Title:

Multi-center Comparison of Lung and Oral Microbiomes of HIV-Infected and HIV-Uninfected Individuals

Authors:

James M. Beck1,2, Patrick D. Schloss3, Arvind Venkataraman3, Homer Twigg III4, Kathleen A. Jablonski5, Frederic D. Bushman6, Thomas B. Campbell1, Emily S. Charlson7, Ronald G. Collman7, Kristina Crothers8, Jeffrey L. Curtis3,10, Kimberly L. Drews5, Sonia C. Flores1, Andrew P. Fontenot1, Mary A. Foulkes5, Ian Frank7, Elodie Ghedin11, Laurence Huang12, Susan V. Lynch12, Alison Morris13, Brent E. Palmer1, Thomas M. Schmidt3, Erica Sodergren14, George M. Weinstock14, Vincent B. Young3, for the Lung HIV Microbiome Project

Affiliations:

1 Department of Medicine, University of Colorado Denver, Aurora, Colorado; 2 Veterans Affairs Eastern Colorado Health Care System, Denver, Colorado; 3 Department of Internal Medicine, University of Michigan, Ann Arbor, Michigan; 4 Department of Medicine, Indiana University, Indianapolis, Indiana; 5 Department of Epidemiology and Biostatistics, George Washington University, Washington, District of Columbia; 6 Department of Microbiology, University of Pennsylvania, Philadelphia, Pennsylvania; 7 Department of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania; 8 Department of Medicine, University of Washington, Seattle, Washington; 10 Veterans Affairs Eastern Colorado Health Care System, Denver, Colorado;
Affairs Ann Arbor Healthcare System, Ann Arbor, Michigan;\textsuperscript{11} Department of
Computational and Systems Biology, University of Pittsburgh, Pittsburgh, Pennsylvania;
\textsuperscript{12} Department of Medicine, University of California San Francisco, San Francisco,
California; \textsuperscript{13} Department of Medicine, University of Pittsburgh, Pittsburgh,
Pennsylvania; \textsuperscript{14} The Jackson Laboratory for Genomic Medicine, Farmington,
Connecticut

**Corresponding Author Information:**

James M. Beck, M.D.

Chief, Medicine Service (111)

Veterans Affairs Eastern Colorado Health Care System

1055 Clermont Street, Denver, Colorado, 80220

Email: james.beck@ucdenver.edu

**Author Contributions:**

Conception and design: J.M.B., P.D.S., A.V., H.T., K.A.J, F.D.B., T.B.C., R.G.C., J.L.C.,
Funding:

Supported by U.S. Public Health Service grants U01HL98961 (J.M.B., J.L.C., V.B.Y.), R01HG005975 (P.D.S.), U01HL098960 (H.T., G.M.W.), U01HL098958 (K.J., M.A.F., P.D.S., K.D.), U01 HL098957 (F.D.B, R.G.C), R01HL090342 (K.C.), U01HL098996 (T.B.C., S.C.F., A.P.F.), U01HL098962 (A.M., E.G.), U01HL098964 (L.H., S.V.L.), NIH/HMP RO100417746 (T.M.S.) and T32 HL007749 (A.V.). Additional NIH support was provided by University of Pittsburgh CTSI UL1 RR024153 and UL1 TR000005, the University of California San Francisco CTSI UL1 TR000004, and the University of Colorado CTSI UL1 TR001082.

Running Head: HIV-Infected and HIV-Uninfected Lung Microbiomes

Descriptor: 10.2 AIDS-Related Lung Disease

Word Count: 3,500

At a Glance Commentary:

Scientific Knowledge on the Subject: Analysis of the lung microbiome is a relatively new field and may suggest novel therapeutic approaches to pulmonary infections and disease processes. Differences in oral and lung microbiomes between HIV-infected and HIV-uninfected subjects require further definition, including possible differences in microbiome composition in HIV-infected subjects who are treatment naïve and those receiving anti-retroviral therapy.

What This Study Adds to the Field: A multi-center study enrolled subjects who were
HIV-negative, HIV-infected and treatment naïve, and HIV-infected and receiving anti-retroviral therapy. We compared oral washes and bronchoalveolar lavages, and used a novel neutral model procedure to exclude environmental contaminants. We demonstrated significant differences among the groups in oral washes, but bronchoalveolar lavages were similar. Each subject group demonstrated differences between oral washes and bronchoalveolar lavages. There was no relationship between CD4 counts and microbiome data.

Online Data Supplement: This article has an online data supplement, which is accessible from this issue’s table of content online at www.atsjournals.org.
Abstract

Rationale: Improved understanding of the lung microbiome in HIV-infected individuals could lead to better strategies for diagnosis, therapy, and prophylaxis of HIV-associated pneumonias. Differences in the oral and lung microbiomes in HIV-infected and HIV-uninfected individuals are not well defined. Whether highly active anti-retroviral therapy influences these microbiomes is unclear.

Objectives: We determined whether oral and lung microbiomes differed in clinically healthy groups of HIV-infected and HIV-uninfected subjects.

Methods: Participating sites in the Lung HIV Microbiome Project contributed bacterial 16S rRNA sequencing data from oral washes and bronchoalveolar lavages (BALs) obtained from HIV-uninfected individuals (n = 86), HIV-infected individuals who were treatment naïve (n = 18), and HIV-infected individuals receiving anti-retroviral therapy (n =38).

Measurements and Main Results: Microbial populations differed in the oral washes among the subject groups (Streptococcus, Actinomyces, Rothia, and Atopobium), but there were no individual taxa that differed among the BALs. Comparison of oral washes and BALs demonstrated similar patterns from HIV-uninfected individuals and HIV-infected individuals receiving anti-retroviral therapy, with multiple taxa differing in abundance. The pattern observed from HIV-infected individuals who were treatment naïve differed from the other two groups, with differences limited to Veillonella, Rothia, and Granulicatella. CD4 cell counts did not influence the oral or BAL microbiome in these relatively healthy, HIV-infected subjects.
Conclusions: The overall similarity of the microbiomes in participants with and without HIV infection was unexpected, since HIV-infected individuals with relatively preserved CD4 cell counts are at higher risk for lower respiratory tract infections, indicating impaired local immune function.

Word Count: 248

Key Words: lung; microbiome; HIV infection; bronchoscopy; bronchoalveolar lavage
Introduction

The diversity and frequency of infections and other pathogenic lung processes in HIV-infected individuals suggests significant impairments in host defense in this population. A better understanding of the respiratory tract microbiome in HIV-infected individuals, including the lung and the upper respiratory tract, could lead to improved strategies for diagnosis, therapy, and prophylaxis of HIV-associated pneumonias in this population.

The lung was long considered to be a sterile compartment, but recent molecular investigations from multiple groups demonstrate that the lung contains bacterial DNA sequences, although at low levels relative to the upper respiratory tract (1). However, clarifying the relationship between the microbiome of the lung and the microbiome of other compartments, including the oropharynx, presents a challenge for lung microbiome analysis (2, 3). Because sampling of the lower airways by bronchoscopy depends upon insertion of a bronchoscope through the nose or mouth, careful methodologic attention to possible contamination from the upper airway is required (4). Environmental sources that may confound the lung data must be evaluated, particularly in low bio-burden samples. While bacterial taxa in the lung closely resemble those of the upper airway from the oropharynx (5), there seem to be differences between the mouth and the lungs in both health and disease (6, 7). A growing body of literature documents changes in the microbiome that occur during pulmonary disease processes, including COPD (8-11), bronchiectasis (12), asthma (13-15), lung transplantation (16-19), bronchopulmonary dysplasia (20, 21), and cystic fibrosis (22, 23).
The Lung HIV Microbiome Project (LHMP) is a multicenter consortium established by the National Heart, Lung, and Blood Institute to understand the respiratory tract microbiome during HIV infection. Previous work from LHMP compared smokers and non-smokers without HIV infection, and demonstrated that the oropharyngeal microbiome in smokers differs from that in non-smokers (6). However, significant differences were not detected in the lung microbiome between smokers and non-smokers. The LHMP then demonstrated that in HIV-infected individuals there is widespread colonization of the lung by the bacterium *Tropheryma whipplei*, and the institution of anti-retroviral therapy decreases the relative abundance of this organism (7). However, the broader bacterial community characteristics of the upper and lower respiratory tract were not addressed in HIV-infected individuals in this study.

In the current study, we carried out a comprehensive analysis of upper and lower respiratory tract bacterial microbiomes from individuals with and without HIV infection. We examined volunteers who were HIV-uninfected, who were HIV-infected but untreated with anti-retroviral therapy, and who were HIV-infected and effectively treated with anti-retroviral therapy. We chose to study clinically healthy individuals, rather than those with active pulmonary infections or other disease processes, to understand baseline changes in microbial populations that occur during HIV infection and treatment. These studies will help to inform further investigations aimed at understanding HIV-associated lung diseases.
Methods

Study Design

The LHMP performed a prospective, multisite cohort study that was approved by the institutional review boards at each participating site. A second level of institutional review was provided by the Data Coordinating Center and the NHLBI, including an Observational Safety Monitoring Board. All subjects provided written informed consent as approved by the review board at each site.

Participants

Participants were healthy individuals with and without HIV infection, recruited from eight cities participating in LHMP. Men and women aged 18 to 80 years were eligible. We enrolled HIV-negative subjects (referred to as “Negative”), HIV-positive subjects who had not been treated with any anti-retroviral medications (referred to as “Naïve”), and HIV-positive subjects currently receiving highly active anti-retroviral medications (referred to as “HAART”).

Subjects completed questionnaires including medical history, use of medications including non-prescription drugs, illicit drugs, tobacco, and/or marijuana. At the time of microbiome sampling, individuals were asymptomatic for upper respiratory or lung infections and had no reported fever, cough, or other acute respiratory symptoms during the previous 4 weeks. Subjects had not used any antibacterial medications in the previous 3 months, and had not used corticosteroids (including inhaled medications) in the previous 6 months. Individuals with known respiratory diagnoses, including COPD and asthma, were ineligible for enrollment.
Sample Collection

Oropharyngeal and lung samples were collected as previously described (6). Smokers were asked to stop smoking at least 12 hours before collection, and all subjects were instructed to avoid food and water from midnight before the procedure. Subjects gargled with 10 ml sterile 0.9% saline at the start of the procedure, and this oral wash sample was collected and iced immediately. To potentially reduce the presence of bacteria in the upper airway, all subjects then gargled with antiseptic mouthwash immediately prior to bronchoscopy. Bronchoscopy was performed as previously described (6). Using minimal sedation, the bronchoscope was inserted through the mouth and was advanced quickly to the wedge position in the right middle lobe or lingula. Bronchoalveolar lavage (BAL) was collected by instillation and aspiration of 0.9% sterile saline to a maximum instillation volume of 300 ml. Subjects from the University of Pennsylvania cohort underwent bronchoscopy with a two-scope method as previously described (5). Samples obtained by this method did not differ systematically from those obtained at the other sites (data not shown), thus data were combined for the current analysis.

Controls obtained at the time of sampling included a neat sample of the sterile saline immediately after the container was opened and 10-50 ml of the same saline aspirated through the suction channel of the designated bronchoscope before the procedure. Controls were processed and sequenced in parallel with the subjects’ samples.

Sample Processing, Sequencing, and Curation

DNA was extracted at each site, using consensus protocols. To assure consistency in sequencing, all samples were sequenced at Washington University. Sequencing was
performed on a Roche 454 FLX Titanium platform using primers for variable regions 1 through 3 (V1-3) and regions 3 through 5 (V3-5). Successful sequencing was substantially more robust for V1-3, compared with V3-5. Additionally, V3-5 primers resulted in less specific amplifications that included more host sequences (data not shown). Therefore, we focused on the V1-3 sequences for this study. The 16S rRNA sequences were assigned to organizational taxonomic units (OTUs) using the mothur software package according to previously described methods (6, 24, 25).

Neutral Model Analysis for Potential Contaminant Sequences

We previously used the neutral model to examine enrichment of oral versus lung samples (6, 26). In this analysis, frequency describes the percentage of samples in which an OTU was detected, and abundance describes the percentage of sequences assigned to an individual OTU. For the present study, we used the neutral model to identify sequences in BAL samples that may have arisen from contamination introduced by the bronchoscope or from reagents used in sample collection or processing. In the implementation of the model, the controls (bronchoscope and reagents) were considered as sources contributing DNA sequences to the BAL samples (Figure 1).

Statistics

Demographic characteristics were compared using Chi-Square tests or Fisher’s exact test for categorical variables, analyses of variance for normally distributed continuous variables, and the Wilcoxon rank test for skewed data (for example, in comparing CD4 counts between groups). Alpha diversity was evaluated by comparing the number of observed OTUs and the Shannon Diversity Index by rarefying each sample to 500 sequences with 1,000 randomizations. The Shannon Diversity Index
combines measurements of evenness (distribution of OTUs across a population) and richness (number of different OTUs in a population) to express diversity numerically. After randomly selecting 500 sequences per sample with 1,000 randomizations, the distances between populations were measured using $\Theta_{YC}$ (Theta$_{YC}$) as previously described (27). $\Theta_{YC}$ measures dissimilarity between the structures of communities. We tested for differences in the average distance between each group of samples by analysis of molecular variance (AMOVA), a statistical tool analogous to ANOVA (28). OTUs were compared by the Wilcoxon signed-rank test for all OTUs with average relative abundances of greater than 1% across all samples. These tools produce similar results to the weighted and unweighted UniFrac distances (data not shown); we chose the $\Theta_{YC}$ distance metric over the UniFrac metrics because the former is based on the relative abundance of individual OTUs while the latter is based on a phylogenetic tree without the use of OTU designations.

Results

Demographics

We enrolled 142 subjects from the clinical LHMP sites (Table 1). Eighty-six subjects were HIV-negative (“Negative”), 18 subjects were HIV positive but had not received anti-retroviral medications (“Naïve”), and 38 subjects were HIV positive and treated with highly active anti-retroviral medications (“HAART”). The groups were not matched for demographic characteristics. The Negative group contained a higher proportion of females compared with both HIV-positive groups, and the HAART group was older than the other groups. The HIV-positive groups contained larger proportions of African-
Americans, and the Naïve group contained more Hispanics.

As expected, plasma HIV RNA was significantly lower in the HAART group compared with the Naïve group. However, the CD4 counts in the Naïve and HAART groups did not differ significantly and reflected relatively preserved immune function in this clinically healthy cohort. There were no significant differences in smoking status among the groups, with each cohort containing non-smokers, current smokers, and former smokers. Analysis by AMOVA revealed no differences in microbiome data by smoking status when the data were corrected for multiple comparisons. Therefore, data were combined for analysis. While we excluded subjects who reported antibacterial use within 3 months of sampling, we also examined more distant antibacterial use. The Naïve group contained a higher proportion of subjects who had used antibacterial medications in the 3-6 months before enrollment, but the difference was not statistically significant.

Neutral Model Application, Sequencing Yield, Observed OTUs and Diversity Indices

We applied the neutral model to the BAL samples to identify and to remove sequences that might have resulted from contamination sources (Figure 1). After curation and removal of sequences potentially derived from control sources, samples containing ≥ 500 sequences were subject to analysis. The increased microbial biomass present in oral wash samples, compared to BAL samples, resulted in improved 16S rRNA sequence amplification and successful sequencing of 136 oral wash samples (Table 2). Despite the relatively low biomass present in BAL, DNA recovery was adequate and sequencing was successful for 111 BAL samples. Sequences were
binned at the 97% identity levels and were assigned OTUs as described in the Methods.

Numbers of observed OTUs in the oral washes and BALs did not differ among groups (see the online data supplement). Similarly, the Shannon Diversity Index did not differ among groups (see the online data supplement).

**Community Structure Comparison by Subject Group**

We compared nonmetric multidimensional scaling (NMDS) plots using all V1-3 sequences from oral washes obtained from subjects in the 3 groups (Figure 2, Panel A). NMDS plots provide two-dimensional visualization of differences between members of a population, with similar members grouping together, and this analysis allows measurement of overall differences in microbial populations among the subject groups. Oral washes from the Negative subjects differed significantly from those obtained from Naïve subjects ($p < 0.01$) and also differed significantly from those obtained from HAART subjects ($p = 0.01$). In comparison of the two HIV-infected groups, there were also significant differences between the oral washes obtained from Naïve and HAART subjects ($p = 0.01$). Thus, oral microbial populations differ significantly between HIV-uninfected and infected subjects, and also differ significantly according to HIV treatment status.

We then compared all sequences from BALs from the subjects in the 3 groups (Figure 2, Panel B). Unlike the oral washes, we did not detect statistically significant differences among the groups. Therefore, our findings do not support differences in the microbial populations sampled in BAL according to HIV infection status or by HIV treatment status.
Community Structure Comparison by Body Site

We next examined the relationships in microbial communities between oral washes and BALs in each subject group. There were significant differences in communities between oral wash and BALs in the Negative group (Figure 3, Panel A; p < 0.01), in the Naïve group (Figure 3, Panel B; p = 0.03), and in the HAART group (Figure 3, Panel C; p < 0.01). These results support our previous investigations (6) and provide further evidence that the lung compartment, as sampled by BAL, does not completely mirror the oropharyngeal compartment.

Differences in Taxa by Subject Group

We examined the taxa corresponding to the most abundant sequences to determine which specific taxa accounted for the differences in community structure described above. Comparison of sequences in all oral washes demonstrated that a *Streptococcus* OTU and two *Actinomyces* OTUs were significantly more abundant in Naïve subjects compared with the other groups, and an Atopobium OTU was significantly more abundant in Naïve and HAART groups compared with Negative subjects (Figure 4). Additionally, a *Rothia* OTU was significantly more abundant in oral washes from the HAART group compared with the other groups. In contrast to the oral washes, there were no significant differences in the most common taxa identified in BALs from the 3 subject groups (Figure 5).

Differences in Taxa by Body Site

Comparison of oral washes and BALs from the Negative subjects demonstrated significant differences in abundances of 11 OTUs, including increased abundances of *Streptococcus, Veillonella, Prevotella, Fusobacterium, Rothia, Porphyromonas,*
Granulicatella, and Gemella OTUs in oral washes compared with BAL (Figure 6A). In contrast, a Veillonella, a Tropheryma, and a Prevotella OTU were found in greater abundance in BAL compared with oral wash. For the Naïve subjects, only Veillonella, Rothia and Granulicatella OTUs were more abundant in oral wash than in BAL (Figure 6B). The pattern of differences in the HAART group resembled the Negative group much more closely, including the increased abundance of Tropheryma in BAL compared with oral wash samples (Figure 6C). Thus, each group of subjects demonstrated a unique pattern in the taxa differing in oral washes and BALs.

Effect of CD4 Count on Microbiome

Finally, we examined whether CD4 count influenced oral wash or BAL microbiome composition in the two groups of HIV-positive subjects. When CD4 count was evaluated as a dichotomous value, there were no significant differences regardless of the nominal value of CD4 used as a cutoff (data not shown). We also determined whether CD4 count, evaluated as a continuous variable, might modulate the microbiome, and again determined that there were no overall differences, including examination of specific OTUs (Figures E1 and E2 in the online data supplement).

Discussion

This large, multi-center study conducted at sites across the United States contributes unique information about the oral and lung microbiomes in healthy HIV-uninfected and HIV-infected individuals and represents the first data directly comparing these subject groups. We did not detect overall differences in numbers of OTUs or in the Shannon Diversity Index. Microbial populations in oral washes differed between
HIV-uninfected and HIV-infected subjects, but BAL populations did not differ significantly. The most surprising finding in our study was the overall similarity of the microbiota in subjects with and without HIV infection. This result was unexpected, since HIV-infected individuals (even with relatively preserved CD4 counts) are at higher risk for lower respiratory tract bacterial infections, including bacterial pneumonia and tuberculosis, indicating significantly impaired local immune function. We did identify greater differences in oral washes than in BALs. It is possible that these differences reflect the greater complexity of the oropharyngeal compartment and/or increased exposure to the environment, but we did observe significant differences in our subject groups related to HIV infection and treatment status. Alterations in the oral microbiome have important implications for disease processes (29), such as periodontal diseases (30), but may precede changes in the lung. Previous work by one of the LHMP collaborators identified increased abundances for a large number of taxa in oral samples from HIV-infected patients with pneumonia, in comparison to HIV-uninfected controls (31). A culture-based study identified abundant *Streptococcus spp.* and *Staphylococcus spp.* in oral samples from HIV-infected subjects, but uninfected subjects were not examined in this study (32). A study of the lingual microbiome associated increased proportions of *Veillonella*, *Prevotella*, *Megasphaera*, and *Campylobacter* with HIV viremia (33).

The current study extends earlier observations from the LHMP cohorts. We previously compared HIV-uninfected smokers and non-smokers, and found that the lung microbiome in HIV-uninfected individuals was largely derived from the oral microbiome, except for some unique lung species such as *Tropheryma whipplei* (6). The mouth
microbiome differed in smokers and nonsmokers, while smoking did not alter the lung microbiome significantly. The LHMP then conducted a targeted analysis of colonization of the lung by *Tropheryma whipplei* in the lungs of HIV-infected individuals and also demonstrated that the use of anti-retroviral therapy decreased the relative abundance of this organism (7). In the current, larger cohort, we adopted a different approach by combining smokers and non-smokers for analysis (as there were no significant differences in smoking status between subject groups), by analyzing all OTUs, and by approaching potential environmental controls conservatively using the neutral model. In comparing relative abundances of OTUs in oral washes and BALs, we found more OTUs that differed in the Negative and HAART subjects compared with the Naïve subjects. Interestingly, we observed differences in *T. whipplei* in Negative and HAART subjects, but not in Naïve subjects.

Our BAL results agree closely with published data obtained from subjects without respiratory disease, and further support the current consensus regarding communities of bacteria identified in the lungs (6, 10, 13, 15, 34). The possible pathogenic, or protective, effects of these taxa require further investigation. However, it is likely that groups of microbes establish lung environments that may encourage or inhibit growth of pathogens or inflammatory reactions. It is important to note that all subject groups in this study were clinically well at the time of enrollment and sampling. Very different results are observed during episodes of clinical pneumonia and with decreased CD4 counts (35).

There are several unique aspects of the current study that advance the literature in this field. First, we collected samples from individuals at eight different clinical sites,
while most of the existing literature examines subjects enrolled from a single center. Previous, small studies have suggested geographic variation, for example in sputa obtained from cystic fibrosis patients in the United States and the United Kingdom (36). We did not detect systematic variation among our sites. The current results support our previous multi-center investigation in which patterns of diversity were not driven by cohort location (7). Next, none of the patients received oral or inhaled corticosteroids. The role of inhaled corticosteroids in modulation of the upper and/or lower airway microbiota remains unclear (14).

Several limitations of the present study require comment. Our groups were not matched demographically, and it is possible that differences influenced the results. However, the statistical power to analyze these factors post-hoc is problematic given the small numbers in the subgroups. We focused on bacteria and did not analyze viruses or fungi. We also did not analyze gut microbiota or the subjects’ diets, but it is likely that the gut and diet have significant influence on the upper and lower airway microbiota (11). In mice, the gut microbiota are intimately associated with pulmonary phenotypes, including allergic and asthma-like conditions (37, 38). Next, the samples were obtained at a single point in time, although several of the LHMP sites are conducting serial sampling. A recent analysis of the Human Microbiome Project (HMP) data demonstrated that samples from oral communities were the least stable over time, in comparison to the relative stability of gut samples (39). We chose oral washes as integrative samples of the mouth, realizing that it was impractical to control accurately for differences in oral hygiene. Thus, small differences in composition of bacterial communities in different regions of the oropharynx may impact the most appropriate
comparator against which to assess BAL specimens (40). Bronchoscopy, while providing the only practical method to sample the lung in large numbers of subjects, runs an inherent risk of contamination from the oropharynx and upper airway during the procedure. The lower biomass of BAL samples from clinically well individuals necessitated the sequencing depth we reported, and future investigations using higher biomass samples should consider deeper sequencing. It also appears that the lung microbiome in individuals who are clinically well and without lung disease is largely derived from the oral microbiome, making it challenging to distinguish carry-over from true members of the community (6). While sampling during surgery can approach a gold standard (9, 10), this methodology is less practical for studies examining large numbers of subjects or for performing serial sampling.

In conclusion, this study compares the oropharyngeal and lung microbiomes in a healthy population of HIV-uninfected and HIV-infected individuals in a large, multi-center cohort. Oral bacterial communities differed significantly depending on HIV status, while lung communities sampled by BAL were similar. We identified specific taxa in oral samples that differed among the subject groups. Within each subject group, we identified a unique signature of taxa differing between the mouth and the lung; the patterns in the Negative and HAART subjects were similar to one another, but differed from the pattern in the Naïve subjects. These LHMP data serve as a basis for further investigations of the role of the lung microbiome during health and disease in HIV-infected individuals (3).
Acknowledgments

The authors thank Sandra Colombini-Hatch, Hannah Peavey, Lis Caler, and James Kiley for NHLBI’s leadership in lung microbiome research. The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of the NIH. Data in this manuscript were collected by the Multicenter AIDS Cohort Study (MACS) with centers at Los Angeles (U01-AI35040): University of California Los Angeles, Schools of Public Health and Medicine: Roger Detels (PI), Otoniel Martínez-Maza (Co-PI), Aaron Aronow, Robert Bolan, Elizabeth Breen, Anthony Butch, Beth Jamieson, Eric N. Miller, John Oishi, Harry Vinters, Dorothy Wiley, Mallory Witt, Otto Yang, Stephen Young, Zuo Feng Zhang; Pittsburgh (U01-AI35041): University of Pittsburgh, Graduate School of Public Health: Charles R. Rinaldo (PI), Lawrence A. Kingsley (Co-PI), James T. Becker, Ross D. Cranston, Jeremy J. Martinson, John W. Mellors, Anthony J. Silvestre, Ronald D. Stall; and the Data Coordinating Center (UM1-AI35043): The Johns Hopkins University Bloomberg School of Public Health: Lisa P. Jacobson (PI), Alvaro Munoz (Co-PI), Alison, Abraham, Keri Althoff, Christopher Cox, Jennifer Deal, Gypsyamber D’Souza, Priya Duggal, Janet Schollenberger, Eric C. Seaberg, Sol Su, Pamela Surkan. The MACS is funded primarily by the National Institute of Allergy and Infectious Diseases (NIAID), with additional co-funding from the National Cancer Institute (NCI). MACS data collection is also supported by UL1-TR000424 (JHU CTSA). Website located at http://www.statepi.jhsph.edu/macs/macs.html. Data in this manuscript were collected by the Women's Interagency HIV Study (WIHS) Collaborative Study Group with centers (Principal Investigators) at The Connie Wofsy Study Consortium of Northern California
(Ruth Greenblatt). The WIHS is funded by the National Institute of Allergy and Infectious Diseases (U01-AI-35004, U01-AI-31834, U01-AI-34994, U01-AI-34989, U01-AI-34993, and U01-AI-42590) and by the Eunice Kennedy Shriver National Institute of Child Health and Human Development (U01-HD-32632). The study is co-funded 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


HIV-infected pneumonia patients is compositionally and functionally distinct from that of San Franciscan patients. *PLos ONE* 2014; 9: e95726.


Figure Legends

Figure 1. Application of the neutral model to assess possible contaminant sequences identified in control samples. The bronchoscope, saline solution for sample collection, and extraction reagents were considered as potential sources of contaminating DNA. Sequences consistent with the model prediction (grey dots) and sequences enriched in the controls (red dots) were considered as potential contaminants and removed from subsequent analyses. Sequences over-represented in BAL samples (green dots) were retained (Panel A). The 10 most abundant sequences excluded from subsequent analysis (Panel B) and retained for analysis (Panel C) are shown. Some genus names are repeated in panels B and C as they reflect different sequences with the same taxonomic classification.

Figure 2. NMDS plots for all subjects showing oral wash (Panel A) and BAL (Panel B). Samples were obtained as detailed from HIV-negative subjects (“Negative”, red circles), HIV-positive but treatment naïve subjects (“Naïve”, blue squares), and HIV-positive subjects treated with highly active anti-retroviral therapy (“HAART”, green triangles). Analysis is based on OTUs from sequences of the V1-V3 regions of the 16S rRNA. Centroids are indicated by crosses. For oral washes, NMDS analysis demonstrated significant differences in OTUs between Negative and Naïve subjects (p < 0.01), and between Negative and HAART subjects (p = 0.01), and between Naïve and HAART subjects (p = 0.01). There were no significant differences in the BALs.

Figure 3. NMDS plots comparing oral washes (open symbols) and BALs (closed symbols) within subject groups. Individual subjects are linked by lines. Samples were
obtained as detailed in Methods and V1-V3 regions were sequenced. There were significant differences in oral wash and BAL samples within Negative (Panel A, p < 0.01), Naïve (Panel B, p = 0.03), and HAART (Panel C, p< 0.01) subjects.

Figure 4. Relative abundances of the most common OTUs identified in oral washes from all subject groups. Samples were obtained as detailed in Methods and V1-V3 regions were sequenced. OTUs were compared by the Wilcoxon signed-rank test for all OTUs with average relative abundances of greater than 1% across all samples, and significant differences are demonstrated by asterisks. *Streptococcus* and *Actinomyces* were more abundant in oral washes from Naïve subjects compared with the other groups. *Rothia* was more abundant in oral washes from HAART subjects compared with the other groups. Some genus designations appear more than once because multiple OTUs have the same consensus taxonomy.

Figure 5. Relative abundances of the most common OTUs identified in BAL from all subject groups. Samples were obtained as detailed and V1-V3 regions were sequenced. OTUs were compared by the Wilcoxon signed-rank test for all OTUs with average relative abundances of greater than 1% across all samples. Overall there were no significant differences. Some genus designations appear more than once because multiple OTUs have the same consensus taxonomy.

Figure 6. Relative abundances of the most common OTUs identified in oral wash versus BAL for each subject group. Samples were obtained as detailed in Methods and V1-V3 regions were sequenced. OTUs were compared by the Wilcoxon signed-rank test for all OTUs with average relative abundances of greater than 1% across all
samples. The OTUs indicated by asterisks demonstrated significant differences between BAL and oral wash in Negative (Panel A), Naïve (Panel B) and HAART (Panel C) subjects. Some genus designations appear more than once because multiple OTUs have the same consensus taxonomy.
Table 1. Characteristics of Participants.

<table>
<thead>
<tr>
<th></th>
<th>HIV negative (n = 86)</th>
<th>HIV/naïve (n =18)</th>
<th>HIV/HAART (N = 38)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sex, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>50 (58)</td>
<td>16 (89)</td>
<td>31 (82)</td>
<td>&lt; 0.01*</td>
</tr>
<tr>
<td>Female</td>
<td>36 (42)</td>
<td>2 (11)</td>
<td>7 (18)</td>
<td></td>
</tr>
<tr>
<td><strong>Age, mean ± SD</strong></td>
<td>44.9 ± 13.0</td>
<td>41.3 ± 7.8</td>
<td>51.7 ± 8.5</td>
<td>&lt; 0.01 †</td>
</tr>
<tr>
<td><strong>Race, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>56 (65)</td>
<td>6 (33)</td>
<td>18 (47)</td>
<td>&lt; 0.01 ‡</td>
</tr>
<tr>
<td>Black</td>
<td>26 (30)</td>
<td>9 (50)</td>
<td>20 (53)</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>4 (5)</td>
<td>3 (17)</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td><strong>Ethnicity, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hispanic</td>
<td>5 (6)</td>
<td>4 (22)</td>
<td>3 (8)</td>
<td>0.02 ‡</td>
</tr>
<tr>
<td>Not Hispanic</td>
<td>81 (94)</td>
<td>13 (72)</td>
<td>35 (92)</td>
<td></td>
</tr>
<tr>
<td><strong>Decline to State</strong></td>
<td>1 (6)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>CD4 count/µl, median (min-max)</strong></td>
<td>ND</td>
<td>668 (290-1,192)</td>
<td>618 (208-1,265)</td>
<td>0.62 †</td>
</tr>
<tr>
<td><strong>Plasma HIV RNA copies/ml, median (min-max)</strong></td>
<td>ND</td>
<td>6,432 (38-88,246)</td>
<td>36 (10-604)</td>
<td>&lt; 0.01 §</td>
</tr>
<tr>
<td><strong>Smoking status, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-smoker</td>
<td>46 (53)</td>
<td>5 (28)</td>
<td>15 (40)</td>
<td>0.25 *</td>
</tr>
<tr>
<td>Current</td>
<td>30 (35)</td>
<td>9 (50)</td>
<td>18 (47)</td>
<td></td>
</tr>
<tr>
<td>Former</td>
<td>10 (12)</td>
<td>4 (22)</td>
<td>5 (13)</td>
<td></td>
</tr>
<tr>
<td>Antibiotics &lt; 6 months, n (%)</td>
<td>7 (8)</td>
<td>4 (22)</td>
<td>5 (13)</td>
<td>0.14 *</td>
</tr>
</tbody>
</table>

ND: not done. * Chi-square test. † ANOVA. ‡ Fisher’s exact test. § Wilcoxon rank test.
<table>
<thead>
<tr>
<th></th>
<th>HIV negative (n = 86)</th>
<th>HIV/naïve (n = 18)</th>
<th>HIV/HAART (n = 38)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oral Wash</td>
<td>80</td>
<td>18</td>
<td>38</td>
</tr>
<tr>
<td>BAL</td>
<td>67</td>
<td>14</td>
<td>30</td>
</tr>
</tbody>
</table>

Samples considered successfully sequenced contained at least 500 sequences after curation and contaminant removal.
A. 

B. 

Sequence classification | Abundance (%) | Frequency (%) |
---|---|---|
Unclassified Actinomycetales | 6.8 | 14.9 | 18 | 20 |
Unclassified Bacillales | 1.7 | 2.1 | 36 | 21 |
*Ralstonia* | 1.4 | 11.3 | 60 | 70 |
Unclassified Rhizobiales | 0.9 | 5.6 | 36 | 42 |
*Ralstonia* | 0.8 | 9.1 | 32 | 29 |
*Rhizobium* | 0.8 | 4.0 | 25 | 39 |
*Pelomonas* | 0.6 | 2.2 | 29 | 29 |
*Streptococcus* | 0.7 | 0.4 | 5 | 5 |
*Streptococcus* | 0.5 | 0.2 | 7 | 6 |
*Bosea* | 0.3 | 2.6 | 25 | 26 |

C. 

Sequence classification | Abundance (%) | Frequency (%) |
---|---|---|
*Streptococcus* | 10.7 | 1.9 | 70 | 25 |
*Veillonella* | 3.2 | 0.3 | 51 | 11 |
*Veillonella* | 2.7 | 0.5 | 70 | 14 |
*Streptococcus* | 2.6 | 0.5 | 65 | 15 |
*Streptococcus* | 1.6 | 0.2 | 55 | 9 |
*Prevotella* | 1.5 | 0.4 | 35 | 8 |
*Trophia* | 1.6 | 0.0005 | 10 | 0.4 |
*Veillonella* | 1.5 | 0.06 | 23 | 4 |
*Prevotella* | 1.5 | 0.4 | 35 | 8 |
*Prevotella* | 1.2 | 0.2 | 27 | 7 |
<table>
<thead>
<tr>
<th>Bacterial Species</th>
<th>Wash Negative</th>
<th>Wash Naive</th>
<th>Wash HAART</th>
</tr>
</thead>
<tbody>
<tr>
<td>Streptococcus</td>
<td>20%</td>
<td>22%</td>
<td>18%</td>
</tr>
<tr>
<td>Veillonella</td>
<td>15%</td>
<td>18%</td>
<td>16%</td>
</tr>
<tr>
<td>Prevotella</td>
<td>10%</td>
<td>12%</td>
<td>8%</td>
</tr>
<tr>
<td>Veillonella</td>
<td>5%</td>
<td>7%</td>
<td>4%</td>
</tr>
<tr>
<td>Fusobacterium</td>
<td>1%</td>
<td>2%</td>
<td>1%</td>
</tr>
<tr>
<td>Neissera</td>
<td>2%</td>
<td>3%</td>
<td>1%</td>
</tr>
<tr>
<td>Actinomyces</td>
<td>2%</td>
<td>3%</td>
<td>1%</td>
</tr>
<tr>
<td>Rothia</td>
<td>1%</td>
<td>2%</td>
<td>1%</td>
</tr>
<tr>
<td>Porphyromonas</td>
<td>1%</td>
<td>2%</td>
<td>1%</td>
</tr>
<tr>
<td>Granulicatella</td>
<td>1%</td>
<td>2%</td>
<td>1%</td>
</tr>
<tr>
<td>Tropheryma</td>
<td>0.5%</td>
<td>1%</td>
<td>0.5%</td>
</tr>
<tr>
<td>Prevotella</td>
<td>0.5%</td>
<td>1%</td>
<td>0.5%</td>
</tr>
<tr>
<td>Gemella</td>
<td>0.5%</td>
<td>1%</td>
<td>0.5%</td>
</tr>
<tr>
<td>Prevotella</td>
<td>0.5%</td>
<td>1%</td>
<td>0.5%</td>
</tr>
<tr>
<td>Megaspheara</td>
<td>0.5%</td>
<td>1%</td>
<td>0.5%</td>
</tr>
<tr>
<td>Actinomices</td>
<td>0.5%</td>
<td>1%</td>
<td>0.5%</td>
</tr>
<tr>
<td>Atopobium</td>
<td>0.5%</td>
<td>1%</td>
<td>0.5%</td>
</tr>
</tbody>
</table>

* indicates statistical significance.
ONLINE DATA SUPPLEMENT

Numbers of observed OTUs in the oral washes and BALs did not differ among groups. Samples from Negative individuals contained 50.32 (43.73, 59.70) OTUs and 46.52 (38.03, 61.16) OTUs in oral washes and BALs, respectively (median OTUs (25th percentile, 75th percentile)). Samples from Naïve individuals contained 52.32 (36.48, 55.09) and 36.72 (28.58, 47.46) OTUs, and samples from HAART individuals contained 44.94 (36.76, 52.92) and 45.90 (37.10, 61.53) OTUs in oral washes and BALs. Overall the p value for subject group (Negative, Naïve, and HAART) was of borderline significance at 0.05, for body site (oral wash and BAL) was 0.40, and for the interaction was 0.25.

Similarly, the Shannon Diversity Index did not differ among groups. The index for samples from Negative individuals was 2.89 (2.66, 3.08) and 2.93 (2.59, 3.14) for oral washes and BALs, respectively (medians, 25th percentile, 75th percentile). Samples from Naïve individuals yielded indices of 2.66 (2.52, 2.81) and 2.73 (2.39, 2.88), and samples from HAART individuals yielded indices of 2.80 (2.67, 2.96) and 2.91 (2.64, 3.12) in oral washes and BALs. The p value comparing subject groups (Negative, Naïve, and HAART) was of borderline significance at 0.05, for body site (oral wash and BAL) was 0.57, and for the interaction was 0.85.

CD4 count, evaluated as a continuous variable, did not modulate the microbiome in our subjects. To illustrate this point, we selected the most abundant OTU for *Streptococcus* (OTU 01) and compared the relative abundance of the sequence to the CD4 count. Comparing oral washes (*Figure E1, Panel A*) and BALs (*Figure E1, Panel
B), there were no differences attributable to CD4 count. *Gemella* (OTU 12) demonstrated the greatest trend of association between CD4 count and relative abundance in oral washes. These data are illustrated and the association was not statistically significant (Figure E2, Panel A). *Veillonella* (OTU 02) demonstrated the greatest trend of association between CD4 count and relative abundance in BALs. These data are illustrated and the association was not statistically significant (Figure E2, Panel B). These data indicate that, in this healthy population, microbial populations do not correlate with host CD4 counts. Of note, our selection of healthy subjects for this study resulted in very few subjects with CD4 counts below 400 cells/mm\(^3\) at the time of enrollment.
FIGURE LEGENDS

Figure E1. Relationship between CD4 count (x-axis) and relative abundance of OTU (y-axis) for Streptococcus in oral wash (Panel A) and BAL (Panel B). There were no overall relationships between CD4 counts and relative abundances of OTUs. This Streptococcus OTU was selected because it was highly abundant in all samples. However, the relationship between relative abundance and CD4 counts was not statistically significant.

Figure E2. Relationship between CD4 count (x-axis) and relative abundance of OTU (y-axis) for Gemella in oral wash (Panel A) and Veillonella in BAL (Panel B). There were no overall relationships between CD4 counts and relative abundances of OTUs. These two OTUs were selected because they demonstrated the strongest relationships with CD4 counts, but the results are not statistically significant.
Relative abundance of Gemella (OTU12) in wash

Relative abundance of Veillonella (OTU02) in BAL

A

Naive
HAART

B

Naive
HAART