Microbiome diversity in the bronchial tract
of patients with chronic obstructive pulmonary disease

Raúl Cabrera-Rubio1, Marian García-Núñez2,3, Laia Setó2,3, Josep M. Antó6,7,8,9, Andrés Moya1,8, Eduard Monsó2,3,4,5,*, Alex Mira1

1 Dept. of Genomics and Health, Center for Public Health Research (CSISP), Avda Cataluña 21, Valencia 46020 (Spain)
2 CIBER de Enfermedades Respiratorias – CIBERes, Bunyola, Mallorca, Illes Balears, Spain
3 Fundació Institut Universitari Parc Taulí, Sabadell, Spain
4 Pulmonary Department, Hospital Universitari Parc Taulí, Sabadell, Spain
5 Departament de Medicina, Universitat Autònoma de Barcelona, Bellaterra, Spain
6 Centre for Research in Environmental Epidemiology (CREAL), Barcelona, Spain
7 IMIM (Hospital del Mar Research Institute), Barcelona, Spain
8 CIBER de Epidemiología y Salud Pública (CIBEResp), Barcelona, Spain
9 Universitat Pompeu Fabra (UPF), Barcelona, Spain

* Corresponding author. E-mail: emonso@tauli.cat

Running title: The Bronchial Tree Microbiome
Abstract

Culturing of bacteria from bronchial secretions in respiratory patients has low sensitivity and does not allow for complete assessment of microbial diversity across the different bronchial compartments. In addition, a significant number of clinical studies are based on sputum samples and it is not known to what extent they describe the real diversity of the mucosa. In order to identify previously unrecognized lower airway bacteria and to investigate the complexity and distribution of microbiota in patients with chronic obstructive pulmonary disease (COPD) we have performed PCR amplification and pyrosequencing of the 16S rRNA gene in patients not showing signs or symptoms of infection. Four types of respiratory samples (sputum, bronchial aspirate, bronchoalveolar lavage and bronchial mucosa) were taken from each individual, obtaining on average over 1,000 16S rRNA sequences per sample. Total number of genera per patient was over 100, showing a high diversity, being Streptococcus, Prevotella, Moraxella, Haemophilus, Acinetobacter, Fusobacterium and Neisseria the most often identified genera. Sputum samples showed significantly lower diversity than the other three sample types. Lower bronchial tree samples -bronchoalveolar lavage and bronchial mucosa-, showed a very similar bacterial composition, different from sputum and bronchoaspirate. Thus, sputum and bronchoaspirate samples are upper bronchial tree samples not representative of the lower bronchial mucosa flora, being bronchoalveolar lavage the sample that shows the closest results to bronchial mucosa. Our data confirms that the bronchial tree is not sterile in COPD patients and support the existence of a different microbiota in its upper and lower compartments.

Introduction

The bronchial tree and the pulmonary parenchyma are considered sterile in healthy subjects, but potentially pathogenic microorganisms (PPM) are often recovered from bronchial secretions in patients with chronic obstructive pulmonary disease (COPD) during periods of clinical stability and, particularly, during exacerbations, when bacterial loads increase significantly (37, 46). When sputum has been used for the identification of bronchial colonization by PPM in stable COPD, positive cultures have
been found in one-fifth to three quarters of the patients, in most cases to a single microorganism (38, 16). Cultures positive for PPM have been found in one-third of the patients with COPD in the absence of symptoms of bronchial infection when lower bronchial secretions have been sampled avoiding the oropharynx through the use of a protected specimen brush for the collection of the specimens under sterile conditions, being *H. influenzae*, *S. pneumoniae* and *M. catarrhalis* the bacteria most often recovered (37), a finding that confirms the presence of these microorganisms in the lower bronchial tree of COPD patients. Bronchial colonization has been also demonstrated in bronchoalveolar lavage samples obtained from COPD patients, a sample that recovered peripheral secretions from the bronchial tree (8, 42, 39). Most PPM cultured from stable COPD patients show low microbial loads (<10000 colony-forming units per mL), however, and are not associated with a neutrophilic inflammatory response (45, 31), with high loads being found almost exclusively when *H. influenzae* is the colonizing bacteria (25, 26). These findings confirm that colonizing PPM may be found in the bronchial tree of some patients with COPD, in most cases as a single culture. The mechanisms behind the recovery of low-load PPM from bronchial secretions in patients with COPD in the absence of signs and symptoms of infection is open to debate, as it may be related to oropharyngeal bacteria that migrated to the bronchial tree or to flora colonizing specifically the lower bronchial tree (30, 44).

Culture-based techniques underdiagnose bronchial colonisation at loads below the detection limit of the sputum culture, and PPM have been identified in one tenth of the culture-negative sputum samples (28). The use of molecular methods like PCR amplification of the 16S rRNA gene followed by cloning and traditional Sanger sequencing in bronchial secretion samples has allowed the identification of bacterial species previously undetected by the selective cultures used for the identification of PPM and more recently, the application of pyrosequencing to PCR-amplified products from human samples has taken the study of microbial diversity to an unprecedented level of detail (7, 14, 17). These approaches have shown that there is a wide diversity of microorganisms in respiratory secretions that remain undetected in culture (15), and have suggested that the bronchial microbiome may be heterogeneous in COPD, with significant differences between bronchial sections (14), a finding previously reported in protected specimen brush samples from patients with respiratory disease (20), and not observed in healthy subjects (7). Accordingly, before the initiation of large studies focusing on the microbiome of the bronchial tree in well characterized COPD patients,
it is necessary to examine the microbial diversity across its different sections, so that the interpretation of the results obtained through sampling of the different bronchial tree compartments is accurate.

The aim of the present study was to identify lower airway bacteria unrecognized through culture in patients with COPD, in the absence of signs and symptoms of bronchial infection, as well as to examine the complexity of microbial flora in these patients by the use of amplification and pyrosequencing of the 16S rRNA gene. In addition, we aim to compare microbial diversity recovered from upper (sputum and bronchoaspirate) and lower (bronchoalveolar lavage and bronchial mucosa) bronchial tree samples simultaneously obtained from the same individuals, under the assumption that these samples represent different bronchial tree compartments.

MATERIALS AND METHODS

Design and participants
A study focused on the microbiome of the bronchial tree in chronic obstructive pulmonary disease (COPD) was performed in stable patients with moderate disease who had not suffered from exacerbations during the previous year and had not been treated with antibiotics during this period. Patients with sputum cultures available and negative for potentially pathogenic microorganisms (PPMs) at enrollment and at least two times in the previous year were recruited between January and June 2010. Gram-negative and Gram-positive bacteria recognised as agents causing respiratory infections, such as Haemophilus influenzae, Haemophilus parainfluenzae, Streptococcus pneumoniae, Moraxella catarrhalis, Pseudomonas aeruginosa, Staphylococcus aureus and Enterobacteriaceae were considered as PPMs (6, 40, 29). Patients with severe lung function impairment (postbronchodilator FEV1 50% of the reference) or requiring regular treatment for chronic respiratory disease were excluded from the study, and additional exclusion criteria were hospitalization within the previous year and any severe disease needing regular therapy. Patients surgically treated for cancer and free of recurrence for a minimum of five years were accepted for inclusion. All patients gave written informed consent and the study protocol was approved by the Regional Ethics Committee. Four types of respiratory samples (sputum, bronchial aspirate, bronchoalveolar lavage and bronchial mucosa) were taken from each individual.
Sociodemographic and clinical measurements

Sociodemographic and clinical data were recorded at enrolment including smoking habits, medical antecedents, respiratory symptoms and treatments. All patients performed forced spirometry and reversibility tests in the morning with a dry rolling-seal spirometer (Spirometrics, Gray, Maine, USA) according to standard techniques (2). Postbronchodilator forced vital capacity (FVC) and forced expiratory volume in 1 second (FEV₁) were measured, and compared with age and height-adjusted reference values obtained from selected volunteers from the Barcelona province (35). Data were analysed using the SPSS statistical software package version 18 (SPSS Inc., Chicago, IL, USA). Results of descriptive statistics for categorical variables are expressed as absolute and relative frequencies. Results for continuous variables are expressed as medians (interquartile range [IQR]).

Sampling procedure

Induced sputum samples were obtained and processed within 60 min at the enrollment visit according to standard methods (32, 33). Briefly, the patient was pretreated with an inhaled b2-agonist 10 min before the nebulization of isotonic saline (0.9%) followed by increasing concentrations of hypertonic saline (3%, 4% and 5%), for 7 min with each concentration. After every induction, the patient attempted to obtain a sputum sample by coughing, and the nebulization procedure was discontinued when the sputum volume collected was equal or higher than 1 mL (1).

The sputum sample was cultured and the determination of microbial typology was carried out by means of culture in selective media, according to standard methods (4), and cultures were considered negative when not growing PPMs. Bronchoscopy was performed under local anaesthesia and sedation, using a flexible videobronchoscope (BF180; Olympus Optical Co, Tokyo, Japan). Local anaesthesia and sedation were achieved using topical lidocaine spray and intravenous midazolam, respectively, in accordance with standard recommendations (35, 5). A broncoalveolar lavage, a bronchial mucosa biopsy and a bronchial aspirate were collected through the working channel of the bronchoscope during the procedure. The bronchoscope, after its usual disinfection procedure (5), was introduced transnasally, passed through the vocal cords without aspiration and wedged in a right middle lobe bronchus for bronchoalveolar lavage, to avoid contamination of the collected sample by oropharingeal flora (7). Fifty
milliliters of saline were instilled, aspirated and discarded, after which 100 ml were
lavaged in the same location and collected, with a recovery of a volume greater than
30% of instilled. Subsequently, a bronchial biopsy was performed in a subsegmentary
bronchus macroscopically normal at white light examination, and a bronchial aspirate
with the tip of the bronchoscope located in the right and left main bronchi were
obtained.

Sample processing and DNA extraction

Samples were collected between April and August 2009 and kept at -80 °C until DNA
e extractions were performed in September 2009. Sputum, bronchial aspirate and
bronchoalveolar lavage samples were treated for 15 min with an equal volume of
Sputasol (Oxoid, Hampshire, UK) followed by centrifugation for 15 min at 13,000 g.
Genomic DNA extraction from sputum, bronchial aspirate and bronchoalveolar lavage
pellets and bronchial mucosa samples was performed using a Qiagen DNA blood Kit
(Qiagen, Crawley, UK) according to the manufacturers’ instructions, with some
modifications. Briefly, samples were treated with 10 μl of an in-house lysis solution as
previously described (11). This stock solution consisted of 10mL of filter-sterilized
Buffer 1 [20mM Tris (pH 8), 2mM EDTA (Sigma) 1.2% Triton X-100], 500 mg of
lysozyme (Sigma, Poole, UK), 50000U of Mutanolysin (Sigma, Poole, UK), and 1000U
of Lysostaphin (Sigma, Poole, UK). The amounts of DNA obtained ranged between 80-
950 ng for sputum, 70-1105 ng for bronchial aspirates, 164-1060 ng for bronchoalveolar
lavages and 70-520 ng for bronchial mucosa samples.

PCR amplification and pyrosequencing

The first 500 bp of the 16S rRNA genes were amplified with the universal eubacterial
primers 27F (5’-AGAGTTTTGATCMTTGCTCAG-3’) and 338R (5’-
GCCTTGCCAGCCCCCTCAGGC-3’) using the high-fidelity AB-Gene DNA
polymerase (Thermo Scientific) with an annealing temperature of 52 °C and 20 cycles
to minimize PCR biases (41). A secondary amplification was performed using the
purified PCR product as a template, in which the universal primers were modified to
contain the pyrosequencing adaptors A and B, and an 8bp “barcode” specific to each
sample, following McKenna and cols. (27). Barcodes were different in at least 3
nucleotides from each other to avoid misclassification in sample assignments. Five
secondary PCRs were performed per sample, pooling their PCR products before
purification, which was done using an Ultrapure PCR purification kit (Roche). The final DNA per sample was measured by picogreen fluorescence in a Modulus 9200 fluorimeter from Turner Biosystems and 12 samples were mixed in equimolar amounts. Each pool of 12 samples was further purified and concentrated by the use of Microcon filters (Millipore) into a final volume of 20 μl and a concentration higher than 100 μg/μl. PCR products were pyrosequenced from the forward primer end only using a GS-FLX sequencer with Titanium chemistry (Roche) at the Center for Genomic Regulation (CRG) in Barcelona, Spain. One sixteenth of a plate was used for each pool of 12 samples.

Sequence Analysis

Reads with an average quality value lower than 20 and/or with more than 4 ambiguities in homopolymeric regions in the first 360 flows were excluded from the analysis. Only reads longer than 200 bp were considered, as it has been shown that taxonomic assignment accuracy decreases dramatically in reads shorter than 200 bp and that the use of short reads inflate rarefaction curves (9). Chimeric sequences were filtered out using the software Ballerophon (19). Sequences with differences in the primer region were excluded from the analysis as well as sequences with more than 4 ambiguities in homopolymeric regions. Sequences were assigned to each sample by the 8-bp barcode and passed through the Ribosomal Database Project classifier (10), where each read was assigned a phylum, class, family and genus, as long as the taxonomic assignment was unambiguous within an 80% confidence threshold, which has been estimated to assign reads with over 95% accuracy at those taxonomic levels. Sequences were deposited in the MG-RAST server under accession numbers # 4481640.3 - 4481663.3. To estimate total diversity, sequences were clustered at 97% nucleotide identity over 90% sequence alignment length using the software CD-HIT (22) and rarefaction curves were obtained with the program Analitic Rarefaction 1.3 (18). For this analysis, sequences over 97% identical were considered to correspond to the same Operational Taxonomic Unit (OTUs), representing a group of reads which likely belong to the same species (47). Rarefaction curves were also obtained using only those sequences assigned to a genus by the RDP classifier, as a conservative estimate of diversity at that taxonomic level.

Statistical Analysis
Principal Component Analyses (PCA) were performed with UNIFRAC (24) using clustering at 97% sequence identity with the weighted analysis option, which compares the 16S-estimated diversity by a phylogenetic approach that takes into account both taxonomically assigned and unassigned reads. For comparison with respiratory tract samples, PCR-amplified sequences from the same region of the 16S rRNA gene in oral samples from supragingival dental plaque were taken (MG-Rast Accession numbers # 4481871.3 - 4481856.3) and included in the PCA. The matrix for performing the PCA is based on the distance between phylogenetic trees corresponding to each sample. This distance is measured in terms of the branches length in the trees that are unique to one sample or the other (24).

Two-way comparisons in bacterial composition were performed using the UniFrac metric (24) in order to measure whether the microbial communities in different sample types were significantly different. A tree including all sequences from each sample type was obtained, and samples are significantly different if the UniFrac value for the real tree is greater than would be expected if the sequences were randomly distributed between the samples. The sample IDs are randomly permuted 1000 times to obtain a p-value representing the fraction of permuted trees that have UniFrac values greater or equal to that of the real tree, using the Bonferroni corrections for multiple comparisons (24).

RESULTS

Participants

Five men and one woman who were former smokers with moderate COPD were included in the study (age median: 71). Four patients reported chronic bronchitis and all patients showed a moderate impairment in their lung function (postbronchodilator FEV1 median: 66 % of the reference) that did not require regular treatment. Four patients had been surgically treated for cancer (3 lung and 1 breast), all them free of recurrence of this disease after surgery for a minimum of five years. Sputum cultures obtained at recruitment and in two occasions during the previous year were negative for PPMs. Clinical data of the study group are shown in Table 1.

Microbial diversity in the bronchial tree
An average of 1033 sequences of the 16S rRNA gene were obtained in the 24 respiratory samples, which correspond to sputum, bronchial aspirate, bronchoalveolar lavage and bronchial mucosa from the six participating COPD patients. When sequences were clustered at 97% sequence identity, a consensus threshold for reads belonging to the same species, rarefaction curves and Chao1 indexes suggested a total diversity of over 500 species per sample.

Rarefaction curves at the genus level reached saturation for the subjects with lower microbial diversity when the sequences of the four sample types from each patient were pooled, and the total number of bacterial genera was found to be between 80-140 per patient, with differences in the bacterial composition among them (Figure 1a). Commonly amplified bacterial phyla were Proteobacteria, Bacteroidetes, Actinobacteria and Firmicutes. *Streptococcus*, *Prevotella*, *Moraxella*, *Haemophilus*, *Acinetobacter*, *Fusobacterium*, and *Neisseria* were the most common bacterial genera amplified, which together account for 60% of the total number of sequences. A complete list of bacterial genera per sample is included in Supplementary Dataset 1. Sequences with maximum identity to *Legionella* and *Mycoplasma* were also identified in the samples. Although the short sequence length of the reads makes species assigning unreliable, further analysis of these sequences against the RDP database suggests that they probably correspond to *Legionella dresdenensis*, a species non-associated to clinical cases isolated from river water and to the non-pathogenic species *Mycoplasma orale* and *M. salivarium*.

**Microbiome differences in bronchial tree compartments**

Bacterial composition was similar for the same sample type (Figure 2) whereas important differences in diversity were observed between upper and lower bronchial samples. Bacterial diversity was lower in sputum, which reached saturation at about 60 bacterial genera, and much higher in bronchoalveolar lavage and biopsy specimens (Figures 1b and 2). Some of the most frequent genera in sputum and in bronchial aspirate correspond to common dwellers of the oral cavity, like *Veillonella*, *Fusobacterium* or *Prevotella*.

In order to test differences in bacterial composition among the four sample types, a Principal Component Analysis (PCA) was performed for all 16S rRNA reads clustered at 97% similarity, giving a higher resolution than genera assignment and taking into account all species-level phylotypes. The two main components accounted for over 83%
of data variation, and the graph shows that bronchial mucosa and bronchoalveolar lavage samples cluster together, whereas sputum and bronchial aspirate samples were distinct from these lower bronchial samples and were different between them (Figure 3).

This was confirmed by two-way comparisons in microbial composition, which showed statistically significant differences between sputum and the other sample types (Unifrac Distance, \( p<0.002 \) in all cases), whereas bronchoalveolar lavage and biopsies were not statistically different (Unifrac Distance, \( p=0.6 \)). The similarity between the microbiome of bronchoalveolar lavage and bronchial mucosa suggest that both samples represent the same bronchial compartment. Bronchial aspirates and sputum are less diverse and contain genera absent from the lower bronchial tree, but which are found in the oral cavity of healthy individuals. Figure 4 shows the genera shared between, and unique to each bronchial tree sample, indicating that 43 genera are shared among all sample types (see Table S1 for a list of genera unique to and shared between samples). Although a large number of genera appear to be unique to each sample type, most of them are detected as single reads in the sequences. When these singletons are excluded (Supplementary Dataset 1), the genera exclusive of each sample type are restricted to Limnobacter in sputum, Arcobacter, Blautia, Emeticicia, and Runella in Bronchial Aspirate and Azonexus, Herbaspirillum, Peredibacter, Simplicispira, Sporolactobacillus and Methylobacillus in Bronchoalveolar Lavage.

**DISCUSSION**

Our study has analyzed the bronchial tree microbiome in stable COPD patients with a moderate impairment in their lung function, through amplification and pyrosequencing of the 16S rRNA gene. We have characterized the microbiology of upper and lower bronchial tree compartments, examining different respiratory samples that included sputum, bronchial aspirate, bronchoalveolar lavage and bronchial mucosa. Rarefaction curves and Chao1 indexes demonstrated a number of 80-140 bacterial genera per patient, with a total estimated diversity of over 500 species in the examined samples. The number of potential species in the respiratory samples could be inflated due to sequencing errors (34) and short sequence length attributable to the coverage of the hypervariable regions 1 and 2 of the 16S rRNA gene by the read. Taxonomic assignment at higher levels like genus or family, however, is highly reliable at this read.
length (23), and confirms the presence of a complex bacterial community. Thus, bacterial diversity in the lower airway, at least in COPD individuals without signs of infection, is much higher than previously anticipated. Frequently amplified phylum in these patients were Proteobacteria, Bacteroidia, Actinobacteria and Firmicutes, being *Streptococcus, Prevotella, Moraxella, Haemophilus, Acinetobacter, Fusobacterium,* and *Neisseria* the most common bacterial genera identified, which together account for 60% of the total number of sequences. Bacterial diversity was lower in sputum, a sample type that reached saturation at about 60 bacterial genera, and much higher in bronchoalveolar lavage and bronchial mucosa specimens. The microbiome of these lower bronchial tree samples showed close similarity, whereas sputum and bronchial aspirate were distinct between them and from the lower samples.

Traditional culture-based studies have described the bronchial tree as sterile in healthy subjects (20, 43, 37), recovering bacteria from bronchial secretions only when the patient suffers from a chronic respiratory disease. In the absence of signs or symptoms of respiratory infection, low-load colonizing PPM are often found in the bronchial tree of COPD patients (37, 25, 26), being the mechanisms behind the recovery of these microorganisms in COPD open to debate. Culture-independent microbiological techniques have demonstrated that the lungs are not sterile during health and have documented changes in the lung microbiome in several chronic lung diseases (3). Charlson and cols. have studied the oropharyngeal and bronchial secretions from healthy subjects, finding close similarities in the microbiologic pattern found in the oropharynx and the bronchial tree, with a lower biomass in bronchial secretions, and concluded that a specific bronchial microbiome in healthy subjects does not exist (7). A different pattern was reported by Hilty and cols in patients with chronic obstructive respiratory diseases (17). The authors compared lower bronchial secretions recovered by bronchial brushing from healthy subjects and patients with COPD and asthma, and demonstrated a frequent recovery of members of the phylum Proteobacteria, that contained important PPM like *Haemophilus* and *Moraxella,* in these patients. Erb-Downward and cols (14) have used a similar approach in smokers without functional abnormalities and with COPD using bronchoalveolar lavage to sample the lower bronchial tree, and found a close similarity in the microbiome of smokers with and without disease, who showed members of the phyla Proteobacteria, Firmicutes, Bacteroidetes and Fusobacteria in over a half of the patients and a similar diversity, with
Streptococcus, Prevotella, Fusobacterium, Pseudomonas, Haemophilus, Veillonella and Porphyromonas as the most commonly identified genera. The same authors have found a different bacterial profile, with an over-representation of the genus Pseudomonas, in severe COPD patients, when explants of the bronchial mucosa obtained from surgical samples were examined (14). An important limitation of these studies performed in COPD patients is that the microbial diversity in the different bronchial compartments has not been assessed, making the interpretation of the results difficult.

Our study has focused in moderate COPD patients, and confirmed that in the absence of signs and symptoms of bronchial infection there is a rich microbiome in the bronchial tree, being the most common amplified bacterial phylum Proteobacteria, Bacteroidia, Actinobacteria and Firmicutes, that have been reported in healthy subjects too (7). In the studied samples common bacterial genera amplified have been Streptococcus, Prevotella Fusobacterium, and Neisseria, also described in the normal population (7), and Moraxella, Haemophilus and Acinetobacter, bacteria that have appeared as overrepresented in patients with COPD (17, 14). Accordingly, our data support the hypothesis that in patients with COPD the bronchial microbiome includes genera present in the healthy subjects, with an additional increased presence of various genera of Proteobacteria that are unusual in the normal population and included well known PPM as Haemophilus and Moraxella.

Clear-cut differences in the microbiome of the upper and lower bronchial tree of moderate COPD patients emerged from our study. The upper respiratory samples, sputum and bronchial aspirate, showed low diversity and the frequent recovery of phyla that are part of the oropharingeal flora of the healthy subject, like Firmicutes and Bacteroidetes (7). Lower bronchial tree samples (bronchoalveolar lavage and bronchial biopsy) showed a more diverse microbiome with a close community profile in both samples, a minor representation of oropharingeal flora, and the recovery of genera that included PPM. These results confirm that the bronchial tree has different compartments with specific characteristics in COPD. The upper bronchial tree has low diversity and an overrepresentation of oropharingeal flora, and lower bronchi show a higher diversity that included genera that are unusual in sputum and aspirates, and only show low prevalence of Firmicutes and Bacteroidetes, frequent in the mouth and pharynx.
The lower level of bacterial diversity found in sputum when compared to other sample types suggests that sputum samples contain a limited fraction of the total bacterial community inhabiting the respiratory tract. In addition, the principal component analysis indicates that sputum samples, which are commonly used for bacterial identification in respiratory tract infections, are not representative of the composition and proportion of bacterial taxa in bronchial mucosa. Bronchoalveolar lavage, on the other hand, showed a similar but slightly higher diversity than mucosa, probably attributable to the wider bronchial surface sampled by lavage, and may be considered a representative substitute of bronchial mucosa samples (Figure 1b). The representativeness of sampling methods is a common problem in human microbiome research. Most studies of microbial diversity in the gastrointestinal tract, for instance, have been performed on stool samples but gut mucosal biopsies have in fact been found to harbor a very different microbial composition when compared with faecal samples of the same individuals (13), questioning the validity of many metagenomic studies. Similarly, the use of inappropriate samples to study the lung’s bacterial diversity can also have important clinical implications. Given that sputum samples are probably the most common lung clinical samples taken because they are readily obtained with non-invasive techniques, it must be born in mind that the microbiota of sputum is not representative of the microbiology of the lower airway. Thus for example, we do not detect *Legionella* in sputum but it appeared in 4 out of 6 bronchoalveolar lavage samples. The role of atypical bacteria such as *Legionella* in the COPD is unclear and should be further studied, as other authors detected *Legionella* in COPD patients by PCR methods (12).

Finally, we want to emphasize that although we cannot discard that the sharing of some bacterial genera between upper bronchial tree samples and oral cavity can be partly attributed to some bacterial biomass contamination at the time of sampling (7), the observed resemblance may also have a biological meaning. The larynx has classically been considered a barrier between the oropharynx and the trachea, which would keep the latter sterile. However, the data presented here suggest that bacteria from the oral cavity and the pharynx are also found in the bronchial tree, and such mechanical barrier does not avoid a regular appearance of oropharyngeal flora in the upper bronchi, that decreases when going further down in the airway, at the level where the bronchoalveolar lavage is done. We hope these results stimulate further characterization of the
respiratory tract microbiota in healthy controls and in individuals with different respiratory diseases.

ACKNOWLEDGMENTS

RCR, AMo and AMi were funded by projects SAF2009-13032-C02-01 and 02 from the Spanish MICINN. AMo is also funded by project BFU2008-04501-E/BMC from Spanish MCINN and Prometeo/2009/092 from Generalitat Valenciana (Spain). We also thank support from Fundació Taulí, SOCAP and CIBERes – Ciber de Enfermedades Respiratorias. CIBERes is an initiative of Instituto de Salud Carlos III.
REFERENCES


Figure 1.
Bacterial diversity in respiratory tract samples. The graph shows rarefaction curves indicating the number of assigned bacterial genera in relation to the number of 16S rRNA sequences, grouped by individual (a) and sampling method (b). Symbols correspond to: Sputum (Sp), Tissue (Ts), Bronchoalveolar Lavage (BAL) and Bronchial Aspirate (BAs).

Figure 2.
Taxonomic assignment of the 24 samples at the level of bacterial classes. Bacterial composition in sputum samples appear to be particularly different from other sample types from the same individual. Symbols correspond to: Sputum (Sp), Tissue (Ts), Bronchoalveolar Lavage (BAL) and Bronchial Aspirate (BAs).

Figure 3.
Principal Component Analysis of the four respiratory tract sample types (n=6 for each sample type) and samples from the oral cavity of healthy individuals (n=16) according to the microbial composition as inferred by pyrosequencing of the 16S rRNA gene. Similar results were obtained using the third component (data not shown). Symbols correspond to: Sputum (Sp), Tissue (Ts), Bronchoalveolar Lavage (BAL), Bronchial Aspirate (BAs) and Supragingival Dental Plaque (Oral) samples.

Figure 4.
Venn diagram showing the bacterial genera unique and/or shared between sample types. A complete list of genera for each location in the diagram is included in Table S1. Symbols correspond to: Sputum (Sp), Tissue (Ts), Bronchoalveolar Lavage (BAL) and Bronchial Aspirate (BAs).
TABLE 1. Characteristics of the studied population

<table>
<thead>
<tr>
<th>Patient</th>
<th>Gender</th>
<th>Smoking History</th>
<th>Cancer Site</th>
<th>Cancer Surgery</th>
<th>COPD Severity(^1)</th>
<th>FEV1%</th>
<th>No. of Sequences(^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Male</td>
<td>Former</td>
<td>Lung</td>
<td>Lobectomy</td>
<td>Moderate</td>
<td>72</td>
<td>4362</td>
</tr>
<tr>
<td>2</td>
<td>Female</td>
<td>Former</td>
<td>Breast</td>
<td>Mastectomy</td>
<td>Moderate</td>
<td>76</td>
<td>2477</td>
</tr>
<tr>
<td>4</td>
<td>Male</td>
<td>Former</td>
<td>Lung</td>
<td>Lobectomy</td>
<td>Moderate</td>
<td>41</td>
<td>2980</td>
</tr>
<tr>
<td>5</td>
<td>Male</td>
<td>Former</td>
<td>None</td>
<td>--------</td>
<td>Moderate</td>
<td>60</td>
<td>6858</td>
</tr>
<tr>
<td>7</td>
<td>Male</td>
<td>Former</td>
<td>None</td>
<td>--------</td>
<td>Moderate</td>
<td>72</td>
<td>5497</td>
</tr>
<tr>
<td>8</td>
<td>Male</td>
<td>Former</td>
<td>Lung</td>
<td>Lobectomy</td>
<td>Moderate</td>
<td>42</td>
<td>2619</td>
</tr>
</tbody>
</table>

\(^1\) according to GOLD (www.goldcopd.org)

\(^2\) Sequences of the 16S rRNA gene obtained after length and quality filtering