

1-2021

## Impaired Host Response and the Presence of *Acinetobacter baumannii* in the Serum Microbiome of Type-II Diabetic Patients

Dasith Perera

Sarah E. Kleinstein

Benjamin Hanson

Hatice Hasturk

Ryan Eveloff

*See next page for additional authors*

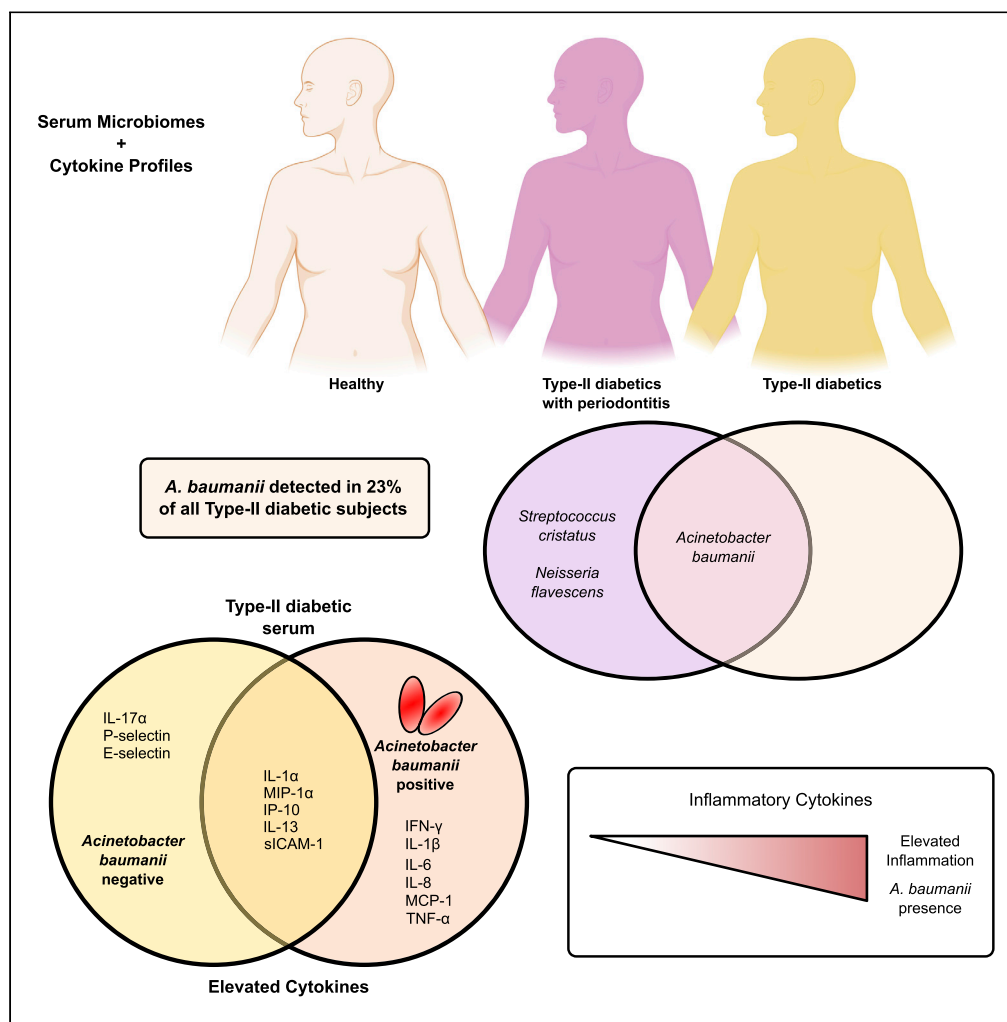
---

**Authors**

Dasith Perera, Sarah E. Kleinstein, Benjamin Hanson, Hatice Hasturk, Ryan Eveloff, Marcelo Freire, and Matthew M. Ramsey

---

## Article

Impaired host response and the presence of *Acinetobacter baumannii* in the serum microbiome of type-II diabetic patients

Dasith Perera,  
Sarah E.  
Kleinstejn,  
Benjamin Hanson,  
Hatice Hasturk,  
Ryan Eveloff,  
Marcelo Freire,  
Matthew Ramsey

mfreire@jcvl.org (M.F.)  
mramsey@uri.edu (M.R.)

**Highlights**

*Acinetobacter baumannii* was found in the serum of 23% type-II diabetic patients

Minor but significant serum microbiome changes between healthy vs type-II diabetic patients

Proinflammatory cytokines were elevated in *A. baumannii*-positive serum samples

*A. baumannii* + serum cytokine profiles were consistent with chronic inflammation

## Article

Impaired host response and the presence of *Acinetobacter baumannii* in the serum microbiome of type-II diabetic patients

Dasith Perera,<sup>1</sup> Sarah E. Kleinstein,<sup>2</sup> Benjamin Hanson,<sup>1</sup> Hatice Hasturk,<sup>3</sup> Ryan Eveloff,<sup>2</sup> Marcelo Freire,<sup>2,4,\*</sup> and Matthew Ramsey<sup>1,5,\*</sup>

## Summary

**Type II diabetes (T2D) affects over 10% of the US population and is a growing disease worldwide that manifests with numerous comorbidities and defects in inflammation. This dysbiotic host response allows for infection of the host by numerous microorganisms. In the course of T2D disease, individuals can develop chronic infections including foot ulcers and periodontitis, which lead to further complications and opportunistic infections in multiple body sites. In this study, we investigated the serum of healthy subjects and patients with T2D with (T2DP) or without periodontitis for both microbiome signatures in addition to cytokine profiles. Surprisingly, we detected the presence of *Acinetobacter baumannii* in the serum of 23% individuals with T2D/T2DP tested. In T2DP, IL-1 $\beta$ , TNF- $\alpha$ , MCP-1, IL-6, IL-8, and IFN- $\gamma$  were significantly elevated in ABC-positive subjects. As an emerging pathogen, *A. baumannii* infection represents a risk for impaired inflammation and the development of comorbidities in subjects with T2D.**

## Introduction

Type-II diabetes (T2D) is a chronic inflammatory disease affecting 10.5% of the US population, with rising prevalence both in the United States and worldwide (Whiting et al., 2011; National Diabetes Statistics Report | Data & Statistics | Diabetes | CDC, 2020). Individuals with T2D typically have numerous comorbidities and are at a greater risk of both opportunistic and nosocomial infections, which can result in significant morbidity, including amputations, and lost quality of life (QOL) measures (Vardakas et al., 2007; Erben et al., 2013; McDonald et al., 2014; Ferlita et al., 2019; Kim et al., 2019). Thus, care must be taken to prevent and monitor the infection status of these individuals, as the development of diabetic mucosal and skin lesions across the body can impact the host immune and diabetic responses. Key infection prevention includes minimizing portals of entry and colonization by numerous pathogens. One example of such a portal of entry is the emergence of diabetic foot ulcers, which can develop diabetic foot infections (DFIs). Roughly 15% of adults with T2D develop foot ulcers and ~14% of these subjects undergo amputation, with the remainder needing extensive debridement, negative pressure treatments, and other therapies including antibiotics, which result in prolonged hospitalization time, lengthy healing duration, and increased treatment costs (Reiber et al., 1998; Driver et al., 2010).

In addition to DFIs, subjects with T2D are ~3 times more likely to develop periodontitis (Emrich et al., 1991; Wu et al., 2020). Periodontitis is an inflammatory oral disease accompanied by a polymicrobial biofilm infection that can lead to tissue damage, chronic lesions, and tooth/bone loss in the oral cavity (Irfan et al., 2001; Genco et al., 2020; Mann et al., 2020). During periodontal infection, the healthy oral microbiome, especially in the subgingival space (along teeth below the gumline, including periodontal ligament, epithelium, alveolar bone, and connective tissues), transitions to one enriched with several opportunistic pathogens that exacerbate inflammation and host tissue damage both locally and systemically (Socransky et al., 1998; Griffen et al., 2012; Hajishengallis and Lamont, 2012; Roberts and Darveau, 2015; Nowicki et al., 2018; Curtis et al., 2020). Although protective, acute inflammation is a helpful immune response, it is characterized by rapid activation of inflammation and subsequent return to homeostasis (resolution). Unresolved inflammation that fails to control the trigger leads to chronic lesions and is a hallmark of both periodontitis and T2D.

<sup>1</sup>The University of Rhode Island, Department of Cell and Molecular Biology, Kingston, RI 02881, USA

<sup>2</sup>J. Craig Venter Institute, La Jolla, CA 92037, USA

<sup>3</sup>The Forsyth Institute, Cambridge, MA 02142, USA

<sup>4</sup>University of California San Diego, School of Medicine, La Jolla, CA 92093, USA

<sup>5</sup>Lead contact

\*Correspondence: mfreire@jvri.org (M.F.), mramsey@uri.edu (M.R.)

<https://doi.org/10.1016/j.isci.2020.101941>



More acute bacteremia can manifest in diseases such as endocarditis or other distal infections by oral microbes (Lockhart and Durack, 1999; Parahitiyawa et al., 2009; Carrizales-Sepúlveda et al., 2018). Various systemic diseases are in fact influenced by microbial metabolism and host interactions. “How” and “why” immune system imbalance fails to control microbial pathogenic transition remains an active area of investigation. Increasing evidence implicates periodontal diseases, especially periodontitis in adults, as a potential risk factor for increased morbidity and mortality from systemic conditions including diabetes, cardiovascular diseases, adverse pregnancy outcomes, and others (Genco et al., 2020; Kleinsteinst et al., 2020b). This is due to two plausible mechanisms: a direct pathway leading to migration of the bacteria to distal organs and an indirect pathway leading to production of microbial metabolites and/or activating inflammatory mediators that activate the immune system locally and systemically. This is likely made possible as periodontitis presents as a chronic inflammatory condition of the tooth-supporting structures occurring with oral microbiome dysbiosis and continuous inflammatory burden. Notably, oral microbes have also been detected (Dowd et al., 2008; Gardner et al., 2013) and isolated from DFIs (Bowler and Davies, 1999; Citron et al., 2007). Thus, we hypothesized that patients with T2D might be at a greater risk of extraoral infection and may exhibit changes in microbiota that transiently occupy their bloodstream. We further hypothesized that these changes would be even more apparent in patients with T2D with periodontitis (T2DP) and that these patients may carry potential DFI-colonizing oral species in their bloodstream. Our initial investigations toward these hypotheses utilized samples derived from previous work on resolution of inflammation in T2D (Freire et al., 2017) and are reported here, with unexpected results.

To investigate the possibility of oral species presence in the bloodstream, we compared the serum microbiomes of subjects with T2D and T2DP versus a healthy cohort. Unexpectedly, we detected *Acinetobacter baumannii* in 23% of the subjects with T2D/P in our study. *A. baumannii* is an emerging infectious pathogen (Villar et al., 2014) notorious for its evolving antibiotic resistances (Gootz and Marra, 2008; Vázquez-López et al., 2020) and nosocomial (Dijkshoorn et al., 2007; Ayoub Moubareck and Hammoudi Halat, 2020) and bloodstream infections (Peleg et al., 2008) and is part of a cluster of *Acinetobacter* species with similar clinical manifestations known as the AB or ABC complex (Gerner-Smidt and Tjernberg, 1993), which are highly similar via ribotyping. ABC complex organisms present several particular risks to subjects with T2D, specifically higher mortality rates for those with bacteremia (Leão et al., 2016) and for those with higher blood glucose concentrations (Leung and Liu, 2019). Patients infected with *A. baumannii* after burns were 9.8 times more likely to develop glucose intolerance, and ~3 times more patients with T2D with burns developed *A. baumannii* infections (27%) than patients without T2D (8.5%) (Furniss et al., 2005). Additionally, *A. baumannii* has also been observed in DFIs, particularly alongside other multidrug-resistant pathogens, presenting a grave concern for individuals with T2D (Castellanos et al., 2019; Henig et al., 2020). In addition to DFIs, ABC complex organisms can also spread through the body and form infections in numerous organs (Peleg et al., 2008; Al-Anazi and Al-Jasser, 2014). *A. baumannii* in a mouse pneumonia model led to increased proinflammatory cytokines, reduced neutrophil infiltration into the lung, and increased extrapulmonary dissemination (Qiu et al., 2009). In T2D, neutrophils present failure in early chemotaxis, but a hyper-inflammatory feedback loop emerges to compensate the initial myeloid cell failure (Kleinsteinst et al., 2020a).

In this work, we first investigated whether or not we could detect signatures of oral bacteria; our findings indicated minor, yet significant, microbiome compositional changes between healthy cohort and patients with T2D and T2DP. We discovered that nearly one-fourth of the subjects with T2DP in our study contained sequence reads identical to that of *A. baumannii*. The specific presence of *A. baumannii* was confirmed by further clone library and 16S rDNA Sanger sequencing. In addition, we assayed the host response by cytokine profiles of sera. When stratifying *A. baumannii*-positive T2D versus *A. baumannii*-negative T2D samples we observed unique cytokine profiles, suggesting a specific microbial impact on systemic inflammation. Although our methodology cannot indicate if *A. baumannii* was transiently present or a subclinical colonizer of our T2DP cohort, we propose that these data reveal a much greater risk of currently “uninfected” individuals with T2D for *A. baumannii* exposure and infection than previously understood and should justify more intensive and thorough investigation into AB complex colonization of patients with T2D.

## Results

### Patient demographics

Serum samples were generated from the population described in Table 1. A total of 81 subjects were investigated in this study, 57 with T2D, of which 29 had periodontitis (T2DP) as defined by the American

**Table 1. Study subject demographics**

	Healthy (N = 24)	Type II diabetes (N = 57)	Total (N = 81)	p value
Age (mean in years $\pm$ SD)	46 $\pm$ 10.72	55 $\pm$ 10.78	52.78 $\pm$ 11.40	<0.001
<b>Gender (no., %)</b>				
Male	13 (54%)	29 (51%)	42 (51.85%)	
Female	11 (46%)	28 (49%)	39 (48.15)	0.45
<b>Ethnicity</b>				
Caucasian	15 (63%)	25 (43.86%)	40 (49.38%)	
Hispanic	2 (8%)	7 (12.28%)	9 (11.11%)	
African American	6 (25%)	23 (40.35%)	29 (35.80%)	
Asian	0	2 (3.51%)	2 (2.47%)	
Other	1 (4%)	0	1 (1.23%)	0.16
<b>Smoking status (no., %)</b>				
Smokers	1 (4%)	5 (8.77%)	6 (7.41%)	
Former smokers	0	21 (36.84%)	21 (25.93%)	
Never smokers	23 (96%)	31 (54.39%)	54 (66.67%)	0.02
BMI (kg/m <sup>2</sup> ; mean $\pm$ SD)	29.34 $\pm$ 2.60	32.93 $\pm$ 6.02	31.87 $\pm$ 5.48	<0.01
Blood cholesterol (mg/dL; mean $\pm$ SD)	148.89 $\pm$ 36.71	211.73 $\pm$ 78.69	194.01 $\pm$ 74.84	<0.001
HbA1c (%; mean $\pm$ SD)	–	7.48 $\pm$ 1.61	–	–
Blood Glucose (mg/dL; mean $\pm$ SD)	99.25 $\pm$ 19.29	190.84 $\pm$ 60.58	163.70 $\pm$ 66.69	<0.0001
<b>Periodontal condition (no., %)</b>				
Healthy	24 (100%)	7 (12.28%)	31 (38.27%)	
Mild	0	15 (26.32%)	15 (18.52%)	
Moderate	0	11 (19.30%)	11 (13.58%)	
Severe	0	12 (21.05%)	12 (14.81%)	
Gingivitis	0	10 (17.54%)	10 (12.35%)	
Stable periodontitis	0	2 (3.51%)	2 (2.47%)	<0.0001
<b>ABC positive (no., %)</b>				
Positive	0	13 (22.81%)	13 (16.05%)	
Negative	19 (79%)	29 (50.88%)	48 (59.26%)	
Not tested	5 (21%)	15 (26.32%)	20 (24.69%)	0.02
Neutrophil count (million cells, mean $\pm$ SD)	74.82 $\pm$ 58.56	158.22 $\pm$ 90.08	131.90 $\pm$ 94.51	<0.0001
Monocyte count (million cells, mean $\pm$ SD)	63.79 $\pm$ 50.35	113.91 $\pm$ 60.09	99.22 $\pm$ 63.28	<0.01

p values calculated by unpaired t tests or  $\chi^2$  (two-sided p values; italicized p < 0.05 significant).

BMI, body-mass index; ABC, *A. baumannii*; %, percentage; mg/dL, milligrams per deciliter; kg/m<sup>2</sup>, kilograms per square meter.

Smoking classification according to CDC National Health Interview Survey (NHIS).

Association of Diabetes ([American Diabetes Association, 2015](#)) and American Association of Periodontology (AAP) ([Armitage, 1999](#)) criteria, respectively. All T2D donors were poorly controlled diabetic subjects who had not taken antimicrobials, non-steroidal anti-inflammatory drugs, or insulin sensitizers within three months. Diagnostic serum glucose measures were taken for all subjects, and HbA1C was measured for diabetic patients to stage disease status. As expected, poorly controlled diabetic subjects (HbA1c% >6.5%) showed significantly elevated glucose levels (p < 0.0001; [Table 1](#)). There were no significant differences in gender or ethnicity between healthy volunteers and diabetic patients. Diabetic patients showed elevated body mass index and tended to be older than healthy controls (p < 0.05), consistent with known disease biology. Periodontal condition was further stratified among the subjects according to the severity of irreversible tissue loss (mild, moderate, severe) and reversible lesions called gingivitis according to AAP ([Armitage, 1999](#)) criteria. Periodontal condition was significantly worse in diabetic compared with healthy

subjects ( $p < 0.0001$ ). All healthy subjects had healthy periodontal condition compared with only 11% diabetic patients, with 15% diabetic subjects suffering from severe periodontitis. Current and former smokers were near uniformly part of the T2D and T2DP groups with only one current smoker in the healthy group. The entire study population had a median age range of  $53 \pm 11$  years, and all subjects were outpatient and thus not part of an extended hospital stay cohort.

### Microbiome analysis controls for low template sequencing

Data generated here was part of a *post hoc* study of serum samples collected for a previous study (Freire et al., 2017) and thus not available in the larger volumes ideal for 16S microbiome dataset generation. Low template concentrations can produce microbiome signatures not derived from that in the biological samples due to amplification of contaminating genomic DNA (gDNA) in sterile buffers and reagents (Kennedy et al., 2014; Salter et al., 2014; Kim et al., 2017), which often manifests as aquatic species signatures, many of which are not compatible with growth in the human host (i.e., many do not even grow at 37°C) (Kim et al., 2017). A description of no-template/reagent-only control samples is provided in the Methods. Data within this study are provided as either minimally filtered (mitochondrial, human, and chloroplast aligned reads removed) or strict filtered (reads removed based on homology with reagent-only controls) and indicated as such throughout the study. All sequence data are made available at NIH SRA Bioproject PRJNA664044, and all scripts used for data generation, filtering, and analyses are provided in the [Supplemental Information](#).

We were able to generate abundant data from these samples with careful consideration paid to controls for amplification bias and background bacterial gDNA contamination of commercial sterile reagents. First, we utilized amplification of a mock standard library of 8 bacterial and 2 fungal species (Figures S1A and S1B) at 10 and 1 ng of total template, which indicated that our library preparation protocol was suitable for species-level analysis where possible by V1-V3 sequence. Second, we utilized 13 separate no-template/reagent-only controls, which were subjected to our sequence analysis pipeline, revealing frequently encountered contaminating taxa. These taxa were filtered from each human-derived sample in the strict-filtered datasets only.

### V1-V3 16S microbiome diversity analyses reveal minor but significant differences between diabetic and nondiabetic groups

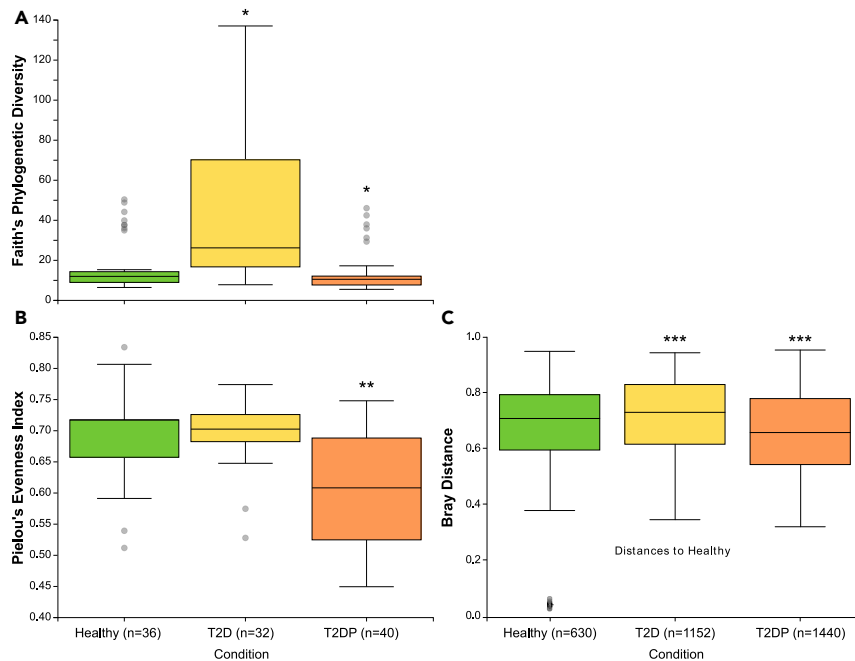
Serum and control samples were subject to bead beater lysis and gDNA purification. Purified gDNA was amplified using the V1-V3-targeted 27F-519R primer pair (Stackebrandt and Goodfellow, 1991; Turner et al., 1999). Quality trimmed reads were subjected to DADA2 analysis (Callahan et al., 2016) and aligned to the SILVA database (Quast et al., 2013) to determine species-level identity where possible. Rarefaction curve analysis was performed to ensure adequate sequencing depth (Figure S2), and resequencing of each Illumina index was performed to generate greater sequencing depth without further template amplification. Run-to-run variability between next-generation sequencing was found to be insignificant (Figure S3). Data analysis and comparisons were performed via QIIME2 (Bolyen et al., 2019) as detailed in the Methods.

We first quantified alpha and beta diversity between the healthy, T2D, and T2DP subject groups for our strict filtered (Figures 1 and S7) and minimally filtered data (Figure S4). For both datasets, species richness for T2D and T2DP groups were significantly different than the healthy cohort, which was also significantly different when compared with T2DP in each dataset for species evenness (T2D more rich, T2DP less rich, and less even). Beta diversity was significantly less for healthy subjects versus T2D and greater when compared with T2DP for either minimal or strict filtered datasets. Thus, we observed minor but significant shifts in microbiome composition between subject groups.

To begin investigating which taxa contributed to this composition difference we first looked at a phylum-level display of our taxonomically assigned data (Figure 2). Figures 2A and 2B, respectively, indicate phylum-level data for strict versus minimal filtered datasets. Similar trends in phyla distribution were evident for either dataset and reveal minor phylum-level changes between subject groups.

### MED and DADA2 analysis reveal the presence of *Acinetobacter baumannii*

Given our initial investigation into whether or not there were differences between T2D and T2DP microbiota and a preliminary hypothesis that oral taxa would be enriched in T2DP serum, we first used a minimal entropy decomposition (MED) analysis identical to the one used previously to characterize oral



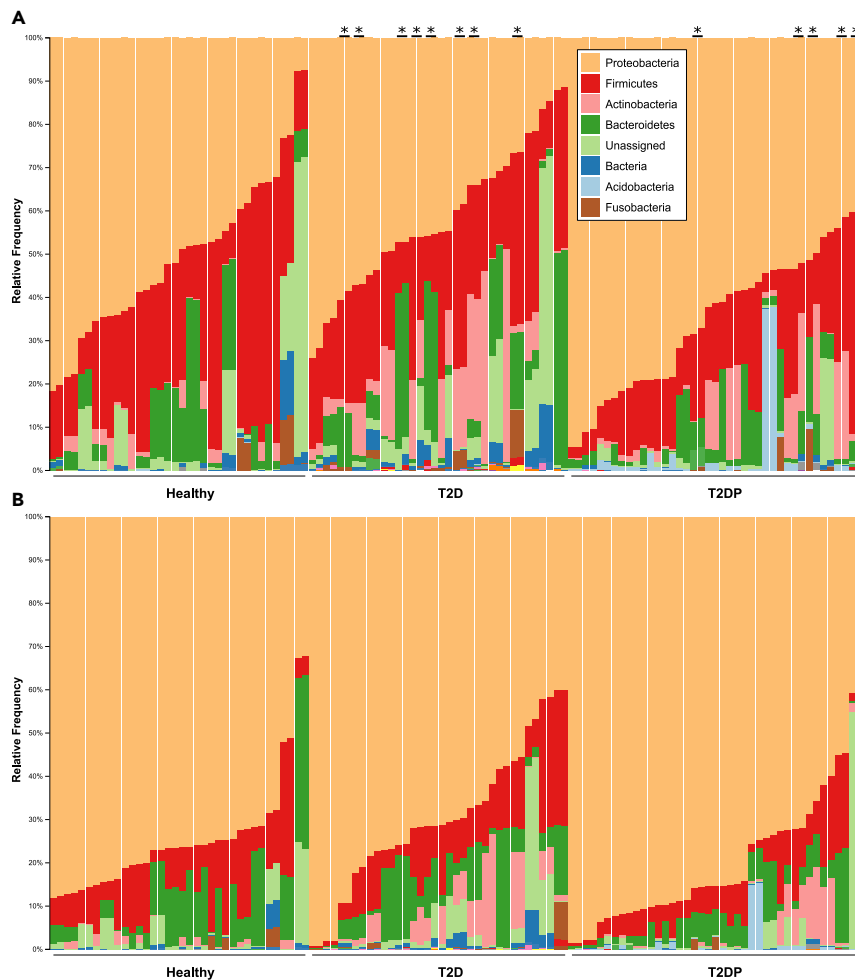
**Figure 1. 16S alpha and beta diversity of serum microbiomes changes with subject status**

(A–C) Species richness is displayed based on Faith's phylogenetic diversity (Faith, 1992) (A), and species evenness is based on Pielou's evenness index (Pielou, 1966) (B) for healthy, type-II diabetic patients (T2D), and type-II diabetic patients with periodontitis (T2DP) samples. Significance was determined by Kruskal-Wallis analysis of variance (Kruskal and Wallis, 1952) for each comparison indicated, and Benjamini-Hochberg correction (Benjamini and Hochberg, 1995) was applied to generate adjusted q-values. All data used were strictly filtered based on no-template control indexes. \*q-value < 0.002 compared with healthy, \*\*q-value < 0.0004 compared with healthy and with T2D. (C) Beta diversity Bray-Curtis distances. Pairwise PERMANOVA of each category versus each (group size of 3, n = 112) was performed in 999 permutations. \*\*\*q-value differs from "healthy" < 0.001. See also Figure S4.

microbiomes (Eren et al., 2014). Oligotyped "nodes" or representative sequences were aligned to both the Human Oral Microbiome (HOMD) (Escapa et al., 2018) and Ribosomal Database Project (RDP) (Cole et al., 2014) databases, and species were assigned based on 98% or higher sequence identity when comparing nodes to reference databases. We next compared species-assigned data between subject groups by linear discriminate analysis effect size measurements (LEfSe) (Segata et al., 2011). Using MED-analyzed data for healthy versus all subjects with T2D (Figure 3) we detected significant enrichments of some oral species (*Streptococcus cristatus*, *Neisseria flavescens*) and also some known reagent contaminants (*Ralstonia picketti*, *Afipia broomeae*) alongside other soil and freshwater microbes that we suspect are also contaminants (*Flectobacillus*, *Stenotrophomonas*, *Sphingobium*, *Arthrobacter*). Surprisingly, we observed *A. baumannii* significantly enriched in the T2D population as a whole. For MED analysis, reads assigned as *A. baumannii* were present from 0.01% to 7.3% of total reads in 8 and 5 subjects with T2D and T2DP, respectively. No reads assigned to *A. baumannii* were detected in any healthy subject. DADA2 amplicon sequence variants (ASVs) are similar to MED nodes, and we observed that one ASV node was a perfect match to the singular MED node assigned to *A. baumannii* based on 100% RDP alignment across its full sequence.

In our DADA2/QIIME2-based analysis we performed taxonomical assignment for each DADA2 ASV node aligned to either the Greenegenes (DeSantis et al., 2006) (not shown) or SILVA database version 132 (Quast et al., 2013). Phylum-level taxonomic assignments of DADA2 data aligned to SILVA were used in Figure 2. ASVs are similar to MED nodes, and we observed that one ASV node was a perfect match to the singular MED node assigned to *A. baumannii* based on 100% RDP alignment across its full sequence. This ASV was assigned only to the *Acinetobacter* genus despite a 100% full-length match to *A. baumannii* sequence at the RDP database and the NCBI non-redundant database (not shown). Based on this, we manually assigned this ASV to *A. baumannii* and found that reads now assigned to *A. baumannii* were present from 0.01% to 9.1% of total reads in 8 and 5 subjects with T2D and T2DP, respectively, with no reads assigned to *A. baumannii* detected in any healthy subjects. This demonstrated agreement between both MED and





**Figure 2. Phylum-level diversity changes between subject status**

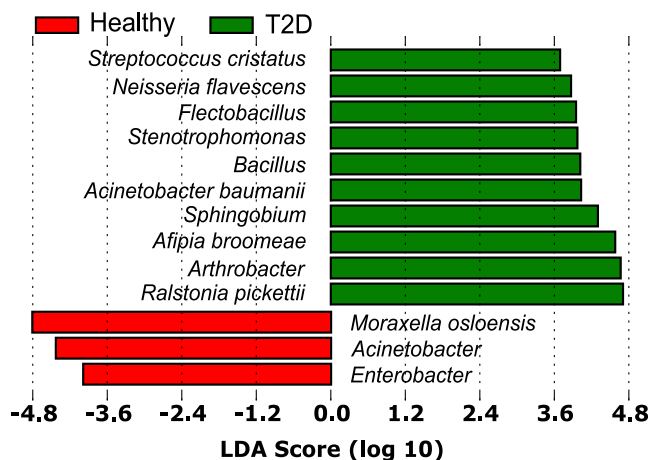
(A and B) DADA2-analyzed data were aligned to the SILVA database for taxonomical assignment via QIIME2 as described in Methods. Data from both sequencing runs were collapsed to phylum-level taxa and displayed for (A) strict filtered and (B) minimally filtered conditions. Asterisks on the top indicate run-run sample pairs that were positive for *A. baumannii*.

DADA2-based 16S data analysis. LefSe analysis was again performed on healthy subjects versus all with T2D, as well as between all three subject groups, and demonstrated again that *A. baumannii* was significantly elevated among all T2D samples (Figure S5).

We next looked at differences in microbiome compositions across subjects by plotting Bray-Curtis distances via non-metric multidimensional scaling (NMDS) while distinguishing subject groups as well as presence or absence of ABC bacteria (Figure 4). Broadly, T2D and T2DP samples had less overlap in composition, whereas healthy samples intersected either group. ABC status did not seem to form a distinct microbiome cluster, suggesting that there was no unique microbiota composition for ABC-positive subjects apart from *A. baumannii* sequence detection.

### Validating the presence of *A. baumannii* in subjects with T2D

The finding of ABC sequences was unexpected, as study participants were not hospitalized or under treatment for any infection(s), and we wished to confirm this result by other techniques. This was especially needed as the V1-V3 portion of the 16S rDNA sequence only differs between *A. baumannii* and *Acinetobacter nosocomialis* by one nucleotide substitution. Using PCR primers for the 16S rDNA region that encompasses nearly all differentiating 16S nucleotides between *Acinetobacter* species we generated amplicons from a portion of our ABC-positive samples alongside negative healthy and no-template control



**Figure 3. Healthy subjects versus all subjects with T2D reveal a potential enrichment of *A. baumannii***

Initial minimal entropy decomposition (MED) analysis of 16S data with taxonomy assigned based on RDP database alignment to generate species-level assignments was performed followed by linear discriminant analysis effect size quantification (LEfSe). This revealed enrichment of several species/genera in subjects with T2D relative to healthy. DADA2 analysis of the same dataset was aligned to the SILVA database and shows similar patterns of enrichment (see also Figure S5).

samples (Figure 5). PCR products were ligated into a TA vector, and individual colonies were propagated and subjected to Sanger sequencing of plasmid inserts. The 1242-bp sequences were then aligned to the NCBI reference RNA sequence database (refseq\_rna). Each sequence (12 clones across 5 samples) was identical and matched the *A. baumannii* 16S sequence across its length. This result validated our initial 16S detection of putative *A. baumannii*.

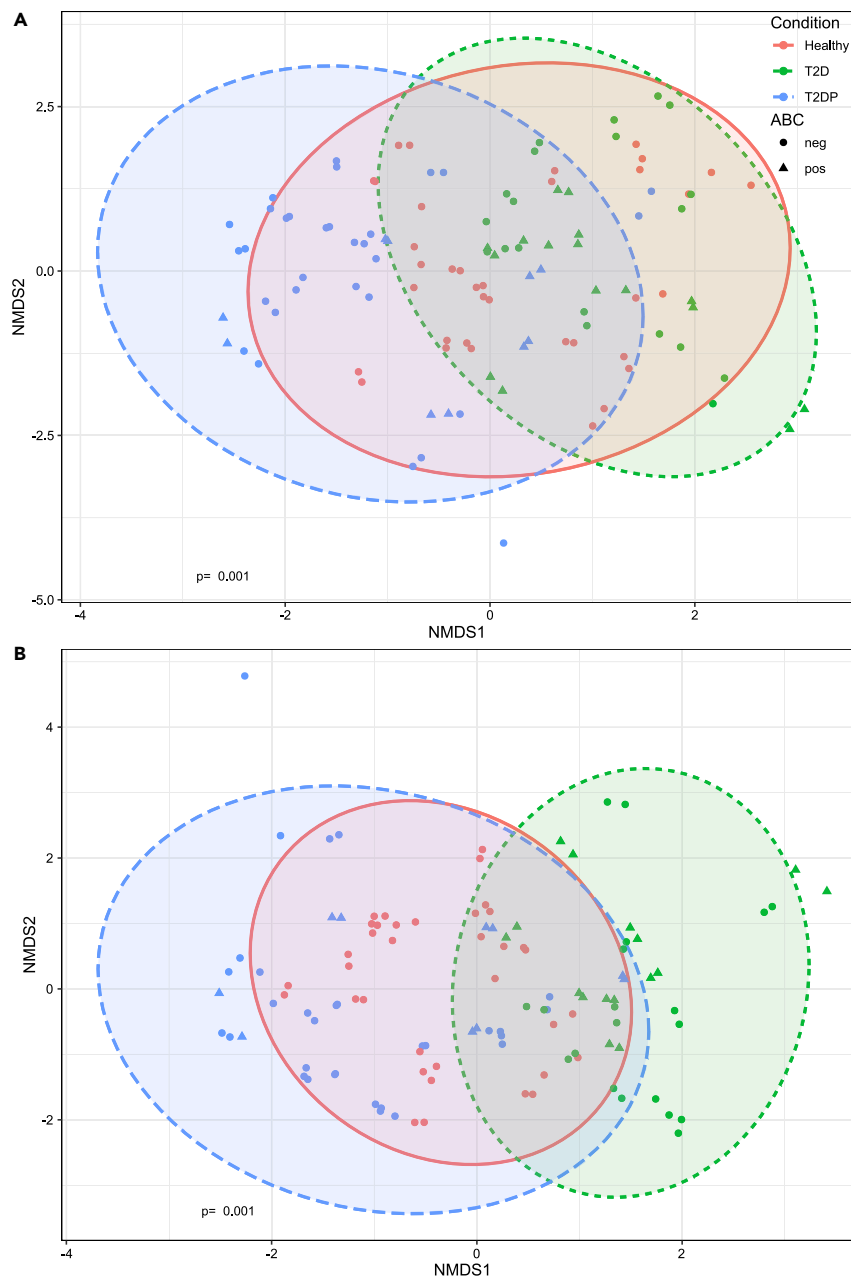
### Host cytokine responses to T2D and T2DP versus health

In addition to serum microbiota, we also quantified 20 different human cytokines in serum samples across all 3 patient groups (Figure 6). Based on relative abundance profiles, we observed that proinflammatory cytokines were significantly elevated in *A. baumannii*-positive samples highlighted in the box (ABC+). MCP-1 and IL-1 $\beta$ , classic proinflammatory cytokines, were increased in diabetic patients. Intriguingly, T2D ABC-negative subjects did not show increase of these markers. ABC+ subjects showed significantly higher expression levels of cytokines IL-1 $\beta$ , IL-6, IL-8, MCP-1, and IFN- $\gamma$  when compared with ABC-negative groups (I and V,  $p < 0.01$ ).

### Host cytokine responses to specific microbiota

Based on the cytokine abundance data and earlier microbiome sequencing, we were able to determine if any specific taxa were significantly correlated with cytokine presence or absence in subjects with T2D (Figure 7). *A. baumannii* was significantly correlated with elevated amounts of IFN- $\alpha$ , IFN- $\gamma$ , IL-12p70, IL-8, and TNF- $\alpha$ . ABC+ subjects presented cytokine correlations consistent with a profile of chronic inflammation. IL-8 is a chemokine that impacts neutrophil phenotype and function, and excessive amounts of this cytokine lead to chronic activation of phagocytes (Moore and Kunkel, 2019). Excessive production of TNF- $\alpha$  increases failure of resolution of inflammation on innate immune cells by controlling a resolution receptor (Freire and Van Dyke, 2013). Here, TNF- $\alpha$  showed a significant increase in ABC+ subjects, suggesting co-existence with ABC in the serum of patients with T2D. We also performed an identical analysis for healthy subjects and saw several genera or species with significant correlations (Figure S6). In healthy subjects unique profiles were found, correlated with lower levels of proinflammatory cytokines and co-clustered with ABC-negative samples.

Although low template concentration microbiome profiling presents many challenges, we were able to utilize numerous control steps to minimize the impact of contaminating taxa. We were also able to demonstrate that there were significant differences between subject groups that include taxa of likely biological origins, and also that specific cytokines significantly correlated with the presence of distinct taxa. Most surprisingly, we were able to detect *A. baumannii* and confirm its presence in serum samples of patients with



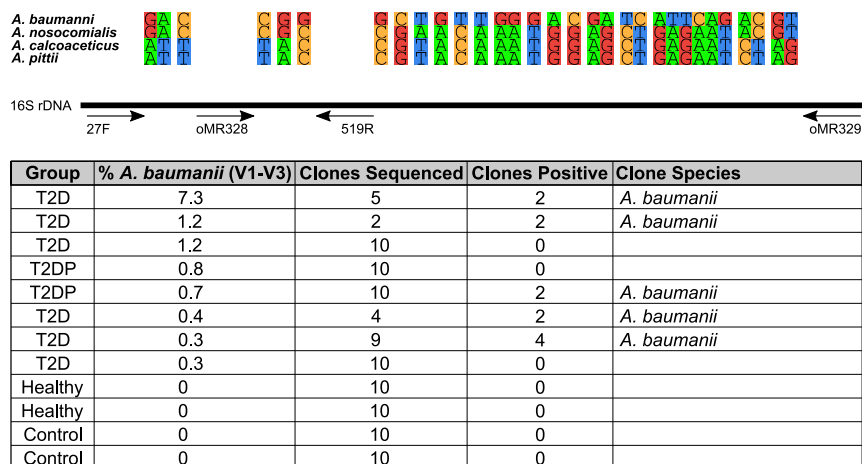
**Figure 4. Beta diversity quantification between subject groups**

(A and B) SILVA-assigned taxonomy of DADA2-analyzed 16S data was used to determine Bray-Curtis dissimilarity between samples for strictly filtered (A) and minimally filtered (B) data. Non-metric multidimensional scaling and statistical analysis for each subject group was performed in R using the *vegan* package as described in Methods. p-value of 0.001 determined by PERMANOVA test via the *adonis2* function.

T2D not known or currently diagnosed with any underlying infection. These data do not *per se* indicate active infection but could also indicate colonization or transient exposure to *A. baumannii* for these subjects compared with healthy subjects.

## Discussion

Issues with amplification-based 16S studies from low template concentrations are well known (Sinha et al., 2015; Kim et al., 2017; Pollock et al., 2018). We began this study as a *post hoc* analysis of existing serum samples from a



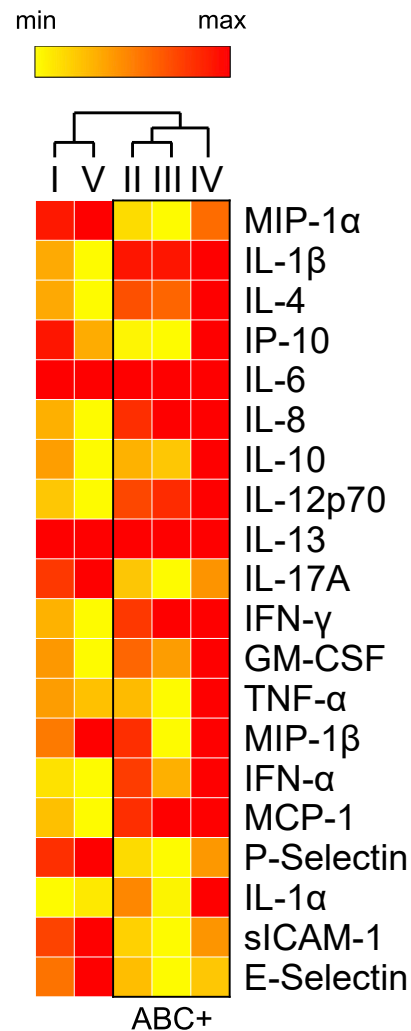
**Figure 5. Clone library design to verify ABC complex member identity**

Amplification with oMR328-329 primers produced an amplicon between 120 and 1306 bp of the *A. baumannii* 16S rDNA sequence. This allowed for Sanger sequencing of information-rich nucleotides indicated above allowing for species differentiation beyond initial V1-V3 sequencing with the 27F-519R primers. The table beneath indicated the total number of individual clones sequenced per sample and which species they aligned to with 100% identity in the NCBI non-redundant database.

previous bank of anonymized specimens (Freire et al., 2017). As sample volumes of serum were mostly limited, we relied on amplification of 16S rDNA, using V1-V3 region primers to improve the likelihood of identifying oral taxa (Eren et al., 2014) that we hypothesized to exist in the diabetic with periodontitis part of our cohort. To minimize effects of amplification, we performed multiple high-throughput sequencing runs of amplified libraries to increase sequence depth as opposed to further amplification and using a higher capacity sequencing platform. Additionally, we also included numerous no-template controls throughout sample preparation and purification to identify as many outside contaminants as we could (Kennedy et al., 2014; Salter et al., 2014; Kim et al., 2017). As a further control, we carried out amplification of mock community libraries (Figure S1) to determine if amplification bias or infidelity affected taxonomic assignment. We also used two separate methods for correction of Illumina-sequenced amplicon errors, DADA2 (Callahan et al., 2016) and minimal entropy decomposition/oligotyping (MED) (Eren et al., 2015), as well as comparing our data from multiple ribosomal databases for taxonomic assignment (DeSantis et al., 2006; Quast et al., 2013; Cole et al., 2014). In addition to Illumina sequencing, we performed a separate amplification of 16S rDNA followed by Sanger sequencing to independently confirm the identity of putative *A. baumannii* sequence presence in a subset of our diabetic samples (Figure 5). Thus, we have made great efforts to validate our findings as far as is feasible and have taken care not to overinterpret data presented here.

We initially hypothesized that oral taxa would be enriched in the bloodstream of subjects with T2DP, and potentially all subjects with T2D, compared with healthy subjects. Our microbiome data revealed changes in composition between our three subject groups (Figures 1, 2, 3, 4, S4, and S5). However, there was only enrichment of few oral taxa when healthy subjects were compared with all subjects with T2D and not with subjects with T2DP alone, specifically *S. cristatus* and *N. flavescens* (Figure 3), which are not associated with periodontitis. Their presence here could indicate further access to the bloodstream by oral microbes in the T2D cohort, although this requires further verification. Given the limited template available for each sample, we cannot rule out that other oral taxa are not elevated in subjects with T2D/P, but more rigorous testing of that hypothesis would require larger serum volumes and using a longitudinal approach. By far the most surprising element of this dataset was the detection of *A. baumannii*- (Figures 3 and S5) assigned reads unique to 13 subjects with T2DP and not present in any control or healthy subject samples.

*A. baumannii* represents a considerable risk to subjects with T2D including higher mortality rates (Leão et al., 2016) and a higher risk of infection in burn wounds, as well as DFIs (Furniss et al., 2005; Castellanos et al., 2019; Henig et al., 2020). To confirm our V1-V3 amplicon data and to determine whether or not these were *A. baumannii* or another ABC species, we used targeted near-full-length 16S rDNA primers that allowed us to sequence the majority of differentiating nucleotides necessary for ribotyping. Sequencing individual clones of these amplicons

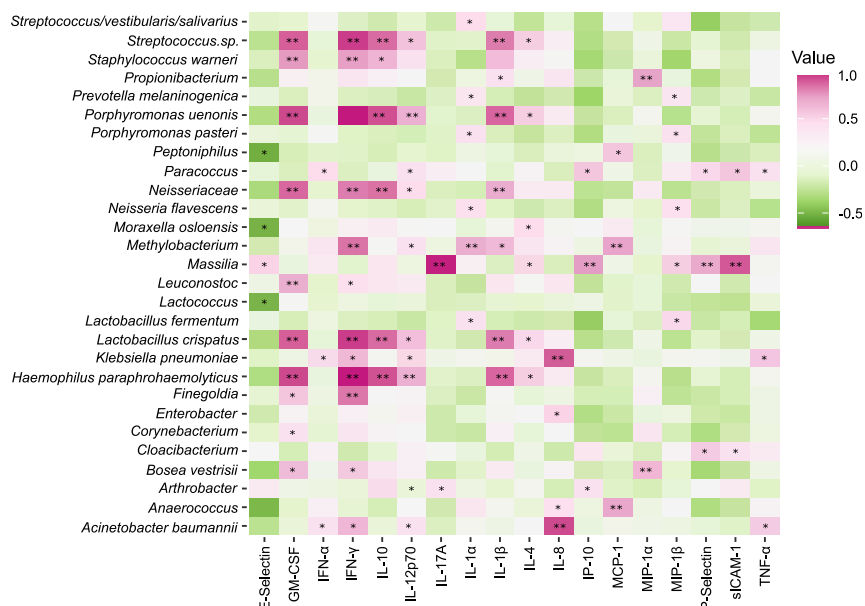


**Figure 6. Cytokine profile quantification reveals distinct differences between subject groupings and *A. baumannii* (ABC) presence**

Heatmap showing immune mediator concentrations derived from serum of diabetic and healthy samples were compared to understand the role of ABC in each sample group. (I) Diabetic/Healthy, (II) All ABC+/All ABC-, (III) ABC+ Diabetic/ABC- Diabetic, (IV) ABC+ Diabetic/Healthy, (V) ABC- Diabetic/Healthy. Data are clustered according to comparison groups (red, highest expression; yellow, lowest expression). Hierarchical clustering was employed by Morpheus from the Broad Institute (Morpheus, <https://software.broadinstitute.org/morpheus>). Observations from hierarchical clustering are shown in a tree, with ABC groups highlighted in the box. ABC+ groups (II, III, and IV) showed significantly different ( $p < 0.01$ ) levels of IL-1β, IL-6, IL-8, MCP-1, IFN-γ. \* $p$  value  $< 0.01$  between groups via unpaired Student's  $t$  test (red, highest expression; yellow, lowest expression).

revealed 100% matches to *A. baumannii*, and no clones ( $n = 50$ ) were *A. baumannii* positive from indicated control samples (Figure 5). This result was surprising, as none of the subjects in our T2D/P cohort were currently under inpatient treatment or had any indication of underlying infection (apart from periodontitis). Given the methodology here, neither can we speculate on the true infection status of these individuals nor can we be certain if the *A. baumannii* present were circulating in the bloodstream or were on the skin and mixed with the sample during a blood draw. No matter the route of entry, we were able to determine significant cytokine profile changes in T2DP *A. baumannii*-positive subjects, as well as broader inflammatory changes in cytokine profiles between subjects with T2DP and T2D and healthy subjects (Figure 6).

Cytokine dysregulation can impact the host locally and systemically, making the subject more prone to severe infections and increased tissue damage (Tisoncik et al., 2012). Individuals who have T2D present



**Figure 7. Host-microbial correlation-resolved type-II diabetic associations with inflammatory cytokines**

HOMD/RDP assigned taxonomy of MED-analyzed 16S data for both T2D and T2DP samples and cytokine concentrations were analyzed via Pearson correlation coefficient in R using the *rcorr* function. Significance was determined using the asymptotic p values generated by *rcorr* with \*p value < 0.05, \*\*p value < 0.01. Data are strictly filtered with taxa present in no-template controls subtracted. See also [Figures S6, S8, and S9](#).

an increased metabolic burden, while often comorbid periodontal disease promotes microbial dysbiosis, with both diseases resulting in a chronic state of inflammation. In the present study, *A. baumannii* was only found in the serum of diabetic subjects, where an increase in proinflammatory cytokines was also observed ([Figures 6 and 7](#)). We utilized an unbiased panel of 20 human cytokines to survey specific immune responses to T2D and T2DP compared with those of healthy individuals. Although there are complications presented by low template sizes in our samples, cytokine profiling from these same volumes were well within desired assay volumes. As mediators of inflammation, cytokine production feeds forward a cascade of signals that can impact tissue, organ, and overall host immunometabolism. In T2D, IL-1 $\beta$ , TNF- $\alpha$ , MCP-1, IL-6, IL-8, and IFN- $\gamma$  were significantly elevated in ABC+ subjects and not in ABC-negative subjects. Of these proinflammatory cytokines, IL-1 $\beta$ , TNF- $\alpha$ , and IL-8 suppression has been investigated as therapeutic targets for clinical management of diabetes and associated chronic lesions ([Peiró et al., 2017](#); [Feng et al., 2018](#)). We previously observed increased TNF- $\alpha$  levels in cell cultured neutrophils from subjects with T2D relative to cells derived from healthy subjects ([Kleinstejn et al., 2020a](#)). Co-occurrence of bacteria in the serum and cytokines also differed among controls. In the current study, co-occurrence among cytokines was more evident in diabetic subjects (100 significant co-occurrences) versus healthy individuals (30 significant co-occurrences) ([Figures 7 and S6](#)). *A. baumannii* especially correlated with neutrophil inflammatory signaling cytokine IL-8, which was not observed in healthy subjects. We have previously shown differences in gene expression in neutrophils of individuals with T2D, which are primed to mount an aberrant inflammatory response when compared with neutrophils of healthy individuals ([Kleinstejn et al., 2020a](#)), consistent with production of inflammatory cytokines observed here. In contrast, IL-17 $\alpha$ , P-selectin, and E-selectin levels were increased in diabetic subjects who were ABC-negative compared with ABC+ diabetic subjects. IL-4 and IL-10 did not co-occur with *A. baumannii* in any of the healthy versus diabetic groups. The cytokine results provide an initial glimpse into the host response to chronic inflammatory diseases, T2DP with and without the presence of *A. baumannii*, which may warrant further investigation in larger studies.

In conclusion, this study has revealed modest serum microbiome compositional changes and significant cytokine profile changes between subjects with T2D with or without periodontitis versus non-T2D individuals, revealing an impaired host response when *A. baumannii* was present. We also present here the potential for lurking *A. baumannii* colonization in asymptomatic subjects with T2D, possibly to

a much larger extent than previously reported. In our opinion, these data present a strong justification for further monitoring of individuals with T2D for *A. baumannii* colonization and support the need for a more robust study to characterize the potentially greater spread of *A. baumannii* among individuals with T2D.

### Limitations of the study

Limitations of this study primarily include small template volumes available for serum samples in this *post hoc* study. Blood and sera microbiome studies will always be more complex due to low bacterial template concentrations inherent to these types of sample. Although we have made utmost efforts to validate our main findings regarding *A. baumannii* presence, there are likely more interesting aspects to our study cohort microbiota that have been missed. These and other minor limitations have been indicated in the article.

### Resource availability

#### Lead contact

Further information and requests for resources, data, code, and reagents should be directed to and will be fulfilled by the Lead Contact, Dr. Matthew Ramsey ([mramsey@uri.edu](mailto:mramsey@uri.edu)).

#### Materials availability

This study did not generate new unique reagents or genetic constructs.

#### Data and code availability

All data is available at NIH – SRA Bioproject: [PRJNA664044](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA664044). All other code used for data analysis has been provided in the [Transparent Methods](#) section (see Scripts) in the [Supplemental Information](#).

### Methods

All methods can be found in the accompanying [Transparent Methods supplemental file](#).

### Supplemental information

Supplemental Information can be found online at <https://doi.org/10.1016/j.isci.2020.101941>.

### Acknowledgments

Thanks to the M. Ramsey and M. Freire lab members for their valuable input and assistance with both experimental work and writing. Thanks to Dr. Janet Atoyán at the Rhode Island Genomics and Sequencing Center for assistance with sequence library preparation and suggestions on sequence data analysis. Thanks to Dr. Thomas E. Van Dyke for helpful discussion and to the Forsyth Center for Clinical and Translational Research Center staff for recruiting patient samples and providing clinical data used in this study. **Funding Sources:** This work was supported by U.S. Public Health Service Grants from the National Institutes of Health/National Institute of Dental and Craniofacial Research, R00 DE0234804 (Freire), K23 DE018917 (Hasturk) and R01 DE027958 (Ramsey) NIGMS/RI-INBRE early career development award (P20GM103430, Ramsey), and the Rhode Island Foundation Medical Research Fund (20164348, Ramsey).

### Author contributions

D.P., B.H., and S.E.K. performed the work. D.P., S.E.K., M.F., and M.R. wrote and edited the manuscript. M.F., H.H., and M.R. provided reagents and samples. D.P., R.E., S.E.K., M.F., and M.R. designed the experiments. M.F. and M.R. secured funding.

### Declaration of interests

The authors declare no competing interests.

Received: September 22, 2020

Revised: October 28, 2020

Accepted: December 10, 2020

Published: January 22, 2021

## References

- Al-Anazi, K.A., and Al-Jasser, A.M. (2014). Infections caused by acinetobacter baumannii in recipients of hematopoietic stem cell transplantation. *Front. Oncol.* 4, 186.
- American Diabetes Association (2015). (2) Classification and diagnosis of diabetes. *Diabetes Care* 38, S8–S16.
- Armitage, G.C. (1999). Development of a classification system for periodontal diseases and conditions. *Ann. Periodontol.* 4, 1–6.
- National diabetes Statistics Report | data & Statistics | diabetes | CDC (2020). <https://www.cdc.gov/diabetes/data/statistics/statistics-report.html>
- Ayoub Moubareck, C., and Hammoudi Halat, D. (2020). Insights into acinetobacter baumannii: a review of microbiological, virulence, and resistance traits in a threatening nosocomial pathogen. *Antibiotics (Basel)* 9, 119.
- Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B (Methodological)* 57, 289–300.
- Bolyen, E., Rideout, J.R., Dillon, M.R., Bokulich, N.A., Abnet, C.C., Al-Ghalith, G.A., Alexander, H., Alm, E.J., Arumugam, M., Asnicar, F., et al. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat. Biotechnol.* 37, 852–857.
- Bowler, P.G., and Davies, B.J. (1999). The microbiology of infected and noninfected leg ulcers. *Int. J. Dermatol.* 38, 573–578.
- Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J., and Holmes, S.P. (2016). DADA2: high-resolution sample inference from Illumina amplicon data. *Nat. Methods* 13, 581–583.
- Carrizales-Sepúlveda, E.F., Ordaz-Farías, A., Vera-Pineda, R., and Flores-Ramírez, R. (2018). Periodontal disease, systemic inflammation and the risk of cardiovascular disease. *Heart Lung Circ.* 27, 1327–1334.
- Castellanos, N., Nakanouchi, J., Yüzen, D.I., Fung, S., Fernandez, J.S., Barberis, C., Tuchscher, L., and Ramirez, M.S. (2019). A study on acinetobacter baumannii and Staphylococcus aureus strains recovered from the same infection site of a diabetic patient. *Curr. Microbiol.* 76, 842–847.
- Citron, D.M., Goldstein, E.J., Merriam, C.V., Lipsky, B.A., and Abramson, M.A. (2007). Bacteriology of moderate-to-severe diabetic foot infections and in vitro activity of antimicrobial agents. *J. Clin. Microbiol.* 45, 2819–2828.
- Cole, J.R., Wang, Q., Fish, J.A., Chai, B., McGarrell, D.M., Sun, Y., Brown, C.T., Porras-Alfaro, A., Kuske, C.R., and Tiedje, J.M. (2014). Ribosomal Database Project: data and tools for high throughput rRNA analysis. *Nucleic Acids Res.* 42, D633–D642.
- Curtis, M.A., Diaz, P.I., and Van Dyke, T.E. (2020). The role of the microbiota in periodontal disease. *Periodontology* 2000, 14–25.
- DeSantis, T.Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E.L., Keller, K., Huber, T., Dalevi, D., Hu, P., and Andersen, G.L. (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl. Environ. Microbiol.* 72, 5069–5072.
- Dijkshoorn, L., Nemeč, A., and Seifert, H. (2007). An increasing threat in hospitals: multidrug-resistant Acinetobacter baumannii. *Nat. Rev. Microbiol.* 5, 939–951.
- Dowd, S.E., Wolcott, R.D., Sun, Y., McKeenan, T., Smith, E., and Rhoads, D. (2008). Polymicrobial nature of chronic diabetic foot ulcer biofilm infections determined using bacterial tag encoded FLX amplicon pyrosequencing (bTEFAP). *PLoS One* 3, e3326.
- Driver, V.R., Fabbri, M., Lavery, L.A., and Gibbons, G. (2010). The costs of diabetic foot: the economic case for the limb salvage team. *J. Am. Podiatr. Med. Assoc.* 100, 335–341.
- Emrich, L.J., Shlossman, M., and Genco, R.J. (1991). Periodontal disease in non-insulin-dependent diabetes mellitus. *J. Periodontol.* 62, 123–131.
- Erben, N., Ozgunes, I., Aksit, F., Doyuk Kartal, E., Colak, E., and Usluer, G. (2013). Healthcare-associated infections and the distribution of causative pathogens in patients with diabetes mellitus. *Eur. J. Clin. Microbiol. Infect. Dis.* 32, 821–825.
- Eren, A.M., Boris, G.G., Huse, S.M., and Mark Welch, J.L. (2014). Oligotyping analysis of the human oral microbiome. *Proc. Natl. Acad. Sci. U S A* 111, E2875–E2884.
- Eren, A.M., Morrison, H.G., Lescault, P.J., Reveillaud, J., Vineis, J.H., and Sogin, M.L. (2015). Minimum entropy decomposition: unsupervised oligotyping for sensitive partitioning of high-throughput marker gene sequences. *ISME J.* 9, 968–979.
- Escapa, I.F., Morrison, H.G., Lescault, P.J., Reveillaud, J., Vineis, J.H., and Sogin, M.L. (2018). 'New insights into human nostril microbiome from the expanded human oral microbiome database (eHOMD): a resource for the microbiome of the human aerodigestive tract'. *Xu, ed. 3. http://msystems/3/6/msys.00187-18. atom, e00187-18.*
- Faith, D.P. (1992). Conservation evaluation and phylogenetic diversity. *Biol. Conserv.* 61, 1–10.
- Feng, S., Yu, H., Yu, Y., Geng, Y., Li, D., Yang, C., Lv, Q., Lu, L., Liu, T., Li, G., and Yuan, L. (2018). Levels of inflammatory cytokines IL-1 $\beta$ , IL-6, IL-8, IL-17a, and TNF- $\alpha$  in aqueous humour of patients with diabetic retinopathy. *J. Diabetes Res.* 2018, 1–6.
- Ferlita, S., Yegiazaryan, A., Noori, N., Lal, G., Nguyen, T., To, K., and Venketaraman, V. (2019). Type 2 diabetes mellitus and altered immune system leading to susceptibility to pathogens, especially Mycobacterium tuberculosis. *J. Clin. Med.* 8, 2219.
- Freire, M.O., and Van Dyke, T.E. (2013). Natural resolution of inflammation. *Periodontology* 2000, 149–164.
- Freire, M.O., Dalli, J., Serhan, C.N., and Van Dyke, T.E. (2017). Neutrophil resolvin E1 receptor expression and function in type 2 diabetes. *J. Immunol.* 198, 718–728.
- Furniss, D., Gore, S., Azadian, B., and Myers, S.R. (2005). Acinetobacter infection is associated with acquired glucose intolerance in burn patients. *J. Burn Care Rehabil.* 26, 405–408.
- Gardner, S.E., Hillis, S.L., Heilmann, K., Segre, J.A., and Grice, E.A. (2013). The neuropathic diabetic foot ulcer microbiome is associated with clinical factors. *Diabetes* 62, 923–930.
- Genco, R.J., Graziani, F., and Hasturk, H. (2020). Effects of periodontal disease on glycemic control, complications, and incidence of diabetes mellitus. *Periodontology* 83, 59–65.
- Gerner-Smidt, P., and Tjernberg, I. (1993). Acinetobacter in Denmark: II. Molecular studies of the Acinetobacter calcoaceticus-Acinetobacter baumannii complex. *APMIS* 101, 826–832.
- Gootz, T.D., and Marra, A. (2008). Acinetobacter baumannii: an emerging multidrug-resistant threat. *Expert Rev. Anti Infect. Ther.* 6, 309–325.
- Griffen, A.L., Beall, C.J., Campbell, J.H., Firestone, N.D., Kumar, P.S., Yang, Z.K., Podar, M., and Leys, E.J. (2012). Distinct and complex bacterial profiles in human periodontitis and health revealed by 16S pyrosequencing. *ISME J.* 6, 1176–1185.
- Hajishengallis, G., and Lamont, R.J. (2012). Beyond the red complex and into more complexity: the polymicrobial synergy and dysbiosis (PSD) model of periodontal disease etiology. *Mol. Oral Microbiol.* 27, 409–419.
- Henig, O., Pogue, J.M., Martin, E., Hayat, U., Ja'ara, M., Kilgore, P.E., Cha, R., Dhar, S., and Kaye, K.S. (2020). The impact of multidrug-resistant organisms on outcomes in patients with diabetic foot infections. *Open Forum Infect. Dis.* 7, ofaa161.
- Irfan, U.M., Dawson, D.V., and Bissada, N.F. (2001). Epidemiology of periodontal disease: a review and clinical perspectives. *J. Int. Acad. Periodontol.* 3, 14–21.
- Kennedy, K., Hall, M.W., Lynch, M.D., Moreno-Hagelsieb, G., and Neufeld, J.D. (2014). Evaluating bias of illumina-based bacterial 16S rRNA gene profiles. *K.E. Wommack, ed. 80, 5717–5722.*
- Kim, D., Hofstaedter, C.E., Zhao, C., Mattei, L., Tanes, C., Clarke, E., Lauder, A., Sherrill-Mix, S., Chehoud, C., Kelsen, J., et al. (2017). Optimizing methods and dodging pitfalls in microbiome research. *Microbiome* 5, 52.
- Kim, E.J., Ha, K.H., Kim, D.J., and Choi, Y.H. (2019). Diabetes and the risk of infection: a national cohort study. *Diabetes Metab. J.* 43, 804–814.
- Kleinstejn, S.E., McCarrison, J., Ahmed, A., Hasturk, H., Van Dyke, T.E., and Freire, M. (2020a). Transcriptomics of type 2 Diabetic and healthy human neutrophils. <https://doi.org/10.21203/rs.2.21951/v1>.



- Kleinstejn, S.E., Nelson, K.E., and Freire, M. (2020b). Inflammatory networks linking oral microbiome with systemic health and disease. *J. Dental Res.* **99**, 1131–1139.
- Kruskal, W.H., and Wallis, W.A. (1952). Use of ranks in one-criterion variance analysis. *J. Am. Stat. Assoc.* **47**, 583–621.
- Leão, A.C.Q., Menezes, P.R., Oliveira, M.S., and Levin, A.S. (2016). *Acinetobacter* spp. are associated with a higher mortality in intensive care patients with bacteremia: a survival analysis. *BMC Infect. Dis.* **16**, 386.
- Leung, C.-H., and Liu, C.-P. (2019). Diabetic status and the relationship of blood glucose to mortality in adults with carbapenem-resistant *Acinetobacter baumannii* complex bacteremia. *J. Microbiol. Immunol. Infect.* **52**, 654–662.
- Lockhart, P.B., and Durack, D.T. (1999). Oral microflora as a cause of endocarditis and other distant site infections. *Infect. Dis. Clin. North Am.* **13**, 833–850.
- Mann, J., Bernstein, Y., and Findler, M. (2020). Periodontal disease and its prevention, by traditional and new avenues. *Exp. Ther. Med.* **19**, 1504–1506.
- McDonald, H.I., Nitsch, D., Millett, E.R.C., Sinclair, A., and Thomas, S.L. (2014). New estimates of the burden of acute community-acquired infections among older people with diabetes mellitus: a retrospective cohort study using linked electronic health records. *Diabetic Med.* **31**, 606–614.
- Moore, B.B., and Kunkel, S.L. (2019). Attracting attention: discovery of IL-8/CXCL8 and the birth of the chemokine field. *J. Immunol.* **202**, 3–4.
- Nowicki, E.M., Shroff, R., Singleton, J.A., Renaud, D.E., Wallace, D., Drury, J., Zirnheld, J., Colleti, B., Ellington, A.D., Lamont, R.J., et al. (2018). Microbiota and metatranscriptome changes accompanying the onset of gingivitis. *mBio* **9**, e00575-18.
- Parahitayawa, N.B., Jin, L.J., Leung, W.K., Yam, W.C., and Samaranyake, L.P. (2009). Microbiology of odontogenic bacteremia: beyond endocarditis. *Clin. Microbiol. Rev.* **22**, 46–64.
- Peiró, C., Lorenzo, Ó., Carraro, R., and Sánchez-Ferrer, C.F. (2017). IL-1 $\beta$  inhibition in cardiovascular complications associated to diabetes mellitus. *Front. Pharmacol.* **8**, 363.
- Peleg, A.Y., Seifert, H., and Paterson, D.L. (2008). *Acinetobacter baumannii*: emergence of a successful pathogen. *Clin. Microbiol. Rev.* **21**, 538–582.
- Pielou, E.C. (1966). The measurement of diversity in different types of biological collections. *J. Theor. Biol.* **13**, 131–144.
- Pollock, J., Glendinning, L., Wisedchanwet, T., and Watson, M. (2018). The madness of microbiome: attempting to find consensus “best practice” for 16S microbiome studies. In Liu, C.-P., ed. *84*, e02627-17.
- Qiu, H., KuoLee, R., Harris, G., and Chen, W. (2009). High susceptibility to respiratory *Acinetobacter baumannii* infection in A/J mice is associated with a delay in early pulmonary recruitment of neutrophils. *Microbes Infect.* **11**, 946–955.
- Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., and Glöckner, F.O. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* **41**, D590–D596.
- Reiber, G.E., Lipsky, B.A., and Gibbons, G.W. (1998). The burden of diabetic foot ulcers. *Am. J. Surg.* **176**, 5S–10S.
- Roberts, F.A., and Darveau, R.P. (2015). Microbial protection and virulence in periodontal tissue as a function of polymicrobial communities: symbiosis and dysbiosis. *Periodontology* **69**, 18–27.
- Salter, S.J., Cox, M.J., Turek, E.M., Calus, S.T., Cookson, W.O., Moffatt, M.F., Turner, P., Parkhill, J., Loman, N.J., and Walker, A.W. (2014). Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol.* **12**, 87.
- Segata, N., Izard, J., Waldron, L., Gevers, D., Miropolsky, L., Garrett, W.S., and Huttenhower, C. (2011). Metagenomic biomarker discovery and explanation. *Genome Biol.* **12**, R60.
- Sinha, R., Abnet, C.C., White, O., Knight, R., and Huttenhower, C. (2015). The microbiome quality control project: baseline study design and future directions. *Genome Biol.* **16**, 276.
- Socransky, S.S., Haffajee, A.D., Cugini, M.A., Smith, C., and Kent, R.L., Jr. (1998). Microbial complexes in subgingival plaque. *J. Clin. Periodontol.* **25**, 134–144.
- E. Stackebrandt, and M. Goodfellow, eds. (1991). *Nucleic Acid Techniques in Bacterial Systematics* (Wiley), pp. 225–236.
- Tisoncik, J.R., Korth, M.J., Simmons, C.P., Farrar, J., Martin, T.R., and Katze, M.G. (2012). Into the eye of the cytokine storm. *Microbiol. Mol. Biol. Rev.* **76**, 16–32.
- Turner, S., Pryer, K.M., Miao, V.P., and Palmer, J.D. (1999). Investigating deep phylogenetic relationships among cyanobacteria and plastids by small subunit rRNA sequence analysis. *J. Eukaryot. Microbiol.* **46**, 327–338.
- Vardakas, K.Z., Siempos, I.I., and Falagas, M.E. (2007). ‘Diabetes mellitus as a risk factor for nosocomial pneumonia and associated mortality’. *Diabetic Medicine. A J. Br. Diabetic Assoc.* **24**, 1168–1171.
- Vázquez-López, R., Solano-Gálvez, S.G., Juárez Vignon-Whaley, J.J., Abello Vaamonde, J.A., Padró Alonzo, L.A., Rivera Reséndiz, A., Muleiro Álvarez, M., Vega López, E.N., Franyuti-Kelly, G., Álvarez-Hernández, D.A., et al. (2020). *Acinetobacter baumannii* resistance: a real challenge for clinicians. *Antibiotics (Basel)* **9**, 205.
- Villar, M., Cano, M.E., Gato, E., Garnacho-Montero, J., Miguel Cisneros, J., Ruiz de Alegría, C., Fernández-Cuenca, F., Martínez-Martínez, L., Vila, J., Pascual, A., et al. (2014). Epidemiologic and clinical impact of *Acinetobacter baumannii* colonization and infection: a reappraisal. *Medicine* **93**, 202–210.
- Whiting, D.R., Guariguata, L., Weil, C., and Shaw, J. (2011). IDF diabetes atlas: global estimates of the prevalence of diabetes for 2011 and 2030. *Diabetes Res. Clin. Pract.* **94**, 311–321.
- Wu, C.-Z., Yuan, Y.H., Liu, H.H., Li, S.S., Zhang, B.W., Chen, W., An, Z.J., Chen, S.Y., Wu, Y.Z., Han, B., et al. (2020). Epidemiologic relationship between periodontitis and type 2 diabetes mellitus. *BMC Oral Health* **20**, 204.

iScience, Volume 24

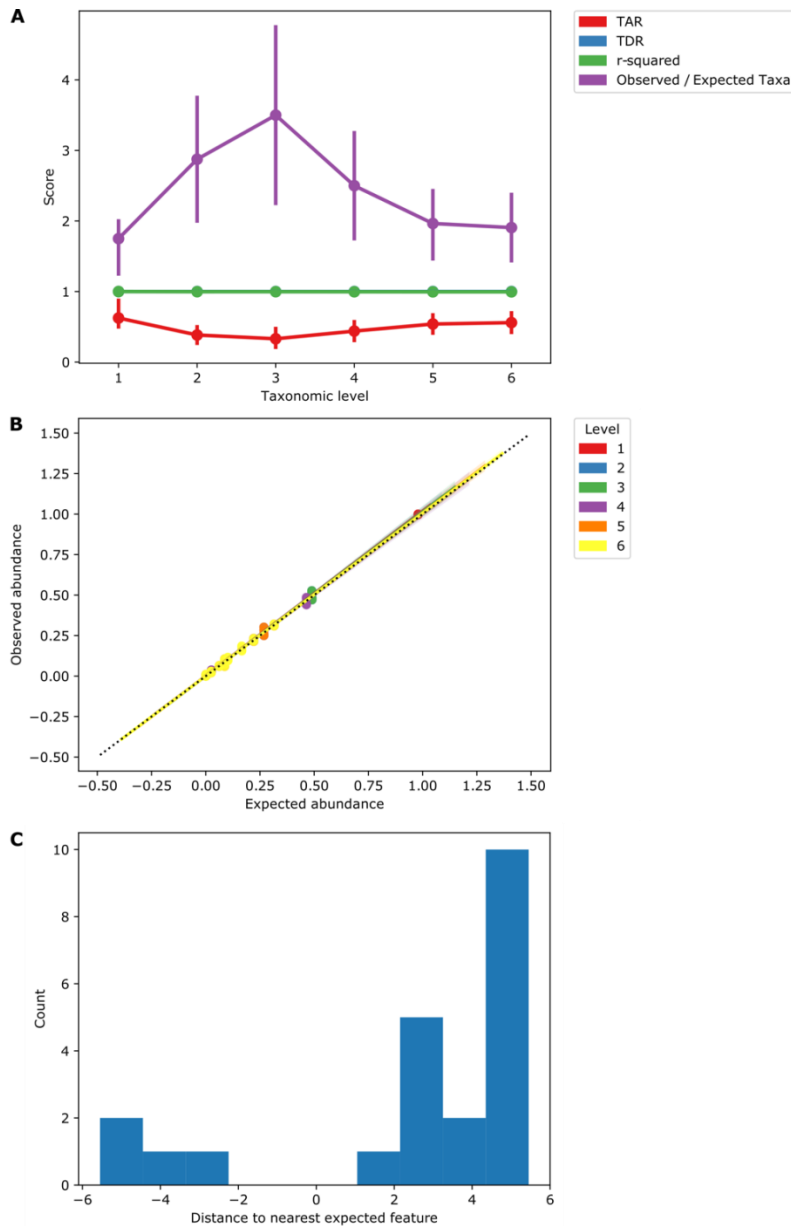
## Supplemental Information

### Impaired host response and the presence of *Acinetobacter baumannii* in the serum microbiome of type-II diabetic patients

Dasith Perera, Sarah E. Kleinstein, Benjamin Hanson, Hatice Hasturk, Ryan Eveloff, Marcelo Freire, and Matthew Ramsey

# 1 Supplemental Information

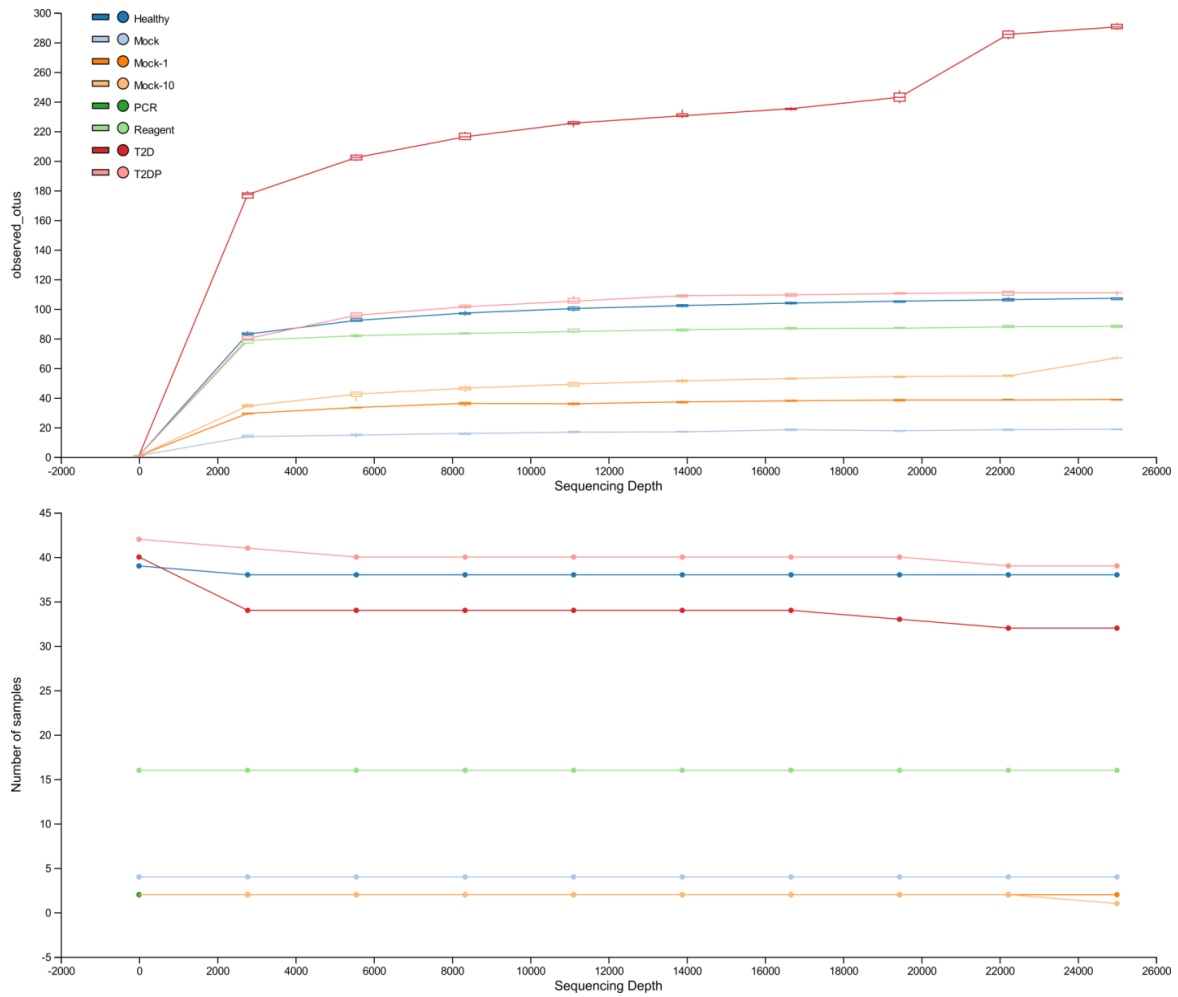
## 2 Supplemental Figures and Legends



3

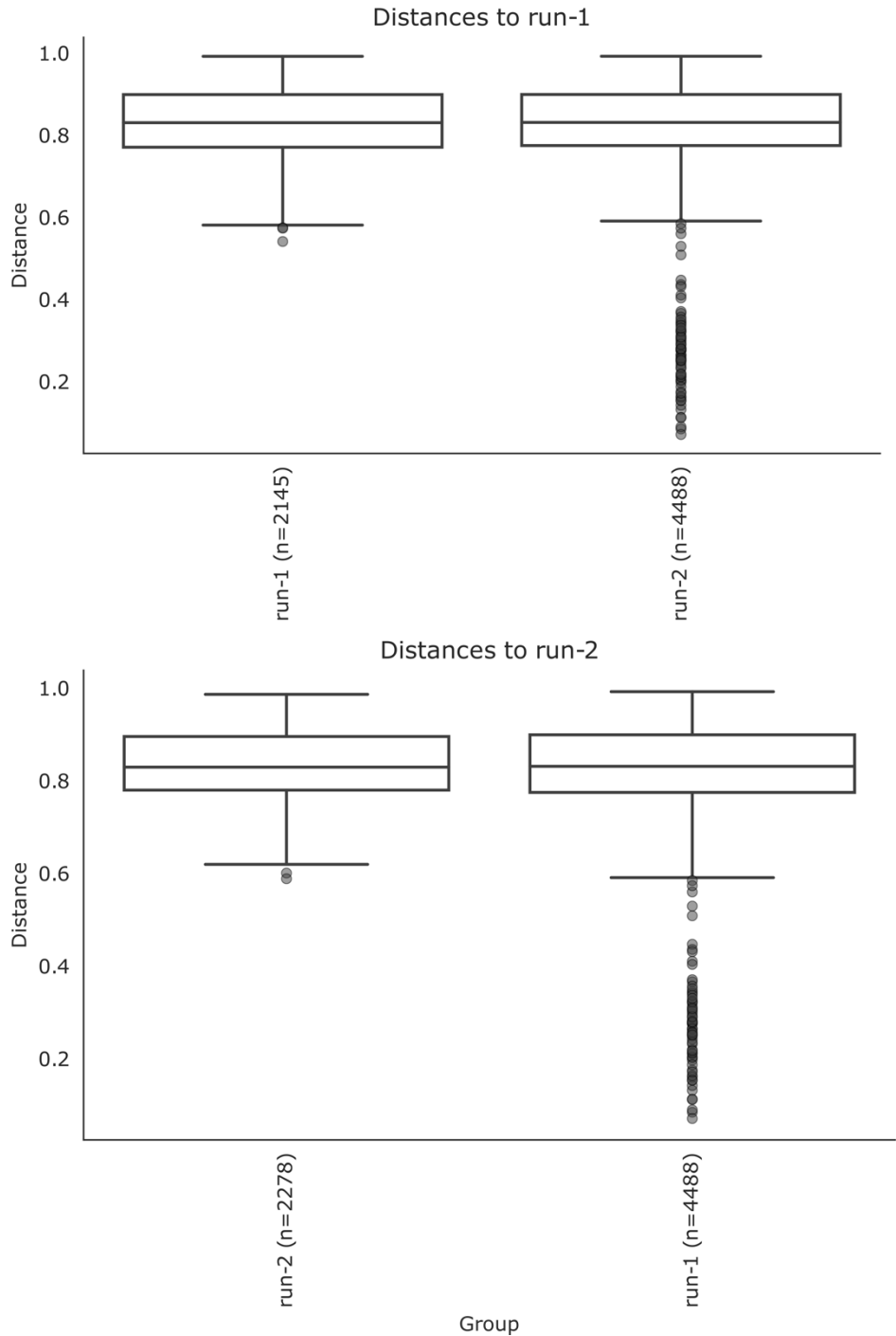
4 **Figure S1. Mock library composition evaluation.** Related to Figure 1. A Zymogen mock  
5 community library was used for sequence library construction at 10 and 1 ng in duplicate  
6 between both sequencing runs. QIIME2 / DADA2 unfiltered data was given taxonomical  
7 assignment via SILVA and assessed via the “q2-quality-control” plugin. (A) Taxon accuracy rate  
8 (TAR), taxon detection rate (TDR) and linear regression scores (r-squared) are plotted at each  
9 taxonomic level. (B) Expected vs observed abundance of each species in the mock community  
10 is plotted at each taxonomic level. (C) Distance between false positive observations vs nearest  
11 expected feature. Misclassifications were unique to environmental contaminant taxa and no  
12 false negative features were observed (not shown).

13



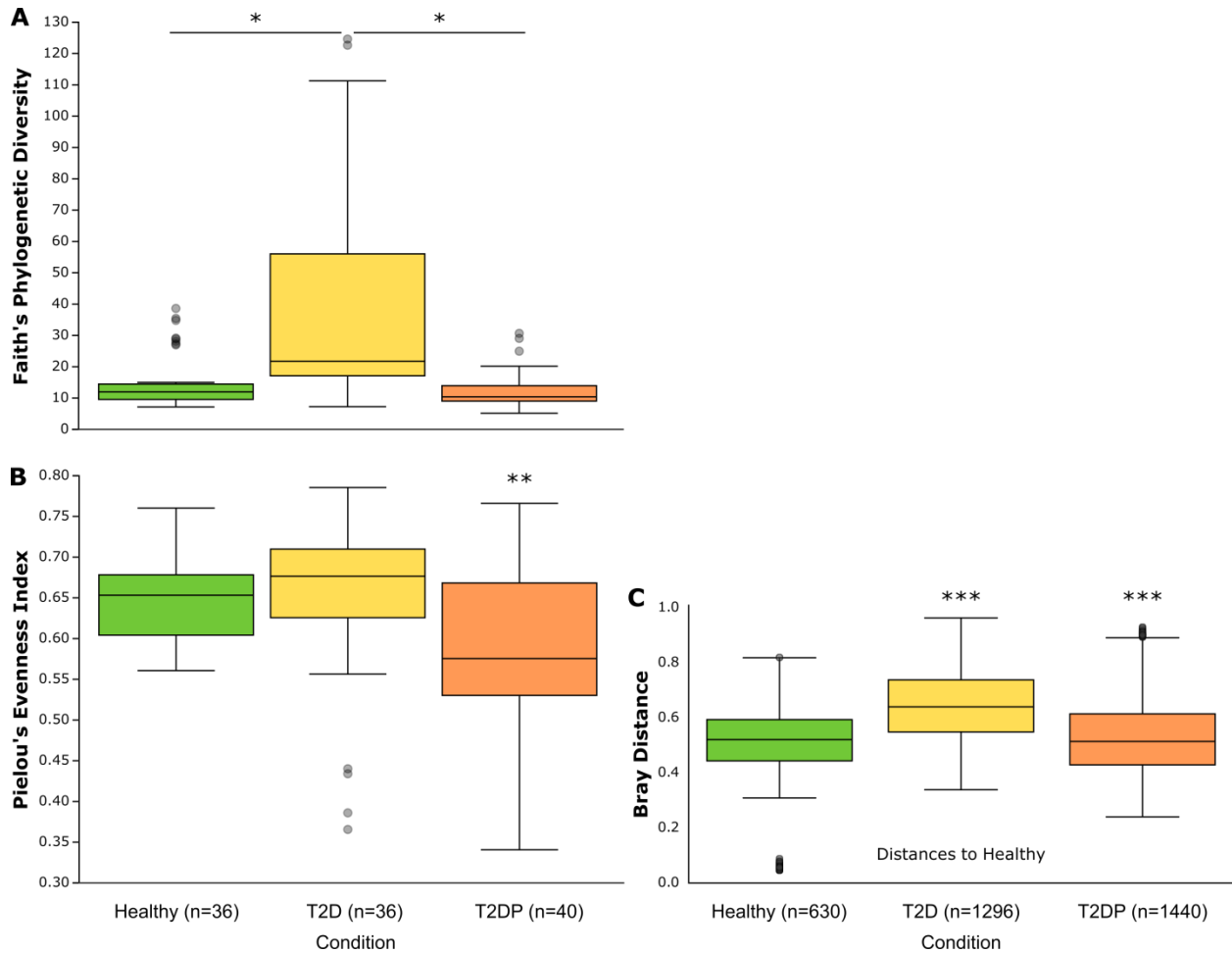
14

15 **Figure S2. Alpha-diversity at different rarefaction levels.** Related to Figure 1. Total observed  
 16 OTUs (here ASVs) are plotted (A) vs sequencing depth and each sample that can  
 17 accommodate that sequencing depth is plotted in (B).  
 18



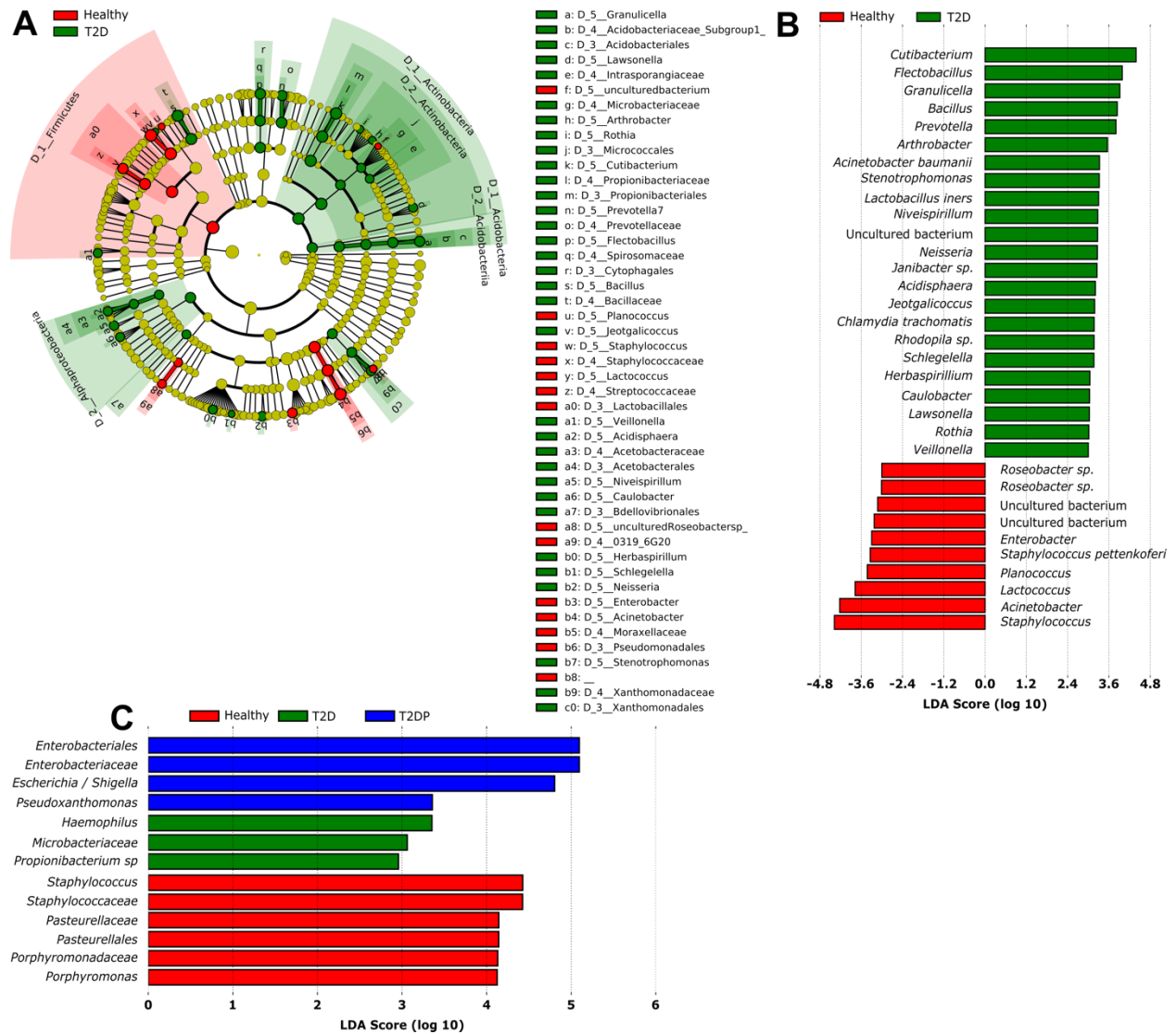
19

20 **Figure S3. Run to Run variability quantification.** Related to Figure 1. A Run vs Run Beta  
 21 diversity comparison was performed for unfiltered data by comparing unweighted Unifrac  
 22 distances as described in the methods. Pairwise PERMANOVA of Run-1 vs Run-2 (group size  
 23 of 2, sample size of n=134 total indexes) was performed in 999 permutations resulting in a  
 24 pseudo-F value of 0.567, p-value of 0.999 and corrected q-value of 0.999. No significant  
 25 differences were observed between each run.  
 26



27

28 **Figure S4. 16S Alpha and Beta diversity of serum microbiomes changes with subject**  
 29 **status.** Related to Figure 1. Species richness is displayed based on Faith's phylogenetic  
 30 diversity (Faith, 1992) (A) and species evenness based on Pielou's evenness index (Pielou,  
 31 1966) (B) for healthy, type-II diabetic (T2D) and type-II diabetics with periodontitis (T2DP)  
 32 samples. Each graph used minimally filtered data for analysis. Significance was determined by  
 33 Kruskal-Wallis analysis of variance (Kruskal and Wallis, 1952) for each comparison indicated  
 34 and Benjamini-Hochberg correction (Benjamini and Hochberg, 1995) was applied to generate  
 35 adjusted q-values. \* indicates q-value < 2e-3 vs healthy, \*\* indicates q-value < 0.024 vs  
 36 healthy, (C) Beta-diversity Bray Curtis distances. Pairwise PERMANOVA of each category vs  
 37 each (group size of 3, n=112) was performed in 999 permutations. \*\*\* indicates q-value differs  
 38 from healthy < 0.001.  
 39

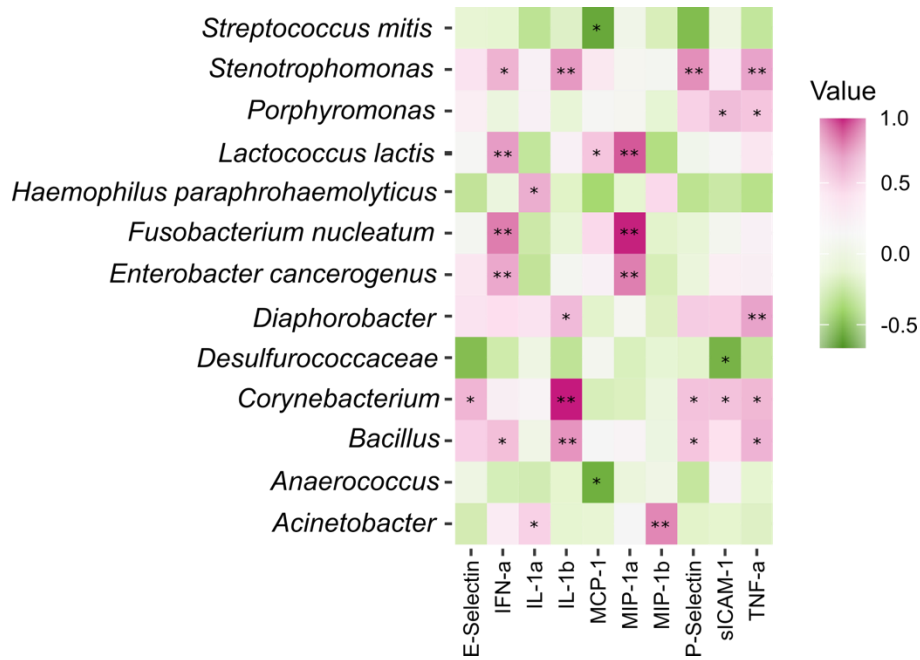


40

41 **Figure S5. LfSe Comparisons reveal taxonomical differences between subject groups.**

42 Related to Figure 3. (A) LfSe Cladogram plots (Segata *et al.*, 2011) reveal general taxonomic  
 43 shifts compared to (B) specifically enriched genera or species in all diabetics vs healthy and (C)  
 44 compared across all 3 subject groups. Data was generated from DADA2 / QIIME2 analyzed  
 45 data aligned to the SILVA database that was then filtered for contaminating taxa abundant in  
 46 no-template control indexes. *Pseudomonas* genus / species level results were also removed for  
 47 clarity.

48



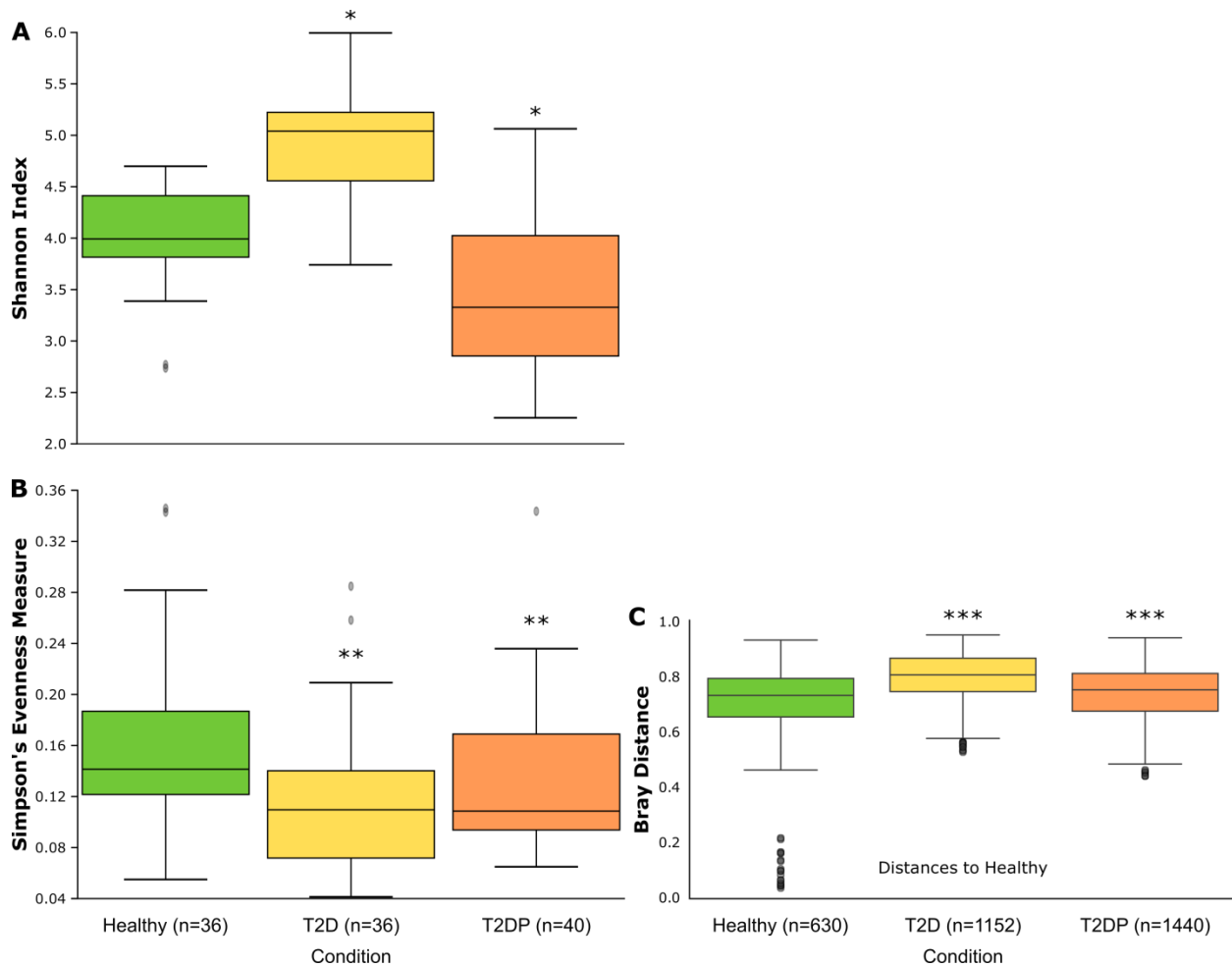
49

50 **Figure S6. Cytokine correlation with specific taxa in healthy subjects.** Related to Figure 7.  
 51 HOMD/RDP assigned taxonomy of MED analyzed 16S data for healthy samples and cytokine  
 52 concentrations were analyzed via Pearson correlation coefficient in R using the *rcorr* function.  
 53 Significance was determined using the asymptotic p-values generated by *rcorr* with \* = p-value  
 54 <0.05 \*\* = p-value <0.01. Data are strictly filtered with taxa present in no-template controls  
 55 subtracted.

56

57



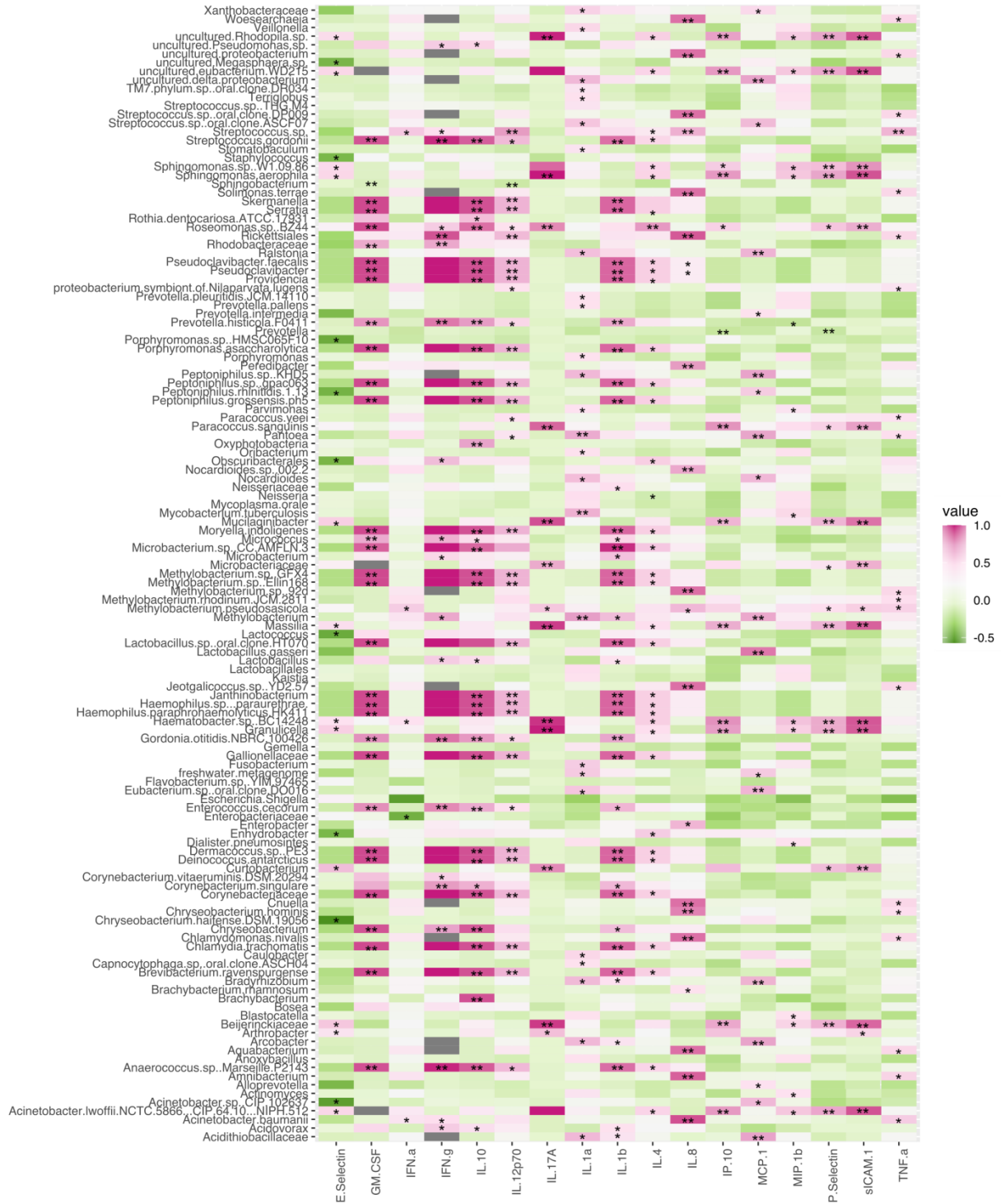


58

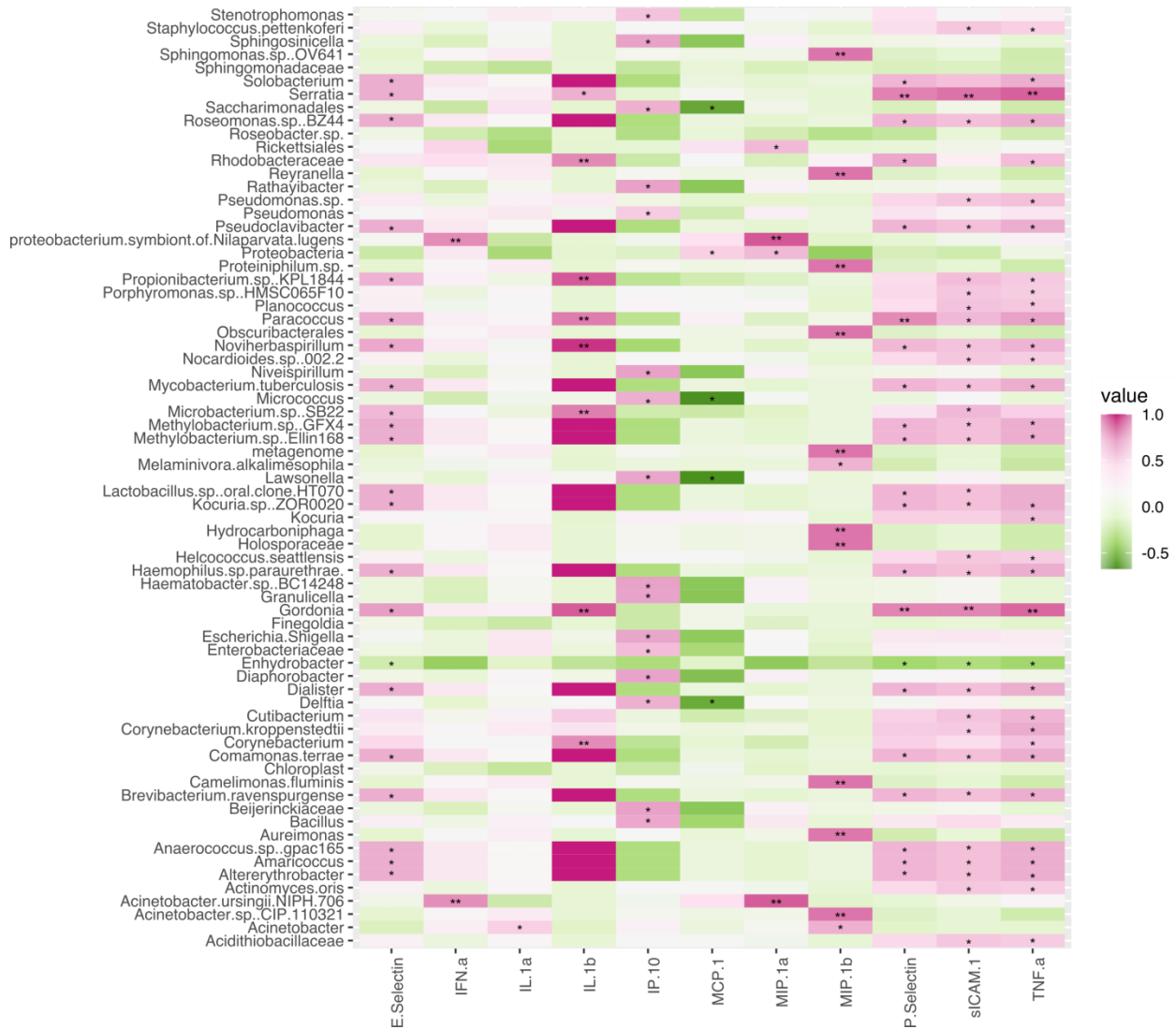
59 **Figure S7. 16S Alpha and Beta diversity of serum microbiomes changes with subject**  
 60 **status.** Related to Figure 1. Alpha diversity is displayed based on Shannon's index (Shannon  
 61 and Weaver, 1975) (A) and species evenness based on Simpson's evenness measure  
 62 (Simpson, 1949) (B) for healthy, type-II diabetic (T2D) and type-II diabetics with periodontitis  
 63 (T2DP) samples. Each graph used minimally filtered data for analysis. Significance was  
 64 determined by Kruskal-Wallis analysis of variance (Kruskal and Wallis, 1952) for each  
 65 comparison indicated and Benjamini-Hochberg correction (Benjamini and Hochberg, 1995) was  
 66 applied to generate adjusted q-values. \* indicates q-value < 1e-5, \*\* indicates q-value < 0.007,  
 67 (C) Beta-diversity unweighted Unifrac distances. Pairwise PERMANOVA of each category vs  
 68 each (group size of 3, n=112) was performed in 999 permutations. \*\*\* indicates q-value differs  
 69 from 'healthy' < 0.02.

70

71



72  
73 **Figure S8. Host-microbial correlation resolved type-II diabetic associations with**  
74 **inflammatory cytokines.** Related to Figure 7. SILVA assigned taxonomy of DADA2 analyzed  
75 16S data for both T2D and T2DP samples and cytokine concentrations were analyzed via  
76 Pearson correlation coefficient in R using the *rcorr* function. Significance was determined using  
77 the asymptotic p-values generated by *rcorr* with \* = p-value < 0.05 \*\* = p-value < 0.01. Data are  
78 strictly filtered with taxa present in no-template controls subtracted.



79  
80  
81  
82  
83  
84  
85  
86

**Figure S9. Cytokine correlation with specific taxa in healthy subjects.** Related to Figure 7. SILVA assigned taxonomy of DADA2 analyzed 16S data for healthy samples and cytokine concentrations were analyzed via Pearson correlation coefficient in R using the *rcorr* function. Significance was determined using the asymptotic p-values generated by *rcorr* with \* = p-value < 0.05 \*\* = p-value < 0.01. Data are strictly filtered with taxa present in no-template controls subtracted.

## 87 **Transparent Methods**

### 88 Subject recruitment, sampling and storage

89  
90 Subject recruitment has been described previously (Freire *et al.*, 2017). Peripheral venous blood  
91 (~60 ml) was collected from patients diagnosed with T2D and from healthy nondiabetic controls.  
92 Blood samples were collected and centrifuged at 2300 rpm, and serum was isolated and frozen  
93 at -80°C until analysis under IRB protocol #13-07. All subjects gave signed informed consent  
94 prior to study evaluations. Clinical periodontal data and peripheral venous blood were collected.  
95 The diagnosis of T2D was made by the subject's primary care physician following American  
96 Association of Diabetes guidelines (American Diabetes Association, 2015). Information was  
97 collected on subject demographics (age, gender, self-reported ethnicity, and self-reported  
98 smoking status), body-mass index (BMI; kg/m<sup>2</sup>), blood total cholesterol, blood glucose (point-of-  
99 care), percent hemoglobin A1C (HbA1c), and periodontal condition (Armitage, 1999). HbA1c  
100 was used to determine the level of glycemic control for diabetic subjects. One T2D individual  
101 lacked HbA1c measurements but fit based on all other diagnostic criteria (blood glucose >200  
102 mg/dl) as well as cytokine profiles in accord with other T2D individuals. Neutrophil and  
103 monocyte cell counts were determined by lab assay (described below). Individuals were  
104 excluded if they were taking insulin sensitizers, nonsteroidal anti-inflammatory drugs, or  
105 antimicrobials within 3 months of study initiation. Smoking status was defined by CDC NHIS  
106 terms  
107 ([https://www.cdc.gov/nchs/nhis/tobacco/tobacco\\_glossary.htm#:~:text=Former%20smoker%3A%20An%20adult%20who,in%20his%20or%20her%20lifetime](https://www.cdc.gov/nchs/nhis/tobacco/tobacco_glossary.htm#:~:text=Former%20smoker%3A%20An%20adult%20who,in%20his%20or%20her%20lifetime)). Site collection before blood  
108 draw: Samples were drawn in clinical phlebotomy settings. From all serum sampled for this  
109 microbiome study, a total of 81 subjects (N=24 healthy, N=57 T2D) were included for analysis,  
110 all of whom were unrelated and over 18 years of age (range: 28-79 years of age).

112

### 113 Sample preparation for 16S analysis

114

#### 115 *DNA Extraction*

116

117 Frozen serum from N=81 samples (N=24 healthy, N=32 T2D and N=25 T2DP) were thawed on  
118 ice and aliquots were separated for DNA extraction. DNA was extracted using the Epicentre  
119 MasterPure Complete DNA and RNA Purification kit (Lucigen, WI, USA). Extraction was carried  
120 out using the manufacturer's instructions, with modifications that enabled bead beating. Briefly,  
121 200µL of Tissue and Cell Lysis Solution (2x), 100µL of nuclease free water and 2µL of  
122 Proteinase K were added to Lysing matrix B (LMB) tubes (MP Biomedicals, Santa Ana, CA).  
123 Thereafter 100µL of serum was added to the tubes and placed in a Beadbeater (Biospec) for 30  
124 seconds, then placed on ice for 3 minutes and then repeated. Each round of DNA extractions  
125 included a no template control, which consisted of 100µL nuclease free water instead of serum.  
126 The samples were then incubated and extracted as described in the manufacturer's instructions.  
127 Total DNA was precipitated using the manufacturer's instructions, however 300 µL of MPC  
128 Protein Precipitation Reagent was used to accommodate the increased volume. DNA was  
129 eluted in 50µL of nuclease free water.

130

#### 131 16s rDNA primers 27F

132 (TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAGAGTTTGTATYMTGGCTCAG) and 519R  
133 (GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGWATTACCGCGGCKGCTG)

134 (Stackebrandt and Goodfellow, 1991; Turner *et al.*, 1999) were used to amplify the V1-V3  
135 regions of 16s rDNA in a 50 µL reaction using 2x Q5 HiFi mastermix (New England Biolabs,  
136 Ipswich, MA) and 23 µL of extracted DNA (35 cycles).

137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152

### *Illumina MiSeq Library preparation*

After the first round of PCR was carried out it was cleaned with Ampure XP beads (Beckman Coulter, Pasadena, CA) and then visualized by agarose gel electrophoresis. Full indices and adapters were added using the Illumina Nextera Index Kit (Illumina, San Diego, CA) by running the second round of PCR (50 ng of template DNA, 5 cycles) in 2x Phusion HF Master Mix (New England Biolabs, Ipswich, MA). PCR were then cleaned with Ampure XP beads and analyzed by agarose gel electrophoresis and using the Agilent BioAnalyzer DNA1000 chip (Agilent Technologies, Santa Clara, CA). Samples were then quantified and normalized prior to pooling using a Qubit fluorimeter (Invitrogen, Carlsbad, CA). The final pooled library was quantified with the KAPA Biosystems Illumina Kit (KAPA Biosystems, Woburn, MA) via qPCR in a Roche LightCycler480. Samples were sequenced (2x250 bp paired-end) on an Illumina MiSeq (Illumina, San Diego, CA) at the Rhode Island Genomics and Sequencing Center (Kingston, RI).

### *Mock Community Standards Preparation*

In order to quantify potential bias due to low template concentrations, PCR amplification, and contamination, we performed mock library assemblies with commercially available bacterial genomic DNA templates (Zymo Research #D6305). We amplified 2 libraries using 10 and 1 ng total starting templates in both sequencing runs for a total of 4 mock libraries synthesized. Total DNA species composition is provided in the manufacturer's instructions.

### 16S clone library resequencing

Primers were made to amplify regions of 16s rDNA that enable differentiation of the *Acinetobacter* species. Primers used were oMR328 (TAGCGGCGGACGGGTGAGTAATGCTTA) and oMR329 (TTCCGACTTCATGGAGTCGAGTTGCAGAC). 50 µL reactions in 2x Q5 mastermix (New England Biolabs, Ipswich, MA) were run for 35 cycles using 23 µL of DNA extracted from serum as the template. The products were then purified using the Qiagen PCR purification kit according to the manufacturer's instructions. Gotaq 2x mastermix (Promega, Madison, WI) was used to generate 'A' overhangs via PCR for 5 cycles using 10 µL of purified DNA as the template. 4µL of the resulting PCR products were then used to generate clones using the TOPO TA Cloning Kit (Invitrogen, Carlsbad, CA) according to the manufacturer's instructions which was modified to use 0.5µL of TOPO vector. Clones were then transformed into NEB5α cells using the manufacturer's instructions and plated on LB plates supplemented with 0.1mM IPTG and 40µg/mL Kanamycin. Additionally, 40µl of 40mg/mL of X-gal was spread onto each plate prior to plating to enable blue-white screening. The plates were incubated for 24 hours at 37°C. White colonies were inoculated into liquid cultures of LB supplemented with 40µg/mL kanamycin. Plasmids were extracted using QIAprep Spin Miniprep kit (Qiagen, Venlo, Netherlands). Inserts were sequenced using the M13 Forward (GTAAAACGACGGCCAGTG) and M13 Reverse (CACAGGAAACAGCTATGACC) primers on an Applied Biosystems 3500xl using the "BigDye" Terminator v3.1 Cycle sequencing kit. Sequences were analyzed using NCBI BLAST (NCBI Resource Coordinators, 2018). A maximum of 10 colonies were screened for each sample or until a sequence matched a species belonging to the *Acinetobacter baumannii-calcoaceticus* complex.

### Cytokine Quantification

185 Frozen (-80°C) subject serum was brought up to room temperature and assayed using the  
186 Invitrogen human inflammation 20-plex ProcartaPlex cytokine panel (Thermo Fisher Scientific,  
187 Waltham, MA) on a Luminex 200 instrument (Luminex, Austin, TX) in universal assay buffer.  
188 Assayed cytokines included: MIP-1 $\alpha$ , IL-1 $\beta$ , IL-4, IP-10, IL-6, IL-8, IL-10, IL-12p70, IL-13, IL-  
189 17A, IFN- $\gamma$ , GM-CSF, TNF- $\alpha$ , MIP-1 $\beta$ , IFN- $\alpha$ , MCP-1, P-Selectin, IL-1 $\alpha$ , sICAM-1, and E-  
190 Selectin. Following manufacturer protocols, all samples were run on a plate with 7 standards  
191 (diluted 1:4) and a control (universal assay buffer only), with all samples, standards, and  
192 controls run in duplicate similar to our methods previously (Pessoa *et al.*, 2019; Kleinstein *et al.*,  
193 2020) .

194  
195 Quality control (QC) steps were conducted according to manufacturer recommendations by  
196 xPONENT 4.2 software (Affymetrix eBioscience, San Diego, USA). Any standards with <70 or  
197 >130 % recovery of beads were invalidated. Samples were required to have a bead count of  
198 >30 beads recovered (all samples had >100 beads recovered and none were excluded at this  
199 step). Following QC, results were reported as average pg/mL for all measured cytokines. The  
200 lower limit of quantification (LLOQ) was determined based on the standard curve (after QC) as  
201 the average value of the lowest validated standards. Values at or below the LLOQ for each  
202 cytokine were reported at the LLOQ. As samples were run in two batches (with similar LLOQs  
203 for each batch), LLOQ for the cytokines (in pg/mL) were averaged across both runs: MIP-  
204 1 $\alpha$ =1.79, IL-1 $\beta$ =5.58, IL-4=23.99, IP-10= 1.17, IL-6=18.89, IL-8=2.48, IL-10=2.01, IL-  
205 12p70=11.72, IL-13=5.63, IL-17A=4.63, IFN- $\gamma$ =11.93, GM-CSF=14.57, TNF- $\alpha$ =9.66, MIP-  
206 1 $\beta$ =7.66, IFN- $\alpha$ =1.32, MCP-1=3.73, P-Selectin=1077.15, IL-1 $\alpha$ =0.74, sICAM-1=442.80, and E-  
207 Selectin=441.00. To compare cytokine levels between study groups, unpaired t-tests were used  
208 with a significance threshold of  $p < 0.05$  and no assumption of consistent standard deviation.  
209

#### 210 Minimum Entropy Decomposition analysis of sequence libraries

211  
212 Poor quality reads were discarded using flexbar (Dodt *et al.*, 2012). Reads were then treated as  
213 single end reads and trimmed for quality and to a length of 160bp using Trimmomatic (Bolger,  
214 Lohse and Usadel, 2014). Minimum Entropy Decomposition (MED) was performed as described  
215 in the oligotyping pipeline (Eren *et al.*, 2015), generating a list of node representative sequences  
216 and their relative abundances. Taxonomic assignment of representative sequences were  
217 generated based on aligning 16s rRNA sequences at 98.5% identity against the eHOMD  
218 database (Escapa *et al.*, 2018) and further evaluated using the Ribosomal Database Project  
219 (RDP) database (Cole *et al.*, 2014). MED nodes were assigned species level taxonomy based  
220 on >98.5% identity matches to eHOMD and/or RDP with eHOMD designations used as 1<sup>st</sup>  
221 priority as our initial aim was to identify oral taxa. Node frequency tables now assigned  
222 taxonomy were then used in association with sample metadata for Linear Discriminate Effect  
223 Size analysis (LEfSe) (Segata *et al.*, 2011).

#### 224 QIIME2 analysis of sequence libraries

225  
226 \*All Scripts at end of this document. All work was performed on the University of Rhode Island  
227 High Performance Computing Bluewaves cluster:

228 <https://web.uri.edu/hpc-research-computing/cluster-specifications/>

229 *Data Import*

230 Fastq sequence data were imported using the **import-fastq.sh** script. Generated demux.qzv  
231 files were viewed using <https://view.qiime2.org/> via the “quality viewer” function to determine  
232 location for sequence quality trimming used in the following step.

### 233 *Fastq Trimming, DADA2 Analysis, Run-run Merging*

234 A second sequencing run of the prepared Illumina libraries was performed to ensure maximum  
235 sequencing depth without further PCR amplification of starting template. At the end of the “D2-  
236 merge.sh” script we then used the **Rarefaction-alpha.sh** script output (Fig. S2) to determine  
237 read depth (22,000) to use in downstream commands.

### 238 *Initial Alpha and Beta Diversity Measurements and Run-run Comparison Testing*

239 Merged libraries were used in the **Alpha-Beta-An.sh** script to determine initial phylogenetics on  
240 unfiltered data using the QIIME2 *core-metrics-phylogenetic* function, Alpha and Beta diversity  
241 calculations via QIIME2 *alpha-* or *beta-group-significance* functions as well as Beta diversity  
242 difference calculation (*beta-group-significance*) between sequencing Run 1 vs. Run 2 (Figs. S3,  
243 S4).

### 244 *Mock Community Standards Quality Control Testing*

245 This section uses the **mocklibQC1.sh** and **mocklibQC2.sh** scripts. Taxonomy files used are  
246 described below. The \*-bar-plots.qzv file generated in the 1<sup>st</sup> script was viewed at  
247 <https://view.qiime2.org/> and the level 6 data from this plot was exported as a .csv file which  
248 provides the reads per each taxa. This data was converted to proportion of total reads per each  
249 taxa and converted into the Zymo\_actual.tsv file. The Zymo\_expected.tsv file was generated by  
250 editing the Zymo\_actual.tsv file with the proportion for each taxa present as provided in the  
251 manufacturers reference:  
252 [https://files.zymoresearch.com/protocols/ d6305 d6306 zymobiomics microbial community\\_d](https://files.zymoresearch.com/protocols/ d6305 d6306 zymobiomics microbial community_d na_standard.pdf)  
253 [na\\_standard.pdf](https://files.zymoresearch.com/protocols/ d6305 d6306 zymobiomics microbial community_d na_standard.pdf)

254 These .tsv inputs were then used in the **mocklibQC2.sh** script to output the visualizations seen  
255 in Figure S1.

### 256 *Classifier Setup and Editing*

257 Initial import of the SILVA 132 classifier was performed by running the following command:

```
258 wget https://data.qiime2.org/2019.7/common/silva-132-99-nb-classifier.qza
```

259 Classifier setup was then performed using the **classifier.sh** script. Upon searching the silva-  
260 taxonomy.qzv file generated, we noted that there were no reads that matched *A. baumannii*  
261 sequences at all. We used the rep-seqs.merge.qzv file output from the **D2merge.sh** script  
262 saved as a FASTA file from within the QIIME2 viewer function and performed a pairwise blast vs  
263 the full length *A. baumannii* 16S sequence from NCBI:

```
264 >A.baumannii 16s
```

```
265 AACGCTGGCGGCAGGCTTAACACATGCAAGTCGAGCGGGGAAGGTAGCTTGCTACCGG  
266 ACCTAGCGGCGGACGGGTGAGTAATGCTTAGGAATCTGCCTATTAGTGGGGGACAACATC  
267 TCGAAAGGGATGCTAATACCGCATACGTCCTACGGGAGAAAGCAGGGGATCTTCGGACCT  
268 TGCGCTAATAGATGAGCCTAAGTCGGATTAGCTAGTTGGTGGGGTAAAGGCCTACCAAGG  
269 CGACGATCTGTAGCGGGTCTGAGAGGATGATCCGCCACACTGGGACTGAGACACGGCCC
```

270 AGACTCCTACGGGAGGCAGCAGTGGGGAATATTGGACAATGGGGGGAACCCTGATCCAG  
271 CCATGCCGCGTGTGTGAAGAAGGCCTTATGGTTGTAAAGCACTTTAAGCGAGGAGGAGGC  
272 TACTTTAGTTAATACCTAGAGATAGTGGACGTTACTCGCAGAATAAGCACCGGCTAACTCT  
273 GTGCCAGCAGCCGCGGTAATACAGAGGGTGCAGCGTTAATCGGATTTACTGGGCGTAAA  
274 GCGTGCGTAGGCGGCTTATTAAGTCGGATGTGAAATCCCCGAGCTTAACTTGGGAATTGC  
275 ATTCGATACTGGTGAGCTAGAGTATGGGAGAGGATGGTAGAATTCCAGGTGTAGCGGTGA  
276 AATGCGTAGAGATCTGGAGGAATACCGATGGCGAAGGCAGCCATCTGGCCTAATACTGAC  
277 GCTGAGGTACGAAAGCATGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCATGCCGTA  
278 AACGATGTCTACTAGCCGTTGGGGCCTTTGAGGCTTTAGTGGCGCAGCTAACGCGATAAG  
279 TAGACCGCCTGGGGAGTACGGTTCGCAAGACTAAACTCAAATGAATTGACGGGGGCCCGC  
280 ACAAGCGGTGGAGCATGTGGTTTAATTCGATGCAACGCGAAGAACCCTTACCTGGCCTTGA  
281 CATACTAGAACTTTCCAGAGATGGATTGGTGCCTTCGGGAATCTAGATACAGGTGCTGCA  
282 TGGCTGTTCGTCAGCTCGTGTTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCT  
283 TTTCCCTTACTTGCCAGCATTTCCGGATGGGAACTTTAAGGATACTGCCAGTGACAACTGGA  
284 GGAAGGCGGGGACGACGTCAAGTCATCATGGCCCTTACGGCCAGGGCTACACACGTGCT  
285 ACAATGGTTCGGTACAAAGGGTTGCTACACAGCGATGTGATGCTAATCTCAAAAAGCCGATC  
286 GTAGTCCGGATTGGAGTCTGCAACTCGACTCCATGAAGTCGGAATCGCTAGTAATCGCGG  
287 ATCAGAATGCCGCGGTGAATACGTTCCCGGGCCTTGACACACCGCCCGTCACACCATGG  
288 GAGTTTGTTCACCAGAAGTAGCTAGCCTAACTGCAAAGAGGGCGGTTACCACGGTGTGG  
289 CCGATGACTGGGGTGAAGT

290 One ASV from the rep-seqs.merge.qzv file (e875355f3179838110485d8f5013d4a6) returned a  
291 100% match homology to the above *A. baumannii* sequence. This ASV was just annotated as  
292 *Acinetobacter* in the SILVA-132-99 taxonomy. We then edited the silva classifier taxonomy file  
293 by first exporting it using the following commands:

294 `module load QIIME2/2019.7`

295 `qiime tools export \`

296 `--input-path silva-taxonomy.qza \`

297 `--output-path silva-taxonomy`

298 This output the file in \*.tsv format which was then opened in a text editor and searched for the  
299 ASV node identity and then modified as indicated here:

300 **before**

301 e875355f3179838110485d8f5013d4a6  
302 D\_0\_\_Bacteria;D\_1\_\_Proteobacteria;D\_2\_\_Gammaproteobacteria;D\_3\_\_Pseudomonadales;D  
303 \_4\_\_Moraxellaceae;D\_5\_\_Acinetobacter 0.9999662516375396

304 **after**

305 e875355f3179838110485d8f5013d4a6  
306 D\_0\_\_Bacteria;D\_1\_\_Proteobacteria;D\_2\_\_Gammaproteobacteria;D\_3\_\_Pseudomonadales;D  
307 \_4\_\_Moraxellaceae;D\_5\_\_Acinetobacter;D\_6\_\_Acinetobacter baumannii  
308 0.9999662516375396

309 The edited file was then saved as taxonomy-1.tsv and imported via the following command:

310 `qiime tools import \`



```
311 --input-path /data/mramseylab/classifiers/silva-taxonomy/taxonomy-1.tsv \  
312 --type 'FeatureData[Taxonomy]' \  
313 --input-format TSVTaxonomyFormat \  
314 --output-path silva-mod-taxonomy.qza  
315 qiime metadata tabulate \  
316 --m-input-file silva-mod-taxonomy.qza \  
317 --o-visualization silva-mod-taxonomy.qzv
```

318 The silva-mod-taxonomy.qza file was used for the rest of our analyses.

### 319 *Data Filtering*

320 Data presented throughout the manuscript is generally described as “Minimally Filtered” or  
321 “Strictly Filtered”. We 1<sup>st</sup> began with all ASV assigned data from the above scripts and used  
322 metadata based filtering to separate out all data from control samples which included no-  
323 template and PCR only control indexes. This was done using the **control-filter.sh** script. Data  
324 output from the script was exported so it could be viewed and then used to subtract taxa from  
325 human-derived datasets in downstream filtering steps. Next we applied the **meta-filter.sh** script  
326 to our data to extract only human derived samples for further analysis and filtering. After this we  
327 applied the **minimal-filter.sh** script. Outputs from this script were used in all “Minimally Filtered”  
328 described data in the manuscript. Minimal filtering included removal of instances of taxa that  
329 appeared in only 2 samples or less, features (ASV sequences) that appeared less than 10 times  
330 across all samples and any taxa with less than 20 reads per sample.

331 Next we took our output files from the above **control-filter.sh** script and used them as part of  
332 the input for the **auto-filter.sh** script to exclude these taxa from the remaining minimally filtered  
333 data. While this removed many spurious taxa from our samples we observed high abundance of  
334 known aquatic microbial contaminant sequences not typically human associated (ex:  
335 *Sphingomonas*, *Ralstonia*). These and other taxa were manually excluded using the **strict-**  
336 **filter.sh** script and further removal of highly abundant *Pseudomonas* sequences were also  
337 removed via the **nopa-filter.sh** script. Data filtered to this extent are referred to as “Strictly”  
338 filtered in the manuscript.

### 339 *Re-analysis of Filtered Data*

340 A reanalysis of filtered data was performed 1<sup>st</sup> for Alpha and Beta diversity measurements via  
341 the **alpha-beta2.sh** script (Fig. 1). Further description and statistical analysis of Beta diversity  
342 differences were performed in R using the *vegan* package via the metaMDS and Adonis  
343 functions primarily as demonstrated in the **MMR20\_ellipses\_NMDS.R** script below.

### 344 *LEfSe Analysis Comparison*

345 Initial LEfSe analysis was performed using ASV output abundance data directly from the initial  
346 MED pipeline (Fig. 3). Further analysis on QIIME2 / DADA2 assigned strictly filtered data (Fig.  
347 S5) was performed first using the **lefse-noTax.sh** script to export and format data for analysis  
348 on the LEfSe Galaxy server (<https://huttenhower.sph.harvard.edu/galaxy/>). This script only  
349 outputs ASV node information without taxonomic assignment. Manual taxonomic assignment  
350 was performed comparing ASV node names to the rep-seqs-merge.qza file from the

351 **D2merge.sh** script. LEfSe analysis was performed using default settings, \*.svg output files were  
352 further edited in Inkscape software for clarity.

353

354 **Supplemental References**

- 355 American Diabetes Association (2015) '(2) Classification and diagnosis of diabetes', *Diabetes*  
356 *Care*, 38 Suppl, pp. S8–S16.
- 357 Armitage, G. C. (1999) 'Development of a classification system for periodontal diseases and  
358 conditions', *Annals of Periodontology*, 4(1), pp. 1–6.
- 359 Benjamini, Y. and Hochberg, Y. (1995) 'Controlling the False Discovery Rate: A Practical and  
360 Powerful Approach to Multiple Testing', *Journal of the Royal Statistical Society: Series B*  
361 *(Methodological)*, 57(1), pp. 289–300.
- 362 Bolger, A. M., Lohse, M. and Usadel, B. (2014) 'Trimmomatic: a flexible trimmer for Illumina  
363 sequence data', *Bioinformatics (Oxford, England)*, 30(15), pp. 2114–2120.
- 364 Cole, J. R. *et al.* (2014) 'Ribosomal Database Project: data and tools for high throughput rRNA  
365 analysis', *Nucleic Acids Research*, 42(Database issue), pp. D633–642.
- 366 Dodt, M. *et al.* (2012) 'FLEXBAR-Flexible Barcode and Adapter Processing for Next-Generation  
367 Sequencing Platforms', *Biology*, 1(3), pp. 895–905.
- 368 Eren, A. M. *et al.* (2015) 'Minimum entropy decomposition: unsupervised oligotyping for  
369 sensitive partitioning of high-throughput marker gene sequences', *The ISME journal*, 9(4), pp.  
370 968–979.
- 371 Escapa, I. F. *et al.* (2018) 'New Insights into Human Nostril Microbiome from the Expanded  
372 Human Oral Microbiome Database (eHOMD): a Resource for the Microbiome of the Human  
373 Aerodigestive Tract', *mSystems*. Edited by J. Xu, 3(6), pp. e00187-18,  
374 /msystems/3/6/msys.00187-18.atom.
- 375 Faith, D. P. (1992) 'Conservation evaluation and phylogenetic diversity', *Biological*  
376 *Conservation*, 61(1), pp. 1–10.
- 377 Freire, M. O. *et al.* (2017) 'Neutrophil Resolvin E1 Receptor Expression and Function in Type 2  
378 Diabetes', *Journal of Immunology (Baltimore, Md.: 1950)*, 198(2), pp. 718–728.
- 379 Kleinstein, S. E. *et al.* (2020) *Transcriptomics of Type 2 Diabetic and Healthy Human*  
380 *Neutrophils*. preprint. In Review.
- 381 Kruskal, W. H. and Wallis, W. A. (1952) 'Use of Ranks in One-Criterion Variance Analysis',  
382 *Journal of the American Statistical Association*, 47(260), pp. 583–621.
- 383 NCBI Resource Coordinators (2018) 'Database resources of the National Center for  
384 Biotechnology Information', *Nucleic Acids Research*, 46(D1), pp. D8–D13.
- 385 Pessoa, L. *et al.* (2019) 'Host-Microbial Interactions in Systemic Lupus Erythematosus and  
386 Periodontitis', *Frontiers in Immunology*, 10, p. 2602.
- 387 Pielou, E. C. (1966) 'The measurement of diversity in different types of biological collections',  
388 *Journal of Theoretical Biology*, 13, pp. 131–144.

- 389 Segata, N. *et al.* (2011) 'Metagenomic biomarker discovery and explanation', *Genome Biology*,  
390 12(6), p. R60.
- 391 Shannon, C. E. and Weaver, W. (1975) *The mathematical theory of communication*. Urbana:  
392 University of Illinois Press.
- 393 Simpson, E. H. (1949) 'Measurement of Diversity', *Nature*, 163(4148), pp. 688–688.
- 394 Stackebrandt, E. and Goodfellow, M. (eds) (1991) *Nucleic acid techniques in bacterial*  
395 *systematics*. Chichester ; New York: Wiley (Modern microbiological methods).
- 396 Turner, S. *et al.* (1999) 'Investigating deep phylogenetic relationships among cyanobacteria and  
397 plastids by small subunit rRNA sequence analysis', *The Journal of Eukaryotic Microbiology*,  
398 46(4), pp. 327–338
- 399

400 **Supplemental Data (Data S1.)** Related to Transparent Methods

401 Scripts used for analyses:

402

403 **import-fastq.sh**

404 module load QIIME2/2019.7

405

406 for d in /data/mramseylab/raw\_reads/2018\_Serum/Run\* ;

407 do

408 SUFX=\${d#\*Run}

409

410 qiime tools import \

411 --type 'SampleData[PairedEndSequencesWithQuality]' \

412 --input-path /data/mramseylab/raw\_reads/2018\_Serum/Run\$SUFX \

413 --input-format CasavaOneEightSingleLanePerSampleDirFmt \

414 --output-path /data/mramseylab/raw\_reads/2018\_Serum/Run\$SUFX/demux-

415 \$SUFX.qza

416

417 qiime demux summarize \

418 --i-data /data/mramseylab/raw\_reads/2018\_Serum/Run\$SUFX/demux-

419 \$SUFX.qza \

420 --o-visualization /data/mramseylab/visualizations/demux-\$SUFX.qzv

421

422 done

423

```
424 D2merge.sh
425
426 module load QIIME2/2019.7
427
428 for d in /data/mramseylab/raw_reads/2018_Serum/Run* ;
429 do
430   SUFFIX=${d#*Run}
431
432   qiime dada2 denoise-single \
433     --i-demultiplexed-seqs
434     /data/mramseylab/raw_reads/2018_Serum/Run${SUFFIX}/demux-${SUFFIX}.qza \
435     --p-trim-left 13 \
436     --p-trunc-len 250 \
437     --o-table /data/mramseylab/raw_reads/2018_Serum/Run${SUFFIX}/denoise-
438     table-${SUFFIX}.qza \
439     --o-representative-sequences
440     /data/mramseylab/raw_reads/2018_Serum/Run${SUFFIX}/rep-seqs-${SUFFIX}.qza \
441     --o-denoising-stats
442     /data/mramseylab/raw_reads/2018_Serum/Run${SUFFIX}/denoising-stats-
443     ${SUFFIX}.qza
444
445 done
446 #Have to MERGE Run1 and Run2 to get pairwise comparison
447 #Must make directory before running below else it will error out.
448
449 mkdir /data/mramseylab/raw_reads/2018_Serum/Merge_runs/
450
451 qiime feature-table merge \
452   --i-tables /data/mramseylab/raw_reads/2018_Serum/Run1/denoise-table-
453   1.qza \
454   --i-tables /data/mramseylab/raw_reads/2018_Serum/Run2/denoise-table-
455   2.qza \
```

```
456     --o-merged-table
457 /data/mramseylab/raw_reads/2018_Serum/Merge_runs/denoise-table-
458 merge.qza
459
460 qiime feature-table merge-seqs \
461     --i-data /data/mramseylab/raw_reads/2018_Serum/Run1/rep-seqs-1.qza \
462     --i-data /data/mramseylab/raw_reads/2018_Serum/Run2/rep-seqs-2.qza \
463     --o-merged-data
464 /data/mramseylab/raw_reads/2018_Serum/Merge_runs/rep-seqs-merge.qza
465
466 qiime feature-table summarize \
467     --i-table /data/mramseylab/raw_reads/2018_Serum/Merge_runs/denoise-
468 table-merge.qza \
469     --o-visualization /data/mramseylab/visualizations/denoise-table-
470 merge.qzv \
471     --m-sample-metadata-file /data/mramseylab/metadata/Serum5.tsv
472
473 qiime feature-table tabulate-seqs \
474     --i-data /data/mramseylab/raw_reads/2018_Serum/Merge_runs/rep-seqs-
475 merge.qza \
476     --o-visualization /data/mramseylab/visualizations/rep-seqs-merge.qzv
477
478 qiime phylogeny align-to-tree-mafft-fasttree \
479     --i-sequences /data/mramseylab/raw_reads/2018_Serum/Merge_runs/rep-
480 seqs-merge.qza \
481     --o-alignment
482 /data/mramseylab/raw_reads/2018_Serum/Merge_runs/aligned-rep-seqs-
483 merge.qza \
484     --o-masked-alignment
485 /data/mramseylab/raw_reads/2018_Serum/Merge_runs/masked-aligned-rep-
486 seqs-merge.qza \
487     --o-tree /data/mramseylab/raw_reads/2018_Serum/Merge_runs/unrooted-
488 tree-merge.qza \
489     --o-rooted-tree
490 /data/mramseylab/raw_reads/2018_Serum/Merge_runs/rooted-tree-merge.qza
```

```
491 Rarefaction-alpha.sh
492 module load QIIME2/2019.7
493
494 rawdir=/data/mramseylab/raw_reads/2018_Serum/Merge_runs/
495 procdir=/data/mramseylab/proc_reads/
496 metadir=/data/mramseylab/metadata/
497 visdir=/data/mramseylab/visualizations/
498
499 qiime diversity alpha-rarefaction \
500   --i-table $rawdir\denoise-table-merge.qza \
501   --i-phylogeny $rawdir\rooted-tree-merge.qza \
502   --p-max-depth 25000 \
503   --m-metadata-file $metadir\Serum4.tsv \
504   --o-visualization $visdir\alpha-rarefaction.qzv
505
506
507
508
509
510
```



```
511 Alpha-Beta-An.sh
512 module load QIIME2/2019.7
513
514 rawdir=/data/mramseylab/raw_reads/2018_Serum/Merge_runs/
515 procdir=/data/mramseylab/proc_reads/
516 metadir=/data/mramseylab/metadata/
517 cmr="core-metrics-results"
518
519 qiime diversity core-metrics-phylogenetic \
520   --i-phylogeny $rawdir\rooted-tree-merge.qza \
521   --i-table $rawdir\denoise-table-merge.qza \
522   --p-sampling-depth 22000 \
523   --m-metadata-file $metadir\Serum5.tsv \
524   --output-dir $procdir$cmr\-initial
525
526 qiime diversity alpha-group-significance \
527   --i-alpha-diversity $procdir$cmr\-initial/faith_pd_vector.qza \
528   --m-metadata-file $metadir\Serum5.tsv \
529   --o-visualization $procdir$cmr\-initial/faith-pd-group-
530   significance.qzv
531
532 qiime diversity alpha-group-significance \
533   --i-alpha-diversity $procdir$cmr\-initial/evenness_vector.qza \
534   --m-metadata-file $metadir\Serum5.tsv \
535   --o-visualization $procdir$cmr\-initial/evenness-group-
536   significance.qzv
537
538 array=( unweighted_unifrac_distance_matrix
539   weighted_unifrac_distance_matrix bray_curtis_distance_matrix )
540
```

```
541 for i in "${array[@]}"
542 do
543
544 qiime diversity beta-group-significance \
545   --i-distance-matrix $procdir$cmr\-$initial/$i.qza \
546   --m-metadata-file $metadir\Serum5.tsv \
547   --m-metadata-column Condition \
548   --o-visualization $procdir$cmr\-$initial/$i.qzv \
549   --p-pairwise
550
551 done
552
553 #Generates data for run1 vs run2 variability to measure batch effect
554 qiime diversity beta-group-significance \
555   --i-distance-matrix $procdir$cmr\-$
556 initial/unweighted_unifrac_distance_matrix.qza \
557   --m-metadata-file $metadir\Serum4.tsv \
558   --m-metadata-column Run \
559   --o-visualization $procdir$cmr\-$initial/unweighted-unifrac-Run-
560 significance.qzv \
561   --p-pairwise
562
```

```
563 Classifier.sh
564 module load QIIME2/2019.7
565
566 rawdir=/data/mramseylab/raw_reads/2018_Serum/Merge_runs/
567 clsdir=/data/mramseylab/classifiers/
568
569 #for the silva release 132 99 .fna file
570
571 qiime feature-classifier classify-sklearn \
572   --i-classifier $clsdir/silva-132-99-nb-classifier.qza \
573   --i-reads $rawdir/rep-seqs-merge.qza \
574   --o-classification $clsdir/silva-taxonomy.qza
575
576 #output the taxonomy table to check for A. baumannii strains in it
577 afterwards.
578
579 qiime metadata tabulate \
580   --m-input-file $clsdir/silva-taxonomy.qza \
581   --o-visualization $clsdir/silva-taxonomy.qzv
582
```

```
583 Control-filter.sh
584 module load QIIME2/2019.7
585
586 tablein=/data/mramseylab/raw_reads/2018_Serum/Merge_runs/denoise-
587 table-merge.qza
588 clsdir=/data/mramseylab/classifiers/
589 metadir=/data/mramseylab/metadata/
590 visdir=/data/mramseylab/visualizations/
591 filtdir=/data/mramseylab/proc_reads/
592 # filter status of input files, "ctrl-filter" is just for taxa
593 belonging to no template controls
594 fil=ctrl-filter
595
596
597 #must make the directory you are filtering to lst or else it will
598 error
599 mkdir $filtdir$fil
600
601
602 qiime feature-table filter-samples \
603   --i-table $tablein \
604   --m-metadata-file $metadir\Serum5.tsv \
605   --p-where "[Source]='Control'" \
606   --o-filtered-table $filtdir$fil/$fil-table.qza
607
608 qiime taxa collapse \
609   --i-table $filtdir$fil/$fil-table.qza \
610   --i-taxonomy $clsdir\silva-mod-taxonomy.qza \
611   --p-level 6 \
612   --o-collapsed-table $filtdir$fil/$fil-collapse-table.qza
```

```
613
614 qiime feature-table relative-frequency \
615   --i-table $filtdir$fil/$fil-collapse-table.qza \
616   --o-relative-frequency-table $filtdir$fil/$fil-relative-collapse-
617   table.qza
618
619 qiime tools export \
620   --input-path $filtdir$fil/$fil-relative-collapse-table.qza \
621   --output-path $filtdir$fil/
622
623 biom convert \
624   -i $filtdir$fil/feature-table.biom \
625   -o $filtdir$fil/$fil-relative-collapse-table.txt \
626   --header-key "taxonomy" \
627   --to-tsv
628
629 #Use above taxa table to filter out based on taxa present in controls
630
631 qiime feature-table filter-features \
632   --i-table $tablein \
633   --m-metadata-file $filtdir$fil/$fil-collapse-table.qza \
634   --o-filtered-table $filtdir$fil/$fil-excluded-table.qza \
635   --p-exclude-ids
636
637 #Use excluded table to generate barplot for checking
638
639 qiime taxa barplot \
640   --i-table $filtdir$fil/$fil-excluded-table.qza \
641   --i-taxonomy $clsdir\silva-mod-taxonomy.qza \
```

```
642 --m-metadata-file $metadir\Serum5.tsv \  
643 --o-visualization $filtdir$fil/$fil-excluded-table.qzv  
644
```

```
645 Meta-filter.sh
646 module load QIIME2/2019.7
647
648 tablein=/data/mramseylab/raw_reads/2018_Serum/Merge_runs/denoise-
649 table-merge.qza
650 clsdir=/data/mramseylab/classifiers/
651 metadir=/data/mramseylab/metadata/
652 visdir=/data/mramseylab/visualizations/
653 filtdir=/data/mramseylab/proc_reads/
654 # filter status of input files, "initial" is the 1st pass no filter at
655 all
656 # "initial-human" is the initial data but for only human samples, no
657 mock or control samples
658 fil=initial-human
659
660 #note must make directoros needed BEFORE running the below command
661 mkdir $filtdir$fil
662
663 qiime feature-table filter-samples \
664   --i-table $tablein \
665   --m-metadata-file $metadir\Serum5.tsv \
666   --p-where "[Source]='Human'" \
667   --o-filtered-table $filtdir$fil/$fil-table.qza
668
```

```
669 Auto-filter.sh
670 module load QIIME2/2019.7
671
672 tablein=/data/mramseylab/proc_reads/minF-hum/table3.qza
673 clsdir=/data/mramseylab/classifiers/
674 metadir=/data/mramseylab/metadata/
675 visdir=/data/mramseylab/visualizations/
676 filtdir=/data/mramseylab/proc_reads/
677
678 # filter status of input files, "ctrl-filter" is just for taxa
679 belonging to no template controls
680 fil=AF-hum
681 #must make the directory you are filtering to lst or else it will
682 error
683 mkdir $filtdir$fil
684
685 #Filtering list from mock samples was generated previously with the
686 control-filter.sh script
687
688 qiime feature-table filter-features \
689   --i-table $tablein \
690   --m-metadata-file /data/mramseylab/proc_reads/ctrl-filter/ctrl-
691   filter-collapse-table.qza \
692   --o-filtered-table $filtdir$fil/$fil-excluded-table.qza \
693   --p-exclude-ids
694
695 #Use excluded table to generate barplot for checking
696 qiime taxa barplot \
697   --i-table $filtdir$fil/$fil-excluded-table.qza \
698   --i-taxonomy $clsdir\silva-mod-taxonomy2.qza \
699   --m-metadata-file $metadir\Serum5.tsv \
```



```
700     --o-visualization $filtdir$fil/$fil-excluded-table.qzv
701
702 #Generate viewable feature table to look for ASVs of interest /
703 filtering stats
704 qiime feature-table summarize \
705     --i-table $filtdir$fil/$fil-excluded-table.qza \
706     --o-visualization $filtdir$fil/$fil-excluded-table2.qzv \
707     --m-sample-metadata-file $metadir\Serum5.tsv
708
```

```
709 Strict-filter.sh
710 module load QIIME2/2019.7
711
712 tablein=/data/mramseylab/proc_reads/AF-hum/AF-hum-excluded-table
713 clsdir=/data/mramseylab/classifiers/
714 metadir=/data/mramseylab/metadata/
715 visdir=/data/mramseylab/visualizations/
716 filtdir=/data/mramseylab/proc_reads/
717
718 # filter status of input files, "ctrl-filter" is just for taxa
719 belonging to no template controls
720 fil=SF-hum
721 #must make the directory you are filtering to lst or else it will
722 error
723 mkdir $filtdir$fil
724
725 #Filtering list from mock samples was generated previously with the
726 auto-filter.sh script
727
728 #Tidying up name here so I can repeatedly use the same input below
729 through the loop
730 cp $tablein.qza /$filtdir$fil/$fil-table.qza
731 #Array here contains all taxa I wish to remove
732 array=( Chloroplast Mitochondria Ralstonia Chryseobacterium
733 Sphingomonas Enhydrobacter Bradyrhizobium Sphingomonadales Rhizobiales
734 Rhodobacterales Sphingobacteriales Halomonadaceae Deinococcales )
735
736 for i in "${array[@]}"
737 do
738
739 qiime taxa filter-table \
```

```
740 --i-table /$filtdir$fil/$fil-table.qza \  
741 --i-taxonomy $clsdir\silva-mod-taxonomy2.qza \  
742 --p-mode contains \  
743 --p-exclude "$i" \  
744 --o-filtered-table /$filtdir$fil/$fil-table.qza  
745  
746 done  
747 #Use excluded table to generate barplot for checking  
748 qiime taxa barplot \  
749 --i-table /$filtdir$fil/$fil-table.qza \  
750 --i-taxonomy $clsdir\silva-mod-taxonomy2.qza \  
751 --m-metadata-file $metadir\Serum5.tsv \  
752 --o-visualization /$filtdir$fil/$fil-table-barplot.qzv  
753  
754 #Generate viewable feature table to look for ASVs of interest /  
755 filtering stats  
756 qiime feature-table summarize \  
757 --i-table /$filtdir$fil/$fil-table.qza \  
758 --o-visualization /$filtdir$fil/$fil-table.qzv \  
759 --m-sample-metadata-file $metadir\Serum5.tsv  
760
```

```
761 Nopa-filter.sh
762 module load QIIME2/2019.7
763
764 tablein=/data/mramseylab/proc_reads/SF-hum/SF-hum-table.qza
765 clsdir=/data/mramseylab/classifiers/
766 metadir=/data/mramseylab/metadata/
767 visdir=/data/mramseylab/visualizations/
768 filtdir=/data/mramseylab/proc_reads/
769
770 # filter status of input files, "ctrl-filter" is just for taxa
771 belonging to no template controls
772 fil=NP-hum
773 #must make the directory you are filtering to lst or else it will
774 error
775 mkdir $filtdir$fil
776
777 #Filtering list from mock samples was generated previously with the
778 strict-filter.sh script
779
780 qiime taxa filter-table \
781   --i-table $tablein \
782   --i-taxonomy $clsdir\silva-mod-taxonomy2.qza \
783   --p-mode contains \
784   --p-exclude "Pseudomonas" \
785   --o-filtered-table $filtdir$fil/$fil-table.qza
786
787 #Use excluded table to generate barplot for checking
788 qiime taxa barplot \
789   --i-table $filtdir$fil/$fil-table.qza \
790   --i-taxonomy $clsdir\silva-mod-taxonomy2.qza \
```

```
791 --m-metadata-file $metadir\Serum5.tsv \  
792 --o-visualization $filtdir$fil/$fil-table-barplot.qzv  
793  
794 #Generate viewable feature table to look for ASVs of interest /  
795 filtering stats  
796 qiime feature-table summarize \  
797 --i-table $filtdir$fil/$fil-table.qza \  
798 --o-visualization $filtdir$fil/$fil-table.qzv \  
799 --m-sample-metadata-file $metadir\Serum5.tsv  
800
```

```
801 Alpha-beta2.sh
802 module load QIIME2/2019.7
803
804 rawdir=/data/mramseylab/raw_reads/2018_Serum/Merge_runs/
805 procdir=/data/mramseylab/proc_reads/
806 metadir=/data/mramseylab/metadata/
807 visdir=/data/mramseylab/visualizations/
808 cmr="core-metrics-results"
809 # filter status of input files, "initial" is the 1st pass no filter at
810 all
811 fil="--initial"
812
813 #note change name of input tables for the 1st command below. Some
814 input tables did not have standardized filename conventions
815 #changed to AF-hum-table.qza and table3.qza to minF-hum-table.qza
816
817 array=( minF AF SF NP )
818
819 for i in "${array[@]}"
820 do
821
822 qiime diversity core-metrics-phylogenetic \
823   --i-phylogeny $rawdir\rooted-tree-merge.qza \
824   --i-table $procdir$i\hum/$i\hum-table.qza \
825   --p-sampling-depth 5000 \
826   --m-metadata-file $metadir\Serum5.tsv \
827   --output-dir $procdir$i\hum/$cmr
828
829 qiime diversity alpha-group-significance \
830   --i-alpha-diversity $procdir$i\hum/$cmr/faith_pd_vector.qza \
```

```
831 --m-metadata-file $metadir\Serum5.tsv \  
832 --o-visualization $procdire{i}\-hum/$cmr/faith-pd-group-  
833 significance.qzv  
834  
835 qiime diversity alpha-group-significance \  
836 --i-alpha-diversity $procdire{i}\-hum/$cmr/evenness_vector.qza \  
837 --m-metadata-file $metadir\Serum5.tsv \  
838 --o-visualization $procdire{i}\-hum/$cmr/evenness-group-  
839 significance.qzv  
840  
841 done  
842  
843 #dual array taking the directories above and then running the next  
844 command on the 3 filenames in array2 for each directory prefix in  
845 array 1  
846 #note different array command syntax from above vs below  
847  
848 array1=( minF AF SF NP )  
849 array2=( unweighted_unifrac_distance_matrix  
850 weighted_unifrac_distance_matrix bray_curtis_distance_matrix )  
851 for indirs in ${array1[@]}  
852 do  
853     for infils in ${array2[@]}  
854     do  
855  
856 qiime diversity beta-group-significance \  
857 --i-distance-matrix $procdire{indirs}\-hum/$cmr/{infils}.qza \  
858 --m-metadata-file $metadir\Serum5.tsv \  
859 --m-metadata-column Condition \  
860 --o-visualization $procdire{indirs}\-hum/$cmr/{infils}.qzv \  
861 --p-pairwise
```

862 done

863 done

864



```

865 MMR20_ellipses_NMDS.R
866 setwd("E:/2019_Diabetic_Serum/Figure Materials/R-Plots/")
867 library(vegan)
868 library(ggplot2)
869 library(dplyr)
870 set.seed(30)
871
872 metadata<-read.csv("Serum5.tsv", header=TRUE, sep="\t",
873 stringsAsFactors = F)
874 featuredf<-read.csv("NP-hum-features.tsv", header=TRUE, sep="\t",
875 stringsAsFactors = F)
876
877 #making a subset of human only metadata and a few other criteria.
878 metadata <- metadata %>% filter(Source == 'Human') %>%
879 select("sample.id", "Condition", "Run", "Source", "ABC")
880
881 #Fixing 1st row to make it row names
882 featuredf <- data.frame(featuredf[,-1], row.names = featuredf[,1])
883
884 #transposing trial1 table to match Evelyns data format
885 featuredf <- as.data.frame(t(featuredf))
886
887 #Brings tables into agreement on matching sample.id values
888 featuredf = featuredf %>% mutate(sample.id = rownames(featuredf))
889
890 #Brings tables into agreement on matching sample.id values
891 table_all = left_join(featuredf, metadata)
892
893 #This gets metadata to agree with row numbers of the features table
894 after splitting out from table_all

```

```

895 metadata <-
896 select(table_all,"sample.id","Condition","Run","Source","ABC")
897
898 #filter the "sample.id" column off the very end of the dataframe and
899 turn it into rownames
900 #NOTE THIS VALUE MUST BE CHANGED FOR EACH INPUT
901 featuredf <- data.frame(featuredf[,-1101], row.names =
902 featuredf[,1101])
903
904 # normalize data by sum of ASVs in each sample (this was from ZP's
905 code)
906 featuredf <- sweep(featuredf,2,colSums(featuredf),`/\`)
907
908 ##Above this was all data input / manipulation, Below this is data
909 analysis and plotting##
910
911 ### The function metaMDS is used to calculate the dissimilarity matrix
912 using the bray curtis distance metrics and at the same
913 #time generates the values from the dissimilarity matrix for an
914 ordination plot.
915 MDS <-metaMDS(featuredf ,distance = "bray", k = 3, trymax = 500)
916
917 ## Next, extract the x and y coordinates from the MDS plot into a new
918 data frame and add the metadata factors to the coordinates the data
919 should be plotted based on.
920 #MMR- added ABC=as.factor and changed [,3] to [,5] to agree with my
921 own metadata file.
922 NMDS_t1=data.frame(NMDS1=MDS$point[,1],NMDS2=MDS$point[,2],
923                   Condition=table_all$Condition,ABC=table_all$ABC)
924
925 ## set theme for following plots
926 theme_set(theme_bw())
927

```

```

928 ## Generate the ordination based on the solution from above and
929 selected grouping factor (Condition)

930 plot.new()

931 ord<-ordiellipse(MDS, table_all$Condition,
932                  display = "sites", kind ="sd", conf = 0.95, label
933 = T)

934 dev.off()

935

936 ## Data frame df_ell_t1 contains values to show ellipses. It is
937 calculated with function veganCovEllipse which is hidden in vegan
938 package.

939 #This function is applied to each level of NMDS (group) and it uses
940 also function cov.wt to calculate covariance matrix.

941 veganCovEllipse<-function (cov, center = c(0, 0), scale = 1, npoints =
942 100)
943 {
944   theta <- (0:npoints) * 2 * pi/npoints
945   Circle <- cbind(cos(theta), sin(theta))
946   t(center + scale * t(Circle %*% chol(cov)))
947 }

948

949 ##Generate ellipse points based on 95% confidence (SD) intervals

950 #Reference : http://stackoverflow.com/questions/13794419/plotting-ordiellipse-function-from-vegan-package-onto-nmgs-plot-created-in-ggplot
951
952

953 #Data frame df_ell contains values to show ellipses. It is calculated
954 with function veganCovEllipse which is hidden in vegan package. This
955 function is applied to each level of NMDS (group) and it uses also
956 function cov.wt to calculate covariance matrix.

957 df_ell_t1 <- data.frame()

958 for(g in levels(NMDS_t1$Condition)){
959   if(g!="" && (g %in% names(ord))){
960     df_ell_t1 <- rbind(df_ell_t1,
961 cbind(as.data.frame(with(NMDS_t1[NMDS_t1$Condition==g,],

```

```

962
963 veganCovEllipse(ord[[g]]$cov, ord[[g]]$center, ord[[g]]$scale)), Condi
964 on=g)}}
```

```

965
966 ## Calculate p-value:
967 adon_t1<-adonis2(featuredf ~Condition, data=metadata,
968 by=NULL, method="bray", k=3)
969
970 NMDSplot_t1<-ggplot(data=NMDS_t1, aes(NMDS1, NMDS2, col=Condition))+
971     #update from Evelyn to add metadata text to plot if wanted
972     #This is useful to identify outlier points if using sample ID text
973     here, Remove for final plot
974     ###
975     geom_text(aes(NMDS1, NMDS2, label=table_all$sample.id), size=2, vjust=0)+
976     # add the p-value in the bottom right corner
977     annotate("text", x=min(NMDS_t1$NMDS1), y=min(NMDS_t1$NMDS2-0.5),
978             label=paste("p= ", adon_t1$`Pr(>F)`[1]), size=3)+
979     # draw the ellipses and define color based on the grouping factor
980     geom_path(data=df_ell_t1, aes(x=NMDS1, y=NMDS2, linetype=Condition),
981 size=1)+
982     #scale_linetype_manual(values=c("4-Con"="dotted", "3-S4"="solid", "2-
983 RI-old"="longdash"))+
984     #scale_colour_manual(values=c("4-Con"="red", "3-S4"="darkgreen", "2-
985 RI-old"="purple"))+
986     # add the points per sample and define shape based on TankLocation
987     geom_point(aes(shape=ABC), size=2) +
988     # Reorder the legend
989     guides(color = guide_legend(order=1), lty= guide_legend(order=1),
990            shape = guide_legend(order=2), legend.position = "bottom")
991
992 # Adding other Aesthetics
```

```
993 NMDS_by_Trial1 <- NMDSplot_t1 + theme(axis.text.y =
994 element_text(size="12", color="black"), axis.title.y =
995 element_text(face="bold",size="12", color="black"))+
996   theme(axis.text.x = element_text(size="12", color="black"),
997 axis.title.x.bottom = element_text(face="bold",size="12",
998 color="black"))+
999   theme(axis.text.x.top= element_text(face="bold",size="12",
1000 color="black")) +
1001   #change name of the legend
1002   theme(legend.title=element_blank())+
1003   theme(legend.position = "right", legend.title =
1004 element_text(colour="black", size=16, face="bold"))
1005
1006
1007 print(NMDS_by_Trial1)
1008 print(NMDSplot_t1)
1009
```

```
1010 Lefse-noTax.sh
1011 module load QIIME2/2019.7
1012
1013 clsdir=/data/mramseylab/classifiers/
1014 metadir=/data/mramseylab/metadata/
1015 visdir=/data/mramseylab/visualizations/
1016 filtdir=/data/mramseylab/proc_reads/
1017 rawdir=/data/mramseylab/raw_reads/2018_Serum/Merge_runs/
1018 procdir=/data/mramseylab/proc_reads/
1019
1020 array=( minF-hum AF-hum SF-hum NP-hum )
1021
1022 for i in "${array[@]}"
1023 do
1024
1025 qiime feature-table relative-frequency \
1026 --i-table $filtdir$i/$i\table.qza \
1027 --o-relative-frequency-table $filtdir$i/$i\table.notax.qza
1028
1029 qiime tools export \
1030 --input-path $filtdir$i/$i\table.notax.qza \
1031 --output-path $filtdir$i/
1032
1033 #note, must use single hashes for -i / -o unlike other qiime commands.
1034 biom convert \
1035 -i $filtdir$i/feature-table.biom \
1036 -o $filtdir$i/$i\notax.table.txt \
1037 --header-key "taxonomy" \
1038 --to-tsv
```

1039

1040 done

1041

1042