Rogers, Geraint B.; van der Gast, Christopher J.; Serisier, David J. 2015. Predominant pathogen competition and core microbiota divergence in chronic airway infection. ISME Journal, 9 (1). 217-225. 10.1038/ismej.2014.124

Copyright © 2015 International Society for Microbial Ecology

This version available http://nora.nerc.ac.uk/507950/

NERC has developed NORA to enable users to access research outputs wholly or partially funded by NERC. Copyright and other rights for material on this site are retained by the rights owners. Users should read the terms and conditions of use of this material at http://nora.nerc.ac.uk/policies.html#access

This document is the author’s final manuscript version of the journal article following the peer review process. Some differences between this and the publisher’s version may remain. You are advised to consult the publisher’s version if you wish to cite from this article.

www.nature.com/

Contact CEH NORA team at noraceh@ceh.ac.uk

The NERC and CEH trademarks and logos (‘the Trademarks’) are registered trademarks of NERC in the UK and other countries, and may not be used without the prior written consent of the Trademark owner.
Predominant pathogen competition and core microbiota divergence in chronic airway infection

Geraint B Rogers1,2*, Christopher J van der Gast3, David J Serisier2,4

1 SAHMRI Infection and Immunity Theme, School of Medicine, Flinders University, Adelaide, Australia
2 Immunity, Infection, and Inflammation Program, Mater Research Institute, University of Queensland, and Translational Research Institute, Woolloongabba, Queensland, Australia
3 NERC Centre for Ecology and Hydrology, Wallingford, OX10 8BB, UK
4 Department of Respiratory Medicine, Mater Adult Hospital, South Brisbane, Australia

* For correspondence, E-mail: geraint.rogers@sahmri.com

Running title: Pathogen dominance in bronchiectasis microbiota

Keywords: chronic infection /microbiota divergence /respiratory microbiome /competitive exclusion /commonness and rarity /interspecific competition

Subject category: Microbial population and community ecology
Abstract

Chronic bacterial lung infections associated with non-cystic fibrosis bronchiectasis represent a substantial and growing healthcare burden. Where *Pseudomonas aeruginosa* is the numerically dominant species within these infections, prognosis is significantly worse. However, in many individuals, *Haemophilus influenzae* predominates, a scenario associated with less severe disease. The mechanisms that determine which pathogen is most abundant are not known. We hypothesised that the distribution of *H. influenzae* and *P. aeruginosa* would be consistent with strong interspecific competition effects. Further, we hypothesised that where *P. aeruginosa* is predominant, it is associated with a distinct 'accessory microbiota' that reflects a significant interaction between this pathogen and the wider bacterial community. To test these hypotheses, we analysed 16S rRNA gene pyrosequencing data generated previously from 60 adult bronchiectasis patients, whose airway microbiota was dominated by either *P. aeruginosa* or *H. influenzae*. The relative abundances of the two dominant species in their respective groups were not significantly different, and when present in the opposite pathogen group the two species were found to be in very low abundance, if at all. These findings are consistent with strong competition effects, moving towards competitive exclusion. Ordination analysis indicated that the distribution of the core microbiota associated with each pathogen, readjusted after removal of the dominant species, was significantly divergent (ANOSIM, \( R = 0.07, P = 0.019 \)). Taken together, these findings suggest that both interspecific competition and also direct and/or indirect interactions between the predominant species and the wider bacterial community, may contribute to the predominance of *P. aeruginosa* in a subset of bronchiectasis lung infections.
Introduction

The World Health Organization has reported that the global burden of diseases is shifting from communicable to non-communicable diseases, with chronic conditions such as heart disease, strokes, and lung diseases now being the chief causes of morbidity and mortality (Lopez et al., 2006). Among the great challenges in studying the causes and treatment of chronic lung infections, is a consequence of Koch’s postulates and the subsequent concept of infection pathogenesis summarized by the expression ‘one microbe, one disease’ (Nelson et al., 2013). Koch’s postulates have shaped our understanding of medical microbiology, as many important microbial diseases conform to them. However, that orthodoxy is being undermined by a growing recognition that diverse important diseases, including chronic lung infections, have a polymicrobial aetiology. This expanding understanding of chronic polymicrobial infections, originating from studies of cystic fibrosis airway microbiota, is beginning to be translated to other chronic lower respiratory diseases, including non-cystic fibrosis bronchiectasis (hereafter referred to as bronchiectasis).

Bronchiectasis is a chronic airway disease characterised by abnormal destruction and dilation of the large airways, bronchi and bronchioles (Cohen and Sahn, 1999). It is associated with chronic and frequently purulent expectoration, multiple exacerbations, and progressive dyspnoea that can become disabling (Ellis et al., 1981; Cohen and Sahn, 1999; Barker, 2002), and represents a substantial and growing healthcare burden. A recent US study demonstrated a marked increased prevalence in older populations varying from 4.2/100000 adults aged 18-34 years to 271.8/100000 aged ≥ 75 years (Weycker et al., 2005). Bronchiectasis often goes unrecognized or is misdiagnosed as asthma or chronic obstructive pulmonary disease (COPD), leading to an underestimated prevalence. Despite this, bronchiectasis is associated with substantial socioeconomic cost due to the frequent use of primary and secondary healthcare resources. An US epidemiological study of bronchiectasis-associated hospitalizations from 1993 to 2006 demonstrated an average annual hospitalization rate of 16.5/100000 population with a significant annual increase of 2.4% in men and 3.0% in women (Seitz et al., 2010), with the cost of managing bronchiectasis appearing to be rising (Joish et al., 2013).

Airway inflammation resulting from chronic bacterial infection is thought to be a significant contributory factor driving disease progression in bronchiectasis (Barker, 2002). The perceived importance of bacterial pathogens in airway disease progression is reflected in the use of antibiotics
(Serisier and Martin, 2011; Serisier et al., 2013a), both as maintenance therapy, and to treat episodes of acute exacerbation. However, our understanding of the mechanisms that underpin relationships between infection by particular bacterial taxa and clinical outcomes is currently poor. This situation undermines the development of rationales for the selection of particular antibiotic treatment regimes (Serisier, 2012) or potentially specific anti-inflammatory therapy (Visser et al., 2012), and achieving better insight into the manner in which treatments achieve beneficial outcomes. While bronchiectasis can result from a variety of recognised aetiologies, it is often considered idiopathic. Recent studies have revealed a substantial and diverse bacterial microbiota (Rogers et al., 2013b; Tunney et al., 2013; Rogers et al., 2014; van der Gast et al., 2014), which are typically dominated by either *Haemophilus influenzae* or *Pseudomonas aeruginosa*. Perhaps unsurprisingly given its colonisation of the upper airways in healthy individuals, *H. influenzae* is detectable in lower airway secretions from almost all bronchiectasis patients and is commonly the numerically dominant species (Rogers et al., 2013b; Rogers et al., 2014). In contrast, *P. aeruginosa* dominated infections occur in a smaller number of patients (Rogers et al., 2014), but, are associated with an accelerated decline in lung function, more frequent pulmonary exacerbations, greater sputum production, and a higher requirement for antibiotic therapy (Evans et al., 1996; Ho et al., 1998; Shoemark et al., 2007; Rogers et al., 2014).

A better understanding of the way in which *P. aeruginosa* interacts with the airway environment could provide important mechanistic insights into chronic infection in this patient group and in chronic respiratory infections more widely. Both the physiochemical characteristics of the airway environment, and the composition of the pre-existing lung microbiota, are likely to influence the likelihood of *P. aeruginosa* infection (Rogers et al., 2013a). Further, where *P. aeruginosa* dominates the infection microbiota, its growth is likely to further affect the composition of airway environment. This impact of colonisation could occur both directly through the metabolomic (Kozlowska et al., 2013) and secretomic (Bergamini et al., 2012) footprint of *P. aeruginosa*, and, in turn, indirectly by stimulating changes in the host immune response (Bergamini et al., 2012) and the activity of other co-colonising species (Bakkal et al., 2010; Tashiro et al., 2013). Whilst the causality in these interactions is difficult to demonstrate, were such relationships to exist, they would result in both an association between *P. aeruginosa* infection and measures of airway disease, and an association between *P. aeruginosa* infection and microbiota composition. The first of these associations has been well documented. However, to our knowledge, there have been no investigations to assess the second.
We hypothesised that (1) the distribution of *H. influenzae* and *P. aeruginosa* in airways samples would be consistent with strong interspecific competition effects; i.e. when *H. influenzae* is the dominant species in a bronchiectasis lung infection, the population size of *P. aeruginosa* will be negatively impacted and vice versa when *P. aeruginosa* is dominant. (2) Where *P. aeruginosa* or *H. influenzae* is dominant species in a bronchiectasis lung infection, they are associated with distinct ‘accessory microbiota’ that reflect a significant interaction between these pathogens and the wider bacterial community. To test these hypotheses, we analysed 16S rRNA gene pyrosequencing data, generated previously, from 60 adult bronchiectasis patients whose airway microbiota was dominated by either *P. aeruginosa* or *H. influenzae*. *H. influenzae*-dominated infections were chosen as a comparator group as they had been shown previously not to differ significantly in total bacterial load, dominant taxon relative abundance, or prior antibiotic burden (intravenous, oral, and combined), with those where *P. aeruginosa* was dominant (Rogers *et al.*, 2014). Further, despite differences in disease course, these patients did not differ significantly in serum C-reactive protein (CRP) levels, or sputum IL-8 and IL-1β levels (Rogers *et al.*, 2014) common markers of systemic and airway inflammation. To limit the potential effect of antibiotic therapy to influence microbiota composition (Serisier, 2013b), a sample set was chosen where there had been a four week period of clinical stability prior to collection, with no supplemental antibiotics administered (Serisier *et al.*, 2013c).

**Materials and methods**

The analysis performed here was based on 16S ribosomal RNA gene pyrosequencing data generated from induced sputum samples from adult patients with bronchiectasis, as part of the BLESS trial (Serisier *et al.*, 2013c). These data are available through the Sequence Read Archive (http://www.ncbi.nlm.nih.gov/sra) under the accession number SRP0356. Details of patient recruitment, sample collection, nucleic acid extraction, PCR amplification pyrosequencing, and bioinformatics processing have been published previously (Rogers *et al.*, 2013b; Serisier *et al.*, 2013c; Rogers *et al.*, 2014) and details of these processes are provided here as Supplementary Methods.

Of the 96 samples previously analysed, 26 were *P. aeruginosa*-dominated and 34 were *H. influenzae*-dominated. It is these 60 samples on which the analysis presented here is based. Patient details for these patients are shown in Table 1. As described previously, the species-level identification of *P. aeruginosa* and *H. influenzae* two species was confirmed using specific PCR-based
assays in all instances (Supplementary Methods), with identification by pyrosequencing treated as presumptive. Where species-level identities were not corroborated by specific assays, identities are presented at the genus level.

Bacterial taxa within each metacommunity were partitioned into core and satellite groups using the Poisson distribution test as previously described (van der Gast et al., 2011; Rogers et al., 2013c). One-way analysis of variance (ANOVA), regression analysis, coefficients of determination ($r^2$), residuals and significance ($P$) were calculated using Minitab software (version 16, Minitab, University Park, PA, USA) as described previously (van der Gast et al., 2011; Rogers et al., 2013c). Canonical correspondence analysis (CCA), analysis of similarity (ANOSIM), similarity of percentages (SIMPER) analysis were performed using the PAST (Palaeontological Statistics, version 2.17) program available from the University of Oslo website link (http://folk.uio.no/ohammer/past) run by Øyvind Hammer. Mann Whitney tests were performed using GraphPad Prism (version 5.01, La Jolla, CA 92037 USA). Where predominant taxa were removed prior to analysis, the remaining relative abundance measures were rescaled and expressed as percentages.

Results and Discussion

*P. aeruginosa* and *H. influenzae* share certain similarities, for example, they are both Gram negative, rod-shaped, facultative anaerobic Gammaproteobacteria. However, these common opportunistic pathogens are associated with very different clinical courses when dominant in bronchiectasis lung infections. Here the relative abundances of the two predominant species in their respective groups were high and not significantly different (ANOVA, $F_{1,58} = 0.096, P = 0.758$; *P. aeruginosa* mean abundance and standard deviation = 87.3 ± 13.4% and *H. influenzae* = 86.0 ± 17.9%) (Fig. 1). When present in the opposite dominated group the two species were found to be in only very low abundances (*P. aeruginosa* = 0.37 ± 1.3% and *H. influenzae* = 0.56 ± 0.77%; ANOVA, $F_{1,58} = 0.436, P = 0.511$). Furthermore, *P. aeruginosa* was not detected in 12 from 34 samples of the *H. influenzae* group, and 2 from 26 for *H. influenzae* in the *P. aeruginosa* group (Fig. 1). This is consistent with strong competition effects between the two species, moving towards competitive exclusion of the inferior competitor species; and more so for *P. aeruginosa* within the *H. influenzae* group. If these patterns of dominance and suppression could be purely explained by the process of interspecies competition then no between group differences in accessory microbiota would be
expected. However, the analyses performed here demonstrate that taxa present in the microbiota associated with *P. aeruginosa* and *H. influenzae* predominance are significantly divergent.

The distribution of the two sets of microbiota, as determined by direct ordination using Bray-Curtis similarity measures, is shown in Fig. 2. Where the dominant taxa (*P. aeruginosa* or *H. influenzae*) were included (Fig. 2a), divergence in the distribution of the microbiota was pronounced and statistically significant (ANOSIM, $R = 1, P < 0.0001$). However, given the high proportion of total bacterial abundance that these predominant taxa would account for within the microbiota (Fig. 1), much of the variation between the two groups will result from their inclusion in the analysis. In order to assess whether accessory microbiota composition differed significantly between the two groups, the predominant taxa were removed and the relative abundances of the remaining taxa redistributed and expressed as percentages. When subjected to ordination analysis, the distribution of the accessory microbiota composition (Fig. 1b) were not found to be significantly divergent (ANOSIM, $R = 0.036, P = 0.11$).

Many accessory microbiota taxa are of low relative abundance. The potential contribution to accessory microbiota of transient bacterial populations within the airways, as opposed to populations of chronically infective bacteria, is therefore high. In order to reduce the effect of these satellite taxa, the core microbiota (composed of non-randomly distributed taxa) in each of the two groups was determined (Fig. 3). This is an approach that has been applied successfully in the analysis of chronic bacterial infections associated with cystic fibrosis (van der Gast *et al.*, 2011; Rogers *et al.*, 2013a). In the *Haemophilus*-dominated group, 9 of the 92 taxa detected were classified as core and 83 as satellite, and in the *Pseudomonas*-dominated group, 8 of the 70 taxa detected were classified as core and 62 as satellite. In each case, the contribution of individual taxa to the core microbiota was assessed by SIMPER analysis (Tables 2 & S1). Ordination analysis was then performed using the core taxa, and again the difference between the distribution of the core microbiota that included the predominant *P. aeruginosa* and *H. influenzae* populations was significant (ANOSIM, $R = 1, P < 0.0001$) (Fig.4a). Moreover, the divergence in the distribution of the core microbiota, readjusted after removal of the dominant species, was also found to be significantly divergent (ANOSIM, $R = 0.07, P = 0.019$) (Fig. 4b). Satellite taxa were not significantly different between the groups (ANOSIM, $R = 0.05, P = 0.06$).
To assess whether significant differences existed in the relative abundances of specific core taxa between the *P. aeruginosa* - and *H. influenzae*-dominated samples, Mann-Whitney tests were used and performed on readjusted core taxa abundance data after *P. aeruginosa* and *H. influenzae* are removed. This process identified *Prevotella* spp. and *Flavobacterium* spp. as being significantly more abundant in the *P. aeruginosa*-dominated samples (*P* < 0.0001 and *P* = 0.003, respectively), while *Neisseria* spp. was significantly more abundant in *H. influenzae*-dominated samples (*P* < 0.0001). *Flavobacterium* is a genera that has been reported previously to contribute to bacterial communities present in chronic lung infections (Rogers et al., 2003; Rogers et al., 2004; van der Gast et al., 2011; Rogers et al., 2013a) although typically present at low relative abundances. In contrast, *Prevotella* spp. have been reported as both common, and often at high abundance in both bronchiectasis and CF lung infections (Tunney et al., 2008; Field et al., 2010; Stressmann et al., 2012), and otitis media (Brook, 2008). *Prevotella* spp. appear to be particularly prevalent when co-colonising with *P. aeruginosa*, a factor that has led to the previous suggestion that a synergistic relationship exists between *P. aeruginosa* and members of this genus (Su and Hassett, 2012). The genus *Prevotella* is composed of species that are obligate anaerobes. While *P. aeruginosa* and *H. influenzae* are both capable of fermentation and growth under anaerobic conditions (Schobert and Jahn, 2010; Langereis and Hermans, 2013), the contributions of thick mucoid secretions that *P. aeruginosa* can produce in the airways (Ma et al., 2012), a trait not demonstrated by *H. influenzae* (Langereis and Hermans, 2013), may contribute to reduced oxygen permeation, leading to the creation of greater opportunities for the growth of strict anaerobes, such as *Prevotella* spp. This model would be consistent with the association observed here between *H. influenzae* and *Neisseria*, a genus of typically aerobic species.

Canonical correspondence analysis was performed next to assess the extent to which variance in the microbiota distribution can be accounted for by variation in measures of disease severity, the presence of comorbidities, and non-antibiotic therapies. The results of these analyses are shown in Table 3, and additionally superimposed onto Figs 2 & 4. In keeping with previous reports (Rogers et al., 2014), the presence of *P. aeruginosa* in samples as the predominant taxon was associated with high pulmonary exacerbation frequency and poor lung function (low FEV₁ percent predicted), with these factors varying significantly with microbiota distribution. However, whilst such clinical measures are associated with the presence of *P. aeruginosa*, these analyses also show that a
significant relationship exists with the wider airway microbiota; a significant relationship was identified
here between the variance in core taxa and Leicester Cough Score (a measure of cough symptom
severity).

Here we observed distributions of predominant taxa consistent with strong interspecific
competition, supporting competitive exclusion in some instances. But we also observed, for instance,
that where \textit{P. aeruginosa} is the numerically dominant species in a bronchiectasis lung infection it is
associated with a distinct accessory microbiota; suggesting in addition to interspecific competition
there are also direct and/or indirect interactions between the predominant species and the core
microbiota. To some extent, such an effect was also observed for \textit{H. influenzae}, but was far less
pronounced. However, several different models could explain such interactions. For example, the
predominant species could influence the accessory microbiota composition through modification of the
airway environment and alteration of its selective properties (and vice versa); here, perhaps the fact
that \textit{H. influenzae} is a common resident of the upper airways means that its presence is less
disruptive to the commonly occurring infective lower airway microbiota. Alternatively, the same change
in the characteristics of the airway could occur through intermediary interaction with the host that
results in an altered inflammatory profile, airway secretion composition, or secretion clearance rate.

Here, a wide array of virulence factors and pro-inflammatory traits possessed by \textit{P. aeruginosa}
(Sadikot \textit{et al.}, 2005) may contribute to the magnitude of the effect of its predominance on the
accessory airway microbiota. Finally, external influences, such as antibiotic therapy, are likely to
contribute to selective pressures in the airway environment. While there was no significant difference
in historical antibiotic burden in the \textit{H. influenzae}- and \textit{P. aeruginosa}-dominated groups, the
contribution of more subtle differences in treatment history cannot be excluded. We suggest that no
single process is responsible for the associations observed, and rather a dynamic interaction between
many different factors give rise to the various types of microbiological scenario seen \textit{in vivo}. The
potential complexity of these interactions makes their elucidation challenging. However, discerning
their basis is important given that a number of important clinical questions arise from the findings we
present here. These questions include; (1) Could accessory microbiota composition predict
subsequent \textit{P. aeruginosa} predominance and its associated poor prognosis? (2) Could intervention
aimed at altering the characteristics of the airway environment, or the composition of the accessory
microbiota, reduce the likelihood of \textit{P. aeruginosa} infection and predominance? (3) What are the
mechanisms to promote the dominance of *H. influenzae*, and as a consequence competitively exclude or suppress *P. aeruginosa*?

In each case, addressing these questions will require *in vitro* competition experiments between the two dominant bacterial species and members of the core microbiota that we have identified as having likely interactions with *P. aeruginosa* and *H. influenzae*. In addition, longitudinal sample sets that span both clinically and microbiologically important time periods will allow us to better understand the *in vivo* mechanisms that lead to predominance of *H. influenzae* or *P. aeruginosa* and its associated worse clinical outcomes. However, obtaining informative longitudinal sample sets in adult bronchiectasis has significant challenges. Intensive, long-term sample collection would be required to span rare and unpredictable events, such as the acquisition of *P. aeruginosa* in a condition that has relatively slow progression (Martínez-García *et al.*, 2007) and is commonly idiopathic (Anwar *et al.*, 2013). Given their potential to provide mechanistic insight into the relationship between recognised airway pathogens, the wider airway microbiota, host immunity, and clinical outcome, such long-term frequent sample collection represents an important next step.

In conclusion, we present evidence supporting the contribution of both interspecific competition, and direct and/or indirect interaction between predominant infective taxa and the wider bacterial community, to determining whether *H. influenzae* or *P. aeruginosa* dominates the chronic lung infections associate with bronchiectasis. Given the prognostic implications of *P. aeruginosa* dominance, these findings provide a basis for identifying the mechanisms that underpin this airway microbial ecology, and perhaps offering novel therapeutic opportunities.

**Acknowledgements**

This study was supported by the Mater Adult Respiratory Research Trust Fund.

**Conflicts of interest**

The authors declare no conflict of interest.

Supplementary information is available at ISMEJ's website


Figure legends

Fig. 1. Relative percentage abundances of *P. aeruginosa* (green circles) and *H. influenzae* (blue) in samples from within *P. aeruginosa* - and *H. influenzae*-dominated groups (P1-P26 and H1-H34, respectively).

Fig. 2. Canonical correspondence biplots for microbiota (a) with and (b) without *Pseudomonas* and *Haemophilus* included. Dots represent microbiota samples from the *Pseudomonas* (denoted with green filled circles) and *Haemophilus* (blue filled squares) groups. In each instance, the 95 % concentration ellipses are given for the *Pseudomonas* (green) and *Haemophilus* (blue) groups. Biplot lines for the clinical variables included in the analyses show the direction of increase for each variable, and the length of each line indicates the degree of correlation with the ordination axes. CCA field labels: “AH” – antihypertensive, “LABA/ICS” – long acting β-agonist, “SABA” – short acting β-agonist, “prior IE” – number of pulmonary exacerbations in the prior twelve months. Percentage of community variation explained by each axis is given in parentheses.

Fig. 3. Distribution and dispersal of bacterial taxa among *Haemophilus*- and *Pseudomonas*-dominated microbiota samples. (a & b) The number of samples for which each detected bacterial taxon (open circles) was observed, plotted against the abundance (log_{10} scale) of that species among all samples within each group ((a) *Haemophilus* group, $r^2 = 0.27$, $F_{1,92} = 33.2$, $P < 0.0001$; and (b) *Pseudomonas* group, $r^2 = 0.33$, $F_{1,68} = 33.5$, $P < 0.0001$). Also given are dispersal plots to identify which bacterial taxa are randomly distributed within the (c) *Haemophilus* and (d) *Pseudomonas* groups; a measure used to assign core versus satellite status. Index of dispersion was calculated as the ratio of variance to mean of abundance for each taxon within each group and plotted for each sample. The line depicts the 2.5 % confidence limit for the $\chi^2$ distribution. Taxa that fall below this line are randomly distributed and were considered satellite taxa, whereas those that are above the line are non-randomly distributed and were considered core taxa. The 97.5 % confidence limit was not plotted, as no taxon fell below that line.

Fig. 4. Canonical correspondence biplots for core microbiota (a) with and (b) without *Pseudomonas* and *Haemophilus* included. Dots represent core microbiota from the *Pseudomonas* (denoted with
green filled circles) and *Haemophilus* (blue filled squares) groups. In each instance, the 95% concentration ellipses are given for the *Pseudomonas* (green) and *Haemophilus* (blue) groups. Biplot lines for the clinical variables included in the analyses show the direction of increase for each variable, and the length of each line indicates the degree of correlation with the ordination axes. CCA field labels: “AH” – antihypertensive, “LCS” – Leicester cough score, “prior IE” – number of pulmonary exacerbations in the prior twelve months. Percentage of community variation explained by each axis is given in parentheses.
Table 1. Clinical, treatment and comorbidity data for patient population. FEV$_1$%* – forced expiratory volume in one second, expressed as a percentage of predicted and measured following administration of a bronchodilator; SGRQ, St George’s Respiratory Questionnaire – range 1-100, lower scores indicate better quality of life. Leicester cough score – lower scores indicate worse cough symptoms.

<table>
<thead>
<tr>
<th>Clinical measures of disease</th>
<th>P. aeruginosa</th>
<th>H. influenzae</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender (male:female)</td>
<td>6:20</td>
<td>16:18</td>
</tr>
<tr>
<td>FEV1%*</td>
<td>30.7-94.2</td>
<td>37.2-114.7</td>
</tr>
<tr>
<td>Duration of bronchiectasis (years)</td>
<td>1-70</td>
<td>10-65</td>
</tr>
<tr>
<td>Pulmonary exacerbations in prior 12 months</td>
<td>2-12</td>
<td>2-7</td>
</tr>
<tr>
<td>Leicester cough score</td>
<td>5.5-20.2</td>
<td>7.7-19.9</td>
</tr>
<tr>
<td>SGRQ total</td>
<td>8.6-79.6</td>
<td>14.9-58.9</td>
</tr>
<tr>
<td>Six minute walk test</td>
<td>291-650</td>
<td>275-710</td>
</tr>
<tr>
<td>C-reactive protein (mg/litre)</td>
<td>0-21</td>
<td>0-19</td>
</tr>
<tr>
<td>Induced sputum IL-8 (ng/ml)</td>
<td>27.4-1053.8</td>
<td>28.8-1326.7</td>
</tr>
<tr>
<td>Induced sputum IL-1β (ng/ml)</td>
<td>0.25-36.2</td>
<td>0.6-115.8</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Treatment</th>
<th>P. aeruginosa</th>
<th>H. influenzae</th>
</tr>
</thead>
<tbody>
<tr>
<td>Short acting β-agonist</td>
<td>14</td>
<td>10</td>
</tr>
<tr>
<td>Inhaled corticosteroid</td>
<td>16</td>
<td>17</td>
</tr>
<tr>
<td>Inhaled corticosteroid + long acting β-agonist</td>
<td>13</td>
<td>10</td>
</tr>
<tr>
<td>Anti-cholinergics</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>Aspirin</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>Beta blocker</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Nasal steroids</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Prednisolone</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Antihypertensive</td>
<td>9</td>
<td>16</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Comorbidity</th>
<th>P. aeruginosa</th>
<th>H. influenzae</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cerebrovascular disease</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Heart disease</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Hypertension</td>
<td>6</td>
<td>14</td>
</tr>
<tr>
<td>Diabetes</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>
**Table 2.** Similarity of percentages (SIMPER) analysis of bacterial community dissimilarity (Bray-Curtis) between core taxa groups without *Pseudomonas* and *Haemophilus*. Given is mean % abundance of sequences for each taxon across the samples they were observed to occupy. Analysis was based on an average of 10647 sequences per sample (SD 5070, range 2395-28916). Also given is the average dissimilarity between samples (overall mean = 61.9%). Percentage contribution is the mean contribution divided by mean dissimilarity across samples. SIMPER analysis with *Pseudomonas* and *Haemophilus* is presented in Table S1.

<table>
<thead>
<tr>
<th>Taxon</th>
<th>% Mean abundance</th>
<th>Samples detected in</th>
<th>Average dissimilarity</th>
<th>% Contribution</th>
<th>Cumulative %</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pseudomonas group</td>
<td>Haemophilus group</td>
<td>Pseudomonas group</td>
<td>Haemophilus group</td>
<td></td>
</tr>
<tr>
<td>Prevotella</td>
<td>31.9</td>
<td>31</td>
<td>23</td>
<td>27</td>
<td>14.96</td>
</tr>
<tr>
<td>Veillonella</td>
<td>20.8</td>
<td>26.7</td>
<td>25</td>
<td>30</td>
<td>11.85</td>
</tr>
<tr>
<td>Streptococcus</td>
<td>13.2</td>
<td>18.3</td>
<td>22</td>
<td>26</td>
<td>10.51</td>
</tr>
<tr>
<td>Moraxella</td>
<td>14</td>
<td>4.9</td>
<td>10</td>
<td>5</td>
<td>8.69</td>
</tr>
<tr>
<td>Neisseria</td>
<td>7.8</td>
<td>13.6</td>
<td>18</td>
<td>25</td>
<td>7.59</td>
</tr>
<tr>
<td>Flavobacterium</td>
<td>8.7</td>
<td>0</td>
<td>17</td>
<td>0</td>
<td>4.35</td>
</tr>
<tr>
<td>Leptotrichia</td>
<td>3.6</td>
<td>2.0</td>
<td>16</td>
<td>18</td>
<td>2.22</td>
</tr>
<tr>
<td>Fusobacterium</td>
<td>0</td>
<td>3.5</td>
<td>0</td>
<td>11</td>
<td>1.72</td>
</tr>
</tbody>
</table>
Table 3. Canonical correspondence analyses for determination of percent variation in lung microbiota from bronchiectasis subjects by clinical variables. LABA, long acting β-agonist; ICS, inhaled corticosteroid; SABA, short acting β-agonist.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Whole microbiota % of variance</th>
<th>Probability</th>
<th>Whole microbiota without P &amp; H % of variance</th>
<th>Probability</th>
<th>Core microbiota % of variance</th>
<th>Probability</th>
<th>Core microbiota without P &amp; H % of variance</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prior Exacerbations</td>
<td>11.35</td>
<td>0.01</td>
<td>2.35</td>
<td>0.02</td>
<td>13.8</td>
<td>0.01</td>
<td>2.7</td>
<td>0.01</td>
</tr>
<tr>
<td>FEV1 % predicted</td>
<td>4.60</td>
<td>0.01</td>
<td>2.41</td>
<td>0.01</td>
<td>5.5</td>
<td>0.01</td>
<td>4.9</td>
<td>0.01</td>
</tr>
<tr>
<td>Hypertension</td>
<td>3.08</td>
<td>0.01</td>
<td>-</td>
<td>-</td>
<td>3.4</td>
<td>0.01</td>
<td>2.7</td>
<td>0.01</td>
</tr>
<tr>
<td>Gender</td>
<td>3.01</td>
<td>0.01</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Anti-Hypertensive</td>
<td>-</td>
<td>-</td>
<td>2.20</td>
<td>0.01</td>
<td>-</td>
<td>-</td>
<td>2.3</td>
<td>0.01</td>
</tr>
<tr>
<td>LABA + ICS</td>
<td>-</td>
<td>-</td>
<td>1.81</td>
<td>0.02</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>SABA</td>
<td>-</td>
<td>-</td>
<td>1.54</td>
<td>0.02</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Leicester cough</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>3.1</td>
<td>0.02</td>
</tr>
</tbody>
</table>
Table S1. Similarity of percentages (SIMPER) analysis of bacterial community dissimilarity (Bray-Curtis) between core taxa groups with *Pseudomonas* and *Haemophilus*. Given is mean % abundance of sequences for each taxon across the samples they were observed to occupy. Analysis was based on an average of 10647 sequences per sample (SD 5070, range 2395-28916). Also given is the average dissimilarity between samples (overall mean = 61.9%). Percentage contribution is the mean contribution divided by mean dissimilarity across samples.

<table>
<thead>
<tr>
<th>Taxon</th>
<th>% Mean abundance</th>
<th>Samples detected in</th>
<th>Average dissimilarity</th>
<th>% Contribution</th>
<th>Cumulative %</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pseudomonas group</td>
<td>Haemophilus group</td>
<td>Pseudomonas group</td>
<td>Haemophilus group</td>
<td></td>
</tr>
<tr>
<td>Pseudomonas</td>
<td>87.3</td>
<td>0.37</td>
<td>26</td>
<td>22</td>
<td>44.44</td>
</tr>
<tr>
<td>Haemophilus</td>
<td>0.56</td>
<td>86</td>
<td>24</td>
<td>34</td>
<td>43.93</td>
</tr>
<tr>
<td>Prevotella</td>
<td>2.61</td>
<td>4.14</td>
<td>23</td>
<td>27</td>
<td>2.51</td>
</tr>
<tr>
<td>Streptococcus</td>
<td>1.51</td>
<td>2.35</td>
<td>22</td>
<td>26</td>
<td>1.69</td>
</tr>
<tr>
<td>Veillonella</td>
<td>1.44</td>
<td>2.8</td>
<td>25</td>
<td>30</td>
<td>1.67</td>
</tr>
<tr>
<td>Flavobacterium</td>
<td>2.38</td>
<td>0</td>
<td>17</td>
<td>0</td>
<td>1.21</td>
</tr>
<tr>
<td>Moraxella</td>
<td>1.18</td>
<td>0.54</td>
<td>10</td>
<td>5</td>
<td>0.83</td>
</tr>
<tr>
<td>Neisseria</td>
<td>0.74</td>
<td>0.97</td>
<td>18</td>
<td>25</td>
<td>0.74</td>
</tr>
<tr>
<td>Leptotrichia</td>
<td>0.57</td>
<td>0.20</td>
<td>16</td>
<td>18</td>
<td>0.35</td>
</tr>
<tr>
<td>Fusobacterium</td>
<td>0</td>
<td>0.45</td>
<td>0</td>
<td>11</td>
<td>0.23</td>
</tr>
</tbody>
</table>
Supplementary Methods

Samples were collected from participants in the Bronchiectasis and Low-dosen Erythromycin Study (BLESS) trial (Serisier et al., 2013). Adult patients aged 20 to 85 years were eligible if they had bronchiectasis documented by high-resolution computed tomographic scan, at least 2 separate pulmonary exacerbations requiring supplemental systemic antibiotic therapy in the preceding 12 months, and daily sputum production. Participants were required to have been clinically stable for at least 4 weeks prior to enrolment (defined as no symptoms of exacerbation, no requirement for supplemental antibiotic therapy, and forced expiratory volume in the first second of expiration [FEV₁] within 10% of best recently recorded value where available). Exclusion criteria included CF, current mycobacterial disease or bronchopulmonary aspergillosis, any reversible cause for exacerbations, maintenance oral antibiotic prophylaxis, prior macrolide use except short-term, changes to medications in the preceding 4 weeks, cigarette smoking within 6 months, and medications or comorbidities with the potential for important interactions with erythromycin. All participants required negative results from sputum mycobacterial cultures prior to randomization. The study was approved by the Mater Health Service human research ethics committee, and all participants provided written, informed consent.

Sample collection

Subjects were instructed to perform their usual chest physiotherapy regime on the morning of the sputum induction procedure. Prior to commencement of hypertonic saline inhalation, any spontaneous sputum expectorated was collected for standard culture. Sputum induction (SI) was performed after inhalation of 400 ug of albuterol, using 4.5% hypertonic saline nebulised from an ultrasonic nebuliser (output >1 mL/min) for 20 minutes in 4 periods of 5 minutes each, according to the standardised protocol recommended by the European Respiratory Society taskforce [Paggiaro et al., 2002] Following mouth-rinsing and expectoration, sputum was collected following each nebulisation period, on each occasion preceded immediately by spirometry. The first sputum sample was refrigerated immediately following collection and frozen at -80 °C within an hour. A cold chain was maintained up until the point of DNA extraction.

DNA extraction protocol
Nucleic acid extractions were performed on 500 μL cell suspensions. Guanidinium thiocyanate–EDTA–sarkosyl (500 μL) and PBS (500 μL), pH 8.0, were added to samples. Cell disruption was achieved using a Fastprep-24 Instrument (MP Biomedicals Europe, Illkirch, France) 6.5 m/s, 60 s, followed by incubation at 90 °C for 1 min and −20 °C for 5 min. Cell debris was pelleted by centrifugation at 12 000 × g for 2 min at 4 °C. Supernatant was transferred to a fresh microfuge tube. NaCl (to a final concentration of 0.5 mol/L) and polyethylene glycol (to a final concentration of 15%) were added and DNA precipitated at 4 °C for 30 min. DNA was pelleted by centrifugation at 12 000 × g for 2 min at 4 °C and resuspended in 300 μL of sterile distilled water. Samples were heated at 90 °C for 30 s and vortexed. Phenol/chloroform (1:1) (300 μL) was added, and samples were vortexed for 20 s before centrifugation at 12 000 × g at 4 °C for 3 min. The upper phase was then transferred to a fresh microfuge tube. Total DNA was then precipitated by the addition of an equal volume of isopropanol, a 0.1-volume 10 mol/L ammonium acetate, and 1 μL of GenElute linear polyacrylamide (Sigma-Aldrich, Gillingham, UK) and incubated at −20 °C for 25 min. DNA was pelleted by centrifugation at 12 000 × g at 4 °C for 5 min. Pelleted DNA was then washed 3 times in 70% ethanol, dried, and resuspended in 50 μL of sterile distilled water. DNA extracts were quantified using the Picodrop Microlitre Spectrophotometer (GRI, Braintree, UK). Negative controls, consisting of sterile water, were included in the PMA treatment, DNA extraction, and PCR amplification steps.

Pyrosequencing

Bacterial tag-encoded FLX amplicon pyrosequencing (bTEFAP) was performed as described previously using Gray28F 5'-TTTGATCNTGGCTCAG-3' and Gray519r 5'-GTNTTACNGCGGCKGCTG-3'). Initial generation of the sequencing library involved a one-step PCR of 30 cycles, using a mixture of Hot Start and HotStar high fidelity Taq DNA polymerase, as described previously (4). Tag-encoded FLX amplicon pyrosequencing analyses utilized Roche 454 FLX instrument with Titanium reagents, titanium procedures performed at the Research and Testing Laboratory (Lubbock, TX) using RTL protocols (www.researchandtesting.com).

Following sequencing, all failed sequence reads, low quality sequence ends and tags and primers were removed. Sequences with ambiguous base calls, sequences with homopolymers > 6bp were removed. Further, any non-bacterial ribosomal sequences and chimeras using B2C2 (Dowd et al., 2008) as described previously (Dowd et al., 2008). To determine the identity of bacterial species in the
remaining sequences, sequences were de-noised, assembled into OUT clusters at 97% identity, and
queried using a distributed .NEt algorithm that utilizes Blastn+ (KrakenBLAST www.krakenblast.com)
against a database of high quality 16S rRNA gene bacterial sequences. Using a .NET and C# analysis
pipeline the resulting BLASTn+ outputs were compiled, data reduction analysis performed, and
sequence identity classification carried out, as described previously (Dowd et al., 2008).

Quantitative PCR

*P. aeruginosa* density was determined using a Taqman assay, in which a 117 bp region between
positions 330 to 447 of the *P. aeruginosa* OprL gene was amplified, as described previously
(Feizabadi et al., 2010). Primers (PsF: 5’-CGAGTACAACATGGCTCTGG-3’, EubR: 5’-
ACCGGACGCTTTACCATA-3’) were used at a concentration of 500 nM each, and the probe
(EubPr: 5’-FAM- CCTGCAGCACCAGGTAGCGC -TAMRA-3’) at a concentration of 250 nM. All
reactions were carried out in a total volume of 20 μl containing primers at a concentration of 500 nM
each, probe concentration of 250 nM, 1 μl of template and LightCycler 480 Probes Master (Roche
Diagnostics GmbH, Mannheim, Germany) at 1x final concentration. Quantitative PCR assays were
carried out using the Rotor-Gene Q (Qiagen, Crawley, UK) with a temperature profile of 95 °C for 5
min, followed by 45 cycles at 95 °C for 15 s and 58 °C for 45 s. Nutrient broth culture of *P. aeruginosa*
(NCTC 12934/ATCC 27853) was incubated at 37 °C for 16 h, with cfu/ml estimated by incubation of
dilutions (n=4) on Nutrient agar at 37° for 24 h, followed by colony counts. DNA was extracted from
tenfold dilutions of the broth culture in the same way as for the sputum samples, and RT-PCR was
carried out as above on the DNA extracts. The standard curve generated using *P. aeruginosa* (qPCR
efficiency = 1.20; R² value = 0.996) was used as reference to allow direct comparisons to be made by
the Rotor Gene Q-series Software (Qiagen, Crawley, UK).

*H. influenzae* density was determined using a Taqman assay, in which a 90-bp region between
positions 518 to 608 of the *H. influenzae* Hel gene was amplified, using primers (HelSF: 5’-
CCGGGTGCAGTAGATTTAATA-3’, EubR: 5’-CTGATTTCAGTGCTTCTTTGC-3’) were used at
a concentration of 100 nM each, and the probe (EubPr: 5’-FAM-
ACAGCCACAACGGTAAAGTTCTACG-TAMRA-3’) (Long, 2011). All PCR reactions were carried
out in a total volume of 20 μl containing primers and probe at a concentration of 1000:500:200 nM
(HelSF: HelSR; HelSPr), 1 μl of template and LightCycler 480 Probes Master (Roche Diagnostics
GmbH, Mannheim, Germany) at 1x final concentration. Quantitative PCR assays were carried out using the Rotor-Gene Q (Qiagen, Crawley, UK) with a temperature profile of 95 °C for 5 min, followed by 45 cycles at 95 °C for 15 s and 60 °C for 60 s. A standard curve was generated by qPCR settings as above using *H. influenzae* DNA extract obtained from the Health Protection Agency (*H. influenza* clinical isolate, chronic respiratory infection, 2012). *H. influenza* load (cfu/ml) was determined by direct comparison with the standard curve (qPCR efficiency = 1.05; R² value = 0.997) using the Rotor Gene Q-series Software (Qiagen, Crawley, UK).

1. Developed in house at the Health Protection Agency London by Dr. Clare Ling, 2011.