ERJ Express. Published on August 9, 2018 as doi: 10.1183/13993003.00810-2018

EUROPEAN RESPIRATORY *journal*

FLAGSHIP SCIENTIFIC JOURNAL OF ERS

Early View

Original article

Evaluation of the airway microbiome in non-tuberculous mycobacteria

Imran Sulaiman, Benjamin G. Wu, Yonghua Li, Adrienne S. Scott, Patrick Malecha, Benjamin Scaglione, Jing Wang, Ashwin Basavaraj, Samuel Chung, Katrina Bantis, Joseph Carpenito, Jose C. Clemente, Nan Shen, Jamie Bessich, Samaan Rafeq, Gaetene Michaud, Jessica Donington, Charissa Naidoo, Grant Theron, Gail Schattner, Suzette Garofano, Rany Condos, David Kamelhar, Doreen Addrizzo-Harris, Leopoldo N. Segal

Please cite this article as: Sulaiman I, Wu BG, Li Y, *et al*. Evaluation of the airway microbiome in non-tuberculous mycobacteria. *Eur Respir J* 2018; in press (https://doi.org/10.1183/13993003.00810-2018).

TITLE:

Evaluation of the Airway Microbiome in Non-Tuberculous Mycobacteria

Imran Sulaiman[1], Benjamin G. Wu[1], Yonghua Li[1], Adrienne S. Scott[1], Patrick Malecha[1], Benjamin Scaglione[1], Jing Wang[1], Ashwin Basavaraj[1], Samuel Chung[1], Katrina Bantis[1], Joseph Carpenito[1], Jose C. Clemente[3,4], Nan Shen[3], Jamie Bessich[1], Samaan Rafeq[1], Gaetene Michaud[1], Jessica Donington[1], Charissa Naidoo[5], Grant Theron[5], Gail Schattner[1], Suzette Garofano[1], Rany Condos[1], David Kamelhar[1], Doreen Addrizzo-Harris[1], Leopoldo N. Segal[1,2,]*

[1] Division of Pulmonary, Critical Care, & Sleep Medicine, New York University School of Medicine, NY
[2] Department of Medicine, New York University School of Medicine, NY
[3] Department of Genetics and Genomic Sciences and Immunology Institute, Icahn School of Medicine at Mount Sinai, NY.
[4] Precision Immunology Institute, Icahn School of Medicine at Mount Sinai, NY
[5] Medicine and Health Sciences, Stellenbosch University, DST/NRF of Excellence for Biomedical Tuberculosis Research, and SA MRC Centre for Molecular and Cellular Biology, Division of Molecular Biology and Human Genetics Tygerberg, ZA

Imran Sulaiman, MD	SheikMohammadImran.Sulaiman@nyumc.org
Benjamin G. Wu, MD	Benjamin.Wu@nyumc.org
Yonghua Li, MD, PhD	Yonghua.Li@nyumc.org
Adrienne Scott MS,	Adrienne.Scott@nyumc.org
Patrick Malecha, MD	Patrick.Malecha@med.nyu.edu
Benjamin Scaglione,	bescaglione@gmail.com
Jing Wang, MD	Jing.Wang3@nyumc.org
Ashwin Basavaraj, MD	Ashwin.Basavaraj@nyumc.org
Samuel Chung MD,	sjc316@gmail.com
Katrina Bantis, BS	kbantis@nyit.edu
Joseph Carpenito	jocarpenito@gmail.com
Jose C. Clemente, PhD	jose.clemente@mssm.edu
Nan Shen, PhD	nan.shen@icahn.mssm.edu
Jamie Bessich, MD	Jamie.Bessich@nyumc.org
Samaan Rafeq, MD	Samaan.Rafeq@nyumc.org
Gaetene Michaud, MD	Gaetane.Michaud@nyumc.org
Jessica Donnington, MD	Jessica.Donington@nyumc.org
Charissa Naidoo, PhD	ccnaidoo@sun.ac.za
Grant Theron, PhD	gtheron@sun.ac.za
Gail Schattner, MD	Gail.Schattner@nyumc.org
Suzette Garofano, MD	Suzette.Garofano@nyumc.org
Rany Condos, MD	Rany.Condos@nyumc.org
David Kamelhar, MD	David.Kamelhar@nyumc.org
Doreen Addrizzo-Harris, MD	Doreen.Addrizzo@nyumc.org
Leopoldo N. Segal, MD[1]	Leopoldo.Segal@nyumc.org

Corresponding Author/ Address for Reprints:
Leopoldo N. Segal, MD[1] Leopoldo.Segal@nyumc.org

NYU School of Medicine
462 First Ave 7W54
New York, NY 10016
Tel: (212) 562-3752
Fax: (212) 263-7445

Financial Disclosure: None
Key words: microbiome
Running Head: Lung Microbiome in NTM

This article has an online data supplement.

Take Home Message (115/117 characters):

16S rRNA gene sequencing is not sensitive to detect Mycobacterium but identifies microbiota signatures associated with inflammation.

Plain language Summary:

Non-tuberculosis mycobacterium (NTM) is increasing in incidence worldwide. However, acquiring this pathogen does not always lead to pulmonary disease and it is unclear what factors are related to the development of disease.

In this study, oral and sputum samples were collected from 106 participants. Just over half the participants had NTM on sputum culture. Additionally, samples from the lower airways through bronchoscopy were collected from 20 participants. Samples were processed using a next generation sequencing.

16S rRNA gene sequencing was not a sensitive technique to detect Mycobacterium. Lower airway samples frequently revealed enrichment with bacteria commonly consider oral commensals. Importantly, enrichment of the lower airways with these oral commensals was associated with an increase in lower airway inflammation in participants with NTM disease. This suggests that aspiration of oral bacteria into the lung is associated with the lower airway inflammatory tone and may be a factor in the development of NTM disease.

Abstract

Background. Aspiration is associated with non-tuberculous mycobacterial (NTM) pulmonary disease and airway dysbiosis is associated with increased inflammation. We examined whether NTM disease was associated with a distinct airway microbiota and immune profile.

Methods. 297 oral wash and induced sputum samples were collected from 106 participants with respiratory symptoms and imaging abnormalities compatible with NTM. Lower airway samples were obtained in 20 participants undergoing bronchoscopy. 16S rRNA gene and a nested mycobacteriome sequencing approaches characterized microbiota composition. Inflammatory profiles of lower airway samples were also examined.

Results. The prevalence of NTM+ cultures was 58%. Few changes were noted in microbiota characteristic or composition in oral wash and sputum samples among groups. Among NTM+ samples, 27% of the lower airway samples were enriched with *Mycobacterium*. A mycobacteriome approach identified *Mycobacterium* in a greater percentage of samples, including some non-pathogenic strains. In NTM+ lower airway samples, taxa identified as oral commensals were associated with increased inflammatory biomarkers.

Conclusions. The 16S rRNA gene sequencing approach is not sensitive in identifying NTM among airway samples which are culture positive. However, associations between lower airway inflammation and microbiota signatures suggest a potential role for these microbes in the inflammatory process in NTM disease.

Introduction

Non-tuberculosis *Mycobacterium* (NTM) disease has an estimated incidence between 1.0 to 7.2 cases per 100,000 persons/year and its incidence is increasing for unknown reasons [1]. Despite wide spread exposure to these organisms, only a minority of exposed individuals will acquire NTM and an even smaller subgroup will develop clinically-evident disease. Importantly, effectiveness of treatments for active NTM disease has been limited due to an absence of antimicrobial agents with low toxicity and good *in vivo* activity against the organism. Thus, treatment for NTM is not recommended for everyone as the clinical course of the pulmonary disease is variable [2].

The current clinically available methods of culturing airway samples are unable to represent interactions that NTM may have with other bacterial organisms present in a complex microbial environment. With improvements in culture-independent techniques, the microbiota-host immune interaction can be examined in further detail [3]. There have been several studies examining the airway microbiota in non-cystic fibrosis bronchiectasis [4-10] using sputum obtained from cohorts where NTM is not prevalent. Description of the lower airway microbiome, however, has been challenging because the bacterial burden in the lung is approximately a million-fold lower than in the gut and a hundred-fold lower than the upper airway [11, 12]. The presence of supraglottic microbes, such as *Veillonella* or *Prevotella*, in the lower airway is common [11-18] and they are associated with increased inflammation [11, 19] supporting the idea that microbiota changes are linked to the host immune phenotype in the airway mucosa. We therefore seek to identify microbial signatures associated with culture-positive NTM that may affect the host immune phenotype. Here we utilized a 16S rRNA gene high throughput sequencing approach in parallel

to a modified “mycobacteriome” sequencing approach on a large cohort of subjects with non-cystic fibrosis bronchiectasis and high prevalence of NTM.

Material and Methods

Study Subjects

This was a prospective observational study of non-HIV-infected patients (n=106) with a diagnosis of non-cystic fibrosis bronchiectasis enrolled over a two-year period at NYU as part of a USA multicentre bronchiectasis cohort (Bronchiectasis Research Registry). All participants signed informed consent and the protocol was approved by the New York University institutional review boards (IRB# S14-01400). See **supplementary methods** for details on patient selection.

Procedures

Oral wash and induced sputum samples were collected from every patient at enrolment and again when sputum was clinically indicated over the two-year study period (**Supplementary Table 1**). In addition, for every induced sputum we collected an oral wash sample prior to sputum induction. A portion of this sputum sample was sent to the clinical laboratory for culture and, based on the epithelial cell count, 90.5% of induced sputum samples met criteria for good quality [20]. Aliquots of oral wash and induced sputum were frozen at -80°C for bacterial DNA sequencing. In order to investigate how reliable induced sputum was to evaluate the lower airway microbiota and to evaluate the inflammatory status of the lower airway mucosae a subset of patients (n=20) underwent bronchoscopy (clinically indicated in 14 while the remaining 6 were done for research only purposes). Sampling included equipment background controls (sterile saline, yankauer and bronchoscope), supraglottic (sampled using yankauer) and two bronchoalveolar lavage (BAL) samples: one from an involved lung segment (predefined based on CT scan) and the other from a non-involved lung segment. No samples were obtained from

participants during an acute exacerbation or recent antibiotic use (<1month). Whole BAL fluid aliquots were frozen at -80°C for bacterial 16S rRNA gene sequencing as well as 16S qPCR.

Details of DNA sequencing are in the **supplementary methods**. In addition to 16S rRNA gene sequencing performed with Illumina MiSeq, we utilized a nested PCR approach in parallel to enrich for *Mycobacterium* DNA coding for the 16S rRNA gene prior to library preparation for sequencing in order to describe the Mycobacteriome as previously published [21]. The obtained 16S rRNA gene sequences were analyzed using the Quantitative Insights into Microbial Ecology (QIIME 1.9) package [22].

Immune profiling was done for all BAL samples from the 20 patients who were part of the bronchoscopy sub-group. *In vivo* inflammation was assessed from acellular bronchoscopy samples by cell count differential and cytokines using Luminex as previously described [23, 24]. *Ex vivo* cytokine production of BAL cells was evaluated during toll like receptor 4 (TLR4) stimulation. (See **supplementary methods** for details)

Statistical Analysis

For association with discrete factors, we used non-parametric tests (Mann-Whitney or Kruskal-Wallis ANOVA). We used the *ade4* package in R to construct Principal Coordinate Analysis (PCoA) based on weighted UniFrac distances [25, 26]. To cluster microbiome communities into exclusive ‘metacommunities’ we used a Dirichlet Multinomial Mixture Model with the R package *DirichletMultinomial* [27, 28]. To evaluate differences between groups of 16S data or inferred metagenomes, we used linear discriminant analysis (LDA) Effect Size (LEfSe) [29]. For tests of

association with continuous variables, we used non-parametric Spearman correlation tests. Co-occurrence between most abundant bacterial genera (>1% relative abundance in at least one sample) were assessed using SparCC [31] with 20 iterations and 500 bootstrap replicates to eliminate correlations where significance was driven by outliers and visualized using Cytoscape v3.0.2 [32]. Only biomarkers that passed FDR correction were used for this analysis. All data is publicly available in Sequence Read Archive (SRA) under accession number PRJNA418131. All codes utilized for the analysis included in this manuscript are available at https://github.com/segalmicrobiomelab/ntm_bronchiectasis_microbiome

Results

Participants

Table 1 shows demographics and clinical characteristics of 106 patients. All participants had imaging abnormalities. Culture data shows that 61/106 participants (58%) had positive NTM sputum cultures at baseline. BMI was lower among NTM+ participants ($p<0.01$). Importantly, patients had variable prevalence of symptoms and radiographic findings.

Comparing Sputum and Oral Wash Microbiome

To evaluate the airway microbiome, we utilized all obtained oral wash and sputum samples (n=297). Oral wash samples had higher α diversity than sputum (Shannon index, **Supplementary Figure 1A**). Further, β diversity analysis showed significant differences between these two sample types (PERMANOVA $p<0.001$, **Supplementary Figure 1B**) although there was greater degree of similarity between samples from same subject than between subjects (**Supplementary Figure 1C**). LEfSe analysis showed that sputum samples were enriched with *Prevotella*, *Veillonella* and *Corynebacterium*, while oral wash samples were enriched with *Streptococcus*, *Rothia* and *Actinomyces* (**Supplementary Figure 1D**).

Comparison of airway microbiota in NTM+ vs. NTM- using sputum and oral wash samples

We then compared differences in microbiota for each of these sample types based on NTM status. The NTM culture status was based on the culture result of the specimen that was sequenced. **Figure 1** and **Supplementary Figure 2** evaluates differences based on NTM culture status at the time of sample collection. In oral wash samples, there was no significant difference in α diversity but a significant difference in β diversity between NTM- and NTM+ samples (PERMANOVA p=0.043). In sputum samples, there were no statistically significant differences in either α diversity (p=0.05) or β diversity (p=0.08) between NTM- and NTM+ samples. Similar negative results were seen when comparing NTM status based on ATS/IDSA diagnostic criteria (NTM culture positivity in $\geq$ 2 sputum or 1 BAL samples)[33] or when only baseline sample were considered (data not shown).

Interestingly, *Mycobacterium* was not found to be enriched in NTM+ samples. Indeed, this genus was only found present in a very small percentage of both oral or sputum samples. We therefore investigated whether more differences in microbiota could be identified by sampling the lower airway.

Microbiota comparison across the airways using bronchoscopic samples

A subgroup of 20 participants from this cohort (40% with culture positive NTM) underwent bronchoscopy (**Supplementary Table 2 and Supplementary Figure 3).** We first compared how representative the sputum was of the lower airway microbiota (where we sampled involved and non-involved lung segments based on

CT). Quantification of 16S rRNA copies qPCR showed that sputum had approximately 2 log higher bacterial load as compared with BAL samples (**Supplementary Figure 4**). The high bacterial load in sputum was comparable with the bacterial load present in oral wash and supraglottic samples. **Figure 2** displays the α and β diversity for all bronchoscopy related samples. There were significant differences in β diversity ($p<0.01$). The degree of similarity between upper airway, sputum and BAL samples was calculated as weighted UniFrac distance. Interestingly, sputum was more similar to oral wash or to supraglottic samples than to BAL (true for both involved or non-involved lung segments, **Figure 2C**). This data suggest that sputum cannot be used as a surrogate for the lower airways to study the airway microbiota in this cohort.

Comparison of airway microbiota in NTM+ vs. NTM- using bronchoscopic samples

Of the twenty patients that underwent bronchoscopy, 12/20 (60%) were NTM- and 8/20 (40%) were NTM+. There were no significant differences in the bacterial load of NTM+ vs. NTM- lower airway samples (p=ns). **Supplementary Figure 5** shows no statistically significant differences in α or β diversity in BAL samples when categorized based on NTM culture status. Similarly, no differences were noted between BAL samples obtained as part of clinically indicated bronchoscopy as compared with research bronchoscopy (data not shown). Even though *Mycobacterium* was enriched in NTM+ BAL samples (**Figure 3**), this taxon was only found present in 27% of these culture positive samples (median relative abundance 0 [0-0.014]). This data suggested that this sequencing approach was not able to detect the “pathogen” identified as responsible for the disease process and is

consistent with prior literature that has shown poor accuracy of 16S rRNA gene sequencing methods to detect *Mycobacterium* [34].

Evaluation of the Mycobacteriome

We therefore utilized an optimized protocol to enrich for *Mycobacterium* DNA coding for the 16S rRNA gene using a recently published DNA isolation method [35] and a nested PCR approach [36] (see **supplement** for more details). We first utilized a mock community of *Mycobacteriun fortuitum* and *Streptococcus* in order to establish the limit of detection for Mycobacterium with this approach (see **Supplementary Results** and **Supplementary Figure 6** for further details). We then utilized this nested mycobacteriome approach in all samples from the participants that underwent bronchoscopy. PCoA plots comparing standard 16S rRNA gene sequencing and nested mycobacteriome approaches of BAL, sputum and supraglottic samples. **Figure 4** shows significant overlap for a large proportion of samples but identified compositional differences in others. Histograms in **Figure 4** also show the relative abundance of different *Mycobacterium* OTUs obtained with both nested mycobacteriome and 16S rRNA gene sequencing approaches. In BAL samples, *Mycobacterium* was detected in four out of 15 (27%) NTM+ samples with standard 16S rRNA gene sequencing approaches but with the nested mycobacteriome approach *Mycobacterium* was detected in all 4 samples plus 3 other NTM+ BAL samples (47%). Blast analysis demonstrate that these sequences matched *Mycobacterium avium*. Further, this nested mycobacteriome approach identified one out of 21 (5%) NTM- sample with *Mycobacterium*. Blast analysis of this OTU was annotated to *Mycobacterium houstonense* (an environmental

Mycobacterium not known to be pathogenic). Use of this nested mycobacteriome approach on background control samples detected a significant amount of *Mycobacterium* reads (>5% relative abundance) in only one control background sample, and fewer reads in only 2/19 background equipment samples. Blast analysis of *Mycobacterium* reads found in background equipment samples were also annotated to a non-pathogenic strain (**Supplementary Figure 7**). Similar results were found when this approach was utilized in oral and sputum samples from those subjects that did not undergo bronchoscopy (See **supplement** for more details).

Lower airway immunological profiling

To evaluate the association of microbial signatures in NTM disease with a distinct mucosal immune phenotype we examined BAL cell differentials, *in vivo* cytokine levels and *ex vivo* cytokine production. In NTM+ participants, BAL samples from involved lung segments had significantly higher neutrophils and fewer macrophages when compared to non-involved lung segments (**Table 2**). In contrast, in NTM- participants, BAL samples from involved lung segments had significantly higher lymphocytes. *In vivo* cytokine levels measured in BAL also showed a different inflammatory profile for NTM+. In NTM+ participants, BAL samples from involved lung segments had significantly higher levels of IFNγ, IL-8, IL-12p70, ITAC, MIP1α, and MIP1Bβ as compared with non-involved lung segments. This pattern was not present in BAL samples from NTM- participants, where involved lung segments had lower levels of MIP3α and IL-17A. Similarly, a distinct inflammatory pattern was observed during *ex vivo* TLR-4 stimulation of BAL cells (**Supplementary Table 3**). In NTM+ participants, BAL cells from involved lung segments had blunted production of

GM-CSF and IFNγ. These differences were not noted in NTM- participants.

We then evaluated the microbiome signatures associated with these inflammatory biomarkers in NTM+ samples and NTM- samples. For this, we utilized a network approach to evaluate taxa that trend to co-occur and were identified as distinct clusters based on DMM clustering (see **Supplement** and **Supplementary Figure 9)**

In BAL of NTM+ participants, oral commensal such as *Prevotella, Veillonella* and *Leptotrichia* tended to co-occur and had significant correlations with neutrophils and several cytokines including IL-6, IL-17, IL-23, and Fractalkine (**Figure 5**). Interestingly, *Mycobacterium* was in a separate co-occurrence cluster and had no significant correlation with inflammatory biomarkers. In BAL of NTM- samples, the relative abundance of oral commensals in the lower airway samples had fewer significant correlations with cytokines and BAL cells (**Supplementary Figure 10**).

Discussion

The purpose of this investigation was to evaluate the airway microbiota using culture-independent techniques in a prospective cohort of patients suspected of having NTM disease. Our analysis showed that using sputum samples, few changes in microbiota composition could be identified between samples with and without NTM identified by culture. Using upper and lower airway samples from a subgroup of participants who underwent bronchoscopy we showed that induced sputum offers a poor representation of the lower airway microbiota in this patient population and more accurately reflects the composition of the oral cavity. Further, the culture-independent approach did not find *Mycobacterium* in a large percentage of samples. We expanded these observations with a *Mycobacterium*-biased nested sequence approach to confirm that in the majority of NTM+ participants, the abundance for this organism was low or not detected in contrast with the many other microbes identified. This data demonstrates the limited sensitivity of these culture independent methods to detect *Mycobacterium* and exemplifies an unrecognized limitation of current universal sequencing methods to study pathogens present at low abundance. Finally, the lower airways of NTM+ participants had a distinct immunological phenotype where levels of several inflammatory biomarkers correlated with the relative abundance of microbes identified as oral taxa and not with the relative abundance of *Mycobacterium*. These data suggest that micro-aspiration and/or failure to clear aspirated oral microbes may contribute to the inflammatory endotype in NTM disease.

Culture-independent techniques have demonstrated that the airways harbor a complex microbiota that has a significant impact on the host immune response [3,

11, 23, 37, 38]. In a recent study involving 76 non-cystic fibrosis bronchiectasis patients from a multicentre European cohort, *Haemophilus influenza*, *Pseudomonas aeurginosa* and *Streptococcus* were found to be the most abundant species in sputum samples [4]. However, this cohort is characterized by a low prevalence of NTM. In the US, NTM is a frequent cause of non-CF bronchiectasis, as recently shown by Aksamit et al. [39] and multiple strains of NTM are associated with bronchiectasis [40]. In our study, the prevalence of NTM was 58%, similar to the US Bronchiectasis Research Registry [39]. Diagnosis of NTM disease is commonly based on induced sputum. We therefore examined microbiota differences in induced sputum as well as oral wash in our cohort. Both sputum and oral wash samples showed non-significant differences in diversity metrics based on NTM status. Moreover, in NTM+ patients, *Mycobacterium* was not enriched in these samples.

To further characterise the lower airway microbiota in NTM disease a subset of patients underwent bronchoscopy with BAL and upper airway (oropharyngeal) sampling. Differences among oropharyngeal microbiota included enrichment with *Streptococcus* and *Rothia* in oral wash and enrichment with *Prevotella* and *Veillonella* in supraglottic samples. Importantly, comparison between induced sputum and both upper and lower airway samples showed that induced sputum is compositionally more similar to the upper airway microbiota (either oral wash or supraglottic) than to the lower airway microbiota. This supports that induced sputum is predominantly influenced by the composition of the upper airway microbiota and offers a poor representation of the lower airway microbiota. Similar to our findings, the use of induced sputum to evaluate the airway microbiota in a cohort of patients with asthma also provides an incomplete reflection of the lower airways and it is mostly influenced by the oral microbiota.[41] This also has implications to our limited

understanding of the lower airway microbiota using non-invasive samples. In our cohort, BAL samples from NTM+ participants were enriched with *Mycobacterium*, and *Oxalobacteraceae* while BAL samples from NTM- participants were enriched with *Porphyromonas*. However, similar to induced sputum, *Mycobacterium* was frequently not identified using 16S rRNA gene sequencing in samples with positive cultures for this organism. This is consistent with prior observations that have been published [34, 35]. As NTM tends to have only one or two 16S rRNA genes per genome, they can be underrepresented in the context of other taxa with more 16S rRNA genes per bacteria using standard methods of 16S sequencing [33, 42]. In BAL, *Mycobacterium* was identified only in 27% of the NTM+ samples using standard 16S rRNA gene sequencing. For this investigation we seek to use a universal sequencing approach that allows us to characterize broadly the bacterial composition of the airway microbiota. Since *Mycobacterium* was rarely present among NTM+ cases using the now broadly accepted 16S rRNA gene sequencing we attempted to improve our sensitivity in identifying this genus by applying a nested amplification approach where the first PCR targeted a *Mycobacterium*-specific region that contains the V4 region of the 16S rRNA gene [34, 43-45]. It is possible that other *Mycobacterium* specific primers would have a better yield at identifying this organism. Nonetheless, with this method of sequencing, we were able to identify *Mycobacterium* in 47% of the BAL samples where NTM grew in culture. In contrast, *Mycobacterium* was only identified in 17% of NTM+ sputum samples, which is likely also related to how representative the sputum is of the lower airway microbiota.

Our bronchoscopic sampling also allowed us to compare the inflammatory phenotype of NTM+ and NTM- participants. It has previously been shown that with mycobacterial infection, through induction of IFNγ, activated macrophages up

regulate the expression of pro-inflammatory cytokines to help protect against mycobacterial infection. These cytokines include IL-6, IL-1β, IL-12, TNFα and nitric oxide [46, 47]**.** In NTM+ BAL samples obtained from involved sites, BAL cells stimulated with LPS showed a significantly blunted IFNγ and GMCSF suggesting important impaired innate immune responses. In a co-occurrence network analysis, significant associations between taxa identified as oral commensals (e.g. *Prevotella, Veillonella* and *Leptotrichia*) and Th17 cytokines were also seen in NTM+ BAL samples. The relative abundance of *Mycobacterium* was not significantly correlated with levels of inflammatory biomarkers suggesting the importance of other microbes on the lower airway inflammatory tone in NTM disease.

There were several limitations in this study. The patients enrolled in this cohort had a mild form of NTM. A large proportion of patients with NTM grew *Mycobacterium avium complex*, consistent with the US Bronchiectasis Research Registry [33, 39]. It is possible that different strains of *Mycobacterium* and different degrees of disease severity would have a different airway microbiota and inflammatory signatures than what was observed in our study. In this investigation, a relatively small proportion of patients underwent bronchoscopy allowing us to evaluate the lower airways [48] and most were done for clinical reasons. Although similar to our larger cohort, these patients may represent a different disease phenotype. However, given the limited representation of the lower airways provided by the microbiota present in sputum samples, further investigation is warranted in a larger cohort to uncover microbiota host interactions that might be relevant in this disease and to define which microbiota signatures present in induced sputum could be used to explore the lower airway microbiota. Further, even though we identified some significant associations between lower airway microbiota signatures and inflammatory biomarkers in the NTM+ group,

we considered these results as exploratory and hypothesis generating. A larger cohort of patients where bronchoscopic samples are obtained will be needed to confirm and expand these findings. Finally, the purpose of this study was to evaluate the microbiome community of patients with NTM related bronchiectasis. Therefore, we did not evaluate the change in microbiome composition with exacerbations nor the effects of treatment for NTM. Importantly, patients on treatment for NTM were excluded from this analysis to avoid this potential confounder to the analysis of differences between NTM+ vs. NTM- groups.

In summary, we identified the limitations of current unbiased culture-independent techniques to identify *Mycobacterium* in patients with culture positivity for NTM, which highlights the need for technical improvements in these methods. In addition, we describe that patients with NTM disease have a distinct inflammatory environment in the lower airways that may be associated with some of the components of the lower airway microbiota including taxa commonly identified as oral commensals. These data suggest a possible role of micro-aspiration or failure to clear upper airway microbes from the lower airways and may explain some of the heterogeneity in presentation and disease progression among participants with culture positive NTM disease. The contribution of the lower airway microbiota to the pathophysiological inflammatory process in NTM disease warrants further investigation in a larger cohort and may have potential therapeutic implications.

Acknowledgments

All listed contributed to this manuscript. IS and LNS were involved in conception and design. BS, JW, AB, YL, AS, SC, KB, JB, JC, SR, GM, JD, RC, DK, DAH, IS and LNS were all involved in acquisition of data. BGW, PM, JCC, NS, CN, GT, IS and LNS were involved in analysis and interpretation of data. BGW, YL, JW, PM, JCC, NS, GT, JB, KB, SC, SM, GM, JD, RC, CN, AB, AS, DK, DAH, IS and LNS were involved in drafting and revising of this manuscript. IS and LNS were involved in the final approval of this manuscript.

BS, JW, AB, YL, AS, SC, KB, JB, JC, SR, GM, JD, RC, DK, DAH, IS and LNS are all affiliated with Division of Pulmonary, Critical Care, & Sleep Medicine, New York University School of Medicine, NY

NS and JCC are affiliated with Department of Genetics and Genomic Sciences and Immunology Institute, Icahn School of Medicine at Mount Sinai, NY

JCC is also affiliated with Precision Immunology Institute, Icahn School of Medicine at Mount Sinai, NY

CN and GT are affiliated with Medicine and Health Sciences, Stellenbosch University, DST/NRF of Excellence for Biomedical Tuberculosis Research, and SA MRC Centre for Molecular and Cellular Biology, Division of Molecular Biology and Human Genetics Tygerberg, ZA

References:

1. Cassidy PM, Hedberg K, Saulson A, McNelly E, Winthrop KL. Nontuberculous mycobacterial disease prevalence and risk factors: a changing epidemiology. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America. 2009;49(12):e124-9.
2. Saleeb P, Olivier KN. Pulmonary Nontuberculous Mycobacterial Disease: New Insights into Risk Factors for Susceptibility, Epidemiology, and Approaches to Management in Immunocompetent and Immunocompromised Patients. Current infectious disease reports. 2010;12(3):198-203.
3. Segal LN, Rom WN, Weiden MD. Lung microbiome for clinicians. New discoveries about bugs in healthy and diseased lungs. Ann Am Thorac Soc. 2014;11(1):108-16.
4. Cox MJ, Turek EM, Hennessy C, Mirza GK, James PL, Coleman M, et al. Longitudinal assessment of sputum microbiome by sequencing of the 16S rRNA gene in non-cystic fibrosis bronchiectasis patients. PLoS One. 2017;12(2):e0170622.
5. Tunney MM, Einarsson GG, Wei L, Drain M, Klem ER, Cardwell C, et al. Lung Microbiota and Bacterial Abundance in Patients with Bronchiectasis when Clinically Stable and during Exacerbation. American Journal of Respiratory and Critical Care Medicine. 2013;187(10):1118-26.
6. Boyton RJ, Reynolds CJ, Quigley KJ, Altmann DM. Immune mechanisms and the impact of the disrupted lung microbiome in chronic bacterial lung infection and bronchiectasis. Clinical and Experimental Immunology. 2013;171(2):117-23.
7. Rogers GB, van der Gast CJ, Cuthbertson L, Thomson SK, Bruce KD, Martin ML, et al. Clinical measures of disease in adult non-CF bronchiectasis correlate with airway microbiota composition. Thorax. 2013;68(8):731-7.
8. Rogers GB, Zain NM, Bruce KD, Burr LD, Chen AC, Rivett DW, et al. A novel microbiota stratification system predicts future exacerbations in bronchiectasis. Ann Am Thorac Soc. 2014;11(4):496-503.
9. Maughan H, Cunningham KS, Wang PW, Zhang Y, Cypel M, Chaparro C, et al. Pulmonary bacterial communities in surgically resected noncystic fibrosis bronchiectasis lungs are similar to those in cystic fibrosis. Pulmonary medicine. 2012;2012:746358.
10. Purcell P, Jary H, Perry A, Perry JD, Stewart CJ, Nelson A, et al. Polymicrobial airway bacterial communities in adult bronchiectasis patients. BMC Microbiology. 2014;14:130-.
11. Segal LN, Alekseyenko AV, Clemente JC, Kulkarni R, Wu B, Chen H, et al. Enrichment of lung microbiome with supraglottic taxa is associated with increased pulmonary inflammation. Microbiome. 2013;1(1):19.
12. Bassis CM, Erb-Downward JR, Dickson RP, Freeman CM, Schmidt TM, Young VB, et al. Analysis of the upper respiratory tract microbiotas as the source of the lung and gastric microbiotas in healthy individuals. MBio. 2015;6(2):e00037.
13. Charlson ES, Bittinger K, Haas AR, Fitzgerald AS, Frank I, Yadav A, et al. Topographical continuity of bacterial populations in the healthy human respiratory tract. Am J Respir Crit Care Med. 2011;184(8):957-63.
14. Erb-Downward JR, Thompson DL, Han MK, Freeman CM, McCloskey L, Schmidt LA, et al. Analysis of the lung microbiome in the "healthy" smoker and in COPD. PLoS One. 2011;6(2):e16384.
15. Morris A, Beck JM, Schloss PD, Campbell TB, Crothers K, Curtis JL, et al. Comparison of the Respiratory Microbiome in Healthy Non-Smokers and Smokers. Am J Respir Crit Care Med. 2013;187(10):1067-75.
16. Pragman AA, Kim HB, Reilly CS, Wendt C, Isaacson RE. The lung microbiome in moderate and severe chronic obstructive pulmonary disease. PLoS One. 2012;7(10):e47305.

17. Sze MA, Dimitriu PA, Hayashi S, Elliott WM, McDonough JE, Gosselink JV, et al. The lung tissue microbiome in chronic obstructive pulmonary disease. Am J Respir Crit Care Med. 2012;185(10):1073-80.
18. Lozupone C, Cota-Gomez A, Palmer BE, Linderman DJ, Charlson ES, Sodergren E, et al. Widespread colonization of the lung by Tropheryma whipplei in HIV infection. Am J Respir Crit Care Med. 2013;187(10):1110-7.
19. Dickson RP, Erb-Downward JR, Prescott HC, Martinez FJ, Curtis JL, Lama VN, et al. Cell-associated bacteria in the human lung microbiome. Microbiome. 2014;2:28.
20. Zsoka Weiszhar IH. Induced sputum analysis: step by step. Breathe. 2013;9(4):300-6.
21. Macovei L, McCafferty J, Chen T, Teles F, Hasturk H, Paster BJ, et al. The hidden 'mycobacteriome' of the human healthy oral cavity and upper respiratory tract. J Oral Microbiol. 2015;7:26094.
22. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. Nat Methods. 2010;7(5):335-6.
23. Segal LN, Clemente JC, Tsay JC, Koralov SB, Keller BC, Wu BG, et al. Enrichment of the lung microbiome with oral taxa is associated with lung inflammation of a Th17 phenotype. Nat Microbiol. 2016;1:16031.
24. Berger KI, Pradhan DR, Goldring RM, Oppenheimer BW, Rom WN, Segal LN. Distal airway dysfunction identifies pulmonary inflammation in asymptomatic smokers. ERJ open research. 2016;2(4).
25. Dray SaD, A.B. The ade4 package: implementing the duality diagram for ecologists. Journal of Statistical Software. 2007;22(4):1-20.
26. Lozupone C, Lladser ME, Knights D, Stombaugh J, Knight R. UniFrac: an effective distance metric for microbial community comparison. The ISME journal. 2011;5(2):169-72.
27. Holmes I, Harris K, Quince C. Dirichlet multinomial mixtures: generative models for microbial metagenomics. PLoS One. 2012;7(2):e30126.
28. Morgan M. DirichletMultinomial: Dirichlet-Multinomial Mixture Model Machine Learning for Microbiome
Data. R package version 1.20.0 ed2017.
29. Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS, et al. Metagenomic biomarker discovery and explanation. Genome Biol. 2011;12(6):R60.
30. Reiner A, Yekutieli D, Benjamini Y. Identifying differentially expressed genes using false discovery rate controlling procedures. Bioinformatics. 2003;19(3):368-75.
31. Friedman J, Alm EJ. Inferring correlation networks from genomic survey data. PLoS Comput Biol. 2012;8(9):e1002687.
32. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome research. 2003;13(11):2498-504.
33. Griffith DE, Aksamit T, Brown-Elliott BA, Catanzaro A, Daley C, Gordin F, et al. An official ATS/IDSA statement: diagnosis, treatment, and prevention of nontuberculous mycobacterial diseases. Am J Respir Crit Care Med. 2007;175(4):367-416.
34. Macovei L, McCafferty J, Chen T, Teles F, Hasturk H, Paster BJ, et al. The hidden ‘mycobacteriome’ of the human healthy oral cavity and upper respiratory tract. Journal of Oral Microbiology. 2015;7:10.3402/jom.v7.26094.
35. Caverly LJ, Carmody LA, Haig SJ, Kotlarz N, Kalikin LM, Raskin L, et al. Culture-Independent Identification of Nontuberculous Mycobacteria in Cystic Fibrosis Respiratory Samples. PLoS One. 2016;11(4):e0153876.
36. Kauppinen J, Mäntyjärvi R, Katila M-L. Mycobacterium malmoense-Specific Nested PCR Based on a Conserved Sequence Detected in Random Amplified Polymorphic DNA Fingerprints. Journal of clinical microbiology. 1999;37(5):1454-8.

37. Segal LN, Clemente JC, Wu BG, Wikoff WR, Gao Z, Li Y, et al. Randomised, double-blind, placebo-controlled trial with azithromycin selects for anti-inflammatory microbial metabolites in the emphysematous lung. Thorax. 2017;72(1):13-22.
38. Shenoy MK, Iwai S, Lin DL, Worodria W, Ayakaka I, Byanyima P, et al. Immune Response and Mortality Risk Relate to Distinct Lung Microbiomes in Patients with HIV and Pneumonia. Am J Respir Crit Care Med. 2017;195(1):104-14.
39. Aksamit TR, O'Donnell AE, Barker A, Olivier KN, Winthrop KL, Daniels MLA, et al. Adult Patients With Bronchiectasis: A First Look at the US Bronchiectasis Research Registry. Chest. 2017;151(5):982-92.
40. Iseman MD, Marras TK. The importance of nontuberculous mycobacterial lung disease. Am J Respir Crit Care Med. 2008;178(10):999-1000.
41. Durack J, Huang YJ, Nariya S, Christian LS, Mark Ansel K, Beigelman A, et al. Bacterial biogeography of adult airways in atopic asthma. Microbiome. 2018;6(1):104.
42. Danielides V, Patrikakos G, Moerman M, Bonte K, Dhooge C, Vermeersch H. Diagnosis, management and surgical treatment of non-tuberculous mycobacterial head and neck infection in children. ORL; journal for oto-rhino-laryngology and its related specialties. 2002;64(4):284-9.
43. Kim BJ, Lee SH, Lyu MA, Kim SJ, Bai GH, Chae GT, et al. Identification of mycobacterial species by comparative sequence analysis of the RNA polymerase gene (rpoB). Journal of clinical microbiology. 1999;37(6):1714-20.
44. Yamada-Noda M, Ohkusu K, Hata H, Shah MM, Nhung PH, Sun XS, et al. Mycobacterium species identification--a new approach via dnaJ gene sequencing. Systematic and applied microbiology. 2007;30(6):453-62.
45. Tirkkonen T, Nieminen T, Ali-Vehmas T, Peltoniemi OA, Wellenberg GJ, Pakarinen J. Quantification of Mycobacterium avium subspecies in pig tissues by real-time quantitative PCR. Acta veterinaria Scandinavica. 2013;55:26.
46. Matsuyama M, Ishii Y, Yageta Y, Ohtsuka S, Ano S, Matsuno Y, et al. Role of Th1/Th17 balance regulated by T-bet in a mouse model of Mycobacterium avium complex disease. Journal of immunology (Baltimore, Md : 1950). 2014;192(4):1707-17.
47. Ito T, Connett JM, Kunkel SL, Matsukawa A. The linkage of innate and adaptive immune response during granulomatous development. Frontiers in immunology. 2013;4:10.
48. Dickson RP, Erb-Downward JR, Freeman CM, McCloskey L, Falkowski NR, Huffnagle GB, et al. Bacterial Topography of the Healthy Human Lower Respiratory Tract. MBio. 2017;8(1).

Figure & Table Legends:

Figure 1: Taxonomic Differences between all oral wash and sputum samples for the NTM+ and NTM- groups. **A**. For oral wash samples LeFSE analysis identified significant taxonomic differences in microbiome enrichment based on NTM status but there was no enrichment with *Mycobacterium* in NTM+ oral wash samples. **B**. For sputum LeFSE analysis detected few taxonomic differences and there was no enrichment with *Mycobacterium* in NTM+ sputum samples.

Figure 2: Evaluation of the lower airway microbiota using bronchoscopic samples. Bronchoscopy samples included: Background (BKG), Nasal Swab, Oral Wash, Sputum, Supraglottic (SUP) and Bronchoalveolar Lavage (BAL). **A**. There were significant differences in α diversity (Shannon Diversity Index) between all samples. **B**. β diversity (based on weighted UniFrac) showed differential clustering based on sample type (PERMANOVA $p<0.01$). **C**. Comparison of the degree of similarities between upper and lower airway samples based on UniFrac distance. Results showed that the microbiota in sputum was more similar to the microbiota of supraglottic and oral wash samples, than to the microbiota in BAL samples (both involved and non-involved).

Figure 3: Taxonomic differences between NTM+ and NTM- in lower airway microbiota. LeFSE analysis showed significant differences in lower airway microbiota composition enrichment based on NTM status. Unlike sputum samples, *Mycobacterium* was enriched in NTM+ BAL samples.

Figure 4: Comparison between sequence data obtained using an unbiased 16S rRNA approach and a biased mycobacteriome approach. Top panels show

differences in β diversity (based on weighted UniFrac distance) between samples processed with unbiased 16S vs. biased mycobacteriome approach. Bar charts in the bottom panels show relative abundance of OTUs annotated to *Mycobacterium* and their annotation based on BLAST on the two datasets. **A**. In BAL, there were significant differences in β diversity between BAL samples processed with unbiased 16S vs. biased mycobacteriome approach (PERMANOVA $p< 0.01$). With the biased mycobacteriome approach, *Mycobacterium avium* was identified in 7 of the NTM+ culture samples while one NTM- sample had a *Mycobacterium* OTU annotated to *M. houstonense* (a strain not known to be pathogenic). **B.** In sputum there were significant differences in β diversity between the sputum samples processed by unbiased16S vs. biased mycobacteriome approach (PERMANOVA $p< 0.01$). With the biased mycobacteriome approach, *Mycobacterium avium* was identified in 3 of the NTM+ culture samples. **C.** In supraglottic there was no significant difference in β diversity between samples processed by unbiased 16S vs. biased mycobacteriome approach. In two samples, a *Mycobacterium* annotated to a non-pathogenic strain was identified.

Figure 5: Associations between taxa and inflammatory biomarkers in the lower airways for NTM+ samples. A network analysis using SparCC, to remove compositional and sparsity effects common in microbiome data, was constructed to identify correlations between taxa at a genus level (round nodes) and both inflammatory biomarkers (green squares). Each node represents a genus, with the size of nodes indicating the log-relative abundance from large (high) to small (low). In addition, taxa identified as marker for lower airway microbiota cluster (based on DMM analysis shown in **Supplementary Figure 5**) was color coded as purple for Cluster 1 and red for Cluster 2. Edges between nodes represent significant

correlations (where blue indicate positive correlations and dashed grey indicate negative correlations), with the length of the edge representing the correlation coefficient strength (shorter edges indicating higher positive correlation).

Table 1: Baseline demographic and clinical information of all patients recruited

VARIABLES	All Patients	NTM Status (-)	NTM Status (+)	p value
N	106	45	61	
Age	67.5 (10.7)	67.6 (10.1)	67.3 (11.3)	0.88
No. Female (%)	95 (89)	37 (82)	58 (95)	0.71
No. Caucasian (%)	92 (87)	34 (76)	58 (95)	0.39
BMI	22.7 (4.3)	24.2 (5.2)	21.5 (3.1)	<0.01
Packs Per Day	1.3 (2.2)	1.2 (1.0)	1.4 (3.0)	0.85
No. Years Smoking	19.2 (13.1)	22.9 (14.5)	15.1 (9.9)	0.04
Symptoms - No. Patients (%)				
Cough	63 (59)	22 (49)	41 (67)	0.07
Sputum	47 (44)	14 (31)	33 (54)	0.15
Hemoptysis	8 (8)	3 (7)	5 (8)	0.76
Shortness of Breath	42 (39)	17 (38)	25 (41)	0.92
Wheeze	14 (13)	7 (16)	7 (11)	0.6
Fatigue	26 (25)	9 (20)	17 (28)	0.6
Postnasal Drip	36 (34)	19 (42)	17 (28)	0.03
Sinusitis	4 (4)	4 (9)	0 (0)	0.03
Current NTM Culture				
MAC	56 (53)	0 (0)	56 (92)	<0.001
M. abscessus	5 (5)	0 (0)	5 (8)	<0.05
Current Bacterial Culture				
Oropharyngeal Flora	15 (14)	4 (9.5)	11 (18)	0.13
Pseudomonas aeruginosa	6 (6)	1 (2)	5 (8)	0.19
MSSA	5 (5)	2 (4)	3 (5)	0.91
Other	10 (9)	4 (9.5)	6 (10)	0.89
Negative Culture	17 (16)	9 (20)	8 (13)	0.34
Lung Physiology*				
FVC (% predicted)	92.7 (20.1)	92.6 (20.4)	92.8 (20.0)	0.97
FEV_1 (% predicted)	83.7 (21.8)	83.4 (22.1)	84.0 (21.8)	0.91
FEV_1/FVC (%)	70.2 (10.6)	69.2 (10.7)	71.0 (10.6)	0.44
RV (% predicted)	121.8 (31.7)	118.4 (32.8)	124.5 (31.0)	0.41
TLC (% predicted)	106.4 (18.5)	105.6 (16.2)	106.9 (20.2)	0.76
DLCO (% predicted)	87.4 (20.7)	90.7 (23.7)	85.4 (18.7)	0.35
HRCT Thorax - No. Patients (%)				
Bronchiectasis	79 (75)	34 (76)	45 (78)	0.45
Thickened Airways	52 (49)	23 (51)	29 (59)	0.73
Mucoid Impaction	54 (51)	20 (44)	34 (74)	0.57
Nodules	70 (66)	29 (64)	41 (67)	0.29
Tree-in-Bud	30 (28)	11 (24)	19 (31)	0.41
Ground Glass Opacification	31 (29)	17 (38)	14 (23)	0.16
Questionnaire Data (n=101)				
SGRQ Total Score	27.5 (19.3)	28.4 (21.1)	26.8 (18.1)	0.68
RSI Total	11.9 (9.2)	13.7 (9.4)	10.7 (8.9)	0.12
FSSG Total	8.5 (8.5)	9.8 (9.7)	7.6 (7.5)	0.22
EAT 10 Total	3.0 (5.8)	3.0 (5.6)	40.8 (20.0)	0.97

* Available in 89 participants. BMI = Body mass index; MAC = *Mycobacterium avium complex*; M.Abcessus = *Mycobacterium abscessus*; *MSSA*= methicillin-sensitive *Staphylococcus aureus*; FVC= forced vital capacity; FEV_1 = Forced expiratory volume in 1 second; RV = Residual volume; TLC = Total lung capacity; DLCO = Diffusion Lung capacity of carbon monoxide; HRCT = High resolution computed tomography; SGRQ = St. George's respiratory questionnaire; RSI = Reflux symptom index; FSSG = Frequency scale for the symptoms of gastro-esophageal reflux disease; EAT = Eating assessment tool.

Table 2: Cell count, and *in vivo* cytokine levels in bronchoalveolar lavage (BAL) of 20 patients from the bronchoscopy cohort.

	NTM- (n=12)			NTM+ (n=8)		
	Involved	Non-Involved	P value	Involved	Non-Involved	P value
Cell Count						
Macrophages	44.2 [28.2-75.7]	75.4 [51.7-80.4]	ns	19.6 [15.75-39.65]	75.4 [66.0-85.0]	*0.02*[¥]
Neutrophils	52.8 [13.4-69.8]	13.2 [10.5-35.05]	ns	79.6 [58.2-82.75]	17.1 [6.5-31.2]	*0.02*[¥]
Lymphocytes	2.8 [2.6-10.9]	11.4 [2.95-12.7]	*0.02*	2.2 [1.7-2.45]	1.8 [0.5-3.8]	ns
Eosinophils	0 [0-0]	0 [0-0]	ns	0 [0-0]	0 [0-0]	ns
In Vivo Cytokine Level (pg/ml)						
ITAC	19.39 [19.4-224.4]*	224.4 [135.6-274.9]	ns	426.4 [372.5-1135.8]*	469.73 [352.8-489.7]	*0.02*[†]
MIP1β	64.875 [60.7-114.2]	83.545 [61.8-109.2]	ns	138.08 [134.1-188.8]	111.83 [71.8-141.4]	*0.02*[†]
IFNγ	21.36 [6.4-30]	8.46 [6.7-13.8]	ns	63.085 [24.8-158.6]	22.095 [17.6-26.5]	*0.03*[†]
IL-8	423.565 [216.9-486.8]*	351.12 [240.7-1810.2]	ns	5927.82 [5927.8-5927.8]*	411.345 [242.8-586.8]	*0.03*[†]
IL-12 p70	3.95 [3.7-4.1]*	3.715 [3.5-4.3]	ns	5.155 [4.2-6.6]*	4.605 [4.3-4.6]	*0.06*
MIP1α	116.88 [66.2-179.7]	95.855 [72.4-128.3]	ns	268.83 [143.4-489.8]	117.975 [78.5-143.1]	*0.06*
MIP3α	49.105 [51.9-86.3]*	86.26 [75.6-254.4]	*0.03*	356.01 [300.8-443.1]*	201.49 [141.7-228.2]	ns
IL-17A	9.65 [10.1-18.2]	18.18 [15-20.7]	*0.06*	11.02 [10.5-13.2]	15.44 [10.4-19.9]	ns
IL-13	7.445 [8.3-19.3]	13.97 [11.3-17.9]	ns	9.73 [8.6-12]	70.87 [9.4-145.9]	ns
IL-2	2.75 [2.8-4.7]*	4.68 [3.6-5.5]	ns	5.07 [3.7-8.3]*	2.80 [2.8-4.3]	ns
Fractalkine	364.12 [364.1-391.3]*	391.255 [360.4-449.6]	ns	488.115 [455.5-544.1]*	558.94 [403.9-558.9]	ns
IL-1β	22.27 [1.9-11.1]	3.94 [2-6.5]	ns	34.345 [3.8-72.4]	2.405 [2.1-65.3]	ns
IL-6	11.275 [10.1-104]*	104.025 [58.9-154.3]	ns	82.735 [60.9-133.4]*	114.99 [69.7-147.3]	ns
IL-21	4.825 [4.5-6.5]	5.95 [4.7-7.3]	ns	6.655 [6-8.8]	7.45 [5.4-9.17]	ns
IL-7	15.935 [15.2-38.4]	24.97 [16.1-36.4]	ns	27.365 [26.7-31.6]	29.545 [22.8-33.6]	ns
IL-5	32.475 [13.4-40.3]	17.03 [12.1-23.2]	ns	26.395 [23.9-31.4]	56.18 [56.2-18.4]	ns
IL-23	158.025 [88.4-160.6]	118.66 [88.4-149.9]	ns	153.865 [126.4-173.7]	252.9 [165.4-359.9]	ns
TNFα	208.485 [36.1-231.5]	44.63 [37.5-105.5]	ns	90.83 [67.3-130.3]	54.845 [39.4-105.2]	ns
GM-CSF	19.225 [18.1-41.2]	29.935 [19.6-40]	ns	35.795 [32.1-42.2]	27.215 [22.5-37.0]	ns
IL-4	29.09 [22-30.5]	25.42 [22.9-32.6]	ns	39.855 [31-50.2]	46.31 [37.1-49.0]	ns
IL-10	61.015 [50.8-84.6]	66.315 [55.1-74]	ns	45.74 [32.1-68.6]	83.5 [68.2-89.1]	ns

Data represented as Median [IQR]. p-value based on Mann Whitney. *Comparing involved sites by NTM status. **Comparing non-involved sites by NTM status. [¥]False Discovery Rate (FDR) <0.05. [†]FDR <0.2.

Figure 1

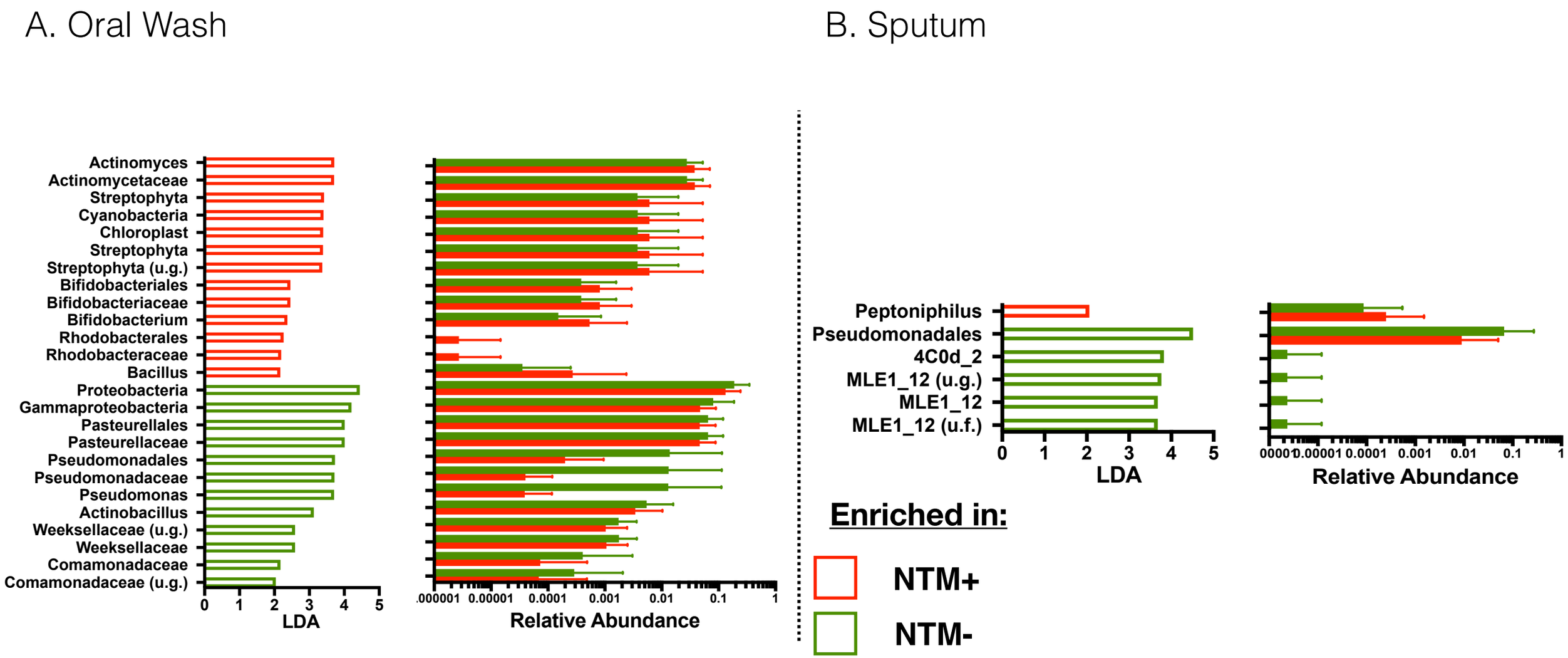

A. Oral Wash
Actinomyces
Actinomycetaceae
Streptophyta
Cyanobacteria
Chloroplast
Streptophyta
Streptophyta (u.g.)
Bifidobacteriales
Bifidobacteriaceae
Bifidobacterium
Rhodobacterales
Rhodobacteraceae
Bacillus
Proteobacteria
Gammaproteobacteria
Pasteurellales
Pasteurellaceae
Pseudomonadales
Pseudomonadaceae
Pseudomonas
Actinobacillus
Weeksellaceae (u.g.)
Weeksellaceae
Comamonadaceae
Comamonadaceae (u.g.)
0 1 2 3 4 5
LDA
.000001 0.00001 0.0001 0.001 0.01 0.1 1
Relative Abundance
B. Sputum
Peptoniphilus
Pseudomonadales
4C0d_2
MLE1_12 (u.g.)
MLE1_12
MLE1_12 (u.f.)
0 1 2 3 4 5
LDA
00001 0.00001 0.0001 0.001 0.01 0.1 1
Relative Abundance
Enriched in:
NTM+
NTM-

Figure 2

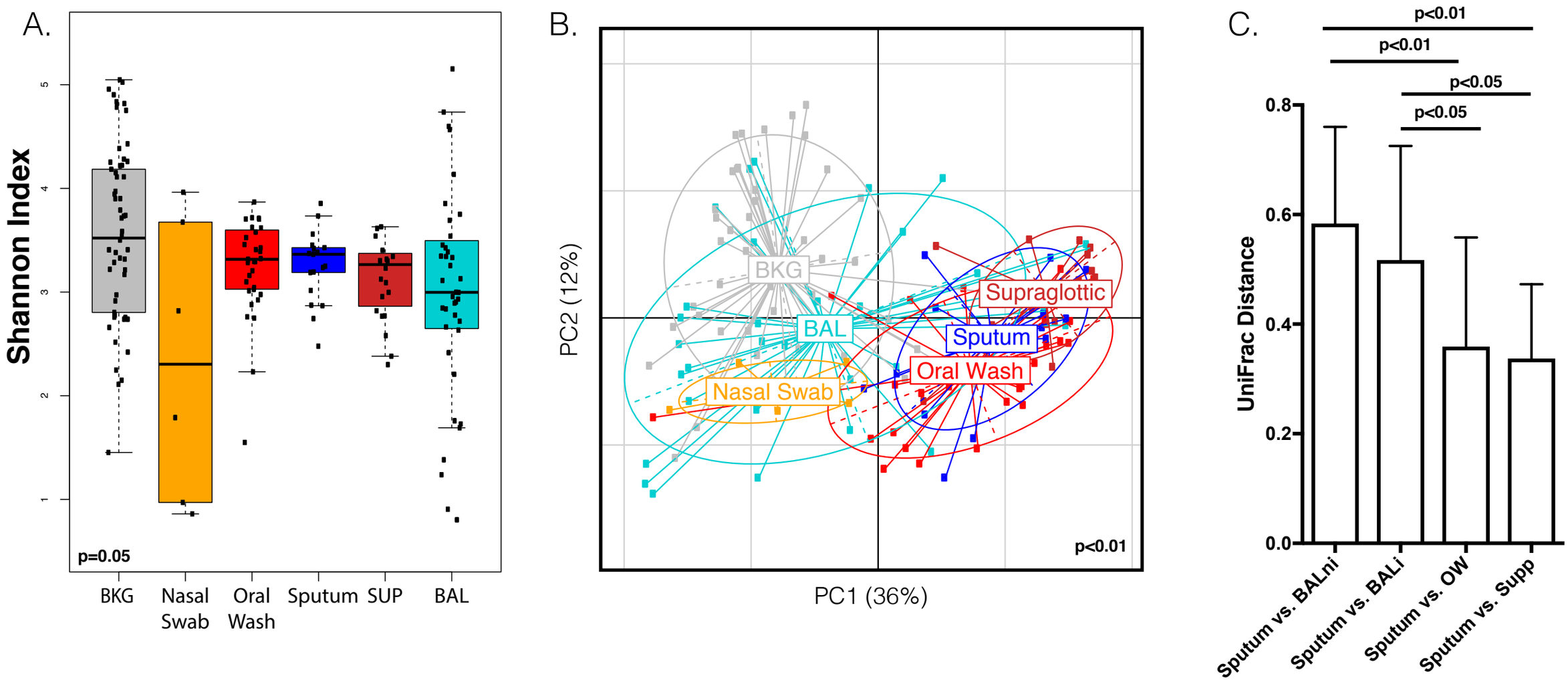

A.
Shannon Index
1
2
3
4
5
p=0.05
BKG
Nasal Swab
Oral Wash
Sputum
SUP
BAL
B.
PC2 (12%)
PC1 (36%)
BKG
BAL
Nasal Swab
Supraglottic
Sputum
Oral Wash
p<0.01
C.
p<0.01
p<0.01
p<0.05
p<0.05
UniFrac Distance
0.0
0.2
0.4
0.6
0.8
Sputum vs. BALni
Sputum vs. BALi
Sputum vs. OW
Sputum vs. Supp

Figure 3

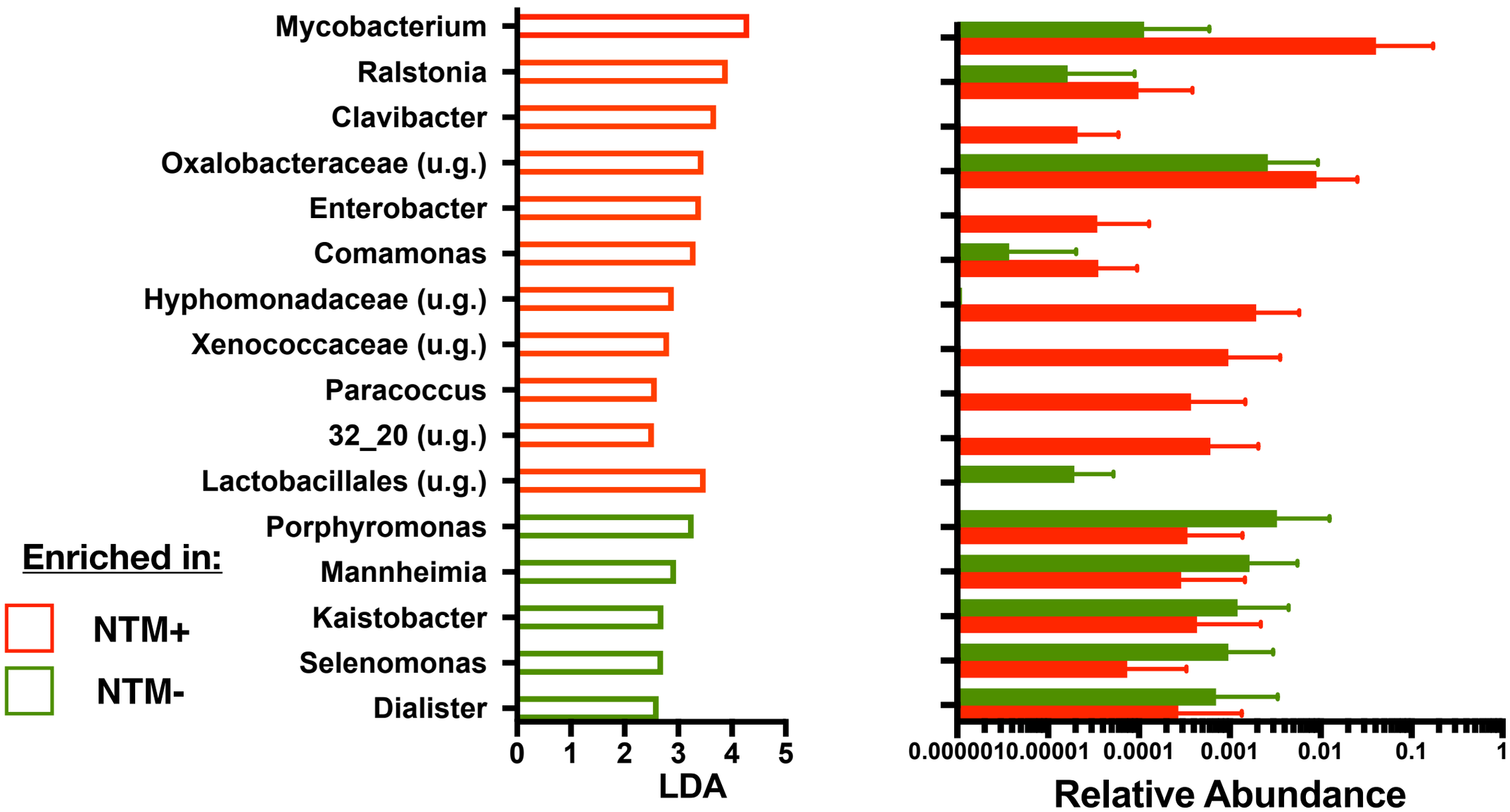

Mycobacterium
Ralstonia
Clavibacter
Oxalobacteraceae (u.g.)
Enterobacter
Comamonas
Hyphomonadaceae (u.g.)
Xenococcaceae (u.g.)
Paracoccus
32_20 (u.g.)
Lactobacillales (u.g.)
Porphyromonas
Mannheimia
Kaistobacter
Selenomonas
Dialister
Enriched in:
NTM+
NTM-
0
1
2
3
4
5
LDA
0.0000010.00001
0.0001
0.001
0.01
0.1
1
Relative Abundance

Figure 4

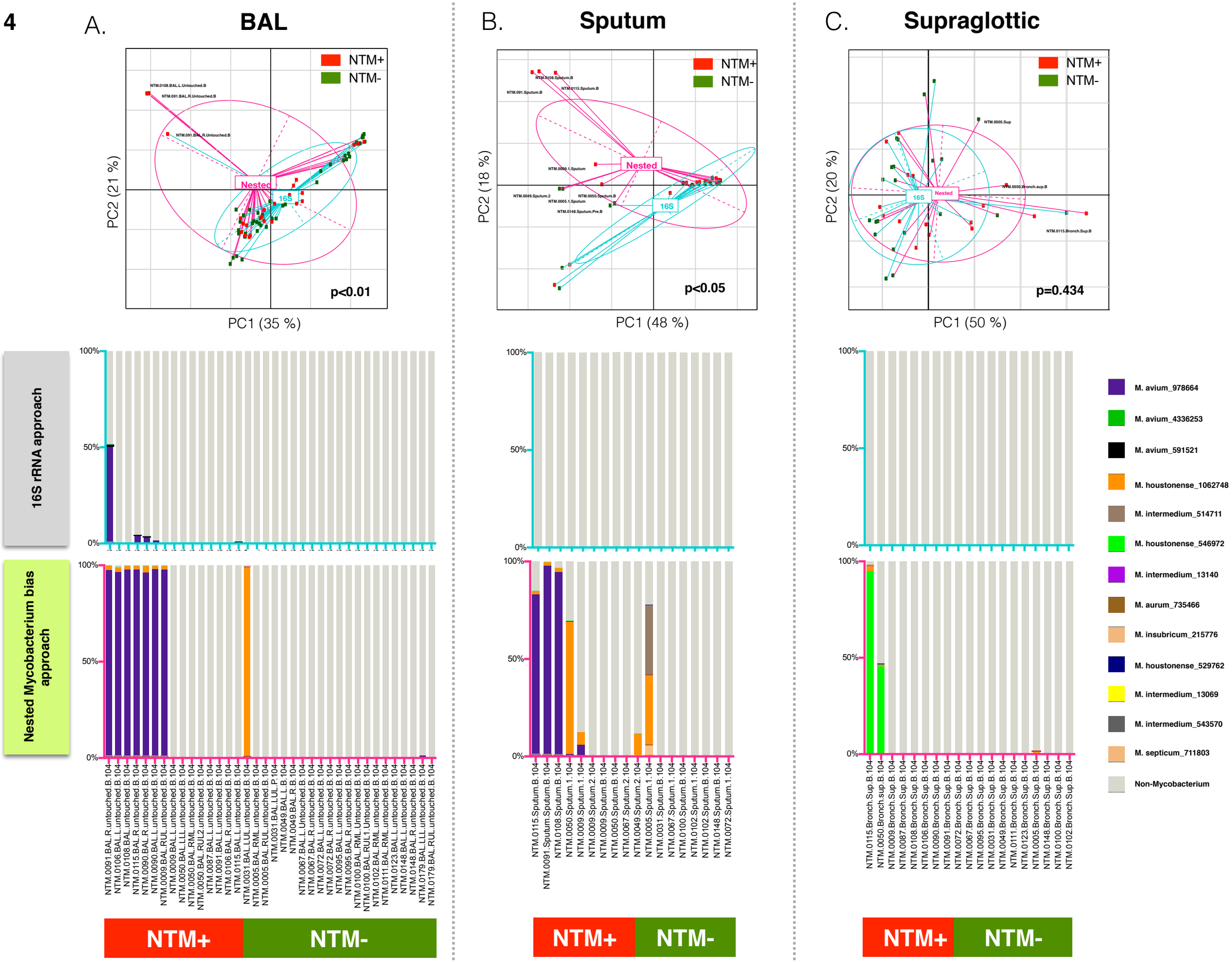
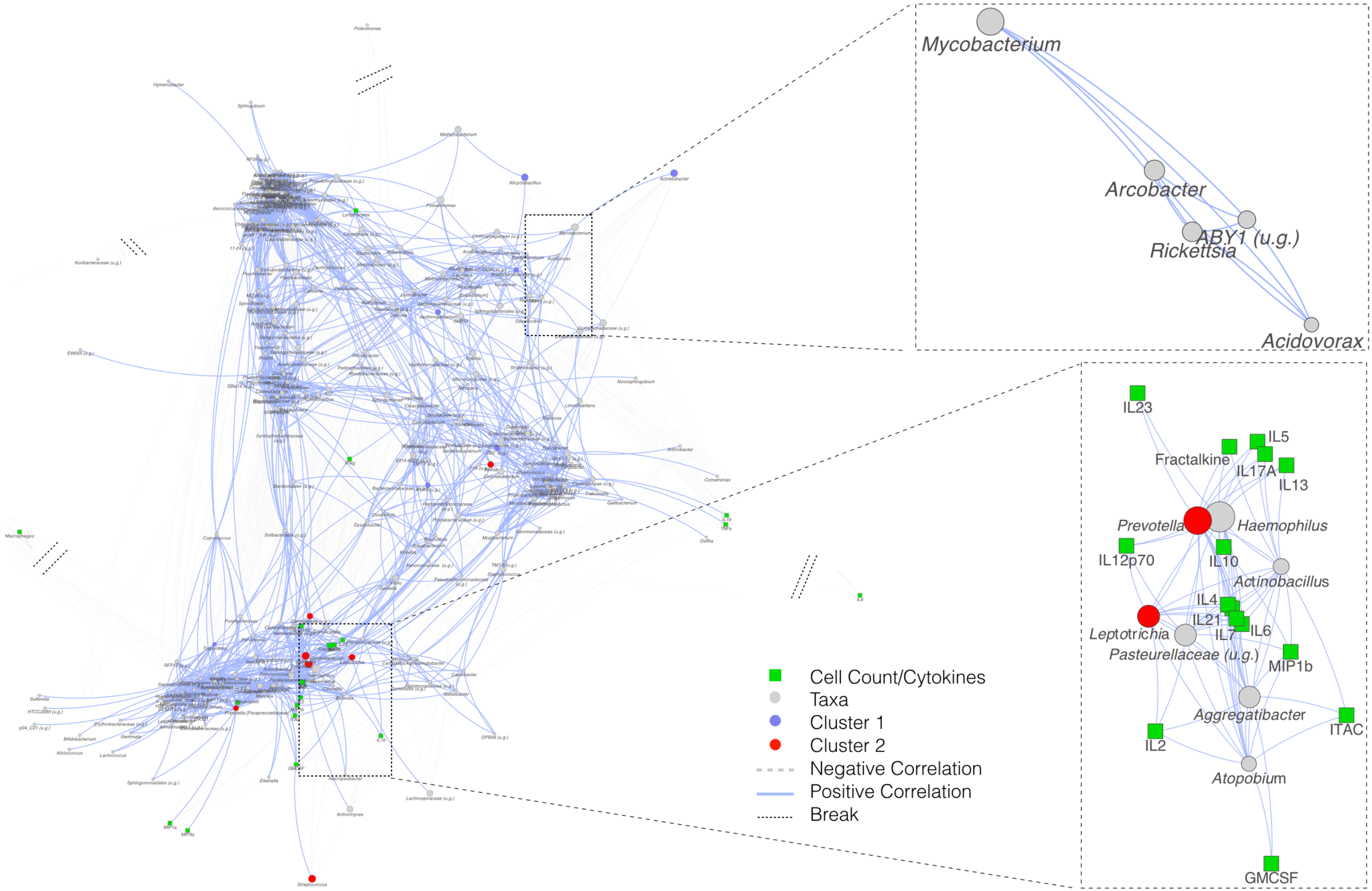

A.
BAL
NTM+
NTM-
Nested
16S
p<0.01
PC1 (35 %)
PC2 (21 %)
B.
Sputum
NTM+
NTM-
Nested
16S
p<0.05
PC1 (48 %)
PC2 (18 %)
C.
Supraglottic
NTM+
NTM-
Nested
16S
p=0.434
PC1 (50 %)
PC2 (20 %)
16S rRNA approach
Nested Mycobacterium bias approach
100%
50%
0%
M. avium_978664
M. avium_4336253
M. avium_591521
M. houstonense_1062748
M. intermedium_514711
M. houstonense_546972
M. intermedium_13140
M. aurum_735466
M. insubricum_215776
M. houstonense_529762
M. intermedium_13069
M. intermedium_543570
M. septicum_711803
Non-Mycobacterium
NTM+
NTM-

Figure 5


Mycobacterium
Arcobacter
ABY1 (u.g.)
Rickettsia
Acidovorax
IL23
IL5
Fractalkine
IL17A
IL13
Prevotella
Haemophilus
IL12p70
IL10
Actinobacillus
IL4
IL21
IL7
IL6
Leptotrichia
Pasteurellaceae (u.g.)
MIP1b
Aggregatibacter
ITAC
IL2
Atopobium
GMCSF
Cell Count/Cytokines
Taxa
Cluster 1
Cluster 2
Negative Correlation
Positive Correlation
Break

Supplementary Methods

Study Design

This was a prospective observational study of 106 patients with a diagnosis of non-cystic fibrosis bronchiectasis enrolled over a two-year period.

Subjects

Subjects were enrolled from a non-HIV, non-cystic fibrosis bronchiectasis cohort at New York University. All subjects signed informed consent to participate in this study and the research protocol was approved by the New York University and Bellevue Hospital Center (New York, NY) institutional review boards (IRB# S14-01400). The inclusion criteria included: CT imaging abnormalities consistent with bronchiectasis (i.e. mucoid impaction) and symptoms consistent with bronchiectasis (i.e. cough). Exclusion criteria included: participants recently on antibiotics and/or steroids (within the last month) and/or a recent history of smoking (within the last year) as cigarette smoke has been shown to be associated with changes in the upper airway microbiota [1]. At the time of recruitment, clinical information and questionnaires were obtained.

Variables Collected

At recruitment clinical information was collected, including age, sex, ethnicity, BMI, symptoms, smoking history, CT thorax imaging reports, lung function, and sputum cultures. Patients were also asked to fill out questionnaires: St. George's Respiratory Questionnaire (SGRQ), Eating Assessment Tool (EAT-10), Frequency Scale for the symptoms of GERD (FSSG), and the Reflux Symptom Index (RSI) [2-5]

Sample Collection

Oral wash and induced sputum samples were collected from every patient at enrolment. In addition, over the two-year period, serial samples were obtained as per clinical need and sent for culture as well as 16S rRNA gene sequencing. Patients were asked to first rinse their mouth and back of throat with 10ml of sterile water, to provide an oral wash sample. Patients were then placed on a 7% hypertonic saline nebuliser for 10 to 15 minutes. Following this, patients would expectorate as much sputum as they could into a sterile cup, to provide a sputum sample. A portion of this sputum sample was sent to the clinical laboratory for sputum culture and sensitivity, acid-fast bacilli testing and mycobacterium culture. All remaining sputum was transferred to our lab on ice for 16S rRNA gene sequencing.

Bronchoscopy

In all patients who consented to the study we obtained induced sputum (paralleled with oral wash) and we offered participation in the bronchoscopy study. A subset of patients from this cohort underwent bronchoscopy (n=20) in order to evaluate whether the sputum was representative of the lower airway microbiota and to assess the immune profile of the lower airway mucosa. As per our protocol, we asked every patient enrolled in this cohort about their interest in participating in a bronchoscopy arm. A few patients agreed to a research bronchoscopy (n=6) [6]. Other 14 patients had a bronchoscopy done as per clinical indication (in general because of difficulties with obtaining three induced sputum or persistence of clinical suspicious of NTM) and agreed to have bronchoscopic samples obtained for this research. All subjects had a CT scan of the chest done prior to bronchoscopy. In all subjects who underwent bronchoscopy we had a similar topographical sampling approach that

included: oral wash samples, supraglottic samples (sampled using Yankauer), background/equipment samples (sterile saline, Yankauer and Bronchoscope), sputum samples and bronchioloalveolar lavage samples (BAL) from involved and non-involved segments (predefined based on CT scan). From the BAL fluid, cell count and differential were obtained. BAL fluid aliquots were frozen at -80°C.

Bacterial 16S rRNA-encoding genes quantification and sequencing

DNA was then extracted from all samples with an ion exchange column (Qiagen). Total bacterial DNA levels were determined by quantitative PCR (qPCR) as previously described.[6, 7] High-throughput sequencing of bacterial 16S rRNA-encoding gene amplicons encoding the V4 region [8] (150bp read length, paired-end protocol) was performed with MiSeq. The V4 region of the bacterial 16S rRNA gene was amplified in duplicate reactions, using primer set 515F/806R, which nearly universally amplifies bacterial and archaeal 16S rRNA genes [8, 9]. Each unique barcoded amplicon was generated in pairs of 25µl reactions with the following reaction conditions: 11µl Polymerase Chain Reaction (PCR)-grade H2O, 10µl Hot MasterMix (5 Prime Cat# 2200410), 2µl of forward and reversed barcoded primer (5µM) and 2µl template DNA. Reactions were run on a C1000 Touch Thermal Cycler (Bio-Rad) with the following cycling conditions: initial denaturing at 94°C for 3 min followed by 35 cycles of denaturation at 94°C for 45 seconds, annealing at 58°C for 1 minute, and extension at 72 C for 90 seconds, with a final extension of 10 min at 72°C.

To evaluate the Mycobacteriome we re-extracted DNA from additional aliquots and performed a nested PCR approach to enrich for *Mycobacterium* DNA template prior to library preparation for sequencing in order to describe the Mycobacteriome as

previously published [10]. Given the concern of inadequate NTM cell lysis using standard DNA isolation methods, we utilized a recently published optimized cell lysis and DNA isolation method as described in Caverly et al. [11]. Briefly, we added zirconium bead beating step followed by DNA isolation. Then, during library preparation, the first amplification was performed with two *Mycobacterium* specific primers (MycF121 and Myc858R) that targeted the 16S rRNA gene. This approach generated a 737bp amplicon that contained the V4 region. We then proceeded with a second PCR using the bar coded 515F/806R primer set as described above to generate the final amplicon product for sequencing. This "Mycobacteriome" approach was performed in parallel with our previously mentioned 16S rRNA gene sequencing approach. These methods were compared using a mock mixture of bacterial DNA (obtained from *Mycobacteriun fortuitum* and *Streptococcus pneumoniae*) and on subject's samples.

Analysis of 16S rRNA gene sequences

The obtained 16S rRNA gene sequences were analyzed using the QIIME package (version 1.9) for analysis of community sequence data [12]. The operational taxonomic unit (OTU) sequence counts were picked based on Greengenes database (version 13-8) and normalized to obtain the relative abundances of the microbiota in each sample. These relative abundances at 97% OTU similarity and each of the 5 higher taxonomic levels (phylum, class, order, family, genus) were tested for univariate associations with clinical variables. To decrease the number of features, we only focused on major taxa and OTUs, defined as those having relative abundance >1% in at least one sample.

Measurement of in vivo cytokines in BAL fluid and Alveolar Macrophages.

In vivo inflammation was assessed by BAL cell count differential and cytokines. Since analytes in the epithelial lining fluid are diluted with sterile saline during BAL, a concentration step was performed via dialysis against Tris 10 mM pH 7.5, EDTA 1 mM and lyophilization, using albumin as an internal control as previously described[13, 14]. For this, the initial volume of acellular BAL fluid was 5mL. After lyophilization at -80 degrees Celsius sample was re-suspended in 60μL of Phosphate-buffered saline. Inflammatory biomarkers were measured using a Human High Sensitivity T Cell Luminex Panel (Millipore HSTCMAG-28SK). Cytokines included: Fractalkine, GM-CSF, IFNγ, IL-1β, IL-2, IL-4, IL-5, IL-6, IL-7, IL-8, IL-10, IL-12 (p70), IL-13, IL-17A, IL-21, IL-23, ITAC, MIP-1α, MIP-1β, MIP-3α, TNF-α. *Ex vivo* cytokine production was assessed using BAL cell supernatant ($10x10^6$ cells in 1 mL of Roswell Park Memorial Institute medium in a 12 well plate) after 18hrs of culture with media alone or 10ng of LPS. *Ex vivo* cytokine production during toll like receptor 4 (TLR4) stimulation was expressed as fold change in levels of biomarkers comparing media alone with LPS.

Statistical Analysis

Since the distributions of microbiome data are non-normal, and no distribution-specific tests are available, we used non-parametric tests of association. For association with discrete factors, we used either the Mann-Whitney test (in the case of 2 categories) or the Kruskal-Wallis ANOVA (in case of > 2 categories). Wilcoxon signed-rank test were used for paired analysis. We used the ade4 package in R to construct Principal Coordinate Analysis (PCoA) based on weighted UniFrac distances [15, 16]. PCoA is a method of dimensionality reduction that uses the

distance between points and plots the variation of these distances across two axes. Therefore, the closer two points are the more similar they are in their microbial composition. Similar methods of analysis were used to examine differences in cell count and cytokines.

To cluster microbiome communities into exclusive ‘metacommunities’ we used a Dirichlet Multinomial Mixture Model [17]. In this method, for each sample, we impute the component most likely to have generated it, thus separating samples into groups it has the highest probability of belonging to. This allows for variable cluster sizes and a more rigorous means of choosing optimal cluster number. The R package DirichletMultinomial was used for this method of analysis [18].

To evaluate differences between groups of 16S data, we used linear discriminant analysis (LDA) Effect Size (LEfSe) [19]. Features significantly discriminating among groups with LDA score > 2.0 were represented as a cladogram, as produced by LEfSe with default parameters. For tests of association with continuous variables, we used non-parametric Spearman correlation tests and false discovery rate (FDR) was used to control for multiple testing [20]. Co-occurrence between bacterial genera with more than one percent relative abundance in any given sample were assessed using SparCC [21] with 20 iterations and 500 boostrap replicates. Significant correlations were selected ($p < 0.05$, $|\square| > 0.4$, two-sided t-test) and visualized with Cytoscape v3.0.2 [22]. The network layout was selected as edge-weighted spring embedded metrics. Correlation of microbial genera with continuous immune markers was estimated using nonparametric Spearman correlation with a cutoff threshold of $p < 0.05$. All data is publicly available in Sequence Read Archive (SRA) under accession number PRJNA418131.

Supplementary Results

Comparing bacterial load of Sputum and Oral Wash Based on NTM status

There were no significant differences in bacterial load based on NTM status (Median [IQR]= 2,616[1,700-42,036] copies/μl vs. 70,846[7,659–100,617] copies/μl for sputum and 8,949[2,180–20,591] copies/μl vs. 13,406[5,169–46,679] copies/μl for oral wash comparing NTM- vs. NTM+ respectively, p=ns).

Evaluation of the Mycobacteriome

To test sensitivity of this approach we first utilized a mock community of *Mycobacterium fortuitum* and *Streptococcus pneumoniae* mixed at various gradient admixtures (Mycobacterium to Streptococcus ratio ranging from 100:1 to 1:1,000,000, **Supplementary Figure 6**). Standard 16S rRNA gene sequencing approach yielded detectable *Mycobacterium* reads up to a *Mycobacterium* to *Streptococcus* ratio of 1:10 but the nested mycobacteriome approach successfully biased the sequencing to a ratio of 1:10,000.

We further validated this approach using a larger number of sputum and oral wash samples (oral wash= 56 samples [52% NTM+], sputum= 54 samples [54% NTM+]) Using our standard 16S rRNA gene sequencing *Mycobacterium* was not abundant (>1% relative abundance) in either sputum or oral wash samples. This approach yielded *Mycobacterium* reads in only 2/56 (4%) oral wash samples (both NTM-) and 5/54 (9%) sputum samples (all NTM+ samples), which was 17% of NTM+ samples, a smaller proportion than that identified in NTM+ BAL samples. Blast analysis was utilized to characterize the *Mycobacterium* species identified by this method. The five NTM+ sputum samples with *Mycobacterium* reads had 100% similarity with *Mycobacterium avium*. In contrast, the two NTM- oral wash samples with

Mycobacterium reads had 100% similarity with *Mycobacterium aurum* (another species that has not been identified as pathogenic).

Clustering of Bronchoscopic Samples

Using DMM two clear clusters were identified within the bronchoscopy samples (**Supplementary Figure 9**) similar to previous description of the lower airway microbiota [6, 13]. A LEfSe analysis of these clusters showed a clear distinction of taxa, where Cluster 1 samples were enriched with *Alicyclobacillus*, *Acinetobacter* and *Bradyrhizobium* whereas Cluster 2 samples were enriched with oral commensals such as *Prevotella*, *Veillonella* and *Streptococcus*.

References:

1. Erb-Downward JR, Thompson DL, Han MK, Freeman CM, McCloskey L, Schmidt LA, et al. Analysis of the lung microbiome in the "healthy" smoker and in COPD. PLoS One. 2011;6(2):e16384.
2. Jones PW, Quirk FH, Baveystock CM. The St George's Respiratory Questionnaire. Respiratory medicine. 1991;85 Suppl B:25-31; discussion 3-7.
3. Belafsky PC, Mouadeb DA, Rees CJ, Pryor JC, Postma GN, Allen J, et al. Validity and reliability of the Eating Assessment Tool (EAT-10). The Annals of otology, rhinology, and laryngology. 2008;117(12):919-24.
4. Kusano M, Shimoyama Y, Sugimoto S, Kawamura O, Maeda M, Minashi K, et al. Development and evaluation of FSSG: frequency scale for the symptoms of GERD. Journal of gastroenterology. 2004;39(9):888-91.
5. Belafsky PC, Postma GN, Koufman JA. Validity and reliability of the reflux symptom index (RSI). Journal of voice : official journal of the Voice Foundation. 2002;16(2):274-7.
6. Segal LN, Alekseyenko AV, Clemente JC, Kulkarni R, Wu B, Chen H, et al. Enrichment of lung microbiome with supraglottic taxa is associated with increased pulmonary inflammation. Microbiome. 2013;1(1):19.
7. Gao Z, Tseng CH, Pei Z, Blaser MJ. Molecular analysis of human forearm superficial skin bacterial biota. Proc Natl Acad Sci U S A. 2007;104(8):2927-32.
8. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, et al. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. The ISME journal. 2012;6(8):1621-4.
9. Walters WA, Caporaso JG, Lauber CL, Berg-Lyons D, Fierer N, Knight R. PrimerProspector: de novo design and taxonomic analysis of barcoded polymerase chain reaction primers. Bioinformatics. 2011;27(8):1159-61.
10. Macovei L, McCafferty J, Chen T, Teles F, Hasturk H, Paster BJ, et al. The hidden 'mycobacteriome' of the human healthy oral cavity and upper respiratory tract. J Oral Microbiol. 2015;7:26094.
11. Caverly LJ, Carmody LA, Haig SJ, Kotlarz N, Kalikin LM, Raskin L, et al. Culture-Independent Identification of Nontuberculous Mycobacteria in Cystic Fibrosis Respiratory Samples. PLoS One. 2016;11(4):e0153876.
12. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. Nat Methods. 2010;7(5):335-6.
13. Segal LN, Clemente JC, Tsay JC, Koralov SB, Keller BC, Wu BG, et al. Enrichment of the lung microbiome with oral taxa is associated with lung inflammation of a Th17 phenotype. Nature microbiology. 2016;1:16031.
14. Berger KI, Pradhan DR, Goldring RM, Oppenheimer BW, Rom WN, Segal LN. Distal airway dysfunction identifies pulmonary inflammation in asymptomatic smokers. ERJ open research. 2016;2(4).
15. Dray SaD, A.B. The ade4 package: implementing the duality diagram for ecologists. Journal of Statistical Software. 2007;22(4):1-20.
16. Lozupone C, Lladser ME, Knights D, Stombaugh J, Knight R. UniFrac: an effective distance metric for microbial community comparison. Isme J. 2011;5(2):169-72.
17. Holmes I, Harris K, Quince C. Dirichlet multinomial mixtures: generative models for microbial metagenomics. PLoS One. 2012;7(2):e30126.
18. Morgan M. DirichletMultinomial: Dirichlet-Multinomial Mixture Model Machine Learning for Microbiome
Data. R package version 1.20.0 ed2017.

19. Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS, et al. Metagenomic biomarker discovery and explanation. Genome biology. 2011;12(6):R60.
20. Reiner A, Yekutieli D, Benjamini Y. Identifying differentially expressed genes using false discovery rate controlling procedures. Bioinformatics (Oxford, England). 2003;19(3):368-75.
21. Friedman J, Alm EJ. Inferring correlation networks from genomic survey data. PLoS Comput Biol. 2012;8(9):e1002687.
22. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome research. 2003;13(11):2498-504.

Supplementary Figures Legends:

Supplementary Figure 1: Comparison of microbial diversity in oral wash and sputum samples. A. Alpha diversity based on Shannon index was higher in oral wash as compared to sputum. **B**. PCoA based on weighted UniFrac distance demonstrates significant differences between oral wash and sputum samples (PERMANOVA $p<0.001$). **C**. Comparison of degree of similarity between oral wash and sputum samples within the same subject vs. between different subjects. **D**. LEfSe analysis was utilized to identify taxa differentially enriched in oral wash and sputum samples. Multiple significant taxonomic differences were observed at different phylogenetic levels as represented in the cladogram, left panel. Bar plots in the right represents Linear Discriminant Analysis (LDA) effect size (left) and differences in relative abundance of differentially enriched taxa at a genus level (LDA>2).

Supplementary Figure 2: Differences in diversity between all oral wash and sputum samples for the NTM+ and NTM- groups. For α diversity Shannon Diversity Index was used, for β diversity weighted UniFrac was used. **A**. For oral wash samples there were no significant differences in a diversity (Mann Whitney p=ns) but significant differences in β diversity were noted (PERMANOVA $p=0.043$). **B**. For sputum samples there were significant differences in α diversity, and a non-significant difference in β diversity.

Supplementary Figure 3: Heat Map of 16S sequencing of all samples obtained during bronchoscopy. Unsupervised hierarchical clustering of most abundant taxa

(relative abundance ≥ 1% in any sample) identified in Background, Nasal Swab, Oral Wash, Sputum, Supraglottic and Bronchoalveolar Lavage (BAL).

Supplementary Figure 4: Comparison of bacterial load in bronchoscopic samples. qPCR for 16S rRNA gene was used to compare bacterial load of background samples, lower airway samples (BAL), upper airway samples (oral wash and supraglottic) and sputum.

Supplementary Figure 5: Differences in diversity between NTM+ and NTM- in lower airway microbiota. **A**. There were no significant differences in α diversity between NTM+ and NTM- groups. **B**. β diversity based on weighted UniFrac showed non-significant differences between NTM+ and NTM- groups.

Supplementary Figure 6: Comparison between an unbiased 16S rRNA and a biased mycobacteriome approach using mock bacterial DNA mixture. *Mycobacterium fortuitum* and *Streptococcus pneumoniae* isolates were used to extract DNA. DNA template was sequenced for each isolate and for a series of mixture ratios of *Mycobacterium*:*Streptococcus*. Mixing ratios started at 100:1 ratio (*Mycobacterium*:*Streptococcus* respectively) to a 1:10,000,000 ratio. Using a standard 16S rRNA sequencing approach, *Mycobacterium fortuitum* was identified until a ratio of 1:10, after which only *Streptococcus* could be identified. Using a nested mycobacterium bias approach, *Mycobacterium* was identified (with a relative abundance close to 100%) even in much lower dilution of its template, up to a ratio of 1:10,000.

Supplementary Figure 7: Comparison between sequence data obtained using an unbiased 16S rRNA approach and a biased mycobacteriome approach for equipment background samples. A. Shows the differences in β diversity (based on weighted UniFrac distance) between samples processed with unbiased 16S vs. biased mycobacteriome approach. **B.** LEFSE analysis showed enrichment of taxa identified through 16S rRNA and those identified through the biased mycobacteriome approach. **C.** Bar charts show relative abundance of OTUs annotated to Mycobacterium and their annotation based on BLAST on the two datasets.

Supplementary Figure 8: Comparison between sequence data obtained using an unbiased 16S rRNA approach and a biased mycobacteriome approach for oral wash and sputum samples. A. In oral wash two samples were enriched with *Mycobacterium aurum*, a non-pathogenic *Mycobacterium* strain. Both of these samples were NTM negative on culture. **B.** In sputum 5 samples were enriched with *Mycobacterium avium*. All 5 samples were NTM positive on culture.

Supplementary Figure 9: Clustering of Bronchoscopy Samples by a Dirichlet Multinomial Model. A. Model fitness was plotted against number of clusters. Lower model fitness indicates best fitness. Two clusters were identified as having the best model fit. **B.** LEfSe analysis identified taxonomic differences in lower airway samples (BAL) between cluster 1 and cluster 2 and represented in Cladogram. **C.** LDA and differences in relative abundance of taxa at genera level found differentially enriched in BAL samples between cluster 1 and cluster 2.

Supplementary Figure 10: Associations between taxa and inflammatory biomarkers in the lower airways for NTM- samples. Correlations seen with taxa identified as oral commensals and Cluster 2 (from DMM) are not seen with NTM- samples (in comparison to Figure 5)

Supplementary Table 1: Oral and induced sputum samples.

	Oral Wash	Sputum	Total
Baseline	106	106	212
2-4 Months	15	13	28
5-7 Months	8	9	17
8-10 Months	4	3	7
11-13 Months	5	3	8
14-16 Months	2	4	6
17-19 Months	6	4	10
20-22 Months	3	3	6
23-24 Months	1	2	3
Total	150	147	297

Supplementary Table 2: Demographic and pulmonary function data of the bronchoscopy cohort

VARIABLES	All Patients	NTM Status (-)	NTM Status (+)	p value
N	20	12	8	
Age	63.8 (12.9)	60.3 (14.0)	69.0 (8.8)	0.46
No. Female (%)	19 (95)	11 (92)	8 (100)	0.33
No. Caucasian (%)	15 (75)	8 (42)	7 (88)	0.56
BMI	22.6 (4.8)	23.7 (5.4)	21.2 (3.6)	0.31
Packs Per Day	1.2 (0.7)	1.1 (0.7)	1.2 (1.1)	0.93
No. Years Smoking	22.1 (12.2)	21.2 (14.3)	25.0 (0.0)	0.73
Lung Physiology*				
FVC (% predicted)	95.5 (10.1)	97.7 (12.3)	93.7 (8.2)	0.50
FEV_1 (% predicted)	92.3 (10.7)	91.7 (13.5)	92.9 (8.7)	0.85
FEV_1/FVC (%)	76.2 (7.8)	74.8 (8.7)	77.4 (7.3)	0.57
RV (% predicted)	116.0 (25.4)	106.0 (22.4)	131.0 (24.5)	0.13
TLC (% predicted)	105.9 (9.0)	101.5 (7.5)	111.2 (8.2)	0.07
DLCO (% predicted)	98.1 (23.5)	105.6 (22.8)	88.8 (24.0)	0.32

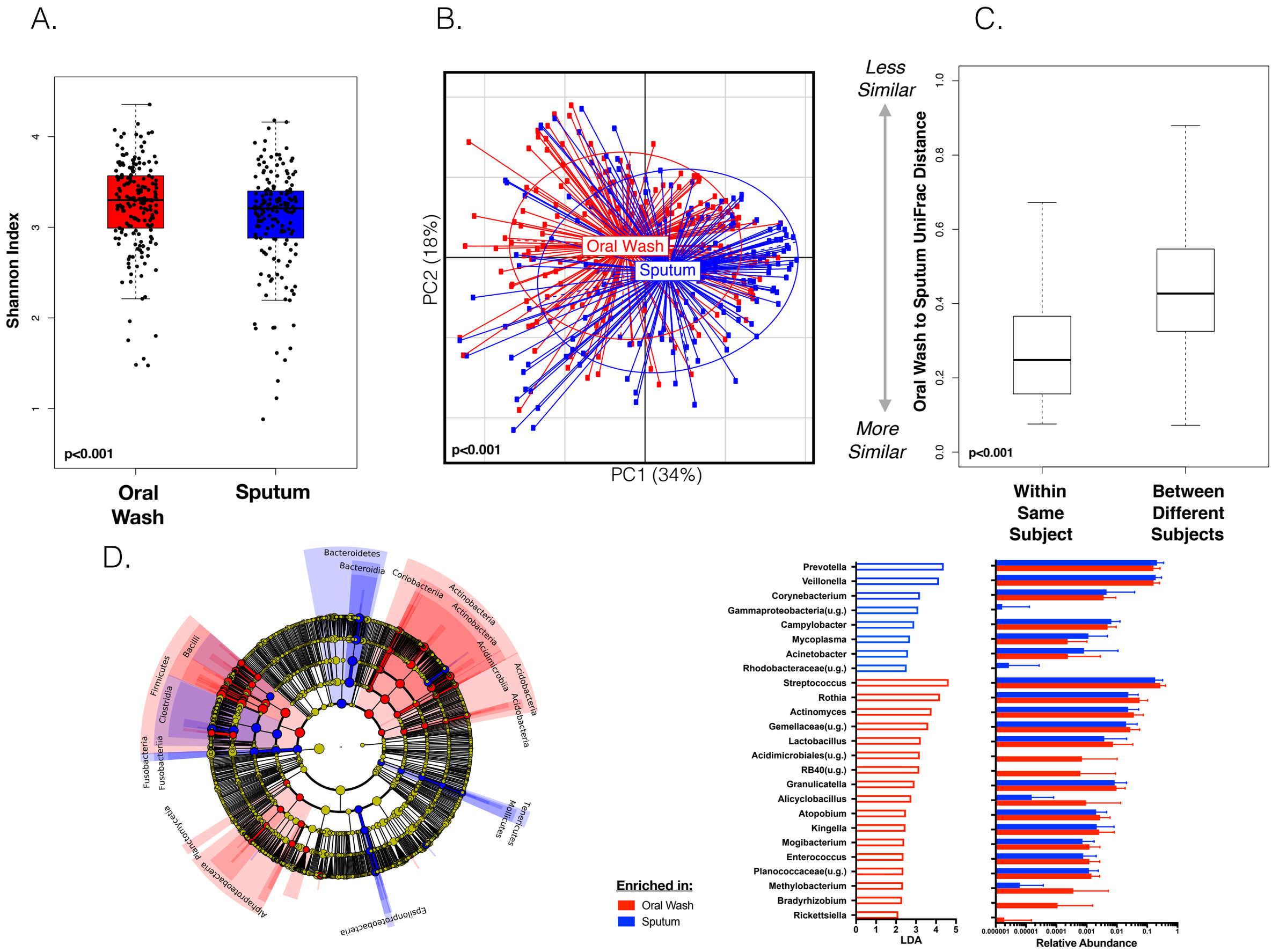
* Available in 13 subjects

Supplementary Table 3: *Ex Vivo* (TLR4 stimulated) cytokine production of BAL cells in the 20 patients from the bronchoscopy cohort

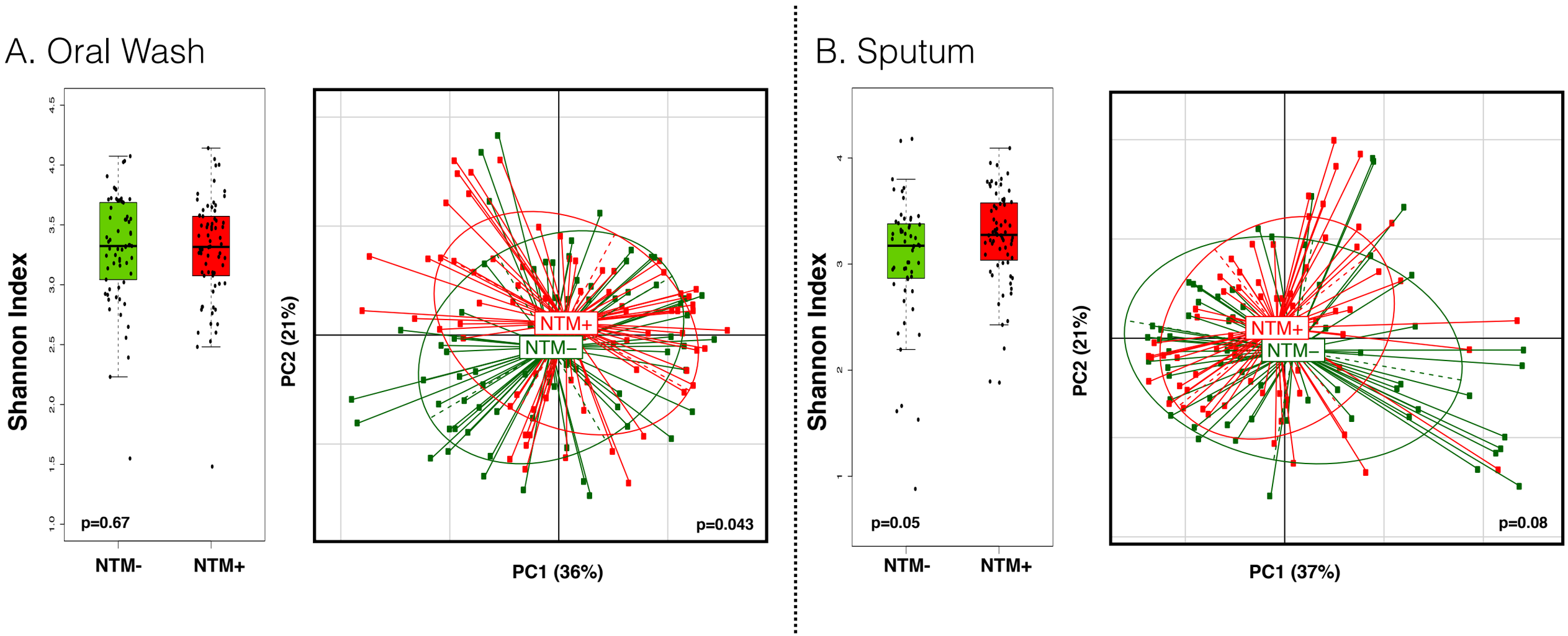
	NTM- (n=12)			NTM+ (n=8)		
	Involved	Non-Involved	P value	Involved	Non-Involved	P value
Ex Vivo Cytokine Production (fold change)						
GM-CSF	12.922 [0-47.6]	89.799 [47.6-123.4]**	ns	8.284 [4.3-9.3]	33.105 [22.7-71.2]**	*0.05*
IFNγ	0.966 [0-1.8]	4.647 [3.3-15.2]**	ns	0.06 [0-0.5]	1.158 [1.1-1.3]**	*0.05*
MIP1β	4.192 [-0.1-4.7]	16.468 [12-74.6]	ns	5.231 [4.1-7.7]	13.354 [11.4-28.7]	ns
IL-23	0.145 [0.1-0.8]	1.948 [1.8-3.8]	ns	1.344 [1.2-1.4]	2.018 [1.7-2.0]	ns
MIP1α	3.134 [0-3.4]	0.2 [0.1-2.2]	ns	6.885 [3.5-7.2]	22.03 [11.4-23.0]	ns
IL-8	0 [0-0.8]	-0.052 [-0.1-0]	ns	16.841 [12.2-81.2]	122.885 [60.9-292.8]	ns
IL-5	0.147 [0-2.8]	4.2 [2.4-4.9]	ns	4.531 [2.3-6.4]	10.982 [9.5-12.7]	ns
MIP3α	3.017 [0.1-12.2]	24.74 [12.5-36.9]	ns	8.414 [6.3-21.3]	41.411 [29.1-51.9]	ns
IL-4	0.29 [0-0.4]	0.832 [0.6-1.1]	ns	0.908 [0.8-1]	0.849 [0.8-1.4]	ns
IL-6	32.907 [0-34.5]	1165.817 [589.7-1380.4]	ns	890.798 [458.9-1171.9]	866.575 [602.9-1109.1]	ns
IL-21	3.141 [-0.2-4.1]	3.719 [1.9-4.3]	ns	2.156 [1.7-2.9]	3.885 [3.6-5.6]	ns
TNFα	9.132 [0.2-10.7]	16.773 [8.4-26]	ns	15.494 [10.9-25.2]	36.424 [27.4-50.8]	ns
Fractalkine	1.982 [-0.1-2.5]	2.127 [1.2-2.8]	ns	1.364 [1.1-2.2]	2.024 [2-2.5]	ns
IL-1β	14.293 [0.1-19.1]	106.462 [54.4-252.6]	ns	11.535 [8.7-15.3]	34.318 [18.9-74.8]	ns
IL-10	9.747 [0.1-18.9]	58.508 [30.7-148.5]	ns	7.229 [5-38.4]	27.244 [16.7-41.5]	ns
IL-2	0.559 [0.5-0.7]	0.376 [0.3-0.5]	ns	-0.036 [-0.2-1.1]	0.031 [0--0.5]	ns
IL-7	3.258 [-0.1-5]	4.475 [2.3-5.6]	ns	4.231 [2.9-5.5]	4.221 [3.7-5.3]	ns
IL-13	1.087 [0.2-1.8]	1.225 [0.8-2]	ns	1.392 [1.1-1.6]	1.921 [1.7-2.2]	ns
IL-12 p70	2.765 [1.9-3.3]	6.118 [3-35.6]	ns	6.462 [4.8-9.3]	7.283 [5.7-7.6]	ns
IL-17A	0.526 [0-0.9]	0.557 [0.5-2.2]	ns	0.723 [0.6-1.3]	1.213 [0.8-1.3]	ns
ITAC	0.224 [0.1-0.4]	0.111 [0.1-1.5]	ns	-0.14 [-0.2-0.2]	0.756 [0.5-0.9]	ns

Data represented as Median [IQR]. p-value based on Mann Whitney. *Comparing involved sites by NTM status. **Comparing non-involved sites by NTM status

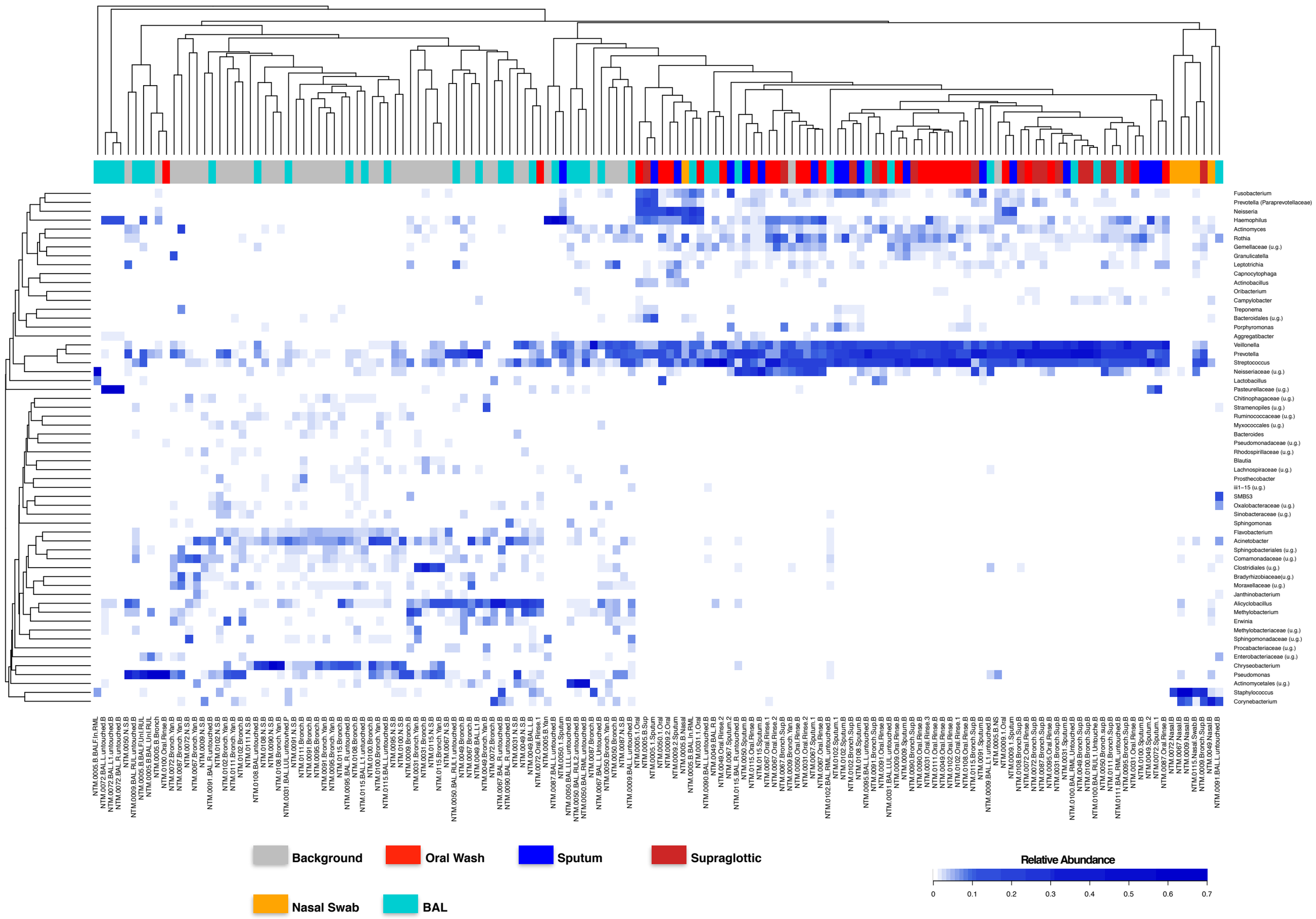
Supplementary Figure 1



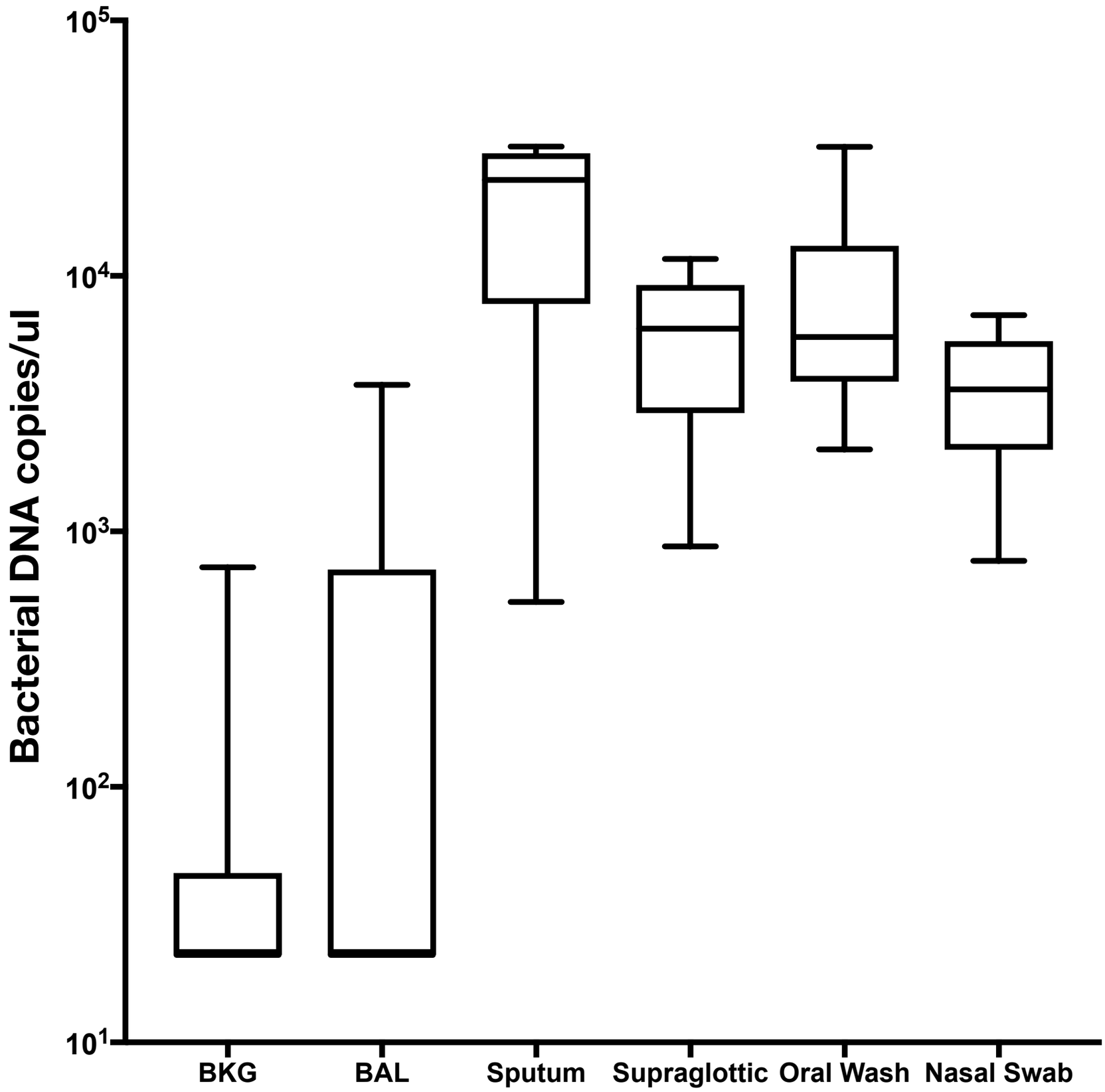
Supplementary Figure 2



Supplementary Figure 3

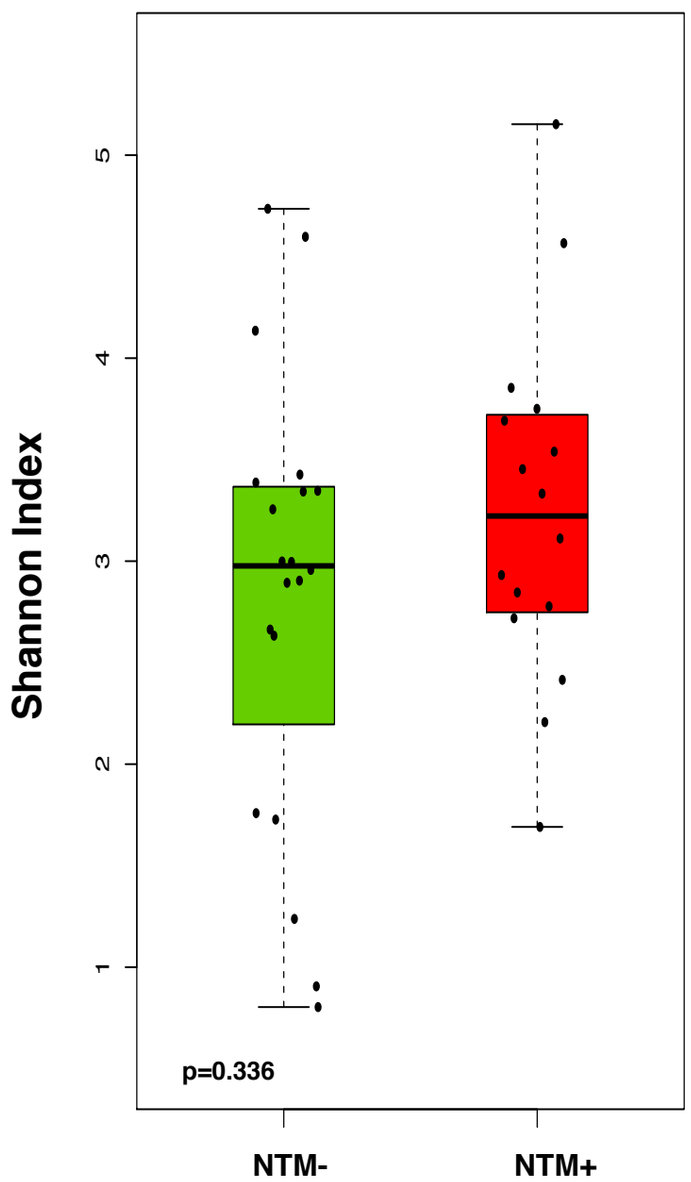


Supplementary Figure 4

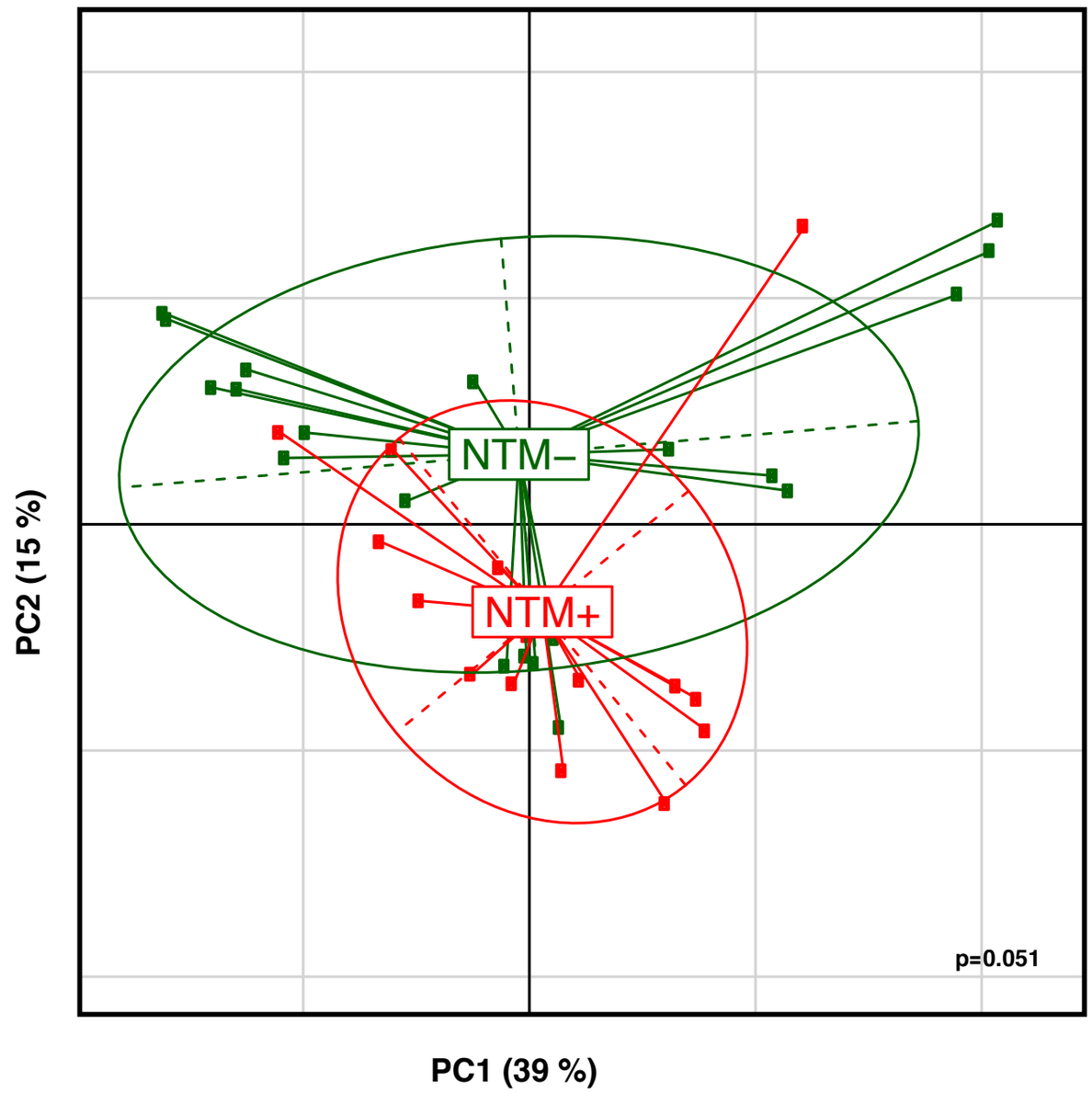


Supplementary Figure 5

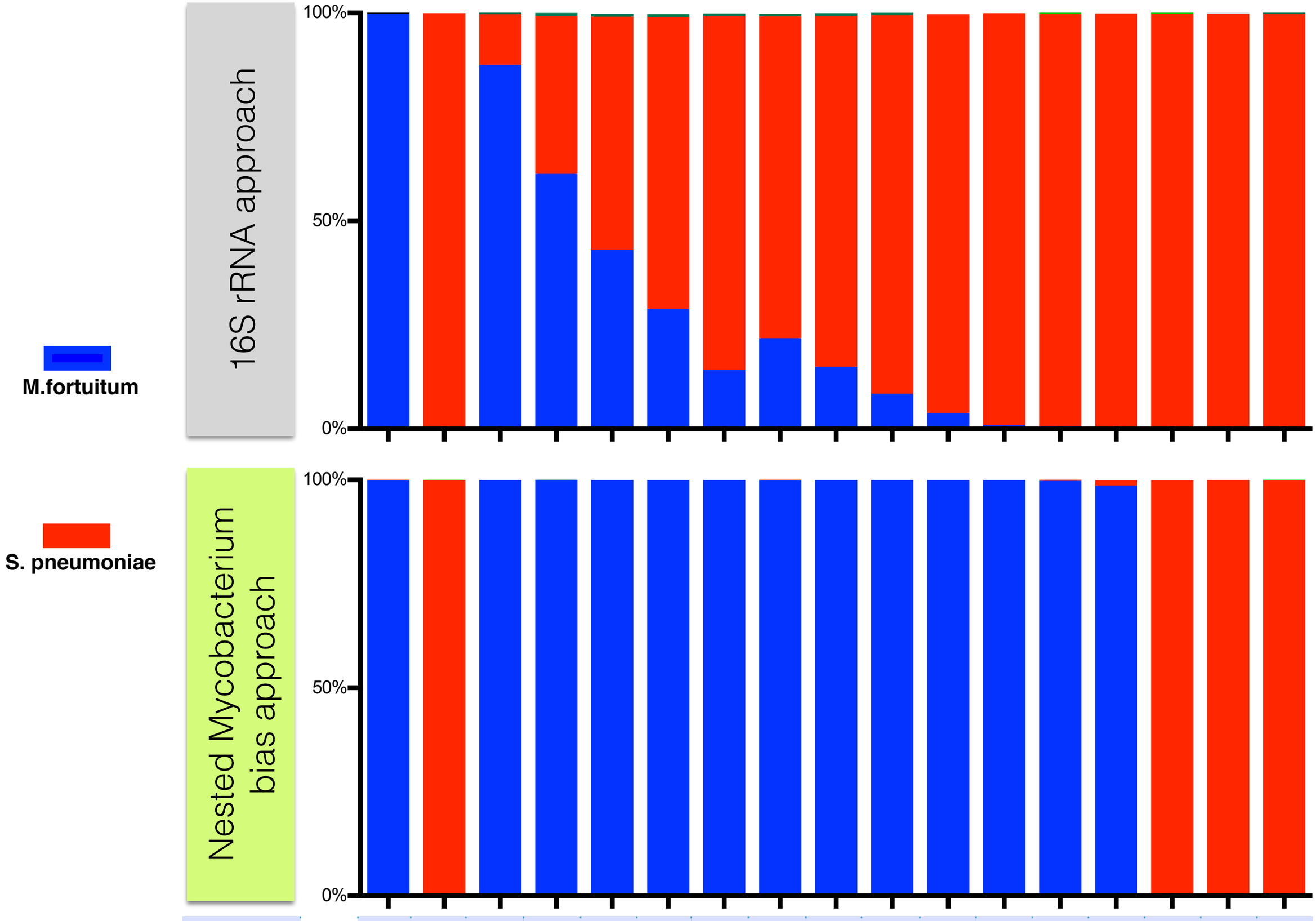
A.



B.



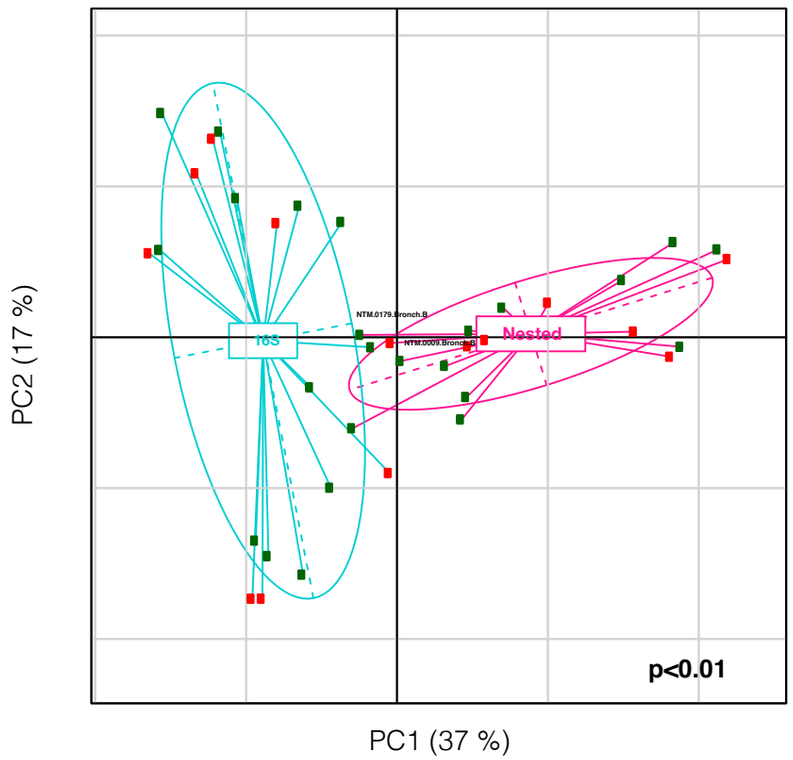
Supplementary Figure 6



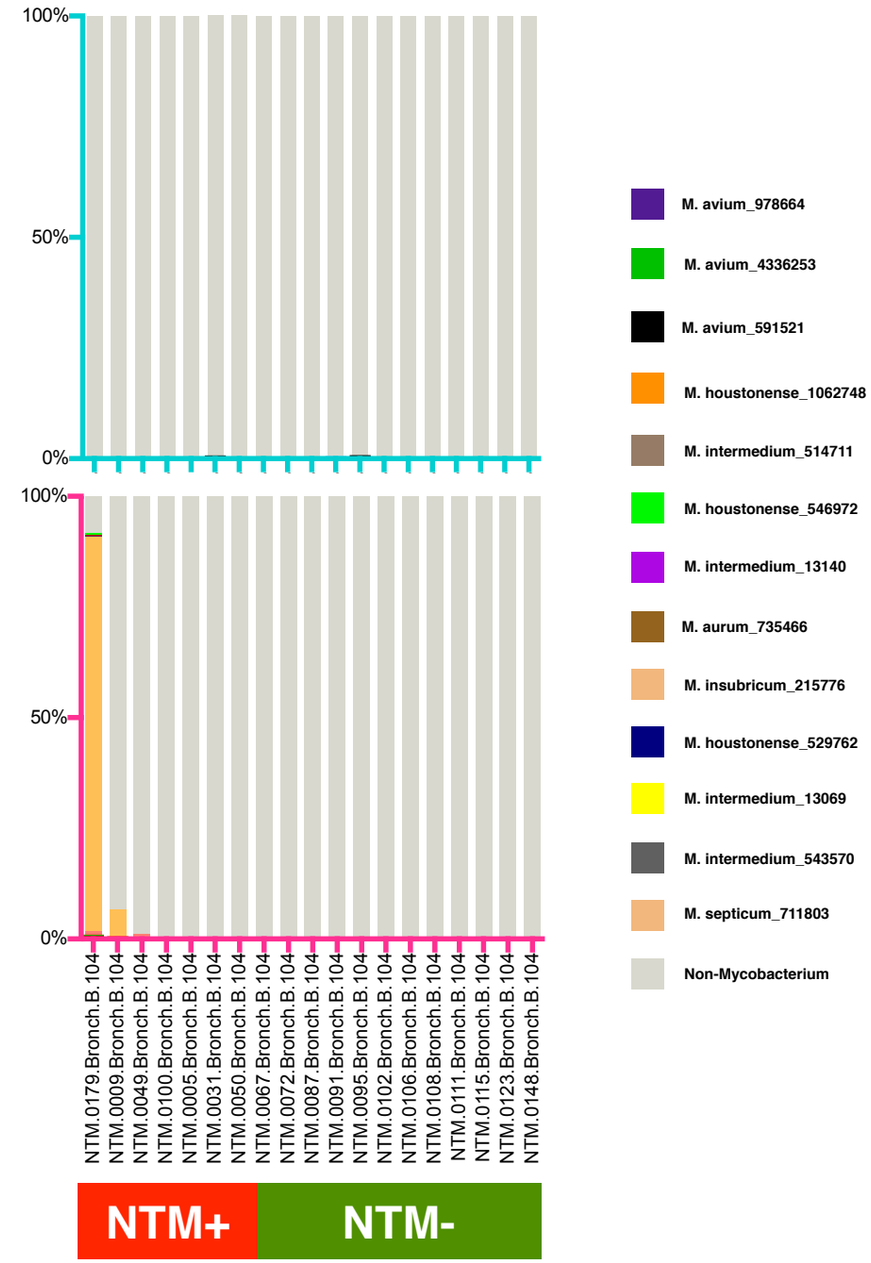
M. fortuitum	100%		10^2	10	4	2	1	1	1	1	1	1	1	1	1	1	1
S. pneumoniae		100%	1	1	1	1	1	1	2	4	10	10^2	10^3	10^4	10^5	10^6	10^7

Supplementary Figure 7

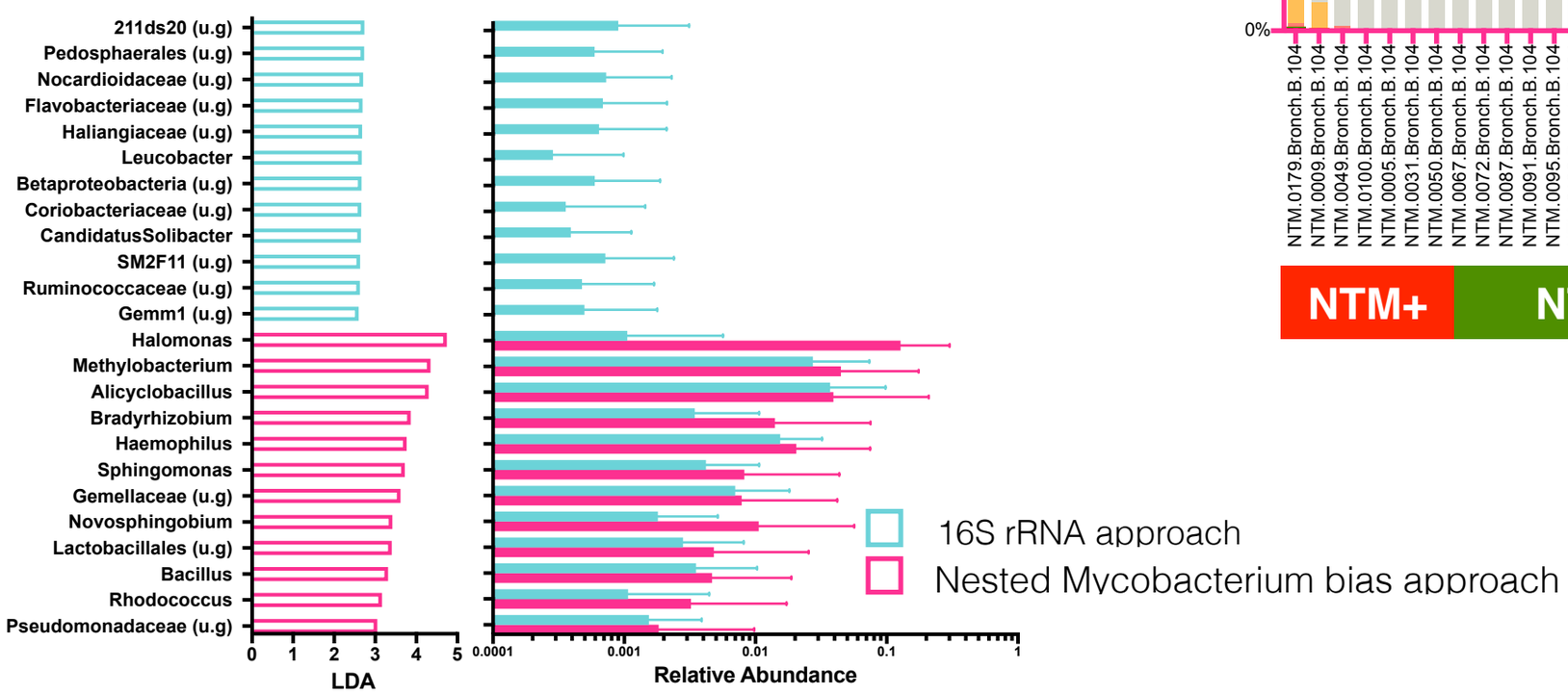
A.



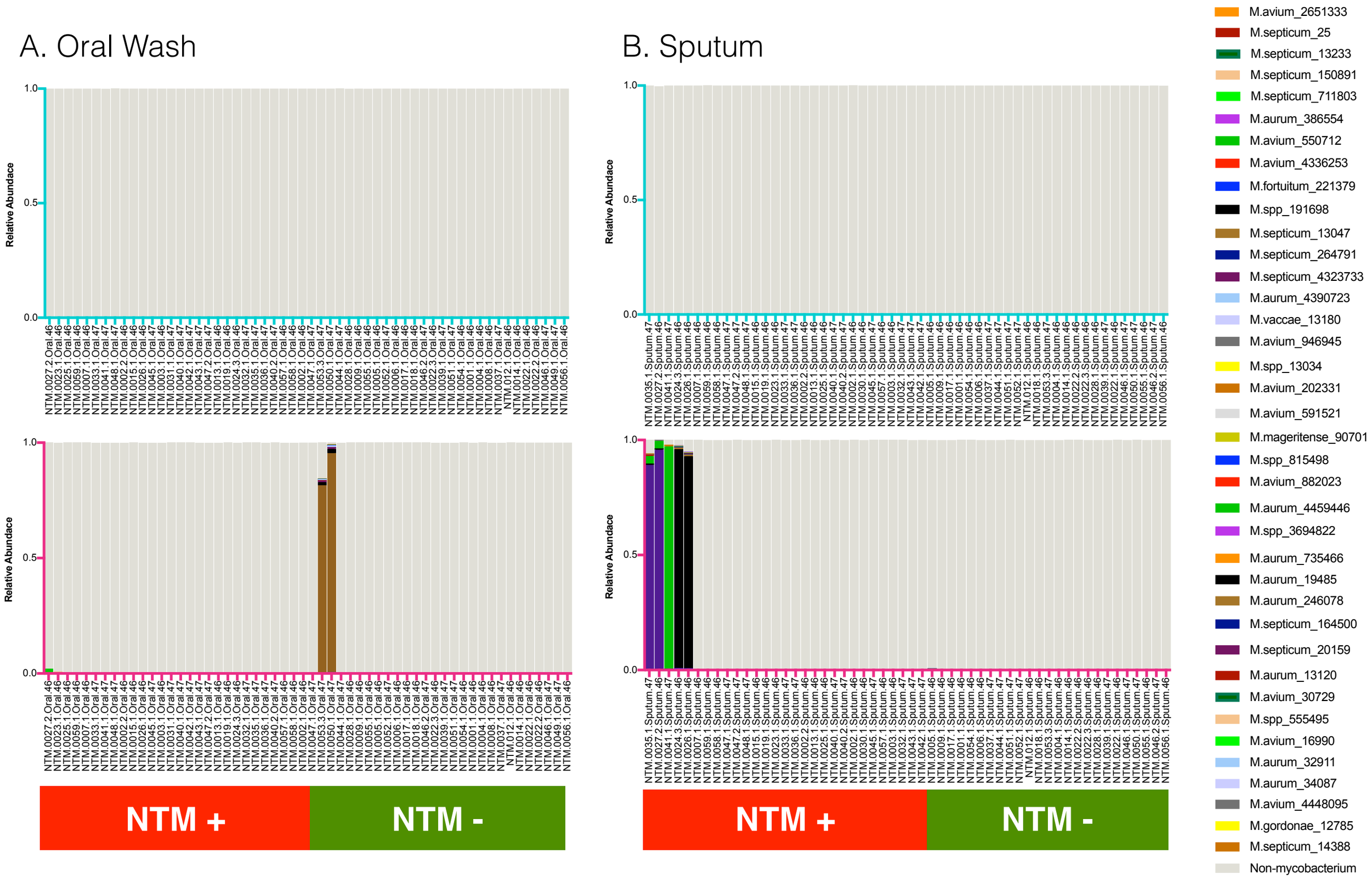
C.



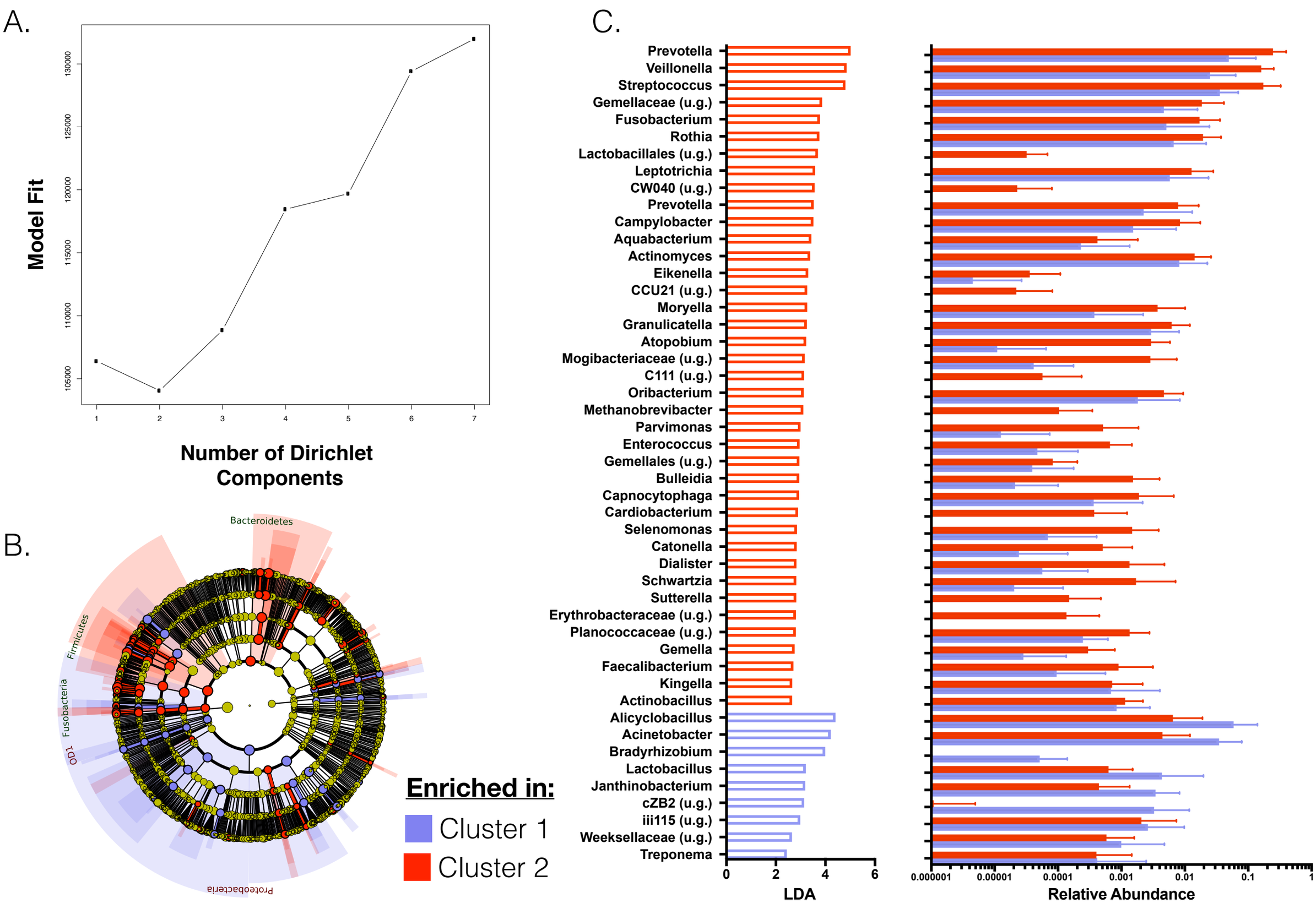
B.



Supplementary Figure 8



Supplementary Figure 9



Supplementary Figure 10