Thorac Soc,* 12**,** 70-8.

Author contributions

Conceptualization: EM, PDR, RW; **Methodology:** EM, PDR, RW; **Software:** RW; **Validation:** EM, PDR, RW; **Formal analysis:** EM, PDR, RW; **Investigation:** RW; **Resources:** EM, PDR; Data **Curation:** EM, PDR, RW; **Writing – original draft preparation:** EM, RW; **Writing – review and editing:** EM, PDR, RW; **Visualization:** RW**; Supervision:** EM, PDR; **Project administration:** EM, PDR, RW; **Funding acquisition:** EM, PDR

Acknowledgements

We would like to thank all the patients and their families who participated in this study, as well as all clinical trial teams for their engagement in the trials (NCT02157922; NCT02453789).

The following investigators and their teams (in alphabetical order) were involved in the clinical trials. Mary Carroll, General Hospital, Southampton, UK; Jane Davies, Royal Brompton and

Harefield NHS Foundation Trust, London, UK (the study was supported by the NIHR Biomedical Research Unit and Clinical Research Facility); Carsten Schwarz and Nico Derichs, Christiane Herzog Zentrum, Charité Berlin, Germany; Damian Downey, Centre for Experimental Medicine, Queen's University, Belfast, UK; Olaf Eickmeier, University Hospital, Frankfurt a.M., Germany; Pal Leyell Finstad, University Hospital, Oslo, Norway; Rainald Fischer, Pneumologische Praxis Pasing, Munich, Germany; Marie Øbro Fossbøl, Rigshospitalet, Copenhagen, Denmark; Marita Gilljam, Sahlgrenska Universitetssjukhuset, Göteborg, Sweden; Charles Haworth, Papworth Hospital, Cambridge, UK; Lena Hjelte, Karolinska Universitetssjukhuset, Stockholm, Sweden; Alan Knox, City Hospital, Nottingham, UK; Silke van Koningsbruggen-Rietschel, University Hospital, Cologne, Germany; Gordon MacGregor, Queen Elizabeth University Hospital, Glasgow, UK; Jann Mortensen, Rigshospitalet, Copenhagen, Denmark; Susanne Nährig, LMU, Munich, Germany; Tacjana Pressler, Rigshospitalet, Copenhagen, Denmark; Joachim Riethmüller, University Hospital, Tübingen, Germany; Felix C. Ringshausen, MHH, Hannover, Germany; Martin Walshaw, Heart and Chest Hospital, Liverpool, UK; qualified person responsible for pharmacovigilance: Hugo Flaten, AlgiPharma AS, Sandvika, Norway; statistician: Nils Meland, Smerud Medical Research International AS, Oslo, Norway. The MCC substudy was performed at three sites: Laura Gow, Bio-Images Research Ltd, Glasgow, UK; Joy Conway, University of Southampton, Southampton, UK; Jann Mortensen, Marie Øbro Fosbøl, Rigshospitalet, Copenhagen, Denmark. Data was evaluated and reviewed by Scott H. Donaldson and William Bennett, UNC Chapel Hill, NC, USA.

Role of the funding source

Funding for this study was received from the Cystic Fibrosis Foundation, Bethesda, MD, USA, with additional funding provided by the study sponsor AlgiPharma AS, Sandvika, Norway. The study sponsor representative (PDR) participated in the study design, data collection, data analysis, data interpretation and writing of the study report. The authors had full access to all

data and had final responsibility for publication. The final decision on content was exclusively retained by the contributing authors. EM and RW acknowledge current funding by the Cystic Fibrosis Foundation (MAHENT20G0).

Declaration of competing interest

PDR reports grants from Cystic Fibrosis Foundation during the conduct of the study and is Chief Scientific Officer at AlgiPharma and holds stock in AlgiPharma AB, outside the submitted work; in addition, he has patents WO 2015/128495 and WO 2016/151051 pending.