

# Comparison of the Respiratory Microbiome in Healthy Nonsmokers and Smokers

Alison Morris<sup>1,2</sup>, James M. Beck<sup>3,4,5</sup>, Patrick D. Schloss<sup>3</sup>, Thomas B. Campbell<sup>5</sup>, Kristina Crothers<sup>6</sup>, Jeffrey L. Curtis<sup>3,4</sup>, Sonia C. Flores<sup>5</sup>, Andrew P. Fontenot<sup>5</sup>, Elodie Ghedin<sup>7,8</sup>, Laurence Huang<sup>9</sup>, Kathleen Jablonski<sup>10</sup>, Eric Kleerup<sup>11</sup>, Susan V. Lynch<sup>9</sup>, Erica Sodergren<sup>12</sup>, Homer Twigg<sup>13</sup>, Vincent B. Young<sup>3,14</sup>, Christine M. Bassis<sup>3,4</sup>, Arvind Venkataraman<sup>15</sup>, Thomas M. Schmidt<sup>15</sup>, and George M. Weinstock<sup>12</sup>; on behalf of the Lung HIV Microbiome Project

<sup>1</sup>Department of Medicine, <sup>2</sup>Department of Immunology, <sup>7</sup>Department of Computational and Systems Biology, and <sup>8</sup>Center for Vaccine Research, University of Pittsburgh, Pittsburgh, Pennsylvania; <sup>3</sup>Department of Internal Medicine and <sup>14</sup>Department of Microbiology and Immunology, University of Michigan, Ann Arbor, Michigan; <sup>4</sup>Veterans Affairs Ann Arbor Healthcare System, Ann Arbor, Michigan; <sup>5</sup>Department of Medicine, University of Colorado Denver, Aurora, Colorado; <sup>6</sup>Department of Medicine, University of Washington, Seattle, Washington; <sup>9</sup>Department of Medicine, University of California San Francisco, San Francisco, California; <sup>10</sup>Department of Epidemiology and Biostatistics, George Washington University, Washington, DC; <sup>11</sup>Department of Medicine, University of California Los Angeles, Los Angeles, California; <sup>12</sup>The Genome Institute, Washington University School of Medicine, St. Louis, Missouri; <sup>13</sup>Department of Medicine, Indiana University, Indianapolis, Indiana; and <sup>15</sup>Department of Microbiology and Molecular Genetics, Michigan State University, East Lansing, Michigan

**Rationale:** Results from 16S rDNA-encoding gene sequence-based, culture-independent techniques have led to conflicting conclusions about the composition of the lower respiratory tract microbiome.

**Objectives:** To compare the microbiome of the upper and lower respiratory tract in healthy HIV-uninfected nonsmokers and smokers in a multicenter cohort.

**Methods:** Participants were nonsmokers and smokers without significant comorbidities. Oral washes and bronchoscopic alveolar lavages were collected in a standardized manner. Sequence analysis of bacterial 16S rRNA-encoding genes was performed, and the neutral model in community ecology was used to identify bacteria that were the most plausible members of a lung microbiome.

**Measurements and Main Results:** Sixty-four participants were enrolled. Most bacteria identified in the lung were also in the mouth, but specific bacteria such as *Enterobacteriaceae*, *Haemophilus*, *Methylobacterium*, and *Ralstonia* species were disproportionately represented in the lungs compared with values predicted by the neutral model. *Tropheryma* was

(Received in original form October 24, 2012; accepted in final form February 22, 2013)

Supported by U.S. Public Health Service grants U01HL098962 (A.M., E.G.), R01HL090339 (A.M.), U01HL98961 (J.M.B., J.L.C., V.B.Y.), R01HG005975 (P.D.S.), U01HL098996 (T.B.C., S.C.F., A.P.F.), R01HL090342 (K.C.), U01HL098964 (L.H., S.V.L.), R01HL090335 (L.H.), K24HL087713 (L.H.), U01HL098958 (K.J.), U01HL098960 (H.T., G.M.W.); NIH/HMP R0100417746 (T.M.S.). The project described was supported by University of Pittsburgh Clinical Translational Science Institute by the National Institutes of Health through Grant Numbers UL1 RR024153 and UL1TR000005, the University of California San Francisco Clinical and Translational Science Institute UL1TR000004, and the Colorado CTSI Grant Number UL1TR000154.

**Author Contributions:** Conception and design: A.M., J.M.B., T.B.C., S.C.F., A.P.F., E.G., L.H., K.J., S.V.L., H.T., V.B.Y., G.M.W. Acquisition of data: A.M., J.M.B., T.B.C., K.C., J.L.C., S.C.F., A.P.F., E.G., L.H., E.K., S.V.L., E.S., H.T., V.B.Y., G.M.W. Analysis and interpretation of data: P.D.S., K.J., C.M.B., A.V., T.M.S. Drafting or revising the article: A.M., J.M.B., P.D.S., T.B.C., K.C., J.L.C., S.C.F., A.P.F., E.G., L.H., K.J., E.K., S.V.L., E.S., H.T., V.B.Y., C.M.B., A.V., T.M.S., G.M.W. Final approval of the manuscript: A.M., J.M.B., P.D.S., T.B.C., K.C., J.L.C., S.C.F., A.P.F., E.G., L.H., K.J., E.K., S.V.L., E.S., H.T., V.B.Y., C.M.B., A.V., T.M.S., G.M.W.

The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of the National Institutes of Health.

Correspondence and requests for reprints should be addressed to Alison Morris, M.D., M.S., Associate Professor of Medicine, Division of Pulmonary, Allergy, and Critical Care Medicine, University of Pittsburgh, 3459 Fifth Avenue, 628 NW MUH, Pittsburgh, PA. E-mail: [morrisa@upmc.edu](mailto:morrisa@upmc.edu)

This article has an online data supplement, which is accessible from this issue's table of contents at [www.atsjournals.org](http://www.atsjournals.org)

Am J Respir Crit Care Med Vol 187, Iss. 10, pp 1067–1075, May 15, 2013

Published 2013 by the American Thoracic Society

Originally Published in Press as DOI: 10.1164/rccm.201210-1913OC on March 14, 2013

Internet address: [www.atsjournals.org](http://www.atsjournals.org)

## AT A GLANCE COMMENTARY

### Scientific Knowledge on the Subject

Definition of the human microbiome is an area of high interest, but studies of the lung microbiome are at early stages. Culture-independent techniques have produced conflicting conclusions about the composition of the lower respiratory tract microbiome. In order to understand changes in the microbial composition of the lung in respiratory disease, characterization of the lung microbiome in healthy individuals is needed.

### What This Study Adds to the Field

We evaluated the respiratory tract microbiome in healthy individuals using standardized techniques in a multicenter study. This study is the largest to date of the lower respiratory tract in a healthy population and the first to apply the neutral theory of community ecology to compare the lung to the oral cavity. We found that lung bacterial communities resemble those in the oral cavity, but specific bacteria appear in significantly higher abundance in the lungs than would be expected if they originated from the mouth, demonstrating that the lung microbiome does not derive entirely from the oral cavity in healthy individuals. Smoking appeared to influence the oral microbiome, but we did not find significant differences in the lungs of smokers and nonsmokers.

also in the lung, but not the mouth. Mouth communities differed between nonsmokers and smokers in species such as *Porphyromonas*, *Neisseria*, and *Gemella*, but lung bacterial populations did not.

**Conclusions:** This study is the largest to examine composition of the lower respiratory tract microbiome in healthy individuals and the first to use the neutral model to compare the lung to the mouth. Specific bacteria appear in significantly higher abundance in the lungs than would be expected if they originated from the mouth, demonstrating that the lung microbiome does not derive entirely from the mouth. The mouth microbiome differs in nonsmokers and smokers, but lung communities were not significantly altered by smoking.

**Keywords:** lung; microbiome; smoking; bronchoscopy; metagenome

Standard microbiological culture-based methods detect only a small proportion of bacterial diversity present in different body

sites, because the great majority of these organisms are to date uncultivated. Culture-independent profiling techniques have allowed detection of complex microbial communities in various body sites, including the skin, oral cavity, and gastrointestinal tract (1, 2). Recent applications of these techniques have led to increased understanding of roles of microbial communities in health and in diseases such as inflammatory bowel disease, obesity, and dental caries (3–5).

Because organisms are rarely cultured from lungs of healthy individuals, it has historically been presumed that the lung is sterile in the normal host. However, because the lungs are constantly exposed to bacteria both from the environment and from the upper airway, the lungs may not be free of microbes. Understanding the microbial composition of the lung in healthy individuals is necessary to understand changes that may occur with respiratory diseases such as chronic obstructive pulmonary disease (COPD) or in immunosuppressed individuals such as those with HIV infection. Culture-independent techniques investigating various respiratory samples have produced conflicting conclusions about the composition of the lower respiratory tract microbiome (6–10). Controversy persists over the existence of distinct organisms in the lung and whether bacteria in the lungs represent microaspiration of oral microbiota or whether upper respiratory contamination of lower respiratory tract samples is caused by passing a bronchoscope through the oral cavity to obtain lung samples. In addition, because of the methodological challenges of these types of studies and the intensity of data analyses, small sample sizes and single-center designs are common, limiting generalization of results. Differences in study populations, geographical location, sample collection techniques, types of respiratory samples examined, and differences in DNA sequencing methodologies all compound the difficulty in comparing existing studies of the lung microbiome.

Optimal techniques to analyze lung microbiome data are also not yet established. As mentioned above, microbes from the mouth are regularly introduced into the lower respiratory tract. Therefore, defining the respiratory microbiome is a two-step process. First, it is necessary to test whether dispersal from the mouth, rather than active environmental selection in the lungs, can sufficiently explain the composition of the microbial community in the lungs. Second, we need to pick out bacterial groups whose distribution in the lungs cannot be explained by dispersal from the mouth, because these are the most plausible members of a lung microbiome. The neutral theory of community ecology is a novel analytic method to assess the respiratory microbiome by performing these two steps. This theory assumes that all species in a community are functionally equivalent (i.e., the observed distribution of species is not a result of active environmental selection) (11), and that dispersal from a source community and randomness in birth and death of microbes are sufficient to explain the observed community structure. Therefore, the neutral theory can be used as the null hypothesis to test whether dispersal from the mouth (source community) can satisfactorily describe the observed microbial distribution in the lung and to pinpoint bacterial groups that significantly deviate from “neutrality,” suggesting that these microbes might have some advantage in the lung.

Smoking may also have a direct impact on the composition of the respiratory microbiome. Prior work has detected bacterial DNA in cigarettes, and immunologic alterations that occur with smoking could lead to changes or shifts in microbial community structures (12, 13). Some studies have found differences in the upper respiratory tract microbial composition of smokers and nonsmokers (14). Others have found few differences in the lower respiratory tract microbiome of smokers and

nonsmokers, but these studies have examined small numbers of subjects (7).

In the present study, our objective was to use a neutral community model to compare the microbiome of the upper and lower respiratory tract in a large number of strictly defined healthy nonsmokers and smokers. These subjects are enrolled in a multicenter cohort, part of the larger Lung HIV Microbiome Project (LHMP) sponsored by the National Heart, Lung, and Blood Institute to characterize the microbiome of the lung. This study is the first project of the group and was designed to establish the lung microbiome in an HIV-uninfected, healthy population. These results represent the most comprehensive evaluation of the normal human lung microbiome to date.

## METHODS

### Study Design

We performed a prospective, multisite, observational cohort study. The study was approved by the institutional review boards at participating sites. Written, informed consent was obtained from all study participants.

### Participants

Participants were otherwise healthy current smokers and nonsmokers who were recruited in eight cities participating in the LHMP. Inclusion criteria for both nonsmokers and smokers were men and women aged 18 to 80 years; no use of antibiotics or immunosuppressive medications in the past 3 or 6 months, respectively; and no evidence of an acute respiratory process defined as no reported fever, cough, or upper respiratory symptoms for the previous 4 weeks. Individuals with a known history of any pulmonary disease (e.g., chronic obstructive pulmonary disease or asthma) or history of abnormal pulmonary function testing were excluded. Detailed inclusion and exclusion criteria are provided in the online supplement.

### Sample Collection

Subjects were asked to fast and refrain from smoking for at least 12 h before sample collection. Oral washes (OW) were performed by having participants gargle with 10 ml sterile 0.9% saline immediately before bronchoscopy. Sterile saline in a sample collection cup was used as an oral wash control. Before bronchoscopy, 10 to 50 ml of sterile 0.9% saline were washed through the bronchoscope and collected as a control for DNA in the bronchoscope. Bronchoalveolar lavage (BAL) was performed according to standardized procedures developed to minimize oral contamination (10). Participants gargled with an antiseptic mouthwash (Listerine) immediately before topical anesthesia. The bronchoscope was then inserted through the mouth and advanced to a wedge position quickly and without use of suction. BAL was performed in the right middle lobe or lingula up to a maximum of 300 ml 0.9% saline.

### Sample Processing and Sequencing

DNA was extracted from samples using standard techniques at each center and shipped to the sequencing center at Washington University (online supplement). Sequencing was performed using Roche 454 FLX Titanium platform using primers for variable regions 1 through 3 (V1–3) and 3 through 5 (V3–5) (online supplement). To verify that DNA processed at different centers produced comparable sequencing results, we also performed a pilot study assessing results of test samples at each center (online supplement).

Because of the low biomass associated with the respiratory tract, we were concerned that reagent-derived contaminants could confound our analysis. Therefore, in addition to BAL and OW samples, DNA was extracted from reagents and from BAL and OW control samples taken before sampling. These control samples were processed in parallel with the actual BAL and OW samples. Ordination of the BAL controls and samples (using V1–3 data) demonstrated that several BAL samples had

community structures resembling those of the control subjects (*see* Figure E1A in the online supplement). These samples were removed from further analyses both for V1–3 and V3–5. Similar analysis with V3–5 data did not show any significant degree of overlap with control subjects (Figure E1B).

### Sequence Curation and Analysis

16S rRNA gene sequences were curated essentially as described previously using the mothur software package (15, 16) (online supplement).

### Comparison of Lung and Oropharynx Populations

A neutral community model (17) was modified and used to distinguish operational taxonomic units (OTUs) that are a result of dispersal from the mouth and those that might be members of a lung microbiome. The neutral model was applied to the 16S rRNA-encoding gene sequence surveys from the lung and mouth using both the V1–3 and V3–5 regions for both nonsmokers and smokers in R (18). We constructed 95% binomial confidence intervals for the neutral model based on the Wilson method (19) with the Hmisc package in R (18). OTUs that fell between the confidence intervals were considered to be present as a result of dispersal from the mouth. OTUs that fell outside the upper bound of the confidence interval on the left were found at disproportionately higher frequencies in the lung than predicted by the neutral model based on their abundance in the mouth. OTUs that fell outside the lower bound of the confidence interval on the right were found less frequently in the lungs than predicted by the neutral model. Among the OTUs that were more frequent in the lung than predicted by the neutral model, those that had a mean relative abundance greater than or equal to 0.5% in the lung were considered the strongest candidates for environmental selection in the lung. Furthermore, a Wilcoxon *t* test was performed on the mean relative abundance of these OTUs between the lung and mouth to determine the statistical significance, using the MASS package in R (18).

Comparisons at the OTU level were also performed between OW and BAL samples from each individual and between OW and BAL in nonsmokers versus smokers. Alpha diversity (a measure of the number of different bacterial sequences in a sample) was measured using the number of observed OTUs, the Shannon and inverse Simpson indices, and phylogenetic diversity (20, 21). Beta-diversity (a measure of the number of different sequences among samples) was measured using  $\Theta_{YC}$  because it provides a balanced representation of abundant and rare populations and is less sensitive to problems of undersampling compared with other metrics (22). Statistical comparisons based on  $\Theta_{YC}$  were performed using nonparametric analysis of variance (23). OTU-by-OTU statistical comparisons were performed using the Kruskal-Wallis rank sum test, and we corrected for multiple comparisons using step-down Bonferroni correction (24). To compare indices between BAL and OW samples, we used linear regression mixed models and assumed an unstructured correlation matrix to control for the repeated measurement per participant.

We then examined several potential confounders of our analyses. Because there were few women smokers, we repeated analyses comparing OW and BAL in nonsmokers and smokers with women excluded. We also examined the effect of degree of smoking by comparing results by participants dichotomized to above and below the median pack-year smoking history. To determine if body mass index (BMI) might confound analyses, we also compared diversity measures and OTU analyses of OWs and BALs categorized by BMI below 18.5 kg/m<sup>2</sup> (underweight), 18.5 to less than 25 (normal weight), 25 to less than 30 (overweight), and greater than or equal to 30 (obese) in the 59 participants with these data available. We also repeated diversity indices analyses comparing OW and BAL at each center with adjustment for smoking status to determine if there were any systematic differences between centers.

## RESULTS

### Cohort

Sixty-four participants were enrolled from eight cities (Table 1). Forty-five were nonsmokers, and 19 were current smokers.

TABLE 1. CHARACTERISTICS OF PARTICIPANTS

Demographics	Nonsmoker (n = 45)	Smoker (n = 19)	P Value
Age, yr	43.1 ± 13.17	43.8 ± 10.57	0.82
Sex			0.038
Male	23 (51.1)	15 (78.9)	
Female	22 (48.9)	4 (21.1)	
Ethnicity			0.31
Hispanic	5 (11.1)	0 (0)	
Not Hispanic	40 (88.9)	19 (100.0)	
Race			0.53
White	35 (77.8)	13 (68.4)	
Other	10 (22.2)	6 (31.6)	

Data are presented as n (%) or mean ± SD.

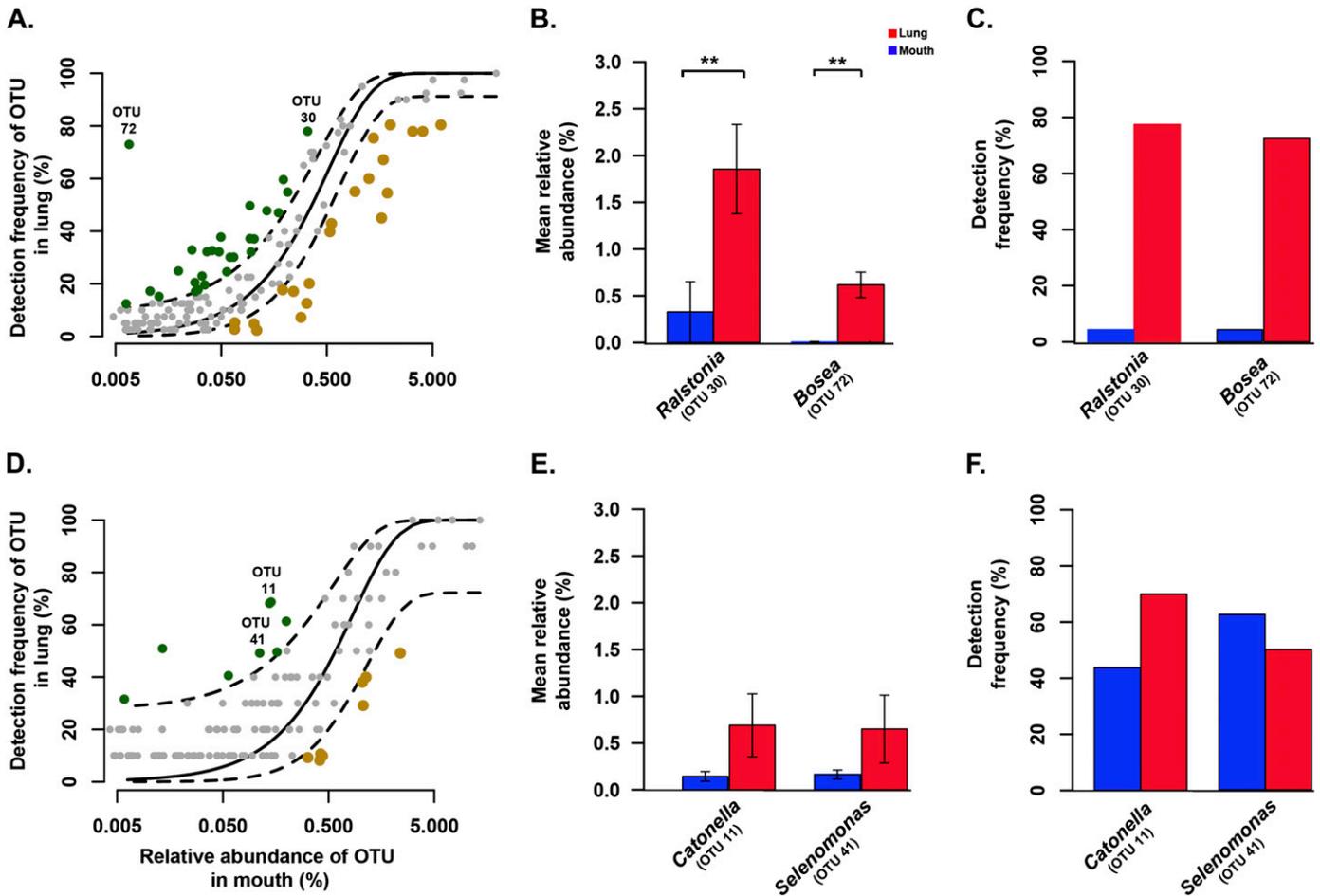
Nonsmokers and smokers were similar in age and ethnicity, but smokers were more likely to be men. Among smokers, median pack-year history was 18 pack-years with a range of 7 to 41 pack-years. Median BMI was 25.7 kg/m<sup>2</sup> (interquartile range, 20.4–31) for the cohort.

### Evidence for a Lung-Specific Microbiome

In our implementation of the neutral model, the mouth was tested as a source for bacterial OTUs found in the lungs. If the composition of the microbial community in the lung was derived through dispersal from the mouth rather than environmental selection in the lung, the mean relative abundance of each OTU in the mouth would dictate the frequency with which that OTU is detected in the lungs. Therefore, a plot of the relative abundance of OTUs in the mouth versus the detection frequency of OTUs in the lung would result in a continuous monotonically increasing curve converging on 1. Accordingly, we found that the neutral model with the mouth as a source explained much of the microbial community observed in the lungs (Figures 1 and 2; Spearman rank correlation coefficient between the model prediction and empirical observation was  $\geq 0.84$ ). This finding suggests that dispersal from the mouth is largely responsible for the composition of the microbial community in the lung at the depth of sequence reads analyzed. However, it is noteworthy that we were also able to identify OTUs that were disproportionately represented in the lung as compared with the values predicted by the neutral model (Figures 1 and 2). For example, based on the V1–3 region in nonsmokers comparing oral wash and BAL, OTUs #30 (classified as *Ralstonia* sp.) and #72 (classified as *Bosea* sp.) were disproportionately represented in the lung (Figure 1). Based on data from the V3–5 region in nonsmokers' oral wash and BAL, OTUs #130 (classified as a *Haemophilus* sp.), #229 (classified as *Enterobacteriaceae* sp.), and #239 (classified as a *Methylobacterium* sp.) were also disproportionately represented (Figure 2). OTU #229 (*Enterobacteriaceae* sp.) was also identified as being disproportionately represented in the lungs of smokers compared with the mouths of smokers (Figure 2).

### OTU-Level Comparisons of Oral and Lung Microbiome

We also used OTU-level analyses to identify particular OTUs that were differentially represented in BAL compared with OW communities (Figure E2). Similar to the neutral model findings, using V3–5 sequence data, an OTU classified as a member of the *Enterobacteriaceae* was significantly more abundant in the BAL samples compared with the OW samples of smokers. There was a similar, but nonsignificant, trend for the comparison of OW to BAL in nonsmokers (Figure E2B). Among the V3–5 data, but not the V1–3 data, there was also an OTU that was classified as *Tropheryma* sp., which was found in more



**Figure 1.** Neutral model for nonsmokers (A, B, C) and smokers (D, E, F) from 16S molecular surveys with variable regions 1 through 3. The solid line is the model prediction and dashed lines represent 95% confidence intervals. Green points are the outlying operational taxonomic units (OTUs) for which the observed frequency is greater than the model prediction and dark golden points are the outlying OTUs for which the observed frequency is less than the model prediction. Mean abundance (B and E) and detection frequency (C and F) of the outlying OTUs that are the strongest candidates for members of a lung microbiome in nonsmokers and smokers, respectively. Error bars in all histograms represent SE (\* $P$  value < 0.05, \*\* $P$  value < 0.01).

BAL samples (26.0%) than in OW samples (1.8%,  $P = 0.006$ ; Figure 3).

#### OTU-Level Comparisons between Nonsmokers and Smokers

In comparisons between OW communities of nonsmokers and smokers using V1–3 sequence data, we measured significant differences in OTUs classified as *Neisseria* sp., *Porphyromonas* sp., and *Gemella* sp. (Figure 4A). Among the V3–5 sequence data, there was a significant difference between the OW communities of smokers and nonsmokers in the relative abundance of an OTU classified as *Porphyromonas* sp. (Figure 4B). There were no significant differences when comparing the BAL of nonsmokers to smokers (Figure E3). Results of OW and BAL comparisons by smoking status were similar when analyses were limited to men only or when examined by pack-year history above or below the median (data not shown).

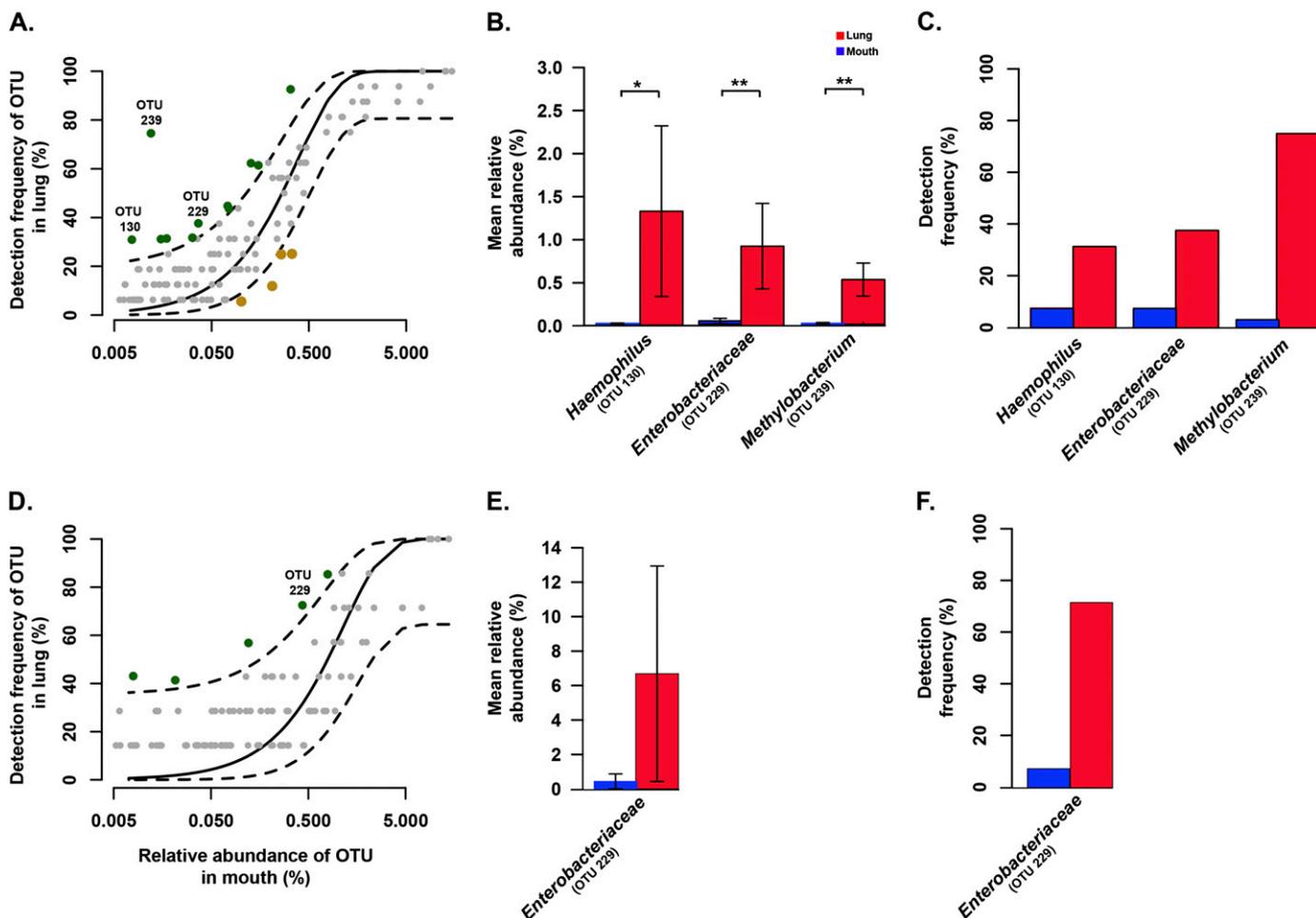
#### Overall Variation in Community Diversity and Structure

To measure the  $\alpha$  diversity of the OW and BAL communities in smokers and nonsmokers, we used the number of observed

OTUs, the Shannon and inverse Simpson indices, and the phylogenetic diversity (Table 2). After accounting for paired samples, there were no significant effects in comparisons of smoking status or OW to BAL on any of the  $\alpha$  diversity measures using the V1–3 dataset, but there were a significantly higher number of OTUs measured by V3–5 in smokers' BAL and OW than nonsmokers' ( $P = 0.02$ ). Nonmetric multidimensional scaling plots also demonstrated differences in comparisons of community structure of OW and BAL in nonsmokers and in smokers, but there was significant overlap between OW and BAL in both nonsmokers and smokers (Figure E4). There were significant differences in the oral communities, but not in the lung communities, of nonsmokers compared with smokers (Figure E5). Results of OW and BAL comparisons by smoking status were similar when analyses were limited to men only (data not shown). There were also no significant differences in comparisons of OW and BAL by BMI strata (data not shown).

#### Center-Level Comparisons

In analyses comparing results from each clinical center, there were no significant differences in the diversity indices (data not shown).



**Figure 2.** Neutral model for nonsmokers (A, B, C) and smokers (D, E, F) from the 16S molecular surveys with variable regions 3 through 5. The *solid line* is the model prediction, and *dashed lines* represent 95% confidence intervals. *Green points* are the outlying operational taxonomic units (OTUs) for which the observed frequency is greater than the model prediction, and *dark golden points* are the outlying OTUs for which the observed frequency is less than the model prediction. Mean abundance (B and E) and detection frequency (C and F) of the outlying OTUs that are the strongest candidates for members of a lung microbiome in nonsmokers and smokers, respectively. Error bars in all histograms represent SE (\**P* value < 0.05, \*\**P* value < 0.01).

## DISCUSSION

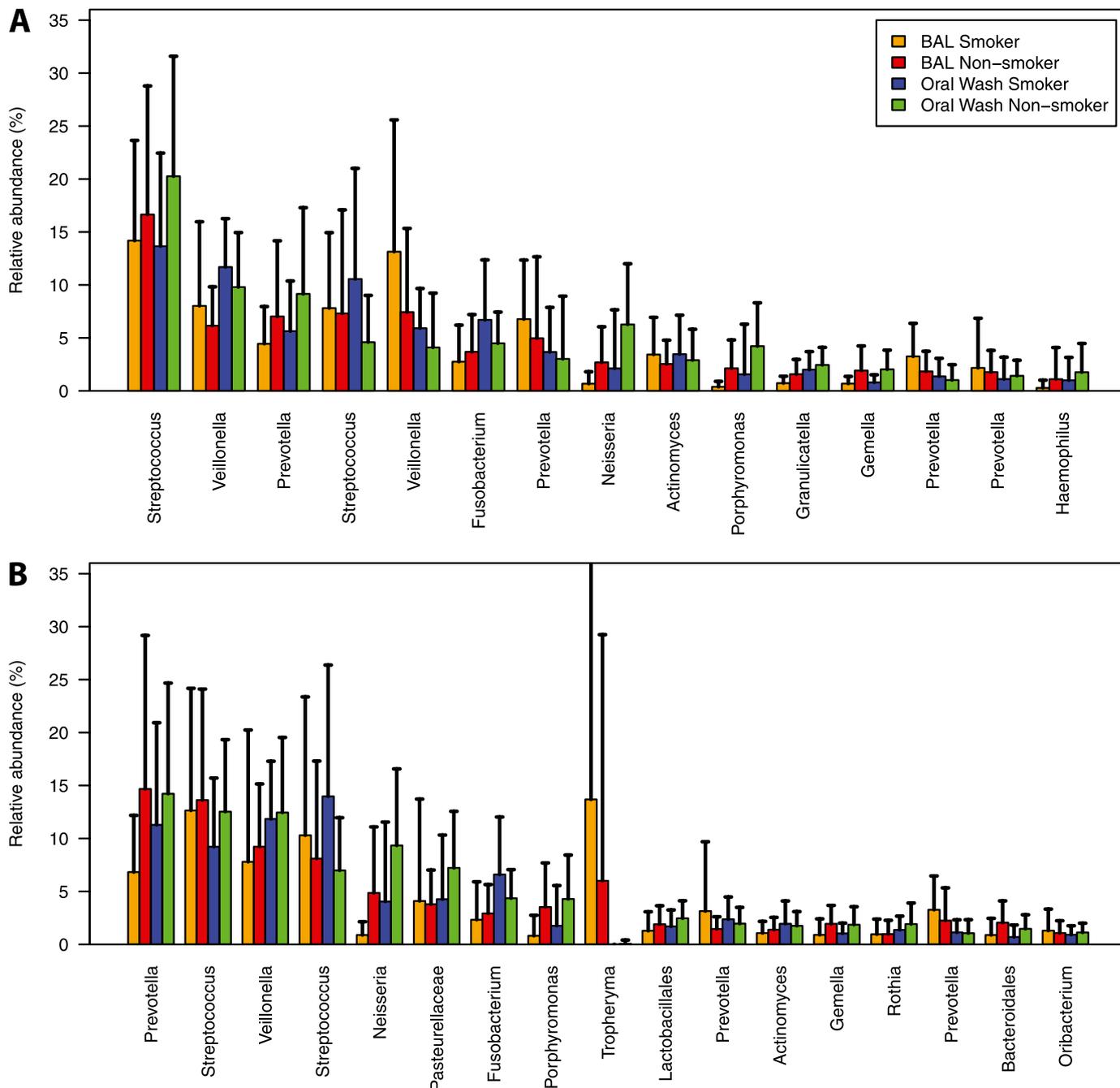
This multicenter study is the largest that examines the microbiome of the lower respiratory tract in healthy nonsmokers and smokers. We found that although the overall bacterial community detected in the lung resembles that found in the oral cavity, there were distinct bacterial species that were overrepresented in the lung. There were also significant differences in the microbiome of the oral cavity of smokers compared with nonsmokers. In contrast, lung populations did not differ significantly in overall analyses between smokers and nonsmokers.

A major controversy has centered on the existence of the “normal” lung microbiome. Previous studies have found that overall, the lung microbiome resembles that of the mouth, but these studies have had limited numbers of subjects (6, 7). The finding that the majority of the organisms in the lung have also been detected in the mouth and upper airways is not surprising, as these sites are anatomically contiguous. The constant transport of microbes between these locations suggests the possibility that the lung microbiome could be in a continual state of flux, with new species being introduced or removed in a stochastic manner. Bronchoscopic contamination or carryover from the upper airway may also occur and influence detection of microbes in

BAL samples, but studies of lung tissue have confirmed the independent presence of bacterial populations (7, 9).

Study of the lung microbiome is a new field, and investigators have not yet reached consensus on how to analyze or present lung microbiome data. Under the null hypothesis of the neutral model, the distribution of microbes in the lung should mirror that seen in the “source” community (i.e., the mouth). Indeed, the presence of many microbes in the lungs was consistent with the neutral model of dispersal from the mouth; however, we also found some OTUs (e.g., *Enterobacteriaceae* sp., *Haemophilus* sp., both common causes of pneumonia) that exhibited significant deviations from the model, implying that these organisms were adapted toward proliferation in the lung environment. An important caveat is that we tested only for the mouth as the source community. Further evidence regarding a lung-specific microbiome needs to be confirmed by considering all possible sources, including the nose, throat, and gastrointestinal tract, and other methods of testing.

In our cohort, the most common genera in both BAL and OW were *Streptococcus*, *Prevotella*, and *Veillonella*. Other groups have found similar organisms in BAL with existence of a core pulmonary microbiome that includes *Pseudomonas*, *Streptococcus*,

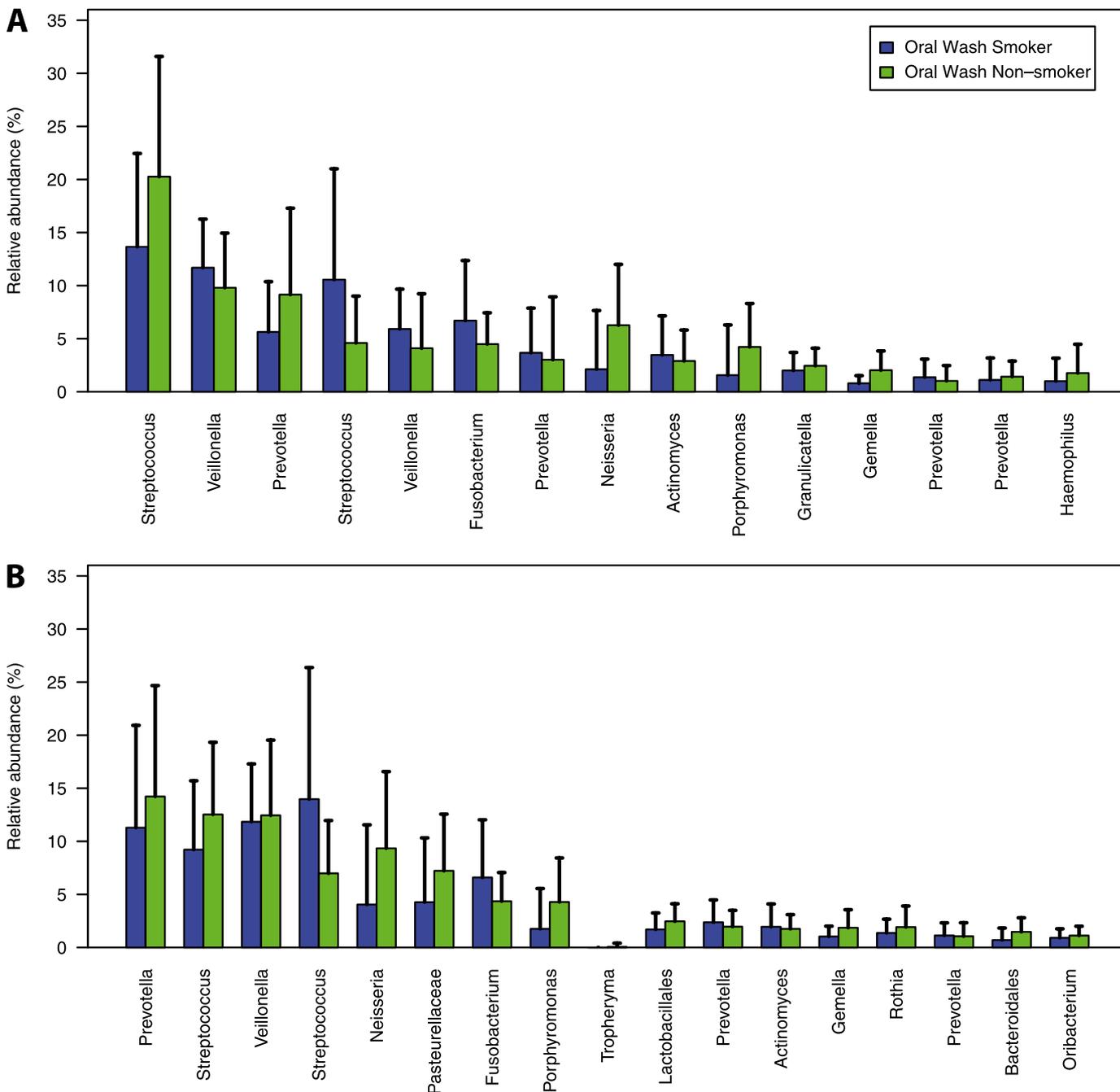


**Figure 3.** Relative abundance of bacterial operational taxonomic units (OTUs) observed in the bronchoalveolar lavage and oral wash samples from nonsmokers and smokers. These bacterial genera represent those populations that averaged more than 1% of the relative abundance across all samples when sequencing variable regions 1 through 3 (A) and variable regions 3 through 5 (B). Some genus names appear twice because multiple OTUs had the same consensus taxonomy.

*Prevotella*, *Fusobacterium*, *Haemophilus*, *Veillonella*, and *Porphyromonas* (7, 8). Organisms found to be more abundant in the lung in the current study included *Haemophilus* and *Enterobacteriaceae*. We also detected *Tropheryma whippelii* in about one-quarter of participants in BAL, but not in OW. This organism was previously reported in BAL from a single healthy individual (6). Our larger sample size, however, likely improved our ability to detect rare bacteria in the lung. These data provide evidence of a lung population that cannot be explained solely by origin in the mouth.

We examined differences in the respiratory tract microbiome in nonsmokers and smokers. Previous work has shown that the

oropharynx of smokers has a more diverse population than that of nonsmokers (14). In addition, other groups have reported enrichment of certain organisms in smokers as well as depletion of organisms, particularly of normal community members (14, 25, 26). We did not find differences in diversity of populations in the oropharynx of smokers compared with nonsmokers, nor did we find enrichment of particular species. We did find decreased relative abundance of *Neisseria*, *Porphyromonas*, and *Gemella* species in smokers. *Gemella* and *Neisseria* are members of the normal oral microbiota, and other studies have reported decreased abundance of *Neisseria* in smokers (6, 14). *Porphyromonas*, a bacteria linked to periodontal disease, is generally increased in



**Figure 4.** Relative abundance of bacterial genera observed in the oral wash samples from nonsmokers and smokers. These bacterial genera represent those populations that averaged more than 1% of the relative abundance across all samples when sequencing variable regions 1 through 3 (A) and variable regions 3 through 5 (B). Some genus names appear twice because multiple operational taxonomic units (OTUs) had the same consensus taxonomy.

smokers who show a decreased inflammatory response to this organism; however, we found that it was depleted in oral washes of smokers (14). Previous studies have generally directly sampled the subgingival region, possibly explaining differences from our results. Overall, these data confirm previous work suggesting that smoking disrupts the normal community structure in the mouth.

The current results provide important insights into the unsettled question of whether the lung microbiome in nonsmokers and smokers differs. In a study examining BAL in three never-smokers and seven healthy smokers, Erb-Downward and colleagues found that there were no significant differences between

these groups (7). In our larger cohort, we also found no significant differences in BAL composition when examined as a whole in the two groups; however, when we specifically examined particular Bacteroidetes and Proteobacteria, we found that these microbes had decreased relative abundance in BAL from smokers. These findings suggest that smoking has a more profound effect on the microbial composition of the upper respiratory tract, but may also impact the community of the lower respiratory tract in more subtle ways.

To compare systematic differences resulting from examination of respiratory samples using primers that amplify different hypervariable regions of the bacterial 16S gene, we sequenced

**TABLE 2. COMPARISON OF ALPHA DIVERSITY METRICS OBSERVED IN NONSMOKERS AND SMOKERS IN BRONCHOALVEOLAR LAVAGE AND ORAL WASH SAMPLES AS CHARACTERIZED USING VARIABLE REGIONS 1 THROUGH 3 AND VARIABLE REGIONS 3 THROUGH 5**

Region	Smoker	Site	Samples	Observed Richness	Shannon Index	Inverse Simpson Index	Phylogenetic Diversity
V1–3	No	BAL	37	58.4 (18.8)	2.90 (0.35)	10.9 (3.9)	2.26 (1.34)
	Yes	BAL	13	63.2 (28.0)	2.89 (0.49)	11.3 (5.5)	2.36 (1.52)
	No	OW	44	57.6 (13.9)	2.77 (0.34)	9.9 (3.4)	2.08 (1.41)
	Yes	OW	18	68.9 (24.0)	2.92 (0.44)	11.6 (4.4)	1.83 (1.67)
V3–5	No	BAL	16	54.0 (17.3)	2.60 (0.70)	8.8 (3.9)	2.75 (2.02)
	Yes	BAL	7	43.5 (18.1)	2.24 (1.01)	7.7 (5.6)	2.13 (1.63)
	No	OW	39	55.2 (12.8)	2.75 (0.30)	10.0 (2.9)	2.21 (1.44)
	Yes	OW	16	65.5 (18.4)	2.86 (0.46)	11.0 (4.7)	2.29 (1.39)

*Definition of abbreviations:* BAL = bronchoalveolar lavage; OW = oral wash; V1–3 = variable regions 1 through 3; V3–5 = variable regions 3 through 5.

All metrics are based on the average of rarefying samples to 1,000 sequences. BAL samples excluded at V1–3 and V3–5 if community structure resembled that of controls. Samples were also excluded at V3–5 if there were insufficient sequences.

samples using primers for both V1–3 and V3–5. There was general agreement between these regions, but more reads were obtained from V1–3, and individual organisms were better detected with V3–5. The differences seen with different primers indicate that caution must be used in comparing results between studies using different 16S regions for amplification.

This study extends previous work in several important ways. We included individuals from multiple geographic locations and performed sequencing at a single center to minimize variability. We had strict inclusion and exclusion criteria to ensure that participants had no underlying medical disorders, including lung disease, that could affect the respiratory microbiome. Smoking status was also rigorously defined to determine whether smoking leads to alterations in lung microbiota.

There are also several limitations. Although this study is the largest to date of the lower respiratory tract, we still may have lacked power to measure significant differences in community structures, and we were unable to adjust for factors such as sex or race. Several borderline analyses might have become significant with more participants. Power calculations have not been standardized for microbiota analyses, as statistics used differ from those in traditional epidemiologic or clinical investigations. Systematic assessment of the mouth was not performed; therefore, we could not determine the impact on periodontal disease or tooth condition on the lung microbiome. Although bronchoscopy protocols were standardized as much as possible, sample collection and processing methods may still have differed between sites, and analyses of replicate extractions or use of different models might have yielded different results (27). We did not detect clustering by study site, demonstrating that respiratory tract samples from multiple investigators can be pooled for central analysis if attention is paid to performing bronchoscopies in a similar fashion. Although we took multiple steps to minimize contamination during the procedure and from the environment and bronchoscope, the possibility of contamination still exists. Importantly, the similarity of our findings to Charlson and colleagues using a two-bronchoscope method (6) suggests that use of a single bronchoscope with appropriate precautions is a feasible approach for lower airway sampling if stringent analytic methods to control for upper airway contamination of lower airway samples are used. There may also be factors such as environmental exposures that influence the microbiome that we were not able to capture. Finally, the culture-independent techniques used for detecting the microbiome do not allow for determination of bacterial viability. However, the use of the neutral model allows us to identify organisms that appear to be enriched for growth in the lung. Although this is an indirect conclusion, we would infer that these bacteria are reproducing in the lung. In addition, even dead bacteria

may be clinically important, as they can still provoke an immune response.

In summary, we have performed the largest study to date of the respiratory microbiome in the healthy host. By applying the novel analytic technique of the neutral community model, we found that lung bacterial populations are similar to those in the oropharynx, but the lung also contains distinctive populations of organisms. Whether these organisms are viable and resident in the lung is difficult to determine, but the current findings suggest that a distinctive lung microbiome exists in some healthy individuals that does not arise solely from contamination or dispersal from the mouth. The role of these organisms and the nature of the immune response to them will be important areas of future research, including study of populations with altered immunity such as HIV-infected individuals.

**Author disclosures** are available with the text of this article at [www.atsjournals.org](http://www.atsjournals.org).

**Acknowledgment:** The authors thank the following individuals for assistance in patient recruitment, DNA sequencing, and data coordination:

**Clinical Centers**—Indiana University: R. B. Day, Q. Dong, X. Gao, N. Gebregziabher, R. Gregory, B. Katz, K. S. Knox, D. Mi, D. Munro, D. E. Nelson, K. Revanna, R. Rong, E. Toh, Y. Ye; University of California San Francisco: E. Auld, F. Calderon, S. Fong, A. Malki, S. Stone, S. Tokman; University of Colorado, Denver: A. Allhouse, J. Clemente, A. Cota-Gomez, M. Dsouza, K. Hammer, A. Hansen, L. Kerwin, M. Li, M. Lin, D. Linderman, B. Putnam, T. Rounds; University of Michigan: T. Ames, C. Freeman, T. Geal, L. McCloskey, A. McCubbrey, A. Myers, M. Reyes, J. Riddell; University of Pittsburgh: M. Busch, D. Camp, J. Dermand, S. Fong, M. P. George, M. Gingo, R. Greenblatt, R. Hoffman, L. Huang, C. Kessinger, E. Kleerup, T. Lawther, N. Leo, L. Lucht, J. Wang.

**Sequencing Centers**—University of California San Francisco: S. Iwai; University of Colorado, Boulder: R. Knight, C. Luzopone, L. Ursell; University of Michigan: G. Huffnagle; University of Pittsburgh: I. Astrovskaya, J. V. DePasse, A. Fitch, J. Paulson, M. Pop, K. Saira; Washington University School of Medicine: B. Herter, M. O’Laughlin, E. Applebaum, K. Mihindukulasuriya.

**Coordinating Center**—George Washington University: M. A. Foulkes, K. L. Drews, L. S. Firrell, Z. Maddipatla, H. Nilakanta, L. Tipton.

**Project Office**—National Institutes of Health Heart, Lung, and Blood Institute: S. Colombini-Hatch, H. Peavy, B. Schmetter.

The authors thank Emily S. Charlson, Ronald Collman, and Frederic Bushman (University of Pennsylvania) for performing DNA extraction on bronchoalveolar lavage samples as part of the DNA validation study. They also thank William T. Sloan (University of Glasgow), Thomas P. Curtis (University of Newcastle upon Tyne), and Mary T. Lunn (University of Oxford) for making their version of the neutral model available and valuable discussions regarding its adaption for our purposes.

A subset of subjects in this cohort was recruited from sites of the Multicenter AIDS Cohort Study (MACS) with centers University of California, Los Angeles (Roger Detels) and University of Pittsburgh (Charles R. Rinaldo, Lawrence Kingsley). The MACS is funded by the National Institute of Allergy and Infectious Diseases, with additional supplemental funding from the National Cancer Institute. UO1-AI-35042, U11-RR025005 (GCRC), UO1-AI-35043, UO1-AI-35039, UO1-AI-35040, UO1-AI-35041. Website located at <http://www.statepi.jhsph.edu/macs/macs.html>. A subset of subjects in this cohort was recruited from the The Connie Wofsy Study Consortium of Northern California (Ruth Greenblatt) of the Women’s Interagency HIV Study (WIHS) Collaborative Study Group.

The WIHS is funded by the National Institute of Allergy and Infectious Diseases (UO1-AI-35004, UO1-AI-31834, UO1-AI-34994, UO1-AI-34989, UO1-AI-34993, and UO1-AI-42590) and by the Eunice Kennedy Shriver National Institute of Child Health and Human Development (UO1-HD-32632). The WIHS is cofunded by the National Cancer Institute, the National Institute on Drug Abuse, and the National Institute on Deafness and Other Communication Disorders. Funding is also provided by the National Center for Research Resources (UCSF-CTSI Grant Number UL1 RR024131).

## References

1. The Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. *Nature* 2012;486:207–214.
2. Huse SM, Ye Y, Zhou Y, Fodor AA. A core human microbiome as viewed through 16S rRNA sequence clusters. *PLoS ONE* 2012;7:e34242.
3. Walker AW, Sanderson JD, Churcher C, Parkes GC, Hudspeth BN, Rayment N, Brostoff J, Parkhill J, Dougan G, Petrovska L. High-throughput clone library analysis of the mucosa-associated microbiota reveals dysbiosis and differences between inflamed and non-inflamed regions of the intestine in inflammatory bowel disease. *BMC Microbiol* 2011;11:7.
4. Turnbaugh PJ, Hamady M, Yatsunenko T, Cantarel BL, Duncan A, Ley RE, Sogin ML, Jones WJ, Roe BA, Affourtit JP, *et al.* A core gut microbiome in obese and lean twins. *Nature* 2009;457:480–484.
5. Yang F, Zeng X, Ning K, Liu KL, Lo CC, Wang W, Chen J, Wang D, Huang R, Chang X, *et al.* Saliva microbiomes distinguish caries-active from healthy human populations. *ISME J* 2012;6:1–10.
6. Charlson ES, Bittinger K, Haas AR, Fitzgerald AS, Frank I, Yadav A, Bushman FD, Collman RG. Topographical continuity of bacterial populations in the healthy human respiratory tract. *Am J Respir Crit Care Med* 2011;184:957–963.
7. Erb-Downward JR, Thompson DL, Han MK, Freeman CM, McCloskey L, Schmidt LA, Young VB, Toews GB, Curtis JL, Sundaram B, *et al.* Analysis of the lung microbiome in the “healthy” smoker and in COPD. *PLoS ONE* 2011;6:e16384.
8. Hilty M, Burke C, Pedro H, Cardenas P, Bush A, Bossley C, Davies J, Ervine A, Poulter L, Pachter L, *et al.* Disordered microbial communities in asthmatic airways. *PLoS ONE* 2010;5:e8578.
9. Sze MA, Dimitriu PA, Hayashi S, Elliott WM, McDonough JE, Gosselink JV, Cooper J, Sin DD, Mohn WW, Hogg JC. The lung tissue microbiome in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2012;185:1073–1080.
10. Iwai S, Fei M, Huang D, Fong S, Subramanian A, Grieco K, Lynch SV, Huang L. Oral and airway microbiota in HIV-infected pneumonia patients. *J Clin Microbiol* 2012;50:2995–3002.
11. Hubbell SP. Neutral theory in community ecology and the hypothesis of functional equivalence. *Funct Ecol* 2005;19:166–172.
12. Sapkota AR, Berger S, Vogel TM. Human pathogens abundant in the bacterial metagenome of cigarettes. *Environ Health Perspect* 2010;118:351–356.
13. Lee J, Taneja V, Vassallo R. Cigarette smoking and inflammation: cellular and molecular mechanisms. *J Dent Res* 2012;91:142–149.
14. Charlson ES, Chen J, Custers-Allen R, Bittinger K, Li H, Sinha R, Hwang J, Bushman FD, Collman RG. Disordered microbial communities in the upper respiratory tract of cigarette smokers. *PLoS ONE* 2010;5:e15216.
15. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, *et al.* Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* 2009;75:7537–7541.
16. Schloss PD, Gevers D, Westcott SL. Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. *PLoS ONE* 2011;6:e27310.
17. Sloan WT, Lunn M, Woodcock S, Head IM, Nee S, Curtis TP. Quantifying the roles of immigration and chance in shaping prokaryote community structure. *Environ Microbiol* 2006;8:732–740.
18. R Development Core Team. A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2012 [accessed 2012 Oct 1]. ISBN 3-900051. Available from: <http://www.R-project.org>
19. Agresti A, Coull BA. Appropriate is better than “exact” for interval estimation of binomial proportions. *Am Stat* 1998;52:199–226.
20. Magurran AE. Measuring biological diversity. Malden, MA.: Blackwell Publishing; 2004.
21. Faith DP. Phylogenetic pattern and the quantification of organismal biodiversity. *Philos Trans R Soc Lond B Biol Sci* 1994;345:45–58.
22. Yue JC, Clayton MK. A similarity measure based on species proportions. *Comm Statist Theory Methods* 2005;34:2123–2131.
23. Anderson MJ. A new method for non-parametric multivariate analysis of variance. *Austral Ecol* 2001;26:32–46.
24. Sokal RR, Rohlf FJ. Biometry: the principles and practice of statistics in biological research, 3rd ed. New York: Freeman; 1995.
25. Shchipkova AY, Nagaraja HN, Kumar PS. Subgingival microbial profiles of smokers with periodontitis. *J Dent Res* 2010;89:1247–1253.
26. Delima SL, McBride RK, Preshaw PM, Heasman PA, Kumar PS. Response of subgingival bacteria to smoking cessation. *J Clin Microbiol* 2010;48:2344–2349.
27. Charlson ES, Bittinger K, Chen J, Diamond JM, Li H, Collman RG, Bushman FD. Assessing bacterial populations in the lung by replicate analysis of samples from the upper and lower respiratory tracts. *PLoS ONE* 2012;7:e42786.