

Online Research @ Cardiff

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository: <https://orca.cardiff.ac.uk/id/eprint/137377/>

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Weiser, Rebecca, Rye, Phillip D. and Mahenthiralingam, Eshwar 2021. Implementation of microbiota analysis in clinical trials for cystic fibrosis lung infection: experience from the OligoG phase 2b clinical trials. *Journal of Microbiological Methods* 181 , 106133. 10.1016/j.mimet.2021.106133 filefile

Publishers page: <http://dx.doi.org/10.1016/j.mimet.2021.106133>
<<http://dx.doi.org/10.1016/j.mimet.2021.106133>>

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies.

See

<http://orca.cf.ac.uk/policies.html> for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.



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

3

4 **Authors: Rebecca Weiser^{1*}, Philip D. Rye², and Eshwar Mahenthiralingam^{1*}**

5

6 **Affiliations:**

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

10

11 ²AlgiPharma AS, Industriveien 33, N-1337 Sandvika, Norway

12

13 *** Corresponding authors:**

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

16

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

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

20

21

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

24

25 **ORCID ID:**

26 0000-0003-3983-3272 (R. Weiser)

27 0000-0001-9014-3790 (E. Mahenthiralingam).

28 0000-0001-7762-3300 (P.D. Rye)

29 **Highlights**

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

35

36

37 **Abstract**

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

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

61

62 **Keywords:** microbiota analysis, cystic fibrosis microbiology, clinical trials, PCR, qPCR and
63 16S rRNA gene sequencing

64

65 **Introduction**

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

74

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

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

105

106 The low molecular weight alginate oligosaccharide OligoG CF-5/20 is a novel, non-antibiotic
107 therapy that has antibiotic potentiation (Khan et al., 2012, Pritchard et al., 2017) and anti-
108 biofilm (Pritchard et al., 2017, Powell et al., 2018) properties, and can alter the visco-elastic
109 properties of sputum (Pritchard et al., 2016). Two phase 2b clinical trials have documented the
110 safety and efficacy of OligoG CF-5/20 in CF patients infected with *P. aeruginosa*
111 (clinicaltrials.gov identifier NCT02157922)(van Koningsbruggen-Rietschel et al., 2020) or Bcc
112 species (clinicaltrials.gov identifier NCT02453789) as the primary CF pathogen in the lung.
113 Lung function, specifically the Forced Expiratory Volume in 1 second (FEV1), is widely used
114 as a pivotal outcome measure in the development of drugs to treat CF (Stanojevic and Ratjen,
115 2016). However, for novel anti-infective agents such as OligoG CF-5/20 which are multimodal
116 in their activity (Pritchard et al., 2017, Powell et al., 2018), understanding which outcome
117 measures are relevant in terms of their efficacy for the treatment of lung infections is
118 challenging. The safety and potential for clinical efficacy of OligoG CF-5/20 within the *P.*
119 *aeruginosa* trial has been recently described (van Koningsbruggen-Rietschel et al., 2020).
120 Here we report the microbiota analysis accompanying both the *P. aeruginosa* and Bcc OligoG

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

124

125 **Methods**

126 **Study design and Patient cohort**

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

142

143

144

145

146

147

148

149 **Table 1. Patient demographics**

Characteristic	Bcc trial	<i>P. aeruginosa</i> 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 Expecterated 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®

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

177

178 **Identification of *Burkholderia* species**

179 To determine the identity of the Bcc species in V1 samples from patients in the Bcc trial, *recA*
180 and *gyrB* gene sequences were PCR amplified from sputum DNA extracts using previously
181 described primers (Spilker et al., 2009)(Table 2). PCR products were purified and sent to
182 Eurofins Genomics for Sanger sequencing of forward and reverse DNA strands. BioEdit (Hall,
183 1999) was used to create consensus sequences for *recA* and *gyrB* genes and the *Burkholderia*
184 species identity determined using the BLASTN tool of the *Burkholderia* Genome Database
185 (Winsor et al., 2008). Full details are given in Supplementary method S2.

186

187 **Quantitative PCR (qPCR)**

188 Quantification of *P. aeruginosa* was performed by targeting the *gyrB* gene using a previously
189 designed TaqMan assay (Anuj et al., 2009). Quantification of *Burkholderia* was achieved using
190 primers targeting the *rpoD* gene (Sass et al., 2013) and a TaqMan probe designed in this
191 study (Table 2). Reaction volumes were 10 µl and comprised 1X Platinum qPCR Supermix-
192 UDG with ROX (Life Technologies), 1.8 µM forward and reverse primers supplied by Eurofins
193 Genomics and 225 nM TaqMan probe (Thermofisher Scientific), 1 µl template DNA and
194 nuclease free water (Severn Biotech Ltd). Reactions were performed in triplicate with negative
195 controls, alongside a standard dilution series of 10²-10⁸ *gyrB* (amplified from *P. aeruginosa*
196 PAO1) or *rpoD* (amplified from *B. cenocepacia* J2315) gene copies per µl (Table 2). The Bio-
197 Rad Chromo4 system was used with the following qPCR cycling conditions: UNG treatment

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

204

205

206 **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					
<i>rpoD</i>	F	GATCTTGCACATCGTCGTC	59	1011	This study
	R	GTTTCGTAACGGAGACGCTG			
<i>gyrB</i>	F	GAGTCGATCACTGTCCGC	58	1186	This study
	R	GCATCTTGTCTGAAGCGCG			
qPCR primers and TaqMan probes					
<i>rpoD</i>	F	GAGATGAGCACCGATCACAC	67	143	(Sass <i>et al.</i>
	R	CCTTCGAGGAACGACTTCAG			2013)
	PROBE	5'FAM-CTGCGCAAGCTGCGTCACC- 3'MGBNFQ			This study
<i>gyrB</i>	F	CCTGACCATCCGTCGCCACAAC	60	220	(Anuj <i>et al.</i>
	R	CGCAGCAGGATGCCGACGCC			2009)
	PROBE	5'FAM- GGTCTGGGAACAGGTCTACCACCACGG- 3'MGBNFQ			

207

208

209

210 **16S rRNA gene sequencing and bacterial diversity analysis**

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

229

230 **Statistical analysis**

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

238 index), total abundance and relative abundance of the key pathogen (*P. aeruginosa* or Bcc)
239 were determined between three start and end points during the trial: Screening (V1) and End
240 (V8), OligoG start (V2 or V5) and OligoG end (V4 or V7), and Placebo start (V2 or V5) and
241 Placebo end (V4 or V7). For Shannon diversity and total abundance, Wilcoxon signed-rank
242 tests (non-parametric) were performed to determine significant differences between paired
243 time points. Differences in relative abundance were analysed using Generalized Additive
244 Models for Location, Scale and Shape (GAMLSS package in R) with a beta inflated family
245 (BEINF; Zero-One inflated beta model), (mu) logit links and other default options as previously
246 recommended for proportional microbiome data (Ho et al., 2019, Peng et al., 2016, Ho et al.,
247 2018) where the variable of interest (relative abundance) may be zero or one (Ospina and
248 Ferrari, 2012). In the models, the response variable was relative abundance, the fixed effect
249 was trial stage (with start as the reference class to which end was compared) and patient ID
250 was specified as the random effect. The estimates (regression coefficients) of the model are
251 the difference in log odds of genus relative abundances between paired start and end groups
252 (Ho et al., 2018).

253

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

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

270

271

272

273 **Results**

274

275 **Patient summary, standardised methods and application of bacterial diversity analysis**

276 The demographics of patients recruited into each trial are given in Table 1, with a total of 666
277 sputum samples collected from 45 individuals (511 samples) within the *P. aeruginosa* trial,
278 and 13 patients (155 samples) within the Bcc trial. Paired sputum samples (designated S1
279 and S2) were obtained for nearly all the 6 time points across both trials (Supplementary Table
280 S1; Supplementary Table S2). Standardised protocols were developed and adhered to for
281 multiple sample processing and analysis (Supplementary methods S1-6). Within the Bcc
282 study, one of the Screening (V1_S1) samples was subjected to *recA* and *gyrB* gene
283 sequencing (26 samples) to determine the identity of the infecting *B. cepacia* complex species.
284 All 155 Bcc trial samples were submitted for 16S rRNA gene sequencing bacterial diversity
285 analysis, and one sample (S1) from each time point (78 samples) was used for *rpoD* qPCR to
286 determine Bcc bacterial load in sputum. Within the *P. aeruginosa* trial, 16S rRNA gene
287 sequencing bacterial diversity analysis was performed for all 511 samples and *P. aeruginosa*
288 load was determined by *gyrB* qPCR for one sample (S1) at each time point (270 samples).
289 Two samples were excluded from the bacterial diversity analysis due to the number of reads
290 falling below the threshold level of 1000 (Patient 27610-011, sample V7_S1 from the Bcc and
291 82601-003, sample V1_S1 from the *P. aeruginosa* trial). Accurate determination of bacterial
292 classification to the genus level was obtained by microbiota analysis, and therefore all

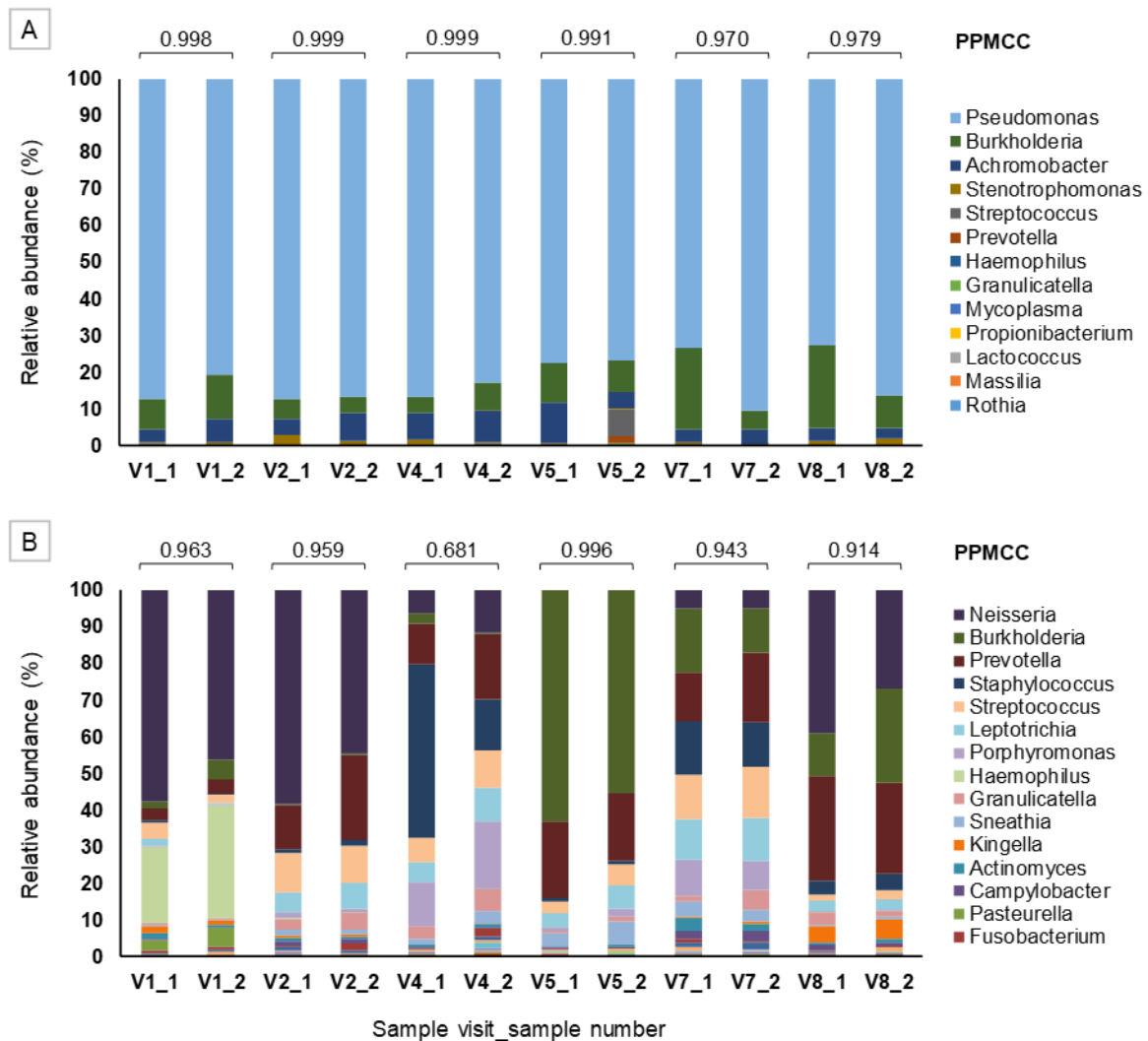
293 subsequent results will be referred to in context of *Pseudomonas*, *Burkholderia* or other
294 genera as appropriate.

295

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

297 Culture-based bacteriology can be highly variable from CF sputum unless rigorous protocols
298 and appropriate standards are employed (Burns and Rolain, 2014). However, the variability of
299 sputum microbiota profiles collected over short time frames such as sampling visits for clinical
300 trials is not known. Therefore, where possible, paired sputum samples were collected within 2
301 hours of each other at each time point during the trials to understand potential sampling
302 variation. In total, 76 and 241 paired samples were collected for the Bcc and *P. aeruginosa*
303 trials, respectively. Microbiota analysis demonstrated that the paired sputum samples
304 collected within 2 hours of each other had concordant bacterial diversity profiles.
305 Representative examples of microbiota profiles indicating levels of similarity between paired
306 samples are shown in Figure 1. Pearson product-moment correlation coefficients (PPMCC)
307 for the *P. aeruginosa* (mean = 0.92, range = 0.07-1.00) and Bcc (mean = 0.98, range = 0.68-
308 1.00) paired samples were also high, corroborating the similarity and stability of the bacterial
309 microbiota signature over the short time frame of 2 hours (Supplementary Figure S1).

310



311

312 **Figure 1. Microbiota stability between paired sputum samples.** Relative abundances of genera in
 313 paired sputum samples (all time points) are shown for: **(A)** a pathogen dominated patient (Bcc trial
 314 patient 27610-007; all genera present shown in key); **(B)** a microbiota diverse patient (Bcc trial patient
 315 27611-006). The top 15 genera for each individual are shown in the key and the Pearson product-
 316 moment correlation coefficient (PPMCC) values are displayed above the stacked bar charts to
 317 demonstrate the level of concordance between paired samples.

318

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

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

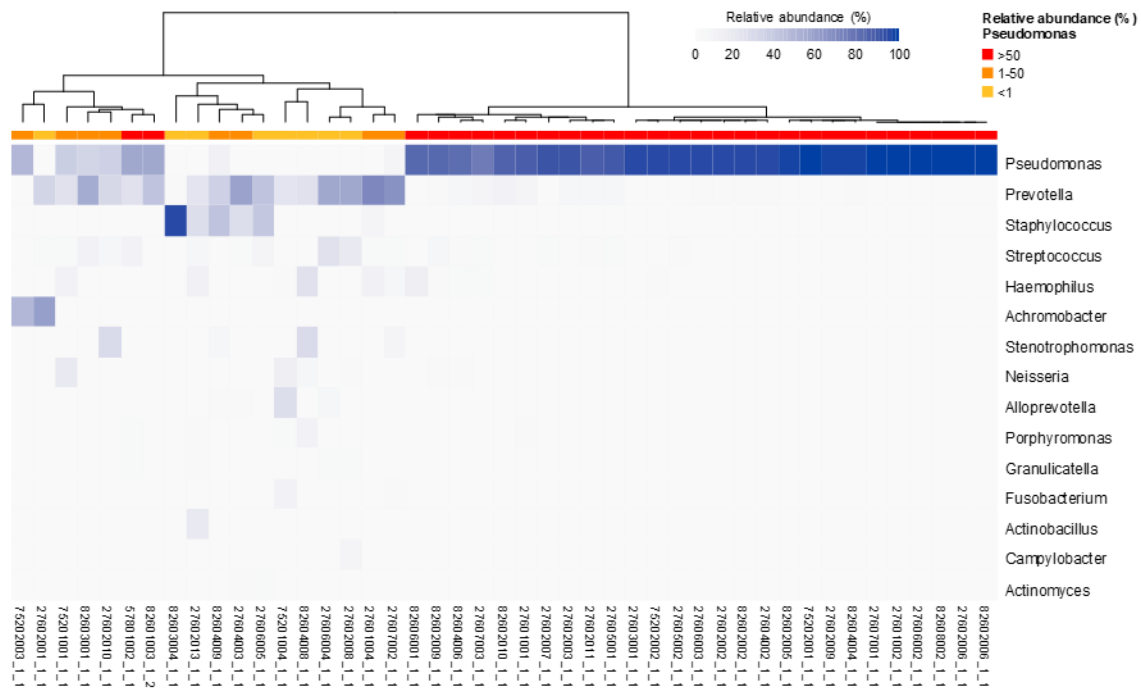
330

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

332 Patients were recruited for the clinical trials on the basis that they were colonised with *P.*
333 *aeruginosa* and not Bcc species for the for *P. aeruginosa* trial, or any Bcc species for the Bcc
334 trial. However, culture-independent molecular diagnostics and bacterial diversity analysis of
335 sputum samples at screening (sample V1) revealed that not all patient samples had
336 substantial evidence of infection with these pathogens. The heterogeneity in bacterial diversity
337 between patients was clearly seen by hierarchical cluster analysis of the 16S rRNA
338 sequencing diversity data. For the *P. aeruginosa* trial, the microbiota profiles split into 2 major
339 groups of individuals, those dominated by the target pathogen in the trial, and those with more
340 diverse microbiota (Figure 2). The same phenomenon was observed for the Bcc trial (Figure
341 5A), and additional microbiota metrics demonstrating the overall variability of the V1 samples
342 for both trials are given in Table 3.

343

344



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;
 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*.

362

363 Furthermore, the *P. aeruginosa* clinical trial had a subset of patients (n=10; 27601-004, 27602-
364 010, 27604-003, 27606-004, 27607-002, 75201-001, 75202-003, 82603-001, 82604-008 and
365 82604-009) where *Pseudomonas* was not the dominant genus (defined as <50% relative
366 abundance in the sample). In these non-dominant *P. aeruginosa* cases, *Prevotella*,
367 *Staphylococcus*, *Achromobacter* and *Stenotrophomonas* were the prevalent genera (>25%
368 relative abundance). The Bcc clinical trial also had a subset of patients where *Burkholderia*
369 was not dominant (n=3; 27610-001, 27610-007 and 27611-006), and *Pseudomonas*,
370 *Streptococcus* and *Neisseria* were instead prevalent. The remaining patients in both trials had
371 either *Pseudomonas* (n=29; 64% of the trial cohort) or *Burkholderia* (n=7; 53% of the trial
372 cohort) as the dominant genus. Overall, the microbiota analysis demonstrated that at
373 screening, despite all participants being unified by a single bacteriology measure of culture
374 positivity for either Bcc or *P. aeruginosa* for the different trials, lung infection in over one third
375 of the trial participants was highly variable.

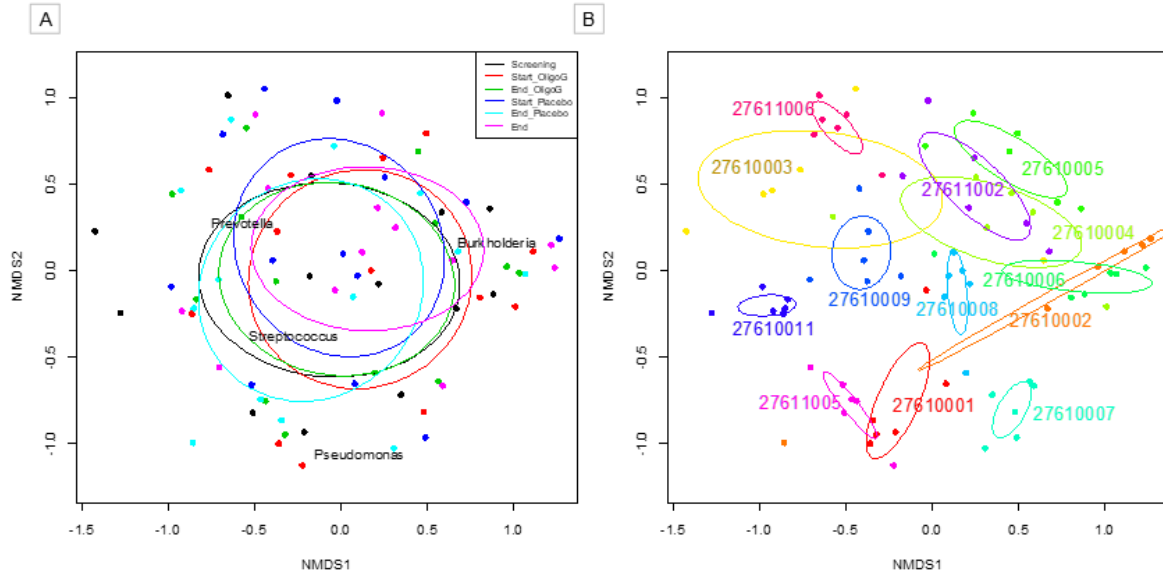
376

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

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

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

392



393

394 **Figure 3. The lung microbiota within the *Bcc* Oligo G trial participants was linked to the**
 395 **individual rather than treatment.** NMDS analysis of Bray-Curtis dissimilarity values for S1 samples
 396 from all 6 time points for the 13 patients in the Bcc trial (except 27610-011 that only had S2 for V7).
 397 Analyses grouped by: **(A)** treatment, and **(B)** patient are shown. Points represent individual samples,
 398 ellipses are standard deviations of points scores for each grouping. The top 7% genera based on
 399 abundance across the dataset are shown in **(A)**. A significant difference was observed between patient
 400 groups (PERMANOVA, $R^2=0.03$ $p=0.043$) but not between treatment groups (PERMANOVA, $R^2=0.03$,
 401 $p=0.98$). For both sample groups (treatment, patients) the group variances were homogeneous,
 402 satisfying the conditions of the PERMANOVA model.

403

404

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

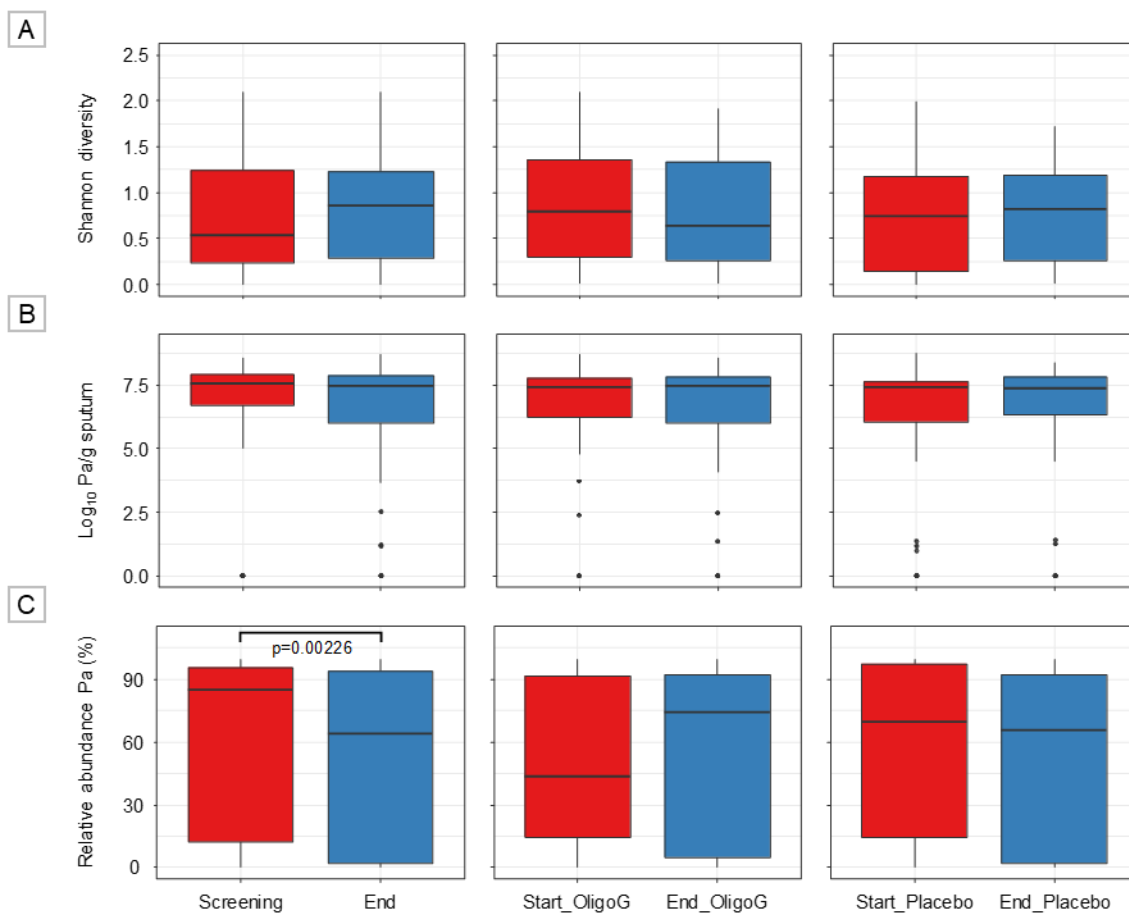
407

Characteristic	Bcc trial	<i>P. aeruginosa</i> 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	<i>Burkholderia</i> , <i>Pseudomonas</i> , <i>Streptococcus</i> , <i>Prevotella</i> , <i>Granulicatella</i>	<i>Pseudomonas</i> , <i>Prevotella</i> , <i>Streptococcus</i> , <i>Haemophilus</i> , <i>Porphyromonas</i>
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	<i>Burkholderia</i> , <i>Pseudomonas</i> , <i>Prevotella</i> , <i>Streptococcus</i> , <i>Granulicatella</i> , <i>Actinomcyes</i> , <i>Leptotrichia</i>	<i>Pseudomonas</i> , <i>Prevotella</i> , <i>Streptococcus</i> , <i>Haemophilus</i> , <i>Granulicatella</i>
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)

408 Footnotes: ^{*}data are given as mean and range of S1 samples (except Bcc 27610-011 V7 and Pa 82601-
409 003 V1 where only S2 samples were available); ^aTop genera were defined as those identified in $\geq 50\%$
410 of V1 samples by 16S rRNA gene sequencing analysis, and had a maximum relative abundance of
411 $\geq 10\%$; ^b FEV1% predicted values were only available for V1-V7 time points for the Bcc trial
412
413
414 Differences in the Shannon diversity (Supplementary Table S3), total (Supplementary Table
415 S4) and relative abundance (Supplementary Table S5) of Bcc and *P. aeruginosa* were also
416 investigated at the three paired start-end points for the trials: Screening and End, OligoG start
417 and OligoG end, and Placebo start and Placebo end. These paired start-end points measured
418 potential short term (28 days; OligoG start to end or Placebo start to end) and longer term

419 changes (112 ± 28 days; Screening to End). In the Bcc trial, no differences were observed
420 between any of the paired time-points with the exception of total *Burkholderia* abundance as
421 measured by qPCR, which significantly decreased ($p=0.02148$) from Placebo start to Placebo
422 end (Supplementary Figure S3). In the *P. aeruginosa* trial, the only difference was the relative
423 abundance of *Pseudomonas* which significantly decreased ($p=0.00226$) from Screening to
424 End (Figure 4). Overall, the microbiota profiles for each chronically infected individual
425 remained stable over the trial period.
426



427

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

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

439

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

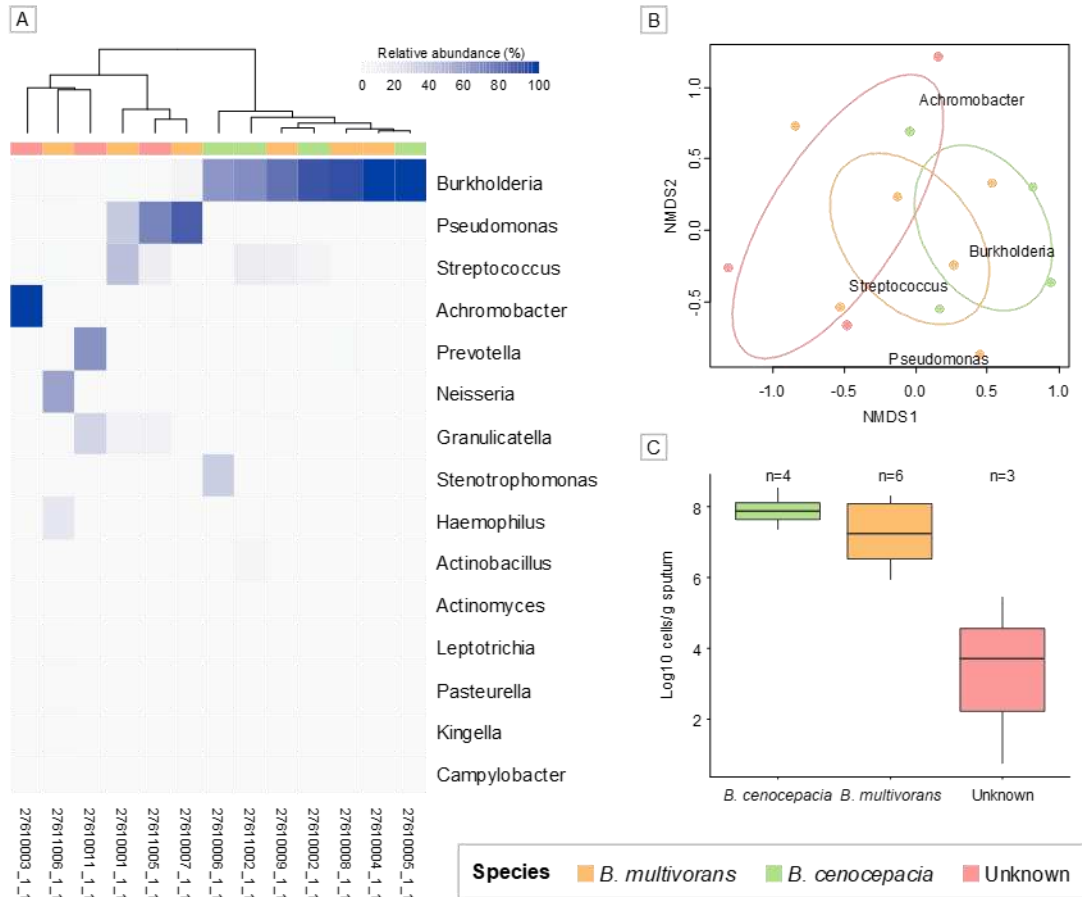
441 The identity of the infecting Bcc species was determined at Screening by *recA* and *gyrB* gene
442 sequence analysis (Supplementary Table S6), and was found to be either *B. cenocepacia*
443 (n=4 patients) or *B. multivorans* (n=6 patients). In three cases, no identification could be made
444 due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

451

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

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

463



464

465

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

476

477

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

488

489 **Discussion**

490 Culture-independent molecular diagnostic methods such as qPCR and microbiota analysis
491 have a number of advantages over culture for CF infection microbiology (Mahboubi et al.,
492 2016). Their application in clinical trials of new anti-infective drugs for CF lung disease has
493 been limited. Here we describe the comprehensive protocols used to incorporate microbiota
494 analysis into the OligoG CF-5/20 Phase 2b clinical trials for *P. aeruginosa* and Bcc. Since
495 OligoG is not a conventional antibiotic and works in a multifactorial manner (Khan et al., 2012,
496 Pritchard et al., 2017, Powell et al., 2018), understanding its effect on both the target pathogen
497 and wider CF lung microbiota was important. OligoG was shown to be safe and have
498 promising clinical efficacy in the *P. aeruginosa* clinical trial (van Koningsbruggen-Rietschel et
499 al., 2020) and the Bcc clinical trial (unpublished data). Our microbiota analysis demonstrated
500 that OligoG CF-5/20 did not cause major changes in the bacterial diversity within the CF lung
501 after short term exposure (28 days). A promising overall signature from the *P. aeruginosa* trial
502 was a significant decrease in the relative abundance of *Pseudomonas* from screening to end
503 (Figure 4), although it was not linked to the OligoG administration periods. The lack of major
504 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 a 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 $112 \pm$
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

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

570

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

588

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

608

609 Overall, to comprehensively evaluate the impact of therapeutics on the CF lung microbiota
610 there are also a variety of other factors that need to be considered. As previously mentioned
611 the choice of study design (longitudinal/cross-sectional) and potential patient stratification
612 might facilitate a greater understanding of different treatment responses in such a
613 heterogeneous group of patients. Furthermore, although not taken into account in this study,
614 antibiotic or therapeutic treatment regimes, patient factors such as age, CF genotype and
615 clinical status need also to be incorporated into analyses to unravel the complexity of
616 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.

- 624 Anuj, S. N., Whiley, D. M., Kidd, T. J., Bell, S. C., Wainwright, C. E., Nissen, M. D. & Sloots,
625 T. P. 2009. Identification of *Pseudomonas aeruginosa* by a duplex real-time
626 polymerase chain reaction assay targeting the *ecfX* and the *gyrB* genes. *Diagnostic*
627 *Microbiology and Infectious Disease*, 63, 127-131.
- 628 Bacci, G., Taccetti, G., Dolce, D., Armanini, F., Segata, N., Di Cesare, F., Lucidi, V., Fiscarelli,
629 E., Morelli, P., Casciaro, R., Negroni, A., Mengoni, A. & Bevivino, A. 2019. The
630 personalized temporal dynamics of microbiome in the airways of cystic fibrosis
631 patients. *bioRxiv*, 609057.
- 632 Bernarde, C., Keravec, M., Mounier, J., Gouriou, S., Rault, G., Ferec, C., Barbier, G. & Hery-
633 Arnaud, G. 2015. Impact of the CFTR-potentiator ivacaftor on airway microbiota in
634 cystic fibrosis patients carrying a G551D mutation. *PLoS One*, 10, e0124124.
- 635 Bevivino, A., Bacci, G., Drevinek, P., Nelson, M. T., Hoffman, L. & Mengoni, A. 2019.
636 Deciphering the Ecology of Cystic Fibrosis Bacterial Communities: Towards Systems-
637 Level Integration. *Trends in Molecular Medicine*, 25, 1110-1122.
- 638 Bjarnsholt, T., Ciofu, O., Molin, S., Givskov, M. & Høiby, N. 2013. Applying insights from biofilm
639 biology to drug development — can a new approach be developed? *Nature Reviews*
640 *Drug Discovery*, 12, 791-808.
- 641 Blainey, P. C., Milla, C. E., Cornfield, D. N. & Quake, S. R. 2012. Quantitative Analysis of the
642 Human Airway Microbial Ecology Reveals a Pervasive Signature for Cystic Fibrosis.
643 *Science Translational Medicine*, 4, 153ra130.
- 644 Burns, J. L. & Rolain, J. M. 2014. Culture-based diagnostic microbiology in cystic fibrosis: can
645 we simplify the complexity? *J Cyst Fibros*, 13, 1-9.
- 646 Carmody, L. A., Zhao, J., Kalikin, L. M., Lebar, W., Simon, R. H., Venkataraman, A., Schmidt,
647 T. M., Abdo, Z., Schloss, P. D. & Lipuma, J. J. 2015. The daily dynamics of cystic
648 fibrosis airway microbiota during clinical stability and at exacerbation. *Microbiome*, 3,
649 12-12.
- 650 Caverly, L. J., Zhao, J. & Lipuma, J. J. 2015. Cystic fibrosis lung microbiome: Opportunities to
651 reconsider management of airway infection. *Pediatric Pulmonology*, 50, S31-S38.
- 652 Davies, J. C., Alton, E. W. F. W. & Bush, A. 2007. Cystic fibrosis. *BMJ (Clinical research ed.)*,
653 335, 1255-1259.
- 654 De Dios Caballero, J., Vida, R., Cobo, M., Máiz, L., Suárez, L., Galeano, J., Baquero, F.,
655 Cantón, R. & Del Campo, R. 2017. Individual Patterns of Complexity in Cystic Fibrosis
656 Lung Microbiota, Including Predator Bacteria, over a 1-Year Period. *mBio*, 8, e00959-
657 17.
- 658 Döring, G., Flume, P., Heijerman, H. & Elborn, J. S. 2012. Treatment of lung infection in
659 patients with cystic fibrosis: Current and future strategies. *Journal of Cystic Fibrosis*,
660 11, 461-479.
- 661 Fajac, I. & De Boeck, K. 2017. New horizons for cystic fibrosis treatment. *Pharmacology &*
662 *Therapeutics*, 170, 205-211.
- 663 Flight, W. G., Smith, A., Paisey, C., Marchesi, J. R., Bull, M. J., Norville, P. J., Mutton, K. J.,
664 Webb, A. K., Bright-Thomas, R. J., Jones, A. M. & Mahenthiralingam, E. 2015. Rapid
665 Detection of Emerging Pathogens and Loss of Microbial Diversity Associated with
666 Severe Lung Disease in Cystic Fibrosis. *J Clin Microbiol*, 53, 2022-9.
- 667 Goddard, A. F., Staudinger, B. J., Dowd, S. E., Joshi-Datar, A., Wolcott, R. D., Aitken, M. L.,
668 Fligner, C. L. & Singh, P. K. 2012. Direct sampling of cystic fibrosis lungs indicates that
669 DNA-based analyses of upper-airway specimens can misrepresent lung microbiota.
670 *Proceedings of the National Academy of Sciences*, 109, 13769.
- 671 Hall, T. A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis
672 program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, 41, 95-98.
- 673 Heirali, A. A., Workentine, M. L., Acosta, N., Poonja, A., Storey, D. G., Somayaji, R., Rabin,
674 H. R., Whelan, F. J., Surette, M. G. & Parkins, M. D. 2017. The effects of inhaled
675 aztreonam on the cystic fibrosis lung microbiome. *Microbiome*, 5, 51.

- 676 Hisert, K. B., Heltshe, S. L., Pope, C., Jorth, P., Wu, X., Edwards, R. M., Radey, M., Accurso,
677 F. J., Wolter, D. J., Cooke, G., Adam, R. J., Carter, S., Grogan, B., Launspach, J. L.,
678 Donnelly, S. C., Gallagher, C. G., Bruce, J. E., Stoltz, D. A., Welsh, M. J., Hoffman, L.
679 R., Mckone, E. F. & Singh, P. K. 2017. Restoring Cystic Fibrosis Transmembrane
680 Conductance Regulator Function Reduces Airway Bacteria and Inflammation in
681 People with Cystic Fibrosis and Chronic Lung Infections. *American journal of*
682 *respiratory and critical care medicine*, 195, 1617-1628.
- 683 Ho, N. T., Li, F., Lee-Sarwar, K. A., Tun, H. M., Brown, B. P., Pannaraj, P. S., Bender, J. M.,
684 Azad, M. B., Thompson, A. L., Weiss, S. T., Azcarate-Peril, M. A., Litonjua, A. A.,
685 Kozyrskyj, A. L., Jaspan, H. B., Aldrovandi, G. M. & Kuhn, L. 2018. Meta-analysis of
686 effects of exclusive breastfeeding on infant gut microbiota across populations. *Nature*
687 *Communications*, 9, 4169.
- 688 Ho, N. T., Li, F., Wang, S. & Kuhn, L. 2019. metamiR: an R package for analysis of
689 microbiome relative abundance data using zero-inflated beta GAMLSS and meta-
690 analysis across studies using random effects models. *BMC Bioinformatics*, 20, 188.
- 691 Jorth, P., Ehsan, Z., Rezayat, A., Caldwell, E., Pope, C., Brewington, J. J., Goss, C. H.,
692 Bencotter, D., Clancy, J. P. & Singh, P. K. 2019. Direct Lung Sampling Indicates That
693 Established Pathogens Dominate Early Infections in Children with Cystic Fibrosis. *Cell*
694 *Reports*, 27, 1190-1204.e3.
- 695 Kenna, D. T. D., Lilley, D., Coward, A., Martin, K., Perry, C., Pike, R., Hill, R. & Turton, J. F.
696 2017. Prevalence of Burkholderia species, including members of Burkholderia cepacia
697 complex, among UK cystic and non-cystic fibrosis patients. *J Med Microbiol*, 66, 490-
698 501.
- 699 Khan, S., Tondervik, A., Sletta, H., Klinkenberg, G., Emanuel, C., Onsoyen, E., Myrvold, R.,
700 Howe, R. A., Walsh, T. R., Hill, K. E. & Thomas, D. W. 2012. Overcoming drug
701 resistance with alginate oligosaccharides able to potentiate the action of selected
702 antibiotics. *Antimicrob Agents Chemother*, 56, 5134-41.
- 703 Lane, D. J. 1991. 16S/23S rRNA Sequencing. In: STACKEBRANDT, E. & GOODFELLOW,
704 M. (eds.) *Nucleic Acid Techniques in Bacterial Systematics*. New York: John Wiley and
705 Sons.
- 706 Lipuma, J. J. 2010. The changing microbial epidemiology in cystic fibrosis. *Clin Microbiol Rev*,
707 23, 299-323.
- 708 Mackenzie, T., Gifford, A. H., Sabadosa, K. A., Quinton, H. B., Knapp, E. A., Goss, C. H. &
709 Marshall, B. C. 2014. Longevity of patients with cystic fibrosis in 2000 to 2010 and
710 beyond: survival analysis of the Cystic Fibrosis Foundation patient registry. *Annals of*
711 *internal medicine*, 161, 233-241.
- 712 Mahboubi, M. A., Carmody, L. A., Foster, B. K., Kalikin, L. M., Vandevanter, D. R. & Lipuma,
713 J. J. 2016. Culture-Based and Culture-Independent Bacteriologic Analysis of Cystic
714 Fibrosis Respiratory Specimens. *Journal of Clinical Microbiology*, 54, 613.
- 715 Muhlebach, M. S., Zorn, B. T., Esther, C. R., Hatch, J. E., Murray, C. P., Turkovic, L.,
716 Ranganathan, S. C., Boucher, R. C., Stick, S. M. & Wolfgang, M. C. 2018. Initial
717 acquisition and succession of the cystic fibrosis lung microbiome is associated with
718 disease progression in infants and preschool children. *PLOS Pathogens*, 14,
719 e1006798.
- 720 Nelson, M. T., Pope, C. E., Marsh, R. L., Wolter, D. J., Weiss, E. J., Hager, K. R., Vo, A. T.,
721 Brittnacher, M. J., Radey, M. C., Hayden, H. S., Eng, A., Miller, S. I., Borenstein, E. &
722 Hoffman, L. R. 2019. Human and Extracellular DNA Depletion for Metagenomic
723 Analysis of Complex Clinical Infection Samples Yields Optimized Viable Microbiome
724 Profiles. *Cell Reports*, 26, 2227-2240.e5.
- 725 Oriano, M., Terranova, L., Teri, A., Sottotetti, S., Ruggiero, L., Tafuro, C., Marchisio, P.,
726 Gramegna, A., Amati, F., Nava, F., Franceschi, E., Cariani, L., Blasi, F. & Aliberti, S.
727 2019. Comparison of different conditions for DNA extraction in sputum - a pilot study.
728 *Multidisciplinary Respiratory Medicine*, 14, 6.
- 729 Ospina, R. & Ferrari, S. L. 2012. A general class of zero-or-one inflated beta regression
730 models. *Computational Statistics & Data Analysis*, 56, 1609-1623.

- 731 Peng, X., Li, G. & Liu, Z. 2016. Zero-Inflated Beta Regression for Differential Abundance
732 Analysis with Metagenomics Data. *Journal of computational biology : a journal of*
733 *computational molecular cell biology*, 23, 102-110.
- 734 Powell, L. C., Pritchard, M. F., Ferguson, E. L., Powell, K. A., Patel, S. U., Rye, P. D.,
735 Sakellakou, S.-M., Buurma, N. J., Brilliant, C. D., Copping, J. M., Menzies, G. E., Lewis,
736 P. D., Hill, K. E. & Thomas, D. W. 2018. Targeted disruption of the extracellular
737 polymeric network of *Pseudomonas aeruginosa* biofilms by alginate oligosaccharides.
738 *npj Biofilms and Microbiomes*, 4, 13.
- 739 Pritchard, M. F., Powell, L. C., Jack, A. A., Powell, K., Beck, K., Florance, H., Forton, J., Rye,
740 P. D., Dessen, A., Hill, K. E. & Thomas, D. W. 2017. A Low-Molecular-Weight Alginate
741 Oligosaccharide Disrupts Pseudomonas Microcolony Formation and Enhances
742 Antibiotic Effectiveness. *Antimicrob Agents Chemother*, 61.
- 743 Pritchard, M. F., Powell, L. C., Menzies, G. E., Lewis, P. D., Hawkins, K., Wright, C., Doull, I.,
744 Walsh, T. R., Onsøyen, E., Dessen, A., Myrvold, R., Rye, P. D., Myrset, A. H., Stevens,
745 H. N. E., Hodges, L. A., Macgregor, G., Neilly, J. B., Hill, K. E. & Thomas, D. W. 2016.
746 A New Class of Safe Oligosaccharide Polymer Therapy To Modify the Mucus Barrier
747 of Chronic Respiratory Disease. *Molecular Pharmaceutics*, 13, 863-872.
- 748 R-Core-Team 2013. R: A Language and Environment for Statistical Computing. Vienna,
749 Austria: R Foundation for Statistical Computing.
- 750 Rogers, G. B., Zain, N. M. M., Bruce, K. D., Burr, L. D., Chen, A. C., Rivett, D. W., Mcguckin,
751 M. A. & Serisier, D. J. 2014. A novel microbiota stratification system predicts future
752 exacerbations in bronchiectasis. *Annals of the American Thoracic Society*, 11, 496-
753 503.
- 754 Salter, S. J., Cox, M. J., Turek, E. M., Calus, S. T., Cookson, W. O., Moffatt, M. F., Turner, P.,
755 Parkhill, J., Loman, N. J. & Walker, A. W. 2014. Reagent and laboratory contamination
756 can critically impact sequence-based microbiome analyses. *BMC Biology*, 12, 87.
- 757 Sass, A. M., Schmerk, C., Agnoli, K., Norville, P. J., Eberl, L., Valvano, M. A. &
758 Mahenthiralingam, E. 2013. The unexpected discovery of a novel low-oxygen-
759 activated locus for the anoxic persistence of *Burkholderia cenocepacia*. *Isme j*, 7,
760 1568-81.
- 761 Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B.,
762 Lesniewski, R. A., Oakley, B. B., Parks, D. H., Robinson, C. J., Sahl, J. W., Stres, B.,
763 Thallinger, G. G., Van Horn, D. J. & Weber, C. F. 2009. Introducing mothur: Open-
764 Source, Platform-Independent, Community-Supported Software for Describing and
765 Comparing Microbial Communities. *Applied and Environmental Microbiology*, 75,
766 7537-7541.
- 767 Shankar, J. 2017. Insights into study design and statistical analyses in translational
768 microbiome studies. *Annals of Translational Medicine*, 5, 2.
- 769 Spilker, T., Baldwin, A., Bumford, A., Dowson, C. G., Mahenthiralingam, E. & Lipuma, J. J.
770 2009. Expanded multilocus sequence typing for *Burkholderia* species. *J Clin Microbiol*,
771 47, 2607-10.
- 772 Stanojevic, S. & Ratjen, F. 2016. Physiologic endpoints for clinical studies for cystic fibrosis. *J*
773 *Cyst Fibros*, 15, 416-23.
- 774 Surette, M. G. 2014. The Cystic Fibrosis Lung Microbiome. *Annals of the American Thoracic*
775 *Society*, 11, S61-S65.
- 776 Terranova, L., Oriano, M., Teri, A., Ruggiero, L., Tafuro, C., Marchisio, P., Gramegna, A.,
777 Contarini, M., Franceschi, E., Sottotetti, S., Cariani, L., Bevivino, A., Chalmers, J. D.,
778 Aliberti, S. & Blasi, F. 2018. How to Process Sputum Samples and Extract Bacterial
779 DNA for Microbiota Analysis. *International journal of molecular sciences*, 19, 3256.
- 780 Van Koningsbruggen-Rietschel, S., Davies, J. C., Pressler, T., Fischer, R., Macgregor, G.,
781 Donaldson, S. H., Smerud, K., Meland, N., Mortensen, J., Fosbøl, M. Ø., Downey, D.
782 G., Myrset, A. H., Flaten, H. & Rye, P. D. 2020. Inhaled dry powder alginate
783 oligosaccharide in cystic fibrosis: a randomised, double-blind, placebo-controlled,
784 crossover phase 2b study. *ERJ Open Research*, 6, 00132-2020.

785 Wang, Q., Garrity, G. M., Tiedje, J. M. & Cole, J. R. 2007. Naive Bayesian classifier for rapid
786 assignment of rRNA sequences into the new bacterial taxonomy. *Applied and*
787 *Environmental Microbiology*, 73, 5261-7.
788 Willner, D., Haynes, M. R., Furlan, M., Schmieder, R., Lim, Y. W., Rainey, P. B., Rohwer, F.
789 & Conrad, D. 2011. Spatial distribution of microbial communities in the cystic fibrosis
790 lung. *The ISME Journal*, 6, 471.
791 Winsor, G. L., Khaira, B., Van Rossum, T., Lo, R., Whiteside, M. D. & Brinkman, F. S. L. 2008.
792 The *Burkholderia* Genome Database: facilitating flexible queries and comparative
793 analyses. *Bioinformatics*, 24, 2803-2804.
794 Zemanick, E. T., Wagner, B. D., Sagel, S. D., Stevens, M. J., Accurso, F. J. & Harris, J. K.
795 2010. Reliability of quantitative real-time PCR for bacterial detection in cystic fibrosis
796 airway specimens. *PloS one*, 5, e15101-e15101.
797 Zlosnik, J. E., Zhou, G., Brant, R., Henry, D. A., Hird, T. J., Mahenthiralingam, E., Chilvers, M.
798 A., Wilcox, P. & Speert, D. P. 2015. *Burkholderia* species infections in patients with
799 cystic fibrosis in British Columbia, Canada. 30 years' experience. *Ann Am Thorac Soc*,
800 12, 70-8.

801

802

803

804

805 **Author contributions**

806

807 **Conceptualization:** EM, PDR, RW; **Methodology:** EM, PDR, RW; **Software:** RW;

808 **Validation:** EM, PDR, RW; **Formal analysis:** EM, PDR, RW; **Investigation:** RW;

809 **Resources:** EM, PDR; **Data Curation:** EM, PDR, RW; **Writing – original draft preparation:**

810 EM, RW; **Writing – review and editing:** EM, PDR, RW; **Visualization:** RW; **Supervision:**

811 EM, PDR; **Project administration:** EM, PDR, RW; **Funding acquisition:** EM, PDR

812

813 **Acknowledgements**

814 We would like to thank all the patients and their families who participated in this study, as well

815 as all clinical trial teams for their engagement in the trials (NCT02157922; NCT02453789).

816 The following investigators and their teams (in alphabetical order) were involved in the clinical

817 trials. Mary Carroll, General Hospital, Southampton, UK; Jane Davies, Royal Brompton and

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

839

840 **Role of the funding source**

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

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

848

849 **Declaration of competing interest**

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

853