Widespread Colonization of the Lung by *Tropheryma whipplei* in HIV Infection

Catherine Lozupone[1], Adela Cota-Gomez[2], Brent E. Palmer[2], Derek J. Linderman[2], Emily S. Charlson[3], Erica Sodergren[4], Makedonka Mitreva[4], Sahar Abubucker[4], John Martin[4], Guohui Yao[4], Thomas B. Campbell[2], Sonia C. Flores[2], Gail Ackerman[1], Jesse Stombaugh[1], Luke Ursell[1], James M. Beck[2,5], Jeffrey L. Curtis[5], Vincent B. Young[5], Susan V. Lynch[6], Laurence Huang[6], George M. Weinstock[4], Kenneth S. Knox[7], Homer Twigg[8], Alison Morris[9], Elodie Ghedin[9], Frederic D. Bushman[3], Ronald G. Collman[3], Rob Knight[1,10], and Andrew P. Fontenot[2]; for the Lung HIV Microbiome Project

[1]Department of Chemistry and Biochemistry and Biofrontiers Institute, University of Colorado, Boulder, Colorado; [2]Department of Medicine, University of Colorado Denver, Aurora, Colorado; [3]Departments of Medicine and Microbiology, University of Pennsylvania, Philadelphia, Pennsylvania; [4]The Genome Institute, Washington University, St. Louis, Missouri; [5]Department of Internal Medicine, University of Michigan, Ann Arbor, Michigan; [6]Department of Medicine, University of California San Francisco, San Francisco, California; [7]Department of Medicine, University of Arizona, Tucson, Arizona; [8]Department of Medicine, Indiana University, Indianapolis, Indiana; [9]Departments of Medicine and Computational and Systems Biology, University of Pittsburgh, Pittsburgh, Pennsylvania; and [10]Howard Hughes Medical Institute, Boulder, Colorado

***Rationale*: Lung infections caused by opportunistic or virulent pathogens are a principal cause of morbidity and mortality in HIV infection. It is unknown whether HIV infection leads to changes in basal lung microflora, which may contribute to chronic pulmonary complications that increasingly are being recognized in individuals infected with HIV.**

***Objectives*: To determine whether the immunodeficiency associated with HIV infection resulted in alteration of the lung microbiota.**

***Methods*: We used 16S ribosomal RNA targeted pyrosequencing and shotgun metagenomic sequencing to analyze bacterial gene sequences in bronchoalveolar lavage (BAL) and mouths of 82 HIV-positive and 77 HIV-negative subjects.**

***Measurements and Main Results*: Sequences representing *Tropheryma whipplei*, the etiologic agent of Whipple's disease, were significantly more frequent in BAL of HIV-positive compared with HIV-negative individuals. *T. whipplei* dominated the community (>50% of sequence reads) in 11 HIV-positive subjects, but only 1 HIV-negative individual (13.4 versus 1.3%; $P = 0.0018$). In 30 HIV-positive individuals sampled longitudinally, antiretroviral therapy resulted in a significantly reduced relative abundance of *T. whipplei* in the lung. Shotgun metagenomic sequencing was performed on eight BAL samples dominated by *T. whipplei* 16S ribosomal RNA. Whole genome assembly of pooled reads showed that uncultured lung-derived *T. whipplei* had similar gene content to two isolates obtained from subjects with Whipple's disease.**

***Conclusions*: Asymptomatic subjects with HIV infection have unexpected colonization of the lung by *T. whipplei*, which is reduced by effective antiretroviral therapy and merits further study for a potential pathogenic role in chronic pulmonary complications of HIV infection.**

AT A GLANCE COMMENTARY

Scientific Knowledge on the Subject

HIV infection is characterized by an impairment of innate and adaptive immunity. The resultant immunodeficiency leads to an increased frequency of pneumonias caused by pathogenic and opportunistic micro-organisms. Currently, it is unknown whether the immunodeficiency associated with HIV infection results in alteration of the lung microbiota.

What This Study Adds to the Field

Tropheryma whipplei, the etiologic agent of Whipple's disease, was significantly more common in the bronchoalveolar lavage of HIV-positive individuals and decreased in relative abundance after successful antiretroviral therapy. Further work will be needed to determine whether *T. whipplei* colonization of the lung leads to adverse clinical outcomes. However, its dominance in the lung of 13% of our HIV-positive cohort indicates that *T. whipplei* should be considered as a potential contributing factor in HIV-positive individuals presenting with lung complications.

(*Received in original form November 30, 2012; accepted in final form January 28, 2013*)

Supported by U.S. Public Health Service grants HL98996 (A.P.F., S.C.F., T.B.C., and R.K.), HL102245 (A.P.F.), DK090285 (C.L.), HL098962 (A.M. and E.G.), HL090339 (A.M.), HL098957 and AI045008 (R.G.C. and F.D.B.), HL098961 (J.M.B., J.L.C., and V.B.Y.), HL087713 (L.H.), HL090335 (L.H.), HL098964 (L.H. and S.V.L.), AI075410 (S.V.L.), HL098960 (H.T. and G.M.W.), and by National Center for Advancing Translational Sciences Clinical and Translational Sciences Institute grants UL1 TR000154 to the University of Colorado Denver and UL1 RR024153 and UL1TR000005 to the University of Pittsburgh.

Author Contributions: C.L., A.C.-G., B.E.P., T.B.C., S.C.F., E.S., G.M.W., R.K., and A.P.F designed the research; J.M.B., V.B.Y., J.L.C., L.H., S.V.L., G.M.W., E.S., H.T., E.G., A.M., E.S.C., F.D.B., and R.G.C. contributed raw sequence data; C.L., A.C.-G., B.E.P., D.J.L., E.S., M.M., S.A., J.M., G.Y., G.A., J.S., L.U., G.M.W., K.S.K., R.K., and A.P.F. performed research; C.L., A.C.-G., M.M., S.A., J.M., G.M.W., R.K., and A.P.F. analyzed data; and C.L., S.C.F., T.B.C., G.M.W., R.K., and A.P.F wrote the manuscript.

Correspondence and requests for reprints should be addressed to Andrew P. Fontenot, M.D., University of Colorado Denver, 12700 East 19th Avenue, B164, Aurora, CO 80045. E-mail: andrew.fontenot@ucdenver.edu

This article has an online supplement, which is accessible from this issue's table of contents at www.atsjournals.org

Am J Respir Crit Care Med Vol 187, Iss. 10, pp 1110–1117, May 15, 2013

Originally Published in Press as DOI: 10.1164/rccm.201211-2145OC on February 7, 2013
Internet address: www.atsjournals.org

Keywords: human; microbiome; metagenome; 16S ribosomal RNA; bronchoalveolar lavage

HIV infection is characterized by impaired innate and adaptive immunity, resulting in an increased frequency of pneumonias caused by pathogenic and opportunistic micro-organisms (1). Antiretroviral therapy (ART) results in decreased viral replication and a concomitant increase in $CD4^+$ T cells (2, 3). Although immune function does not completely normalize with ART, a decreased incidence of pneumonia and other opportunistic infections is seen. However, increased frequencies of noninfectious respiratory complications, such as chronic obstructive pulmonary disease (COPD), lung cancer, and pulmonary hypertension,

occur during chronic HIV infection (4). Whether the development of chronic, noninfectious lung diseases during chronic HIV infection is related to an alteration in basal lung microbiota remains unknown.

The introduction of culture-independent techniques, such as 16S ribosomal RNA (rRNA) gene sequencing, has called into question the dogma that the healthy lung is a sterile environment. For instance, one study suggested the existence of a core pulmonary bacterial microbiome based on 16S rRNA gene sequencing of bronchoalveolar lavage (BAL) fluid (5). In nonsmoking healthy subjects, however, the bacterial sequences in BAL significantly overlap with those in the oropharynx (6, 7), suggesting that the bacterial sequences detected in BAL may arise from either microaspirated oropharyngeal contents and/or contamination of the bronchoscope during passage through the high microbial biomass–containing upper airway. With the development of structural lung disease, such as asthma, COPD, and cystic fibrosis, however, an increased frequency of *Proteobacteria* has been seen in the lower airway (7) and lung tissue (8).

In the present study, we tested whether subjects infected with HIV harbored different bacterial communities in the lung compared with HIV-negative individuals using 16S rRNA targeted sequencing of BAL and paired oral wash samples. Preliminary analysis of a patient cohort from the University of Colorado identified a subject whose BAL 16S rRNA community was unexpectedly dominated by sequences identical to *Tropheryma whipplei*, the agent of Whipple's disease, which decreased in relative abundance after ART. To investigate the relationship between *T. whipplei* and HIV infection, we evaluated carriage and relative abundance of *T. whipplei* across 82 HIV-positive and 77 HIV-negative individuals from 8 cohorts of the Lung HIV Microbiome Project (LHMP) and in a separate cohort of 29 individuals sampled longitudinally before and after ART. We also generated the first complete genome sequence of uncultivated *T. whipplei* by direct metagenomic sequencing of whole BAL specimens and assembly of pooled reads.

METHODS

Subjects and Sample Collection

Data were collected at six different research sites of the LHMP, a consortium whose objective is to characterize the lung microbiota in HIV infection. A total of 82 HIV-positive (ART-treated and naive) and 77 HIV-negative subjects were enrolled across sites. The demographics of the study population are shown in Table E1 in the online supplement. HIV-negative subjects from all cohorts were excluded if they had pneumonia or *Mycobacterium tuberculosis* infection. Information on current and prior history of lung disease was available for the majority of cohorts (Table E1). Informed consent was obtained from each subject, and the study protocol was approved by the Human Subject Institutional Review Boards at each participating institution. Bronchoscopy and BAL were performed as described in the online supplement. Oral washes and, in some cases, posterior pharyngeal swab were collected at the time of BAL to control for potential contamination from the upper respiratory tract.

DNA Extraction and 16S rRNA Amplification

DNA was isolated from swabs, brushes, and BAL samples using the PowerSoil DNA isolation kit (MoBio, Carlsbad, CA), per the manufacturer's instructions, in a BSL2+ hood after treatment with DNAOff (MP Biomedicals, Solon, OH), as previously described (9), for all except the Indiana (DNeasy Blood and Tissue kit; QIAGEN, Valencia, CA) and UCSF cohorts (AllPrep kit; QIAGEN) (Table E2). Bacterial 16S rRNA genes were amplified as detailed in the online supplement.

Sequence Analysis

Raw sequence files were obtained from each site. Deidentified data associated with each sample were also collected in the MIMARKS standard format (10). For all studies, we quality filtered the 16S rRNA sequences using the default values in QIIME version 1.4.0 (11). *See* the online supplement for methodological details. The 16S rRNA sequence libraries and metadata are available for download from the QIIME database (www.microbio.me/qiime) in projects entitled "LHMP_" followed by the research site.

T. whipplei Quantitative PCR

We quantified *T. whipplei* from the BAL samples by species-specific quantitative PCR (qPCR) using primers that amplify a 503-bp fragment of the *hsp65* gene of *T. whipplei* (12). *See* the online supplement for methodological details.

Shotgun Metagenomics Analysis

Eight BAL samples were sequenced on a lane on the Illumina HiSeq instrument to produce 100-bp end reads. *See* the online supplement for methodological details.

RESULTS

T. whipplei Dominates the Lung Microbiota of an Individual Chronically Infected with HIV and Decreases with ART

To study the relationship between HIV infection and the lung microbiota, we initially sequenced bacterial 16S rRNA from BAL of 5 HIV-positive and 11 HIV-negative individuals recruited in Colorado. Given the high prevalence of oropharyngeal-related phylotypes in BAL and sputum samples (6, 13), we were interested in identifying phylotypes highly enriched in the lung compared with the mouth. The highest enrichment for a single phylotype occurred in the lung microbiota of an ART-naive male smoker with chronic HIV infection (Table 1). This phylotype, which had a 16S rRNA gene sequence identical to *T. whipplei*, had a relative abundance of 67–99% in BAL fluid from two different lung sites, regardless of whether unfractionated BAL or the cell-free bacterial pellet was surveyed (Table 1). In contrast, *T. whipplei* 16S rRNA gene sequences were not found in the oral wash sample, and were only present in low abundance in a posterior pharyngeal swab sample (Table 1). Although *T. whipplei* is most frequently a gut pathogen, targeting the gastrointestinal tract before translocating to other body sites (14), we did not detect *T. whipplei* in the stool of this subject (Table 1).

The relative abundance of the *T. whipplei* phylotype in the lung of this individual significantly decreased after 6.5 months of successful ART ($P = 0.027$; paired t test; Table 1), associated with a decreased plasma HIV-1 RNA viral load from 136,945 to 40 copies/ml and an increased $CD4^+$ T cell count from 386 to 547 cells/μl. Despite a decrease in relative abundance after ART, *T. whipplei* sequences remained highly enriched in lung samples compared with the upper respiratory tract (Table 1; *see* also Figure E1 in the online supplement).

Overall Patterns in Microbial Diversity Are Not Driven by Research Cohort

The dominance of the *T. whipplei* phylotype in this subject before ART initiation and the decrease with ART, as well as a known association between Whipple's disease and immune dysfunction (14–21), led us to hypothesize that *T. whipplei* would be more prevalent in lungs of individuals with HIV infection. Therefore, we examined lung samples collected from 82 HIV-positive and 77 HIV-negative subjects from 8 LHMP cohorts.

Data were collected using similar, but distinct, research protocols (Table E2). To determine whether the different experimental

TABLE 1. PERCENTAGE OF SEQUENCES WITHIN A 97% IDENTITY THRESHOLD TO *Tropheryma whipplei* 16S RIBOSOMAL RNA FROM A HUMAN IMMUNODEFICIENCY VIRUS-INFECTED SUBJECT BEFORE AND AFTER ANTIRETROVIRAL THERAPY

	Before ART % (n)*	After ART % (n)
Medial NeatBAL	99 (31,156)†	34 (46,350)
Medial BALpellet	93 (31,508)	58 (47,352)
Lateral NeatBAL	67 (5,060)	11 (85,151)
Lateral BALpellet	ND	39 (82,028)
Posterior pharyngeal swab	0.045 (35,548)	0.0026 (38,079)
Oral wash	0 (25,730)	0 (45,108)
Stool	0 (42,010)	ND

Definition of abbreviations: ART = antiretroviral therapy; BAL = bronchoalveolar lavage; BALpellet = cell-free bacterial pellet; ND = no data; NeatBAL = unfractionated BAL.

*Percentages are based on nonrarefied data.

† The number of sequences evaluated in each case is in parentheses.

protocols employed across sites, such as DNA extraction, primer selection, or sequencing platform, biased the total observed bacterial diversity, we clustered samples from all sites using unweighted UniFrac and principal coordinate analysis (22). The first principal coordinate axis showed clear separation between oral wash samples and methodological controls, whereas BAL samples fell between these two sample types (Figure 1A). This pattern was stronger than clustering by research cohort, despite the differences in experimental protocols employed (Figure 1B). However, pairwise UniFrac distances from within-cohort comparisons of lung samples were significantly smaller than those for between-cohort comparisons, indicating that the overall cohort had some effect on the observed diversity.

Higher Prevalence and Relative Abundance of *T. whipplei* in BAL in HIV-Positive Individuals

T. whipplei was detected in BAL samples from all eight cohorts (Table 2). Despite cohort effects on the overall observed diversity, *T. whipplei* relative abundance was not significantly different between cohorts within HIV-negative (Kruskal-Wallis test; $P = 0.6098$) or HIV-positive ($P = 0.7182$) individuals. Because the number of 16S rRNA sequences per sample varied widely (Table E2), and deeper sequencing results in higher detection rates of rare species, we measured carriage using all available data and using 500 randomly selected sequences from a single lung sample per subject.

The estimated carriage rate (using all 16S rRNA gene sequences) in HIV-negative individuals ranged from 12 to 40% within individual cohorts, and was 23.4% when considering all 77 individuals (Table 2). Carriage in HIV-positive individuals ranged from 20 to 70% within individual cohorts, and was 53.7% when considering all 82 individuals. Carriage was significantly higher in HIV-positive individuals when combining data from the different cohorts (31.7 versus 13.0% when standardized at 500 sequences from a single lung sample; G test for independence $P = 0.002$) (Table 2). Individuals infected with HIV also had a significantly higher relative abundance of *T. whipplei* compared with HIV-negative subjects when combining data from all cohorts ($P = 0.002$ *t* test; Table 2).

Of the eight cohorts, four had samples from both HIV-positive and HIV-negative individuals, but no cohort showed significant *T. whipplei* differences with HIV infection status in isolation (Table 2). To exclude the possibility that a significant overall result was driven by systematic differences induced by variability in the methodology across the different cohorts, we analyzed the *P* values from the four cohorts that had samples from HIV-positive and HIV-negative individuals using Fisher's method, which combines the results from several independent tests bearing upon the same overall hypothesis. The combined *P* value indicated a significantly greater relative abundance of the *T. whipplei* phylotype in HIV-positive compared with HIV-negative individuals ($P = 0.035$), but the difference in carriage rate was not significant ($P = 0.1$; Table 2).

Of particular interest are individuals in whom *T. whipplei* dominated the lung microbiota. *T. whipplei* represented the majority of sequences (>50% of assigned sequence reads) in at least 1 lung sample in 11 of 82 HIV-positive subjects (13.4%), and comprised over 90% of sequences in at least 1 BAL sample in 8 HIV-positive individuals (9.8%). Conversely, *T. whipplei* comprised the majority of assigned reads in HIV-negative subjects only rarely (1 of 77 subjects), and the incidence of *T. whipplei* dominance significantly differed between the HIV-positive and HIV-negative groups ($P = 0.0018$ G test of independence).

The relative abundance of *T. whipplei* was significantly higher in HIV-positive subjects when controlling for cohort using a linear regression model ($P < 0.05$). In addition, a significant positive correlation was observed between *T. whipplei* relative abundance and BAL leukocyte counts ($r = 0.33$, $P = 0.004$). Although smokers were significantly enriched in our HIV-positive population compared with the HIV-negative population ($P = 0.00071$ G test of independence), the relative abundance of *T. whipplei* was significantly higher in HIV-positive subjects when controlling for smoking using a linear regression model ($P = 0.020$). In the HIV-negative population, however, smokers had a trend toward a higher relative abundance of *T. whipplei* compared with nonsmokers (0.038 versus 0.0015; $P = 0.08$ *t* test).

None of the individuals in whom *T. whipplei* dominated the population (made up >50% of the sequence reads in at least one lung sample) reported any serious lung disorders at the time of sampling. One individual who reported *M. tuberculosis* infection had *T. whipplei* accounting for 25% of the sequence reads.

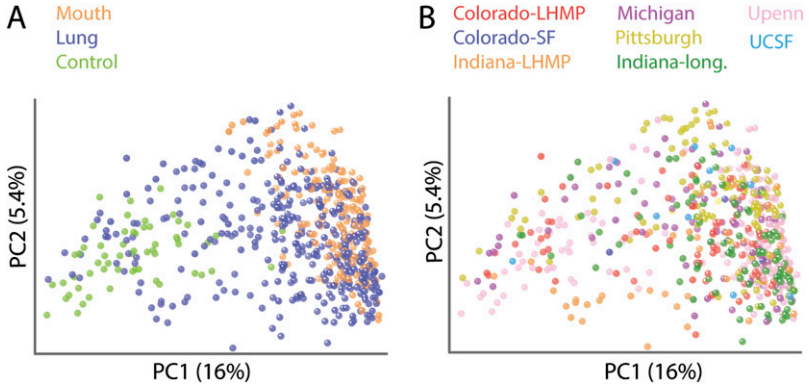

Figure 1. Clustering of data from all study sites. The result of applying principal coordinate analysis to a matrix of unweighted UniFrac distances, showing the relationship between the overall collection of bacterial sequences detected in samples from all of the Lung HIV Microbiome Project (LHMP) cohorts, and that the overall pattern is not driven by methodological differences that existed across cohorts. (*A*) and (*B*) are the same plot, but the points are colored by sample type in (*A*) and by cohort in (*B*). The control group in (*A*) is a range of methodological controls, as detailed in the online supplement. PC = principal coordinate.

TABLE 2. CARRIAGE AND RELATIVE ABUNDANCE OF THE *TROPHERYMA WHIPPLEI* PHYLOTYPE IN LUNG

	Carriage*									
	HIV$^+$			HIV$^-$				Average Relative Abundance†		
Site	n	All %	Rare %	n	All %	Rare %	*P* Value (*G* Test)‡	HIV$^+$	HIV$^-$	*P* Value (*t* Test)
Colorado–LHMP	5	40.0	40.0	11	36.4	18.2	0.20	0.20 ± 0.44	0.015 ± 0.05	0.20
Colorado–San Francisco	5	40.0	40.0	0	ND	ND	ND	0.20 ± 0.44	ND	ND
Indiana–LHMP	10	70.0	50.0	15	40.0	20.0	0.12	0.096 ± 0.22	0.006 ± 0.013	0.11
Indiana-long.	30	63.3	43.3	0	ND	ND	ND	0.087 ± 0.24	ND	ND
Michigan	0	ND	ND	25	12.0	4.0	ND	ND	0.0013 ± 0.006	ND
Pittsburgh	15	40.0	33.3	20	20.0	15.0	0.11	0.20 ± 0.40	0.037 ± 0.15	0.09
Pennsylvania	12	50.0	16.7	6	16.7	16.7	0.50	0.021 ± 0.066	0.001 ± 0.002	0.16
University of California, San Francisco	5	20.0	20.0	0	ND	ND	ND	0.0004 ± 0.0009	ND	ND
Total	82	53.7	31.7	77	23.4	13.0	0.0022	0.11 ± 0.28	0.014 ± 0.078	0.002
Fisher's combined probability test							0.10			0.035

Definition of abbreviations: LHMP = Lung HIV Microbiome Project; ND = no data.

*Carriage is the percentage of individuals surveyed in which *Tropheryma whipplei* was detected. Estimates are given for both all data (all) and with 500 randomly selected sequences from one lung sample (rarefied; rare).

† Average relative abundances are the average (±1 SD) fraction of sequence reads that were assigned to *T. whipplei* using 500 randomly selected reads in a single lung sample.

‡ The *P* value for carriage was calculated using the *G* test of independence and rarefied data.

Validation of *T. whipplei* 16S rRNA Sequences Using hsp65-targeted qPCR

Although 16S rRNA sequences were clustered with *T. whipplei* with 97% sequence identity, we wished to confirm the presence of *T. whipplei* with an independent assay targeting a distinct genetic region. Therefore, we interrogated samples from the Colorado–LHMP, Indiana–LHMP, and Pennsylvania cohorts using qPCR targeting *hsp65*, which is highly conserved across all *T. whipplei* serotypes (12). There was no measurable amplification with these primers from purified *Escherichia coli* genomic DNA or from any of the methodological controls (Figure 2A). The sensitivity of the assay was reliably one copy per reaction; therefore, we could reliably detect *T. whipplei* in samples in which it was not very abundant. When normalized to copies of 16S rDNA, the abundance of *T. whipplei* ranged from less than 0.01% to nearly 100% (in a few patients) of all bacteria present (Figure 2B). qPCR measurements were significantly correlated with relative abundance calculated from 16S rRNA sequences (Colorado: $r = 0.90$, $P < 0.0001$; Indiana: $r = 1.0$, $P < 0.00001$; and Pennsylvania: $r = 0.50$, $P = 0.002$), thus validating our *T. whipplei* 16S rRNA sequencing data.

Prevalence of *T. whipplei* in the Upper Respiratory Tract of HIV-Positive and HIV-Negative Individuals

Because *T. whipplei* has been previously detected in mouth samples (23–25), and the upper respiratory tract is the major source of overall diversity in our lung samples (Figure 1), we compared the relative abundance of *T. whipplei* in matched oral wash and BAL samples in individuals in whom *T. whipplei* was detected ($n = 27$) and found a significant enrichment of *T. whipplei* sequences in the lung ($P = 0.0041$, paired *t* test). *T. whipplei* was never detected in oral wash/posterior pharyngeal swab samples when 500 sequences per sample were evaluated (Table E3) and, in contrast to the lung, it was neither more frequently detected nor more abundant in oropharyngeal samples from HIV-positive than HIV-negative individuals.

Effects of ART on *T. whipplei* Abundance in the Lung

We next explored whether *T. whipplei* decreased in HIV-positive individuals after ART. In a cross-sectional analysis, *T. whipplei* relative abundance and carriage were not significantly higher in ART-naive ($n = 44$) compared with ART-treated ($n = 33$) individuals (Table E4), nor was it significantly correlated with CD4 count or HIV viral load, giving an initial indication that ART did not help to control *T. whipplei* in the lung. Incomplete control of *T. whipplei* after ART is further supported by the observation that the relative abundance of *T. whipplei* was significantly higher in HIV-positive individuals than HIV-negative individuals when only ART-treated individuals were considered ($P = 0.01$, *t* test).

However, when comparing BAL samples collected longitudinally from 29 HIV-positive individuals before treatment and 4 weeks, 1 year, and 3 years after ART, treatment led to a significant reduction of the *T. whipplei* phylotype within individuals. We detected the *T. whipplei* phylotype in BAL samples from at least one time point in 19 of the 29 (66%) individuals. In six individuals, over 10% of the sequence reads belonged to the *T. whipplei* phylotype in at least one of the BAL samples (Figure 3). There was a significant decrease in the relative abundance of *T. whipplei* in BAL after 6 months to 3 years, but not 4 weeks of ART therapy. We measured significance with paired *t* tests comparing the average relative abundance of the *T. whipplei* phylotype at baseline and 4-week time points to the average from any 6-month, 1-year, and 3-year post-ART samples available. Significance was achieved when all 20 carriers (19 from

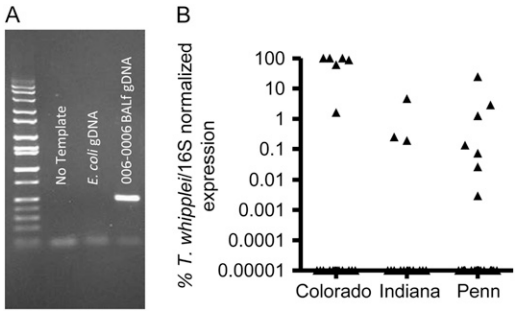

Figure 2. Quantitative PCR analysis with the *hsp65* gene. (*A*) Representative PCR of *Tropheryma whipplei*–specific *hsp65* showing a base pair amplicon for a bronchoalveolar lavage (BAL) fluid sample (006–0006) from an individual with high *T. whipplei* relative abundance based on 16S ribosomal RNA (rRNA) in the Colorado–San Francisco (SF) cohort. (*B*) Copy numbers normalized to 16S rRNA expression in the Colorado, Indiana, and Pennsylvania cohorts is shown.

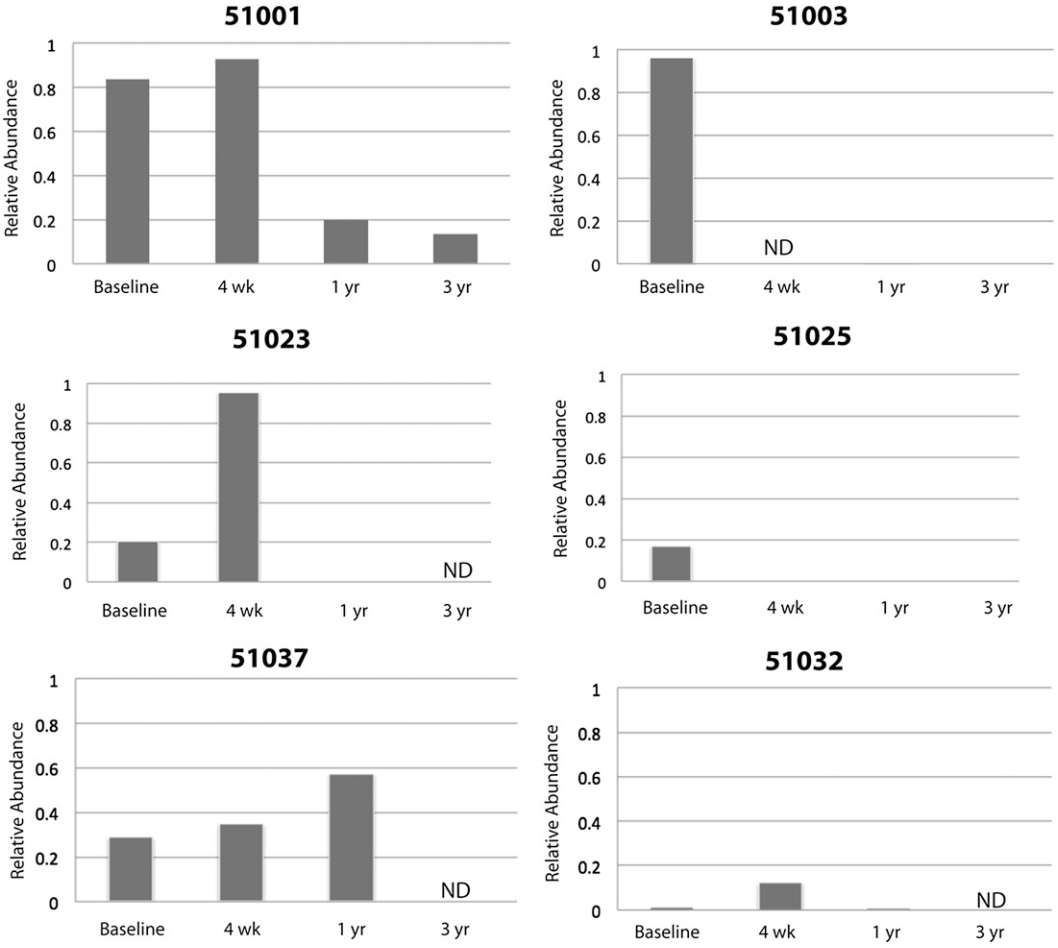

Figure 3. Longitudinal analysis of *Tropheryma whipplei* relative abundance in individuals on antiretroviral therapy (ART) treatment. The relative abundance of *T. whipplei* is plotted for the six individuals in whom the *T. whipplei* phylotype exceeded 10% of the sequences in at least one sample. Columns denoted with "ND" had no data, either because the 3-year time point was not available or because the sample did not achieve a sequencing depth of at least 225 reads.

Indiana-longitudinal and 1 from Colorado) were considered ($P = 0.032$, paired t test), when only 7 individuals with high *T. whipplei* relative abundance were considered (>10% in at least one sample; $P = 0.029$, paired t test), and when only 9 low-abundance individuals were considered (<0.1% in all samples; $P = 0.018$, paired t test).

Similar to the cross-sectional analysis, successful ART did not always result in a decrease in the relative abundance of the *T. whipplei* phylotype. For example, subject 51,037 had no decrease after 1 year of treatment (Figure 3), despite a decrease in plasma HIV-1 RNA viral load from 10,600 to 400 copies/ml and an increase in $CD4^+$ T cell count from 182 to 400 cells/μl during this period. Thus, although ART decreases the overall relative abundance of *T. whipplei* within individuals, ART does not completely restore the ability of HIV-positive individuals to control *T. whipplei* in the lung.

Genomic Analysis of *T. whipplei* in Lung Samples

To determine whether *T. whipplei* in lung resembled isolates cultured from patients with Whipple's disease (Twist, isolated from a cardiac valve [26] and TW08/27, isolated from cerebrospinal fluid [27]), we applied shotgun metagenomic sequencing directly to BAL (eight samples from six individuals) with a high relative abundance of the *T. whipplei* phylotype (Table E5). Although the majority of the sequenced reads aligned to the human genome, a comparison of 69,271,986 kb of nonhuman reads to nearly 6,000 reference genome sequences showed *T. whipplei* to be the most abundant bacterium (Table E6). Thus, the abundance of *T. whipplei* that was inferred from sequences of the 16S rRNA and *hsp65* genes was confirmed at the whole-genome level.

The sequences that aligned to the *T. whipplei* TW08/27 reference genome were assembled either for each individual when coverage of *T. whipplei* was sufficient (≥2×) or from pooling of all eight samples. The best coverage (20×) and most contiguous assembly came from the pooled sequences (Table E7), which were used in all subsequent analyses. Alignment of the assembled contigs to two *T. whipplei* reference genomes showed 90 and 92% coverage for TW08/27 and Twist, respectively (Figure 4; Table E8). A total of 27 potential genes were novel to the lung isolates (Table E9). Similar to previous *T. whipplei* genome comparisons (26, 28), the novel genes included three members of the highly variable WiSP family of surface proteins, suggesting their potential role in host interactions and immune evasion. In addition, 20 and 27 genes in Twist and TW08/27 genomes were not identified in the pooled assembly from lung isolates (Figure 4; Table E8), including several WiSP proteins, although we cannot determine whether these were truly absent in the lung strains or not detected due to the level of coverage.

DISCUSSION

We found that approximately half of the individuals with HIV infection harbor *T. whipplei* bacterial sequences in their lungs, often at very high relative abundance, and that relative abundance decreases with effective ART. The identity of this organism was confirmed by extensive cross-validation using 16S rRNA sequencing, *hsp65*-targeted qPCR, and shotgun metagenomic whole-genome sequencing. This is the first description of widespread lung colonization in asymptomatic HIV infection, and the first application of emerging microbiome tools to identify, within

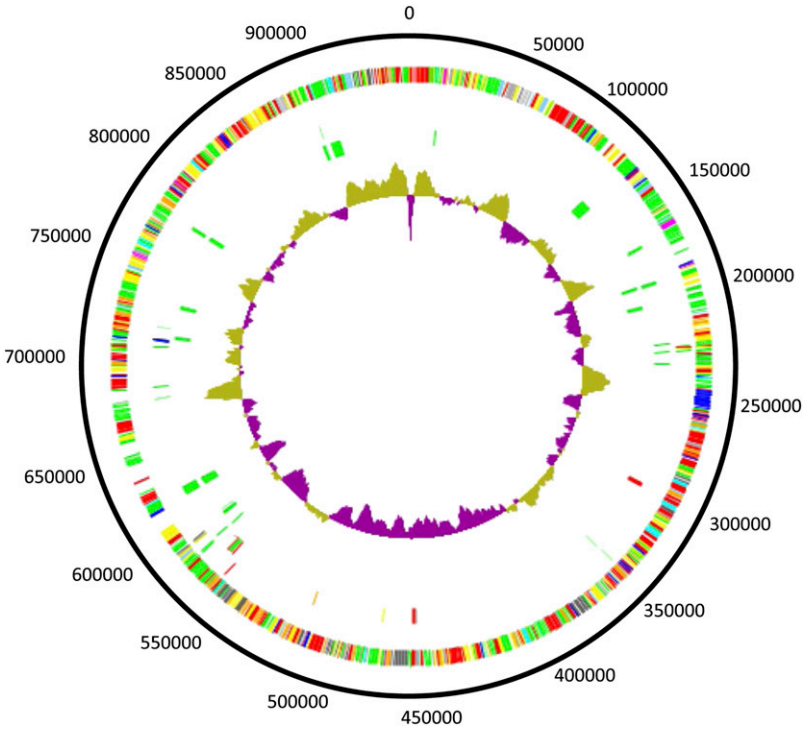

Figure 4. Gene content of lung *Tropheryma whipplei* strains deduced from metagenomic data. The *outer ring* shows the TW08/27 genome, the *next inner circle* shows the genes missing in Twist, followed by the genes missing in lung pooled assembly. The *inner circle* shows the 14 WiSP loci in the TW0827 reference. The gene content plot shows above-average regions in *olive* and the below-average regions in *purple*. The color code for the *outer ring* indicates functional categories for gene: *dark blue*, pathogenicity or adaptation; *black*, energy metabolism; *red*, information transfer; *dark green*, membranes or surface structures; *cyan*, degradation of macromolecules; *purple*, degradation of small molecules; *yellow*, central or intermediary metabolism; *light blue*, regulators; *orange*, conserved hypothetical; *pale green*, unknown; and *brown*, pseudogenes.

the lower respiratory tract, a specific organism shared among individuals in the absence of clinical respiratory tract disease.

T. whipplei is the causative agent of Whipple's disease, a rare systemic infectious disease. Classic Whipple's disease typically involves the gastrointestinal tract, and is associated with diverse clinical manifestations, including weight loss, diarrhea, and joint, neurological, and cardiac and/or pulmonary involvement (14). Localized infections without histological digestive lesions have also been described (14). Pulmonary manifestations of Whipple's disease include dyspnea, pleuritic chest pain, chronic cough, reduced lung volumes, and pleural adhesions (29). Whipple's disease is often associated with defects of innate immune activation, resulting in an inability to control the organism (14–21). However, classic Whipple's disease has seldom been described in HIV-positive subjects (29–34). Nevertheless, our whole bacterial genome sequences from pooled reads in shotgun metagenomic data derived directly from BAL indicates that *T. whipplei* in lung represents strains extremely similar to those that cause Whipple's disease.

Our results indicate a high frequency of carriage of *T. whipplei* in the lung of HIV-positive individuals; however, it was also identified in a minority of HIV-negative individuals. A potential niche for *T. whipplei* in the healthy lung has previously been suggested (6). *T. whipplei* has also been identified in the stool and mouth of healthy individuals, but asymptomatic carriers have significantly lower *T. whipplei* loads than patients with Whipple's disease (23). Estimates of *T. whipplei* carriage in the gastrointestinal tract and mouth of healthy subjects vary considerably across studies (0.6–35% for saliva), which may reflect age, geography, and environmental exposures and/or methodology (23–25, 35). Our estimates of *T. whipplei* relative abundance in the lung, however, did not significantly differ across cohorts, despite the geographical distances and differences in DNA extraction and sequencing technologies. Genotyping of 39 *T. whipplei* DNA samples from patients and 10 from asymptomatic carriers revealed that *T. whipplei* genetic diversity is unrelated to bacterial pathogenicity (36).

Our data reveal a new ecological niche within the human body for *T. whipplei* colonization, and suggest that a lack of immunologic control due to HIV coinfection increases prevalence. Although *T. whipplei* carriage in the mouth (23–25, 35) and a specific niche for *T. whipplei* in the subgingival and gingival sulcus plaque has been suggested (35), the dominance of *T. whipplei* in the lung compared with matched mouth samples suggests that the true niche of *T. whipplei* detected in saliva may often be the lung. Alternately, the oral cavity or small bowel may serve as a reservoir for *T. whipplei*, and HIV-positive individuals may have a reduced ability to clear this microbe once introduced into the lung. A niche in the lung, however, is consistent with a high abundance in the lung of macrophages, the main target cell for *T. whipplei* elsewhere in the body, and the correlation between *T. whipplei* prevalence and BAL leukocyte counts.

Genomic sequencing has been performed on two isolates of *T. whipplei* from patients with Whipple's disease: Twist (26) and TW08/27 (27). Sequenced after prolonged *in vitro* culture, these genomes were highly similar to one another, displaying 99% identity at the nucleotide sequence level (26). This remarkable degree of conservation was confirmed by comparative genomic hybridization of 16 clinical isolates to the Twist strain (28). Both studies, however, observed variability specifically in the WiSP membrane protein family, which was hypothesized to be associated with changes in surface-exposed bacterial proteins (26, 28). Our analysis, based on metagenomic sequences obtained directly from primary human material without *in vitro* culture, similarly identified members of the WiSP family among both the putative lung-specific genes and genes absent in our pooled assembly. The strong genomic similarity that we observed between the *T. whipplei* in our BAL samples and isolates from patients with Whipple's disease, such as a reduced genome size, lack of mobile DNA elements, lack of genes encoding essential metabolic capabilities, and an elaboration of mechanisms for the variation of surface structures potentially for immune evasion, are consistent with a highly host-restricted and host-adapted intracellular lifestyle (26, 28).

The clinical consequences of *T. whipplei* growth in the lung of HIV-positive individuals are not known. Individuals infected with HIV have a much higher prevalence of noninfectious lung

diseases, including pulmonary hypertension and COPD, compared with control subjects (4). Whipple's disease is associated with pulmonary manifestations (37), and a complete resolution of pulmonary hypertension after antibiotic treatment has been reported in an individual with Whipple's disease (37, 38). Another study using 16S rRNA gene sequencing reported that *T. whipplei* comprised 66% of sequences from a subject with interstitial lung disease (39). Similarly, *T. whipplei* was the only identified bacterium in BAL from an immunocompromised patient who presented with community-acquired pneumonia and septic shock (40). In addition, respiratory symptoms can precede the development of gastrointestinal manifestations of Whipple's disease, suggesting that the primary route of infection is not always the gastrointestinal tract (41). The possibility that *T. whipplei* overgrowth could lead to lung disease is supported by the pulmonary manifestations of Whipple's disease and the finding that *T. whipplei* is abundant in the lung of subjects with other pulmonary diseases. Because *T. whipplei* is resistant to growth with standard culture techniques (42), its importance in HIV-associated lung disease may be underappreciated. Alternative methods to cultural techniques for detecting *T. whipplei* in clinical samples have been extensively developed (14), but are not typically used for patients with acquired immune deficiency syndrome.

Further clinical and immunologic studies are needed to elucidate the functional consequences of *T. whipplei* expansion in the lung of individuals infected with HIV, and its relationship to long-term pulmonary sequelae. Because progression to classic systemic Whipple's disease rarely occurs in HIV-positive individuals (29), HIV-associated *T. whipplei* overgrowth may have unique, but still important, clinical implications.

Author disclosures are available with the text of this article at www.atsjournals.org.

Acknowledgment: A subset of subjects in this cohort was recruited from the University of California, Los Angeles (Roger Detels) and the University of Pittsburgh (Charles R. Rinaldo, Lawrence Kingsley), as part of the Multicenter AIDS Cohort Study (AI35042, UL1-RR025005, AI35043, AI35039, AI35040, AI35041). A subset of subjects in this cohort was recruited from the Connie Wofsy Study Consortium of Northern California (Ruth Greenblatt) of the Women's Interagency HIV Study Collaborative Study Group (AI35004, AI31834, AI34994, AI34989, AI34993, UO1AI42590 and HD32632). The WIHS Collaborative Study Group was cofunded by the National Cancer Institute, the National Institute on Drug Abuse, and the National Institute on Deafness and Other Communication Disorders. Funding was also provided by National Center for Research Resources University of California San Francisco–CTSI grant UL1 RR024131.

References

1. Huang L, Quartin A, Jones D, Havlir DV. Intensive care of patients with HIV infection. *N Engl J Med* 2006;355:173–181.
2. Hammer SM, Squires KE, Hughes MD, Grimes JM, Demeter LM, Currier JS, Eron JJ Jr, Feinberg JE, Balfour HH Jr, Deyton LR, *et al.* A controlled trial of two nucleoside analogues plus indinavir in persons with human immunodeficiency virus infection and CD4 cell counts of 200 per cubic millimeter or less. AIDS Clinical Trials Group 320 Study Team. *N Engl J Med* 1997;337:725–733.
3. Staszewski S, Morales-Ramirez J, Tashima KT, Rachlis A, Skiest D, Stanford J, Stryker R, Johnson P, Labriola DF, Farina D, *et al.* Efavirenz plus zidovudine and lamivudine, efavirenz plus indinavir, and indinavir plus zidovudine and lamivudine in the treatment of HIV-1 infection in adults. *N Engl J Med* 1999;341:1865–1873.
4. Crothers K, Huang L, Goulet JL, Goetz MB, Brown ST, Rodriguez-Barradas MC, Oursler KK, Rimland D, Gibert CL, Butt AA, *et al.* HIV infection and risk for incident pulmonary diseases in the combination antiretroviral therapy era. *Am J Respir Crit Care Med* 2011;183:388–395.
5. 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.
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. 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.
8. 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.
9. Costello EK, Lauber CL, Hamady M, Fierer N, Gordon JI, Knight R. Bacterial community variation in human body habitats across space and time. *Science* 2009;326:1694–1697.
10. Yilmaz P, Kottmann R, Field D, Knight R, Cole JR, Amaral-Zettler L, Gilbert JA, Karsch-Mizrachi I, Johnston A, Cochrane G, *et al.* Minimum information about a marker gene sequence (mimarks) and minimum information about any (x) sequence (mixs) specifications. *Nat Biotechnol* 2011;29:415–420.
11. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena AG, Goodrich JK, Gordon JI, *et al.* Qiime allows analysis of high-throughput community sequencing data. *Nat Methods* 2010;7:335–336.
12. Morgenegg S, Dutly F, Altwegg M. Cloning and sequencing of a part of the heat shock protein 65 gene (hsp65) of "*Tropheryma whippelii*" and its use for detection of "*T. Whippelii*" in clinical specimens by PCR. *J Clin Microbiol* 2000;38:2248–2253.
13. Goddard AF, Staudinger BJ, Dowd SE, Joshi-Datar A, Wolcott RD, Aitken ML, Fligner CL, Singh PK. Direct sampling of cystic fibrosis lungs indicates that DNA-based analyses of upper-airway specimens can misrepresent lung microbiota. *Proc Natl Acad Sci USA* 2012;109:13769–13774.
14. Desnues B, Al Moussawi K, Fenollar F. New insights into Whipple's disease and *Tropheryma whipplei* infections. *Microbes Infect* 2010;12:1102–1110.
15. Kalt A, Schneider T, Ring S, Hoffmann J, Zeitz M, Stallmach A, Persing DH, Marth T. Decreased levels of interleukin-12p40 in the serum of patients with Whipple's disease. *Int J Colorectal Dis* 2006;21:114–120.
16. Marth T, Kleen N, Stallmach A, Ring S, Aziz S, Schmidt C, Strober W, Zeitz M, Schneider T. Dysregulated peripheral and mucosal Th1/Th2 response in Whipple's disease. *Gastroenterology* 2002;123:1468–1477.
17. Marth T, Neurath M, Cuccherini BA, Strober W. Defects of monocyte interleukin 12 production and humoral immunity in Whipple's disease. *Gastroenterology* 1997;113:442–448.
18. Marth T, Roux M, von Herbay A, Meuer SC, Feurle GE. Persistent reduction of complement receptor 3 alpha-chain expressing mononuclear blood cells and transient inhibitory serum factors in Whipple's disease. *Clin Immunol Immunopathol* 1994;72:217–226.
19. Moos V, Kunkel D, Marth T, Feurle GE, LaScola B, Ignatius R, Zeitz M, Schneider T. Reduced peripheral and mucosal *Tropheryma whipplei*–specific Th1 response in patients with Whipple's disease. *J Immunol* 2006;177:2015–2022.
20. Moos V, Schmidt C, Geelhaar A, Kunkel D, Allers K, Schinnerling K, Loddenkemper C, Fenollar F, Moter A, Raoult D, *et al.* Impaired immune functions of monocytes and macrophages in Whipple's disease. *Gastroenterology* 2010;138:210–220.
21. Schinnerling K, Moos V, Geelhaar A, Allers K, Loddenkemper C, Friebel J, Conrad K, Kuhl AA, Erben U, Schneider T. Regulatory T cells in patients with Whipple's disease. *J Immunol* 2011;187:4061–4067.
22. Lozupone C, Knight R. Unifrac: a new phylogenetic method for comparing microbial communities. *Appl Environ Microbiol* 2005;71:8228–8235.
23. Fenollar F, Laouira S, Lepidi H, Rolain JM, Raoult D. Value of *Tropheryma whipplei* quantitative polymerase chain reaction assay for the diagnosis of Whipple disease: usefulness of saliva and stool specimens for first-line screening. *Clin Infect Dis* 2008;47:659–667.
24. Rolain JM, Fenollar F, Raoult D. False positive PCR detection of *Tropheryma whipplei* in the saliva of healthy people. *BMC Microbiol* 2007;7:48.
25. Street S, Donoghue HD, Neild GH. *Tropheryma whippelii* DNA in saliva of healthy people. *Lancet* 1999;354:1178–1179.
26. Raoult D, Ogata H, Audic S, Robert C, Suhre K, Drancourt M, Claverie JM. *Tropheryma whipplei* twist: a human pathogenic actinobacteria with a reduced genome. *Genome Res* 2003;13:1800–1809.

27. Bentley SD, Maiwald M, Murphy LD, Pallen MJ, Yeats CA, Dover LG, Norbertczak HT, Besra GS, Quail MA, Harris DE, *et al.* Sequencing and analysis of the genome of the Whipple's disease bacterium *Tropheryma whipplei*. *Lancet* 2003;361:637–644.
28. La MV, Crapoulet N, Barbry P, Raoult D, Renesto P. Comparative genomic analysis of *Tropheryma whipplei* strains reveals that diversity among clinical isolates is mainly related to the WiSP proteins. *BMC Genomics* 2007;8:349.
29. Patel SJ, Huard RC, Keller C, Foca M. Possible case of CNS Whipple's disease in an adolescent with AIDS. *J Int Assoc Physicians AIDS Care (Chic)* 2008;7:69–73.
30. Jankovic J. Whipple's disease of the central nervous system in AIDS. *N Engl J Med* 1986;315:1029–1030.
31. Autran B, Gorin I, Leibowitch M, Laroche L, Escande JP, Hewitt J, Marche C. AIDS in a Haitian woman with cardiac Kaposi's sarcoma and Whipple's disease. *Lancet* 1983;1:767–768.
32. Maiwald M, Meier-Willersen HJ, Hartmann M, von Herbay A. Detection of *Tropheryma whippelii* DNA in a patient with AIDS. *J Clin Microbiol* 1995;33:1354–1356.
33. Bargiacchi O, Salerno AM, Brondolo R, Rizzo G, Brustia D, Garavelli PL. Whipple's disease in an AIDS patient with Kaposi's sarcoma: report of a case. *Infez Med* 2010;18:198.
34. Guthikonda B, Rouah E, Krishnan B, Powell SZ, Goodman JC, Gopinath SP, Simpson RK. Whipple disease of the central nervous system: an unusual occurrence in association with acquired immune deficiency syndrome. *J Neurosurg* 2010;112:983–989.
35. Zinkernagel AS, Gmur R, Fenner L, Schaffner A, Schoedon G, Schneemann M. Marginal and subgingival plaque—a natural habitat of *Tropheryma whipplei*? *Infection* 2003;31:86–91.
36. Li W, Fenollar F, Rolain JM, Fournier PE, Feurle GE, Muller C, Moos V, Marth T, Altwegg M, Calligaris-Maibach RC, *et al.* Genotyping reveals a wide heterogeneity of *Tropheryma whipplei*. *Microbiology* 2008;154: 521–527.
37. Symmons DP, Shepherd AN, Boardman PL, Bacon PA. Pulmonary manifestations of Whipple's disease. *Q J Med* 1985;56:497–504.
38. Najm S, Hajjar J, Nelson RP Jr, Moorthy RS, Wolf KM, Lahm T. Whipple's disease–associated pulmonary hypertension with positive vasodilator response despite severe hemodynamic derangements. *Can Respir J* 2011;18:e70–e72.
39. Harris JK, De Groote MA, Sagel SD, Zemanick ET, Kapsner R, Penvari C, Kaess H, Deterding RR, Accurso FJ, Pace NR. Molecular identification of bacteria in bronchoalveolar lavage fluid from children with cystic fibrosis. *Proc Natl Acad Sci USA* 2007;104:20529–20533.
40. Bousbia S, Papazian L, Auffray JP, Fenollar F, Martin C, Li W, Chiche L, La Scola B, Raoult D. *Tropheryma whipplei* in patients with pneumonia. *Emerg Infect Dis* 2010;16:258–263.
41. Urbanski G, Rivereau P, Artru L, Fenollar F, Raoult D, Puechal X. Whipple disease revealed by lung involvement: a case report and literature review. *Chest* 2012;141:1595–1598.
42. Raoult D, Fenollar F, Birg ML. Culture of *T. whipplei* from the stool of a patient with Whipple's Disease. *N Engl J Med* 2006;355:1503–1505.