2021-08-17

# Inflammation-type dysbiosis of the oral microbiome associates with the duration of COVID-19 symptoms and long-COVID

John P. Haran
*University of Massachusetts Medical School*

*Et al.*

JCI insight

# Inflammation-type dysbiosis of the oral microbiome associates with the duration of COVID-19 symptoms and long-COVID

John P. Haran, … , Beth A. McCormick, Vanni Bucci

*JCI Insight.* 2021. https://doi.org/10.1172/jci.insight.152346.

Research In-Press Preview COVID-19 Inflammation

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) caused the pandemic Coronavirus Disease 2019 (COVID-19) and now many face the burden of prolonged symptoms—long-lasting COVID-19 symptoms or "long-COVID". Long-COVID is thought to be linked to immune dysregulation due to harmful inflammation, with the exact causes being unknown. Given the role of the microbiome in mediating inflammation, we aimed to examine the relationship between the oral microbiome and the duration of long-COVID symptoms. Tongue swabs were collected from patients presenting with symptoms concerning for COVID-19. Confirmed infections were followed until resolution of all symptoms. Bacterial composition was determined by metagenomic sequencing. We used random forest modeling to identify microbiota and clinical covariates that associated with long-COVID symptoms. Of the patients followed, 63% (17/27) developed ongoing symptomatic COVID-19 and 37% (10/27) went on to long-COVID. Patients with prolonged symptoms had significantly higher abundances of microbiota that induce inflammation, such as members of the genera *Prevotella* and *Veillonella*. Of note are species that produce lipopolysaccharides and the similarity of long-COVID patients' oral microbiome to those of patients with chronic fatigue syndrome. All together, we our findings suggest an association with the oral microbiome and long-COVID revealing the possibility that dysfunction of the oral microbiome may contribute to this draining disease.

# Inflammation-Type Dysbiosis of the Oral Microbiome Associates with the Duration of COVID-19 symptoms and Long-COVID

**Authors:**

John P. Haran, MD, PhD[1,2,3*§], Evan Bradley MD, PhD[1,3*], Abigail L. Zeamer BS[2,3], Lindsey Cincotta BS[1], Marie-Claire Salive BA[1], Protiva Dutta BS[1], Shafik Mutaawe BS[1], Otuwe Anya BS[1], Mario Meza-Segura PhD[2], Ann Moormann PhD, MPH[4], Doyle V. Ward PhD[2,3], Beth A. McCormick PhD[2,3] and Vanni Bucci PhD[2,3§]

**Affiliations:**

[1]Department of Emergency Medicine, University of Massachusetts Medical School, 55 Lake Avenue North, Worcester, MA 01655

[2]Department of Microbiology and Physiological Systems, University of Massachusetts Medical School, 55 Lake Avenue North, Worcester, MA 01655

[3]Program in Microbiome Dynamics, University of Massachusetts Medical School, 55 Lake Avenue North, Worcester, MA 01655

[4]Department of Medicine, University of Massachusetts Medical School, 55 Lake Avenue North, Worcester, MA 01655

* These authors contributed equally to the manuscript and are joint first authors.

§ Joint senior authors.

**Corresponding Author:** John P. Haran, MD, Ph.D., 55 Lake Avenue North, Worcester, MA 01655; Phone: 508-450-8688; email: john.haran@umassmed.edu; fax: 508-421-1490

**Conflict of Interest Statement:** The authors have declared that no conflict of interest exists.

**ABSTRACT**

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) caused the pandemic Coronavirus Disease 2019 (COVID-19) and now many face the burden of prolonged symptoms—long-lasting COVID-19 symptoms or “long-COVID”. Long-COVID is thought to be linked to immune dysregulation due to harmful inflammation, with the exact causes being unknown. Given the role of the microbiome in mediating inflammation, we aimed to examine the relationship between the oral microbiome and the duration of long-COVID symptoms. Tongue swabs were collected from patients presenting with symptoms concerning for COVID-19. Confirmed infections were followed until resolution of all symptoms. Bacterial composition was determined by metagenomic sequencing. We used random forest modeling to identify microbiota and clinical covariates that associated with long-COVID symptoms. Of the patients followed, 63% (17/27) developed ongoing symptomatic COVID-19 and 37% (10/27) went on to long-COVID. Patients with prolonged symptoms had significantly higher abundances of microbiota that induce inflammation, such as members of the genera *Prevotella* and *Veillonella*. Of note are species that produce lipopolysaccharides and the similarity of long-COVID patients’ oral microbiome to those of patients with chronic fatigue syndrome. All together, we our findings suggest an association with the oral microbiome and long-COVID revealing the possibility that dysfunction of the oral microbiome may contribute to this draining disease.

**Key Words:** Oral Microbiome, COVID-19, SARS-CoV-2, symptom duration.

## INTRODUCTION

The oral cavity holds the second largest microbial community in the human body, after the gut, with over 1,000 species of commensal bacteria residing in the oral cavity (1). Dysbiosis or disrupted homeostasis caused by an imbalance in the microflora in the oral cavity has been linked to many other systemic inflammatory or infectious diseases (2). There is mounting evidence that links oral bacterial species to systemic diseases including pneumonia (1, 3, 4). Bacteria in the oral cavity may promote respiratory infections either directly via aspiration or indirectly by enzyme production that may hinder pathogen clearance, promote lung colonization or alter respiratory epithelial immune responses (5).

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is responsible for the current coronavirus disease 2019 (COVID-19) pandemic. This pandemic began in early 2020 and has seen over half a million deaths in the US alone (6). Building upon the body of evidence that the microbiome plays a role in the regulation of innate and adaptive immunity to viral infections (7, 8) studies done early in the pandemic have demonstrated a connection with an altered gut microbiome and the severity of COVID-19 disease (9, 10). Additionally, among COVID-19 patients there has been a large number of coinfection cases with organisms that originate from the oral cavity (11). Recently, decreased oral microbiome diversity and increased dysbiotic species abundances have been identified as predictive of COVID-19 disease (12). This has raised the possibility of using the oral microbiome to diagnose SARS-CoV-2 infection, however studies linking the observed dysbiotic oral microbiota to disease outcomes have been lacking. Also lacking is evidence that this COVID-related microbiome, which occurs early in the disease process, is predictive of key outcomes such as symptom duration.

Most hospitalized patients have persistent long-lasting symptoms that can take weeks to resolve (13) and negatively impact health-related quality of life (14). Symptoms persisting greater than 4 weeks after an acute infection are called ongoing symptomatic COVID-19, as characterized by The British National Institute for Health and Care Excellence (NICE) (15). Symptoms lasting even longer, 8-12 weeks or greater (16) and characterized by symptoms of fatigue, headache, dyspnea, and anosmia (17, 18), are termed long-lasting COVID-19 symptoms (long-COVID). Long-COVID does not currently have a strict definition (19). At the 10-week mark after SARS-CoV-2 infection, more than 50% of long COVID patients suffer profound fatigue (20). Increasing age, body mass index, and female gender are known to associate with long-COVID (16). It is currently unknown why most people recover fully within two to three weeks and others experience symptoms for weeks or months longer (21). There is evidence, however, of persistently perturbed inflammatory pathways long after the acute SARS-CoV-2 infection has subsided (22).

Given the emerging associations between the human microbiome and SARS-CoV-2 infection and the unknown driver for COVID-19 patients suffering from long lasting symptoms, we sought to explore if oral microbiome dysbiosis associates with ongoing symptoms among post-hospitalized COVID-19 patients. Accordingly, we enrolled a cohort of SARS-CoV-2 PCR positive COVID-19 patients from one US Emergency Department, collecting oral swabs early in the disease course, and followed them for 4- and 10-week symptom resolution outcomes. We analyzed oral microbiome composition by shotgun metagenomic sequencing. Our findings uniquely describe how dysbiosis of the oral microbiome may play a pivotal role in lengthening symptom duration leading to the long-COVID syndrome.

## RESULTS

### Patient Population

From a prospective sampling of 164 patients presenting with COVID symptoms over a 9-month period, 84 (51.2%) tested positive by PCR for SARS-CoV-2. Of these patients 27 were successfully contacted for follow-up at both 4 and 10 weeks (**Figure 1**). Average age was 62.6 (sd 12.5) with 70.4% men, 66.7% white, 7.4% African American and 25.9% Hispanic. Among the cohort for high-risk medical comorbidities 16 (59.3%) had hypertension, 8 (29.6%) diabetes, and 5 (18.5%) chronic obstructive pulmonary disease. Neither of these medical comorbidities nor the patients' Charlson Comorbidity Index

(CCI) scores differed by symptom duration outcome (**Table 1**). None of these patients lived in the same household. All these patients were admitted to the hospital with 4 (14.8%) admitted to the ICU. The average hospital length of stay was 8.3 (sd 7.7) days with 85.2% requiring oxygen and 25.9% getting advanced oxygen delivery by high flow or positive airway pressure. Two patients were intubated with an endotracheal tube.

**Symptom Duration**

The average length of symptom duration was 45.8 days (sd 30.4) with 14 patients (51.9%) experiencing continuation of symptoms after 4 weeks from disease onset, and 10 patients (37.0%) experiencing symptoms longer than 10 weeks. The symptoms that lasted the longest were respiratory in nature (81.5% cough or short of breath) followed by fatigue (55.6%), gastrointestinal symptoms (14.8%), confusion or "brain fog" (22.2%) and ageusia or anosmia (14.8%). Brain fog is a symptom more recently linked to long-COVID characterized by lack of clear memory or ability to focus (23, 24). There were no significant differences in demographics, medical history, or hospital treatments among the 2 outcomes categories (**Table 1**). However, among patients with symptoms lasting longer than 10 weeks, fatigue and brain fog were the most prominent symptoms that lasted the longest duration.

**Oral Microbiome Composition Predicts Ongoing Symptomatic COVID-19**

We set out to explore the associations of oral microbiome composition with the symptoms of ongoing symptomatic COVID-19 disease. To do this we profiled the oral microbiome of subjects with acute COVID-19 infection using shotgun metagenomic sequencing (See Methods). Microbial species abundances were determined by running Metaphlan3 (25). We estimated microbiome alpha diversity by calculating Shannon diversity index (26). We started by applying unsupervised learning methods, such as Principle Coordinate analysis (PCoA) and t-Distributed Stochastic Neighbor Embedding (t-SNE) and, as expected, found that interindividual variability overwhelmingly accounted for the majority of the information in the data (Figure S1). PERMANOVA analysis on samples classified according to COVID-

19 symptoms duration was not statistically significant (p-value <0.05). We then applied random-forest classification (RFC) (27, 28) to identify microbiome and clinical features associated with ongoing disease. Feature selection was performed using the Boruta algorithm on five-fold cross-validated data and then running RFC using the union of the Boruta selected features on the same five-fold cross-validated data to estimate model performance (29). We compared classification accuracy for different models that were trained (i) only on demographics + clinical data, (ii) only on microbiome species abundances, (iii) only on Shannon Diversity, (iv) on demographics + clinical data + Shannon Diversity, (v) on demographics + clinical data + microbiome species + Shannon Diversity and, (vi) on clinical data + microbiome species + Shannon Diversity (Figure 2A). Each model was run starting from 10 different random seeds to calculate appropriate performance statistics. The mean F1 score, the harmonic mean of precision and recall, was used to select the top performing model for a given outcome. The best model—clinical data + microbiome species + Shannon Diversity—performed with a mean F1 score of 0.751 (**Figure 2A**).

Specific microbial members had the greatest contribution to correctly classifying samples. We detected both bacterial and eukaryotic organisms in the oral microbiome analysis with only bacteria demonstrating associations with the outcomes. We examined the 19 bacterial species whose abundances were associated with ongoing symptomatic COVID-19 disease and two clinical covariates based on their median RFC-estimated permutated importance score over the 10 RFC pipeline iterations (**Figure 2B, C**). The model finds both viral load and Shannon Diversity to be of moderate importance, while specific microbiome members contributed most to correct sample prediction. In particular, two of the three top predictors (*Veillonella dispar* and *Veillonella infantium*) as well as 2 other species associated with ongoing symptomatic COVID-19 disease belong the genus *Veillonella*. Members of this genus are gram-negative anaerobic coccus that can cause infection in humans(30). Specifically, *V. infantium* has been found in the bronchoalveolar lavage fluid of the COVID-19 patients suggesting it is a significant co-infectious agent (31). Other pathobionts (organisms that can co-exist or cause disease under certain circumstances) such as *Solobacterium moorei* (32, 33), *Streptococcus infantis*(34), and *Rothia dentocariosa* (35) were in higher abundances in ongoing symptomatic COVID-19 disease patients. Interestingly, *S. infantis* has been found to be enriched in fecal samples from COVID-19 patients(9) and *R. dentocariosa* was predictive of SARS-CoV-2 presence in hospital rooms (36).

In addition to being implicated in co-infection, the *Veillonella* species are also known to produce a large amount of lipopolysaccharides (LPS) (37). Another pattern from this data that emerges is the higher abundances of other LPS-producing species are predictive of ongoing symptomatic COVID-19 disease. Five members of the *Prevotella* genus are positively associated with ongoing symptomatic disease in our analysis. *Prevotella* exhibits increased inflammatory properties (38) and has been thought to be a clinically important pathobiont involved in promoting chronic inflammation (39, 40). Other pro-inflammatory species such as *Leptotrichia wadei* (12) also are in higher abundances in patients with a longer symptom duration.

**Dysbiotic Inflammatory Type Oral Microbiome Associates with the Development of Long-COVID-19 Syndrome**

We repeated our machine learning-based analysis described above to predict long-COVID outcome from microbial abundance and clinical covariates. RFC was not able to capture any signal in the data for models that lacked microbiome information (i.e. i, iii, and iv in **Figure 2A**). The top performing RFC for long-COVID was the one trained on data on clinical data + microbiome species, resulting in an F1 score on 0.615 (**Figure 3A**). From the modeling we identified 29 different bacterial species whose abundances were associated with long-COVID (**Figure 3B**). Similar to ongoing symptomatic COVID, multiple *Veillonella* species were associated with long-COVID. Several of the top predicting species (4 out of 29) belong to the genus *Actinomyces*. *Actinomyces* cause actinomycosis, a rare infectious disease in which bacteria can spread to the respiratory tract causing inflammation (41). As with ongoing symptomatic COVID-19, multiple *Prevotella* species (38) are associated with long-COVID. *Prevotella* species are overrepresented in COVID-19 patients and are thought to produce proteins that can promote SARS-CoV-

2 infection and increase clinical severity of COVID-19 disease (42). Additional species known to cause infections such as *Streptococcus anginosus* group bacteria that have been reported to be particularly important in the pathogenesis of respiratory infections (43) and *Gemella sanguinis,* which has been shown to cause bloodstream infections in COVID-19 patients (44) were also found to be associated with long-COVID.

**Inflammatory Metabolic Pathways Associate with Ongoing Symptomatic and Long-COVID Disease States**

Building upon the taxonomy analysis, we explored the metabolic pathways and their association with ongoing symptomatic and long-COVID disease states using HUMAnN3(45). For each outcome we again performed RFC analysis and compared classification accuracy for different trained models: (i) demographics + clinical data + relative pathway abundances and (ii) only relative pathway abundances. For both ongoing symptomatic COVID and long-COVID, the top performing model was (ii), producing an F1 score of 0.814 and 0.689, respectively (**Figure 4A, 5A**). We identified >40 metabolic gene pathways whose abundances were associated with both ongoing symptomatic and long-COVID-19 disease (**Figure 4B, 5B**). The top 15 predictors indicate a striking pro-inflammatory pattern.

For ongoing symptomatic COVID, there are 5 pathways involved in the biosynthesis of branched amino acids that are reduced in patients with longer symptoms (**Figure 4B, C**). These include the superpathway of L-isoleucine I (MetaCyc PWY-3001), L-isoleucine biosynthesis III (PWY-5103), superpathway of branched amino acids (BRANCHED-CHAIN-AA-SYN-PWY), L-valine (VALSYN-PWY), and L-isoleucine (ILEUSYN-PWY) biosynthesis pathways(46) (**Figure 4C**). Branched amino acid have been shown to act as anti-inflammatory agents (47, 48) with orally administered L-isoleucine and L-leucine exhibiting anti-inflammatory activities (49). Four out of 15 of the top pathways involve synthesis of molecules with anti-inflammatory effects and are lower in ongoing symptomatic COVID patients. These include the top predictor, Polyisoprenoid(50), whose biosynthesis has also been identified as significantly decreased in inflammatory conditions such as Crohn's disease (51). Tetrapyrrole (52) and, farnesol (53) also have anti-inflammatory effects. Conversely, three pathways for biosynthesis of pro-inflammatory molecules are increased in ongoing symptomatic COVID patients: dTDP-L-rhamnose (DTDPRHAMSYN-PWY)(54), pyrimidine (PWY-6545) (55) and purine (P164 PWY) (56) deoxyribonucleotides. Finally, O-antigen building block biosynthesis (OANTIGEN-PWY), an important step in the lipopolysaccharide (LPS) biosynthetic pathway (57), and the superpathway of phospholipid biosynthesis (PHOSLIPSYN-PWY), important in LPS production (58, 59), are both higher among patients with ongoing symptomatic COVID.

Similar patterns emerge with the long-COVID analysis with 6 predictors shared with those for ongoing symptomatic COVID analysis. Pro-inflammatory molecule synthesis is higher among long-COVID patients relative to those without as well as reduced branch-chain amino acid and anti-inflammatory molecule biosynthesis (**Figure 5C**). Additional pro-inflammatory molecule biosynthesis are noted with chorismite (PWY-6163) (60), colanic acid (COLANSYN-PWY) (61), and NAD biosynthesis (PWY-241) (62) all being higher among the long-COVID patients.

## DISCUSSION

Many patients recovering from SARS-CoV-2 infection have symptoms that last long after the acute infection has run its course and our study highlights this same phenomenon. Over 1/3 of our cohort had symptoms lasting longer than 10 weeks and thus enter the long-COVID disease stage. Fatigue and "brain fog" were the longer lasting, most prominent symptoms among these patients. In an attempt to better understand both ongoing symptomatic and long-COVID patients, we investigated potential clinical and microbiome associations with these disorders. Our modeling identified: 1) microbial associations that are known to promote inflammation via LPS production or other mechanisms, 2) reduction of anti-inflammatory metabolic pathways, 3) pathobionts known to cause pulmonary infections, and 4) microbiota previously shown to have associations with COVID-19. Thus, our work begins to shed light on the hypothesis that the oral microbiome composition may influence the duration of COVID-19 disease symptoms.

**Patients with longer COVID-19 symptoms have dysbiotic, inflammatory-type oral microbiome**

The oral microbiome has been shown to closely associate with SARS-CoV-2 co-infections in the lungs (11) and the oral-lung aspiration axis is a key factor leading to many respiratory infectious processes (63). We hypothesized that the oral microbiome might associate with the duration of post-acute infection symptoms presented in ongoing symptomatic and long-COVID disease states (64). Our findings extend previous work demonstrating how specific member of the genera *Prevotella* and *Veillonella,* were distinctive in the oral microbiota of COVID-19 patients (65). *Prevotella* species have been overrepresented in COVID-19 patient populations (42) while both members of the *Prevotella* and *Veillonella* genera have been found in the bronchoalveolar lavage fluid of the COVID-19 patients (31). Members of the *Prevotella* genus are thought to produce proteins that can promote SARS-CoV2 infection and increase clinical severity of COVID-19 (42) and have previously been tied to systemic diseases, including low-grade systemic inflammation (38). The increased abundances of these two genera on the tongue have also been associated with an increased risk of death due to pneumonia in older, frail patients (66, 67). Finally, both genera induce inflammatory responses. *Veillonella* species have shown a strong capacity to induce IL-6 (68) while *Prevotella* strains primarily activate toll-like receptor 2 and enhance the expression of inflammatory cytokines, including IL-23 and IL-1 (69, 70). Other pro-inflammatory microbiota were identified in our analysis that also associated with longer disease symptoms such as *L. wadei* (12), *S. moorei* (71), and multiple *Actinomyces* species (41).

Metabolic pathways associated with the production of pro-inflammatory molecules were increased in abundance while pathways associated with production of anti-inflammatory molecules were decreased in patients presenting ongoing and long-COVID symptoms. One of the top predictors and thus demonstrating the strongest association in our data with both ongoing symptomatic and long-COVID disease was polyisoprenoid biosynthesis. Polyisoprenoid expresses anti-inflammatory activity (50) and is significantly decreased in inflammatory conditions such as Crohn's disease (51). Among the top predictors in our analysis was reduced abundance of genes involved in the production of branched amino acids. Branched amino acids have long been shown to act as anti-inflammatory agents (47, 48). Evidence is accumulating to support the hypothesis that systemic chronic inflammation contributes to the symptomatic progression to long-COVID (22, 72). Given that changes in the microbiome composition can result in chronic inflammation and metabolic dysfunction (73), it is possible that the pro-inflammatory, microbiome profiles we observe here could play a pivotal role in this disease process.

**Lipopolysaccharide-producing bacteria may promote inflammation and drive COVID-19 symptom duration**

Lipopolysaccharides (LPS) is an outer-membrane component of gram-negative bacteria and can also be released in vesicles (74). Vesicle-associated LPS can have proinflammatory effects on host immune systems (75). Microbiome-derived LPS causes systemic inflammation (76, 77) and can even induce cognitive impairment and neuroinflammation (78, 79). Increases in lipopolysaccharide-producing bacteria, such as *Leptotrichia*, have been demonstrated in the oral cavity of COVID-19 patients and are thought to be involved in the inflammatory response (12). Our analysis reveals higher abundances of many LPS-producing bacteria in patients with longer lasting symptoms. For example, *Veillonella* species, known to produce large amounts of LPS (37), are present in increased abundances in our COVID-19 patients with longer lasting symptoms. Increases in species such as *V. dispar*, *V. infantium*, and *V. atypica* are top predictors of ongoing symptomatic COVID while *V. infantium* is found in higher abundances among long-COVID patients. Other LPS producing species such as *L. wadei* (12) and *M. micronuciformis* (80) are also found to be in increased abundances. Additionally, our metabolic pathway analysis revealed an association with important steps in LPS biosynthesis and ongoing symptomatic and long-COVID disease states. It is possible that LPS production may be a marker of other risk factors rather than a direct causal contributor. This would be critical to investigate in future work, however this evidence points towards the important association of inflammation and long symptom disease states.

**Myalgic encephalomyelitis/chronic fatigue syndrome linking to long-term COVID-19 symptoms through oral microbiome dysbiosis**

There has been a growing concern that COVID-19 patients with long-term sequelae resembling patients with myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS) (81). These two conditions share some of the same symptoms, especially fatigue and cognitive impairment (17, 82). ME/CFS is a condition characterized by chronic fatigue, lasting at least 6 months, that impairs one's ability to perform daily activities and typically has additional impairments in memory and concentration (83). This syndrome is also linked closely to chronic inflammation as the driver of these patients symptoms (84). The link to long-term symptoms is not unique to COVID-19 disease as patients with both SARS-CoV1 and Middle East respiratory syndrome have also suffered from long-term sequelae in the previous epidemics (85).

ME/CFS has been hypothesized to be linked to infectious agents and microbiome dysbiosis has specifically been described in this syndrome either through the presence of pathobionts or microbial species that promote chronic inflammation (86). The gut microbiome has been shown to have reduced diversity and altered composition in ME/CFS patients (87) and viral-induced microbiome changes are also thought to play a pivotal role (88). Clinical trials targeting the gut microbiome have shown promise in treating ME/CFS (89). Interestingly, ME/CFS patients have been shown to have altered dysbiotic oral microbiomes characterized by increased abundances in the genera *Leptotrichia*, *Prevotella*, and *Fusobacterium* (90). Using whole genome sequencing, we have shown many species belonging to these genera are increased abundance in both ongoing symptomatic and long-COVID patients. Specifically, top predicting species *L. wadei*, *P. sp F0091*, *P. denticola*, *P. nigrescens*, *P. histicola*, and *P. oulorum* in the ongoing symptomatic COVID group and *P. denticola*, P. *melaninogenica*, *P. jejuni*, *P. nigrescens* and *F. nucleatum* in the long-COVID group were all present in higher abundances in patients suffering from longer lasting symptoms. These finding add intriguing evidence of a possible link between ME/CFS and COVID-19 patients suffering from longer lasting symptoms related to inflammation in the oral microbiome.

## Strengths and Limitations

This study has several notable strengths and limitations. This study is limited in the number of patients enrolled and followed for symptom duration outcomes. A more robust cohort would allow deeper investigation of preexisting medical conditions and medications which might shape the oral microbiome composition. Larger cohorts would also include a more diverse patient set involving those treated as outpatients and more intensive care unit admissions. Generalization of our findings would need to be performed in a more diverse patient population. This limitation is balanced by our application of whole genome sequencing, which provide greater resolution than 16S rRNA gene sequencing used in many of the previous microbiome investigations (91). We also applied random forest classification which enable us to include both clinical and microbiome data in our modeling (27, 28). This modeling approach has significant advantages compared to traditional classification techniques, as it is agnostic to model structure (e.g. non-parametric regression), it does not need to meet common assumptions underlying classical regression techniques, and is able to intrinsically perform permutated ranked feature selection (29). We also have the advantage of collecting samples at the time of diagnosis before medical treatments that may alter the microbiome composition.

## Conclusions

In conclusion, the oral microbiome of patients with prolonged symptoms falling under the ongoing symptomatic or long-COVID disease states demonstrates a dysbiotic pattern with increased pathobionts, increases in inflammation-inducing and LPS-producing microbiota, and reduction of metabolic pathways known to have anti-inflammatory properties. This work needs further validation however it supports the tenet that the microbiome may play a role in prolonging symptom duration among COVID-19 through promotion of inflammation. The microbiome may therefore hold the key to better understanding the post infection prolonged syndromes now facing patients after they recover from acute infection and provide a way to predict and subsequently act upon and prevent the development of long-COVID.

## MATERIALS and METHODS

### Study Setting and Population

This prospective cohort consists of patients presenting to one Emergency Department located in central Massachusetts from April 2020 through February 2021. We enrolled patients who presented with symptoms consistent with a COVID-19 infection but analyzed only those with a positive SARS-CoV-2 PCR whom we could contact for follow-up. We defined symptoms of COVID-19 based off of the Centers for Disease Control and Prevention guidelines (92).

### Data Collection

We collected baseline factors that included demographics, medical history, and presenting disease duration and symptomatology. Comorbidity was assessed at baseline using the Charlson Comorbidity Index (CCI), a widely used instrument designed to measure the burden of medical diseases and predict mortality (93). Patients were then followed through their hospital course for treatment types and length of stay. After discharge from the hospital subsequent healthcare visits were recorded through the medical record. Patients were contacted by phone after 4 weeks of total symptoms after discharge and then again, a second time, if they were experiencing ongoing symptoms, after 10 weeks. Patients were categorized as symptoms >4 weeks and symptoms >10 weeks for analysis. Patients were also queried as to the type of symptoms that lasted the longest. Patients were excluded from follow-up if they died, were unable to communicate in English, had severe dementia, were in hospice or withdrew themselves from the study.

### Sample Collection and Processing

Oropharyngeal samples were collected using OMNIgene•ORAL collection kits (OMR-120, DNAgenotek). Briefly, the posterior oropharynx was swabbed for 30 seconds and then the swab was inserted into a tube with a DNA/RNA stabilization buffer. Samples were heated to 65-70 $^{o}$C for one hour to inactivate SARS-CoV-2 virus (94) and stored frozen. Nucleic acids were extracted by first thawing samples and then treating with 5ul Proteinase K (P8107S, New England Biolabs) for 2 hours at 50C. DNA and RNA was then extracted using ZymoBIOMICS DNA/RNA Miniprep Kits (R2002, Zymo Research) as per manufacture protocol.

**Sequence Processing and Analysis**

Metagenomic DNA sequencing libraries were constructed using the Nextera XT DNA Library Prep Kit (FC-131-1096, Illumina) and sequenced on a NextSeq500 Sequencing System as 2 x 150 nucleotide paired-end reads. Shotgun metagenomic reads were first trimmed and quality filtered to remove sequencing adapters and host contamination using Trimmomatic (95) and Bowtie2 (96), respectively, as part of the KneadData pipeline (https://bitbucket.org/biobakery/kneaddata). As in our previous work (28, 97), metagenomic data was profiled for microbial taxonomic abundances and microbial metabolic pathways using Metaphlan3 (98) and HUMAnN3 (45), respectively. The total number of microbial and contaminant reads recovered as presented in Supplemental Table 1.

**SARS-CoV-2 viral load quantification**

PCR was performed using the ViiA 7 Real-Time PCR System (Applied Biosystems) and the GoTaq® Probe 1-Step RT-qPCR System (Promega, A6120). The primer-probe set N1 (2019-nCoV_N1-F: 5′-GAC CCC AAA ATC AGC GAA AT-3′; 2019-nCoV_N1-R: 5′-TCT GGT TAC TGC CAG TTG AAT CTG-3′; 2019-nCoV_N1-P: 5′-FAM-ACC CCG CAT TAC GTT TGG ACC-BHQ1-3′) designed by the Centers for Disease Control and Prevention were obtained from Integrated DNA Technologies (IDT, 10006713) and used at concentrations of 500 nM and 125 nM, respectively (99). 5 µl of eluted RNA were used to prepare 20 µl PCR reactions. Cycling conditions were as indicated by the Centers for Disease Control and

Prevention: 45°C for 15 min, 95°C for 2 min, followed by 45 cycles of 95°C for 3 s and 55°C for 30 s (99). Cycle threshold (Ct) values were converted into viral RNA copies based on a standard curve prepared from 4-fold serial dilutions of known quantities (1.0 × 106 to 2.44 x 102 viral copies) of a SARS-CoV-2_N positive control plasmid (IDT, 10006625). The lower limit threshold for positive detection in our study was 244 viral copies per reaction. Viral load was calculated as number of genome copies per milliliter of transport media to resuspend tongue swabs. The assay was run in triplicate for each sample and three non-template wells were included as negative controls.

**Statistical and Computational Analysis**

To determine similarity in oral microbiome samples among the COVID-19 patients and to associate microbiome features to duration of symptom outcomes, we started by performing traditional unsupervised correspondence analysis (Principal Coordinate Analysis and t-Distributed Stochastic Neighbor Embedding). As most of the signal from the unsupervised analysis was accounted by inter-individual variability, we then decided to run supervised machine learning models. We built a random forest classification (RFC) pipeline to predict either ongoing symptomatic COVID or long-COVID from a given data subset. One sample failed the sequencing run and thus 26 samples were included in our modeling. The first step of our pipeline used the feature selection algorithm Boruta on five-fold cross-validated data to estimate model performance (29). The permutated variable importance from each RFC was also calculated. Each model was run starting from ten different random seeds to calculate performance metrics. F1 score, the harmonic mean of precision and accuracy, was used to select the top performing model for each outcome.

**Study Approval**

This prospective cohort study was approved by the Institutional Review Board at the University of Massachusetts Medical School. Written informed consent was received from all study participants prior to inclusion in the study.

## AUTHOR CONTRIBUTIONS

JPH, BAM, AM, and EB conceived and led the study. JPH, EB, CT supervised the conduct of the study and data collection. LC, MMS, SM, CT, and PD managed the clinical data, including quality control. LC and MMS handled the sample collection and storage. DW managed sample extraction and sequencing and performed metagenomic profiling. AZ and VB provided statistical advice on study design and analyzed the data. JPH and EB wrote the manuscript with input from all authors. JPH composed the first draft of the majority of the manuscript and was responsible for incorporation of all authors edits. Accordingly, JPH was assigned the first author slot.

## ACKNOWLEDGEMENTS

We would like to thank the UMass Memorial Medical Center Emergency Department Staff, especially the nursing and resident physicians for making it possible to collect biological samples from COVID-19 patients with acute and sometimes severe symptoms within in the Emergency Department. Thank you to the Human Subjects Institutional Review Board at the University of Massachusetts Medical School and especially A. Blodgett for their guidance in helping to design and implement the human subject protocol early in the pandemic. Thank you to The Society for Academic Emergency Medicine as well as the Dean of University of Massachusetts Medical School and the many donors who through their financial support made this work possible. We would also like to thank the NIH for funding that provided salary support for this work (1RF1AG067483-01).

## DATA AVAILABILITY

Data relating to the metagenomic sequencing that support the findings of this study have been uploaded to the NCBI BioProject (https://www.ncbi.nlm.nih.gov/bioproject/) and are available for download via the accession number PRJNA735193 under the title Oral Microbiome associated with Coronavirus disease 2019 (COVID-19).

**BIBLIOGRAPHY**


1. Dewhirst FE, Chen T, Izard J, Paster BJ, Tanner AC, Yu WH, et al. The human oral microbiome. *J Bacteriol.* 2010;192(19):5002-17.
2. Willis JR, and Gabaldón T. The Human Oral Microbiome in Health and Disease: From Sequences to Ecosystems. *Microorganisms.* 2020;8(2).
3. Seymour GJ, Ford PJ, Cullinan MP, Leishman S, and Yamazaki K. Relationship between periodontal infections and systemic disease. *Clin Microbiol Infect.* 2007;13 Suppl 4:3-10.
4. Awano S, Ansai T, Takata Y, Soh I, Akifusa S, Hamasaki T, et al. Oral health and mortality risk from pneumonia in the elderly. *J Dent Res.* 2008;87(4):334-9.
5. Scannapieco FA. Role of oral bacteria in respiratory infection. *J Periodontol.* 1999;70(7):793-802.
6. CDC COVID Data Tracker. https://covid.cdc.gov/covid-data-tracker/#datatracker-home. *CENTERS FOR DISEASE CONTROL AND PREVENTION (CDC).*Accessed 3/24/21.
7. Trompette A, Gollwitzer ES, Pattaroni C, Lopez-Mejia IC, Riva E, Pernot J, et al. Dietary Fiber Confers Protection against Flu by Shaping Ly6c(-) Patrolling Monocyte Hematopoiesis and CD8(+) T Cell Metabolism. *Immunity.* 2018;48(5):992-1005.e8.
8. Belkaid Y, and Harrison OJ. Homeostatic Immunity and the Microbiota. *Immunity.* 2017;46(4):562-76.
9. Zuo T, Liu Q, Zhang F, Lui GC, Tso EY, Yeoh YK, et al. Depicting SARS-CoV-2 faecal viral activity in association with gut microbiota composition in patients with COVID-19. *Gut.* 2021;70(2):276-84.
10. Yeoh YK, Zuo T, Lui GC, Zhang F, Liu Q, Li AY, et al. Gut microbiota composition reflects disease severity and dysfunctional immune responses in patients with COVID-19. *Gut.* 2021;70(4):698-706.
11. Bao L, Zhang C, Dong J, Zhao L, Li Y, and Sun J. Oral Microbiome and SARS-CoV-2: Beware of Lung Co-infection. *Frontiers in microbiology.* 2020;11:1840.

12. Ren Z, Wang H, Cui G, Lu H, Wang L, Luo H, et al. Alterations in the human oral and gut microbiomes and lipidomics in COVID-19. *Gut.* 2021.

13. Tenforde MW, Kim SS, Lindsell CJ, Billig Rose E, Shapiro NI, Files DC, et al. Symptom Duration and Risk Factors for Delayed Return to Usual Health Among Outpatients with COVID-19 in a Multistate Health Care Systems Network - United States, March-June 2020. *MMWR Morb Mortal Wkly Rep.* 2020;69(30):993-8.

14. Garrigues E, Janvier P, Kherabi Y, Le Bot A, Hamon A, Gouze H, et al. Post-discharge persistent symptoms and health-related quality of life after hospitalization for COVID-19. *J Infect.* 2020;81(6):e4-e6.

15. Context | COVID-19 rapid guideline: managing the long-term effects of COVID-19. *National Institute for Health and Care Excellence.*Accessed 4/3/21.

16. Sudre CH, Murray B, Varsavsky T, Graham MS, Penfold RS, Bowyer RC, et al. Attributes and predictors of long COVID. *Nature medicine.* 2021;27(4):626-31.

17. Marshall M. The lasting misery of coronavirus long-haulers. *Nature.* 2020;585(7825):339-41.

18. Baig AM. Chronic COVID syndrome: Need for an appropriate medical terminology for long-COVID and COVID long-haulers. *J Med Virol.* 2021;93(5):2555-6.

19. Brodin P. Immune determinants of COVID-19 disease presentation and severity. *Nature medicine.* 2021;27(1):28-33.

20. Townsend L, Dyer AH, Jones K, Dunne J, Mooney A, Gaffney F, et al. Persistent fatigue following SARS-CoV-2 infection is common and independent of severity of initial infection. *PloS one.* 2020;15(11):e0240784.

21. Greenhalgh T, Knight M, A'Court C, Buxton M, and Husain L. Management of post-acute covid-19 in primary care. *Bmj.* 2020;370:m3026.

22. Doykov I, Hällqvist J, Gilmour KC, Grandjean L, Mills K, and Heywood WE. 'The long tail of Covid-19' - The detection of a prolonged inflammatory response after a SARS-CoV-2 infection in asymptomatic and mildly affected patients. *F1000Res.* 2020;9:1349.

23. Koralnik IJ, and Tyler KL. COVID-19: A Global Threat to the Nervous System. *Annals of neurology.* 2020;88(1):1-11.

24. Stefano GB. Historical Insight into Infections and Disorders Associated with Neurological and Psychiatric Sequelae Similar to Long COVID. *Med Sci Monit.* 2021;27:e931447.

25. Beghini F, McIver LJ, Blanco-Míguez A, Dubois L, Asnicar F, Maharjan S, et al. Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. *Elife.* 2021;10.

26. DeJong TM. A Comparison of Three Diversity Indices Based on Their Components of Richness and Evenness. *Oikos.* 1975;26(2):222-7.

27. Hajjem A, Bellavance F, and Larocque D. Mixed-effects random forest for clustered data. *Journal of Statistical Computation and Simulation.* 2014;84(6):1313-28.

28. Haran JP, Bhattarai SK, Foley SE, Dutta P, Ward DV, Bucci V, et al. Alzheimer's Disease Microbiome Is Associated with Dysregulation of the Anti-Inflammatory P-Glycoprotein Pathway. *mBio.* 2019;10(3).

29. Wipperman MF, Bhattarai SK, Vorkas CK, Maringati VS, Taur Y, Mathurin L, et al. Gastrointestinal microbiota composition predicts peripheral inflammatory state during treatment of human tuberculosis. *Nature communications.* 2021;12(1):1141.

30. Cobo F, Pérez-Carrasco V, García-Salcedo JA, and Navarro-Marí JM. Bacteremia caused by Veillonella dispar in an oncological patient. *Anaerobe.* 2020;66:102285.

31. Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, et al. A new coronavirus associated with human respiratory disease in China. *Nature.* 2020;579(7798):265-9.

32. Liu WJ, Xiao M, Yi J, Li Y, Kudinha T, and Xu YC. First case report of bacteremia caused by Solobacterium moorei in China, and literature review. *BMC infectious diseases.* 2019;19(1):730.

33. Pedersen RM, Holt HM, and Justesen US. Solobacterium moorei bacteremia: identification, antimicrobial susceptibility, and clinical characteristics. *Journal of clinical microbiology.* 2011;49(7):2766-8.

34. Bek-Thomsen M, Tettelin H, Hance I, Nelson KE, and Kilian M. Population diversity and dynamics of Streptococcus mitis, Streptococcus oralis, and Streptococcus infantis in the upper respiratory tracts of adults, determined by a nonculture strategy. *Infection and immunity.* 2008;76(5):1889-96.

35. Sadhu A, Loewenstein R, and Klotz SA. Rothia dentocariosa endocarditis complicated by multiple cerebellar hemorrhages. *Diagn Microbiol Infect Dis.* 2005;53(3):239-40.

36. Marotz C, Belda-Ferre P, Ali F, Das P, Huang S, Cantrell K, et al. Microbial context predicts SARS-CoV-2 prevalence in patients and the hospital built environment. *medRxiv.* 2020:2020.11.19.20234229.

37. Delwiche EA, Pestka JJ, and Tortorello ML. The veillonellae: gram-negative cocci with a unique physiology. *Annual review of microbiology.* 1985;39:175-93.

38. Larsen JM. The immune response to Prevotella bacteria in chronic inflammatory disease. *Immunology.* 2017;151(4):363-74.

39. Bellocchi C, and Volkmann ER. Update on the Gastrointestinal Microbiome in Systemic Sclerosis. *Current rheumatology reports.* 2018;20(8):49.

40. Kim D, and Kim WU. Editorial: Can Prevotella copri Be a Causative Pathobiont in Rheumatoid Arthritis? *Arthritis Rheumatol.* 2016;68(11):2565-7.

41. Valour F, Sénéchal A, Dupieux C, Karsenty J, Lustig S, Breton P, et al. Actinomycosis: etiology, clinical features, diagnosis, treatment, and management. *Infection and drug resistance.* 2014;7:183-97.

42. Khan AA, and Khan Z. COVID-2019-associated overexpressed Prevotella proteins mediated host-pathogen interactions and their role in coronavirus outbreak. *Bioinformatics.* 2020;36(13):4065-9.

43. Noguchi S, Yatera K, Kawanami T, Yamasaki K, Naito K, Akata K, et al. The clinical features of respiratory infections caused by the Streptococcus anginosus group. *BMC pulmonary medicine.* 2015;15:133.

44. Bonazzetti C, Morena V, Giacomelli A, Oreni L, Casalini G, Galimberti LR, et al. Unexpectedly High Frequency of Enterococcal Bloodstream Infections in Coronavirus Disease 2019 Patients Admitted to an Italian ICU: An Observational Study. *Crit Care Med.* 2021;49(1):e31-e40.

45. Franzosa EA, McIver LJ, Rahnavard G, Thompson LR, Schirmer M, Weingart G, et al. Species-level functional profiling of metagenomes and metatranscriptomes. *Nature methods.* 2018;15(11):962-8.

46. Caspi R, Billington R, Keseler IM, Kothari A, Krummenacker M, Midford PE, et al. The MetaCyc database of metabolic pathways and enzymes - a 2019 update. *Nucleic acids research.* 2020;48(D1):D445-D53.

47. Lee JH, Park E, Jin HJ, Lee Y, Choi SJ, Lee GW, et al. Anti-inflammatory and anti-genotoxic activity of branched chain amino acids (BCAA) in lipopolysaccharide (LPS) stimulated RAW 264.7 macrophages. *Food Sci Biotechnol.* 2017;26(5):1371-7.

48. Da Silva MS, Bigo C, Barbier O, and Rudkowska I. Whey protein hydrolysate and branched-chain amino acids downregulate inflammation-related genes in vascular endothelial cells. *Nutr Res.* 2017;38:43-51.

49. Saxena RN, Pendse VK, and Khanna NK. Anti-inflammatory and analgesic properties of four amino-acids. *Indian J Physiol Pharmacol.* 1984;28(4):299-305.

50. Pronin AV, Danilov LL, Narovlyansky AN, and Sanin AV. Plant polyisoprenoids and control of cholesterol level. *Arch Immunol Ther Exp (Warsz).* 2014;62(1):31-9.

51. Klaassen MAY, Imhann F, Collij V, Fu J, Wijmenga C, Zhernakova A, et al. Anti-inflammatory Gut Microbial Pathways Are Decreased During Crohn's Disease Exacerbations. *J Crohns Colitis.* 2019;13(11):1439-49.

52. Jelić D, Tatić I, Trzun M, Hrvačić B, Brajša K, Verbanac D, et al. Porphyrins as new endogenous anti-inflammatory agents. *Eur J Pharmacol.* 2012;691(1-3):251-60.

53. Ku CM, and Lin JY. Anti-inflammatory effects of 27 selected terpenoid compounds tested through modulating Th1/Th2 cytokine secretion profiles using murine primary splenocytes. *Food Chem.* 2013;141(2):1104-13.

54. van der Beek SL, Zorzoli A, Çanak E, Chapman RN, Lucas K, Meyer BH, et al. Streptococcal dTDP-L-rhamnose biosynthesis enzymes: functional characterization and lead compound identification. *Mol Microbiol.* 2019;111(4):951-64.

55. Dimitrova P, Skapenko A, Herrmann ML, Schleyerbach R, Kalden JR, and Schulze-Koops H. Restriction of de novo pyrimidine biosynthesis inhibits Th1 cell activation and promotes Th2 cell differentiation. *Journal of immunology.* 2002;169(6):3392-9.

56. Barnes VM, Teles R, Trivedi HM, Devizio W, Xu T, Mitchell MW, et al. Acceleration of purine degradation by periodontal diseases. *J Dent Res.* 2009;88(9):851-5.

57. Han W, Wu B, Li L, Zhao G, Woodward R, Pettit N, et al. Defining function of lipopolysaccharide O-antigen ligase WaaL using chemoenzymatically synthesized substrates. *The Journal of biological chemistry.* 2012;287(8):5357-65.

58. Rothfield L, and Pearlman M. The role of cell envelope phospholipid in the enzymatic synthesis of bacterial lipopolysaccharide. Structural requirements of the phospholipid molecule. *The Journal of biological chemistry.* 1966;241(6):1386-92.

59. Emiola A, Andrews SS, Heller C, and George J. Crosstalk between the lipopolysaccharide and phospholipid pathways during outer membrane biogenesis in Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America.* 2016;113(11):3108-13.

60. Rodrigues-Vendramini FAV, Marschalk C, Toplak M, Macheroux P, Bonfim-Mendonça PS, Svidzinski TIE, et al. Promising New Antifungal Treatment Targeting Chorismate Synthase from Paracoccidioides brasiliensis. *Antimicrobial agents and chemotherapy.* 2019;63(1).

61. Hanna A, Berg M, Stout V, and Razatos A. Role of capsular colanic acid in adhesion of uropathogenic Escherichia coli. *Appl Environ Microbiol.* 2003;69(8):4474-81.

62. Gerner RR, Klepsch V, Macheiner S, Arnhard K, Adolph TE, Grander C, et al. NAD metabolism fuels human and mouse intestinal inflammation. *Gut.* 2018;67(10):1813-23.

63. Mammen MJ, Scannapieco FA, and Sethi S. Oral-lung microbiome interactions in lung diseases. *Periodontol 2000.* 2020;83(1):234-41.

64. Dani M, Dirksen A, Taraborrelli P, Torocastro M, Panagopoulos D, Sutton R, et al. Autonomic dysfunction in 'long COVID': rationale, physiology and management strategies. *Clin Med (Lond).* 2021;21(1):e63-e7.

65. Iebba V, Zanotta N, Campisciano G, Zerbato V, Di Bella S, Cason C, et al. Profiling of oral microbiota and cytokines in COVID-19 patients. *bioRxiv.* 2020:2020.12.13.422589.

66. Asakawa M, Takeshita T, Furuta M, Kageyama S, Takeuchi K, Hata J, et al. Tongue Microbiota and Oral Health Status in Community-Dwelling Elderly Adults. *mSphere.* 2018;3(4).

67. Kageyama S, Takeshita T, Furuta M, Tomioka M, Asakawa M, Suma S, et al. Relationships of Variations in the Tongue Microbiota and Pneumonia Mortality in Nursing Home Residents. *J Gerontol A Biol Sci Med Sci.* 2018;73(8):1097-102.

68. van den Bogert B, Meijerink M, Zoetendal EG, Wells JM, and Kleerebezem M. Immunomodulatory properties of Streptococcus and Veillonella isolates from the human small intestine microbiota. *PloS one.* 2014;9(12):e114277.

69. Segal LN, Clemente JC, Tsay JC, Koralov SB, Keller BC, Wu BG, et al. Enrichment of the lung microbiome with oral taxa is associated with lung inflammation of a Th17 phenotype. *Nat Microbiol.* 2016;1:16031.

70. Segal LN, Alekseyenko AV, Clemente JC, Kulkarni R, Wu B, Gao Z, et al. Enrichment of lung microbiome with supraglottic taxa is associated with increased pulmonary inflammation. *Microbiome.* 2013;1(1):19.

71. Hiranmayi KV, Sirisha K, Ramoji Rao MV, and Sudhakar P. Novel Pathogens in Periodontal Microbiology. *J Pharm Bioallied Sci.* 2017;9(3):155-63.

72. Inciardi RM, Solomon SD, Ridker PM, and Metra M. Coronavirus 2019 Disease (COVID-19), Systemic Inflammation, and Cardiovascular Disease. *J Am Heart Assoc.* 2020;9(16):e017756.

73. Sommer F, and Bäckhed F. The gut microbiota--masters of host development and physiology. *Nature reviews Microbiology.* 2013;11(4):227-38.

74. Farhana A, and Khan YS. *StatPearls*. Treasure Island (FL): StatPearls Publishing
Copyright © 2021, StatPearls Publishing LLC.; 2021.

75. Bonnington KE, and Kuehn MJ. Protein selection and export via outer membrane vesicles. *Biochim Biophys Acta.* 2014;1843(8):1612-9.

76. Noailles A, Maneu V, Campello L, Lax P, and Cuenca N. Systemic inflammation induced by lipopolysaccharide aggravates inherited retinal dystrophy. *Cell Death Dis.* 2018;9(3):350.

77. Salguero MV, Al-Obaide MAI, Singh R, Siepmann T, and Vasylyeva TL. Dysbiosis of Gram-negative gut microbiota and the associated serum lipopolysaccharide exacerbates inflammation in type 2 diabetic patients with chronic kidney disease. *Exp Ther Med.* 2019;18(5):3461-9.

78. Zhao J, Bi W, Xiao S, Lan X, Cheng X, Zhang J, et al. Neuroinflammation induced by lipopolysaccharide causes cognitive impairment in mice. *Scientific reports.* 2019;9(1):5790.

79. Lee JW, Lee YK, Yuk DY, Choi DY, Ban SB, Oh KW, et al. Neuro-inflammation induced by lipopolysaccharide causes cognitive impairment through enhancement of beta-amyloid generation. *J Neuroinflammation.* 2008;5:37.

80. Shetty SA, Marathe NP, Lanjekar V, Ranade D, and Shouche YS. Comparative genome analysis of Megasphaera sp. reveals niche specialization and its potential role in the human gut. *PloS one.* 2013;8(11):e79353.

81. Wostyn P. COVID-19 and chronic fatigue syndrome: Is the worst yet to come? *Med Hypotheses.* 2021;146:110469.

82. Long COVID: let patients help define long-lasting COVID symptoms. *Nature.* 2020;586(7828):170.

83. Yancey JR, and Thomas SM. Chronic fatigue syndrome: diagnosis and treatment. *Am Fam Physician.* 2012;86(8):741-6.

84. Komaroff AL. Inflammation correlates with symptoms in chronic fatigue syndrome. *Proceedings of the National Academy of Sciences of the United States of America.* 2017;114(34):8914-6.

85. O'Sullivan O. Long-term sequelae following previous coronavirus epidemics. *Clin Med (Lond).* 2021;21(1):e68-e70.

86. Proal A, and Marshall T. Myalgic Encephalomyelitis/Chronic Fatigue Syndrome in the Era of the Human Microbiome: Persistent Pathogens Drive Chronic Symptoms by Interfering With Host Metabolism, Gene Expression, and Immunity. *Frontiers in pediatrics.* 2018;6:373.

87. Giloteaux L, Goodrich JK, Walters WA, Levine SM, Ley RE, and Hanson MR. Reduced diversity and altered composition of the gut microbiome in individuals with myalgic encephalomyelitis/chronic fatigue syndrome. *Microbiome.* 2016;4(1):30.

88. Newberry F, Hsieh SY, Wileman T, and Carding SR. Does the microbiome and virome contribute to myalgic encephalomyelitis/chronic fatigue syndrome? *Clin Sci (Lond).* 2018;132(5):523-42.

89. Rao AV, Bested AC, Beaulne TM, Katzman MA, Iorio C, Berardi JM, et al. A randomized, double-blind, placebo-controlled pilot study of a probiotic in emotional symptoms of chronic fatigue syndrome. *Gut pathogens.* 2009;1(1):6.

90. Wang T, Yu L, Xu C, Pan K, Mo M, Duan M, et al. Chronic fatigue syndrome patients have alterations in their oral microbiome composition and function. *PloS one.* 2018;13(9):e0203503.

91. Ranjan R, Rani A, Metwally A, McGee HS, and Perkins DL. Analysis of the microbiome: Advantages of whole genome shotgun versus 16S amplicon sequencing. *Biochemical and biophysical research communications.* 2016;469(4):967-77.

92. https://www.cdc.gov/coronavirus/2019-ncov/symptoms-testing/symptoms.html. Symptoms of COVID-19. *CENTERS FOR DISEASE CONTROL AND PREVENTION (CDC).*Updated 2/22/2021.

93. Austin SR, Wong YN, Uzzo RG, Beck JR, and Egleston BL. Why Summary Comorbidity Measures Such As the Charlson Comorbidity Index and Elixhauser Score Work. *Med Care.* 2015;53(9):e65-72.

94. Rabenau HF, Cinatl J, Morgenstern B, Bauer G, Preiser W, and Doerr HW. Stability and inactivation of SARS coronavirus. *Med Microbiol Immunol.* 2005;194(1-2):1-6.

95. Bolger AM, Lohse M, and Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 2014;30(15):2114-20.

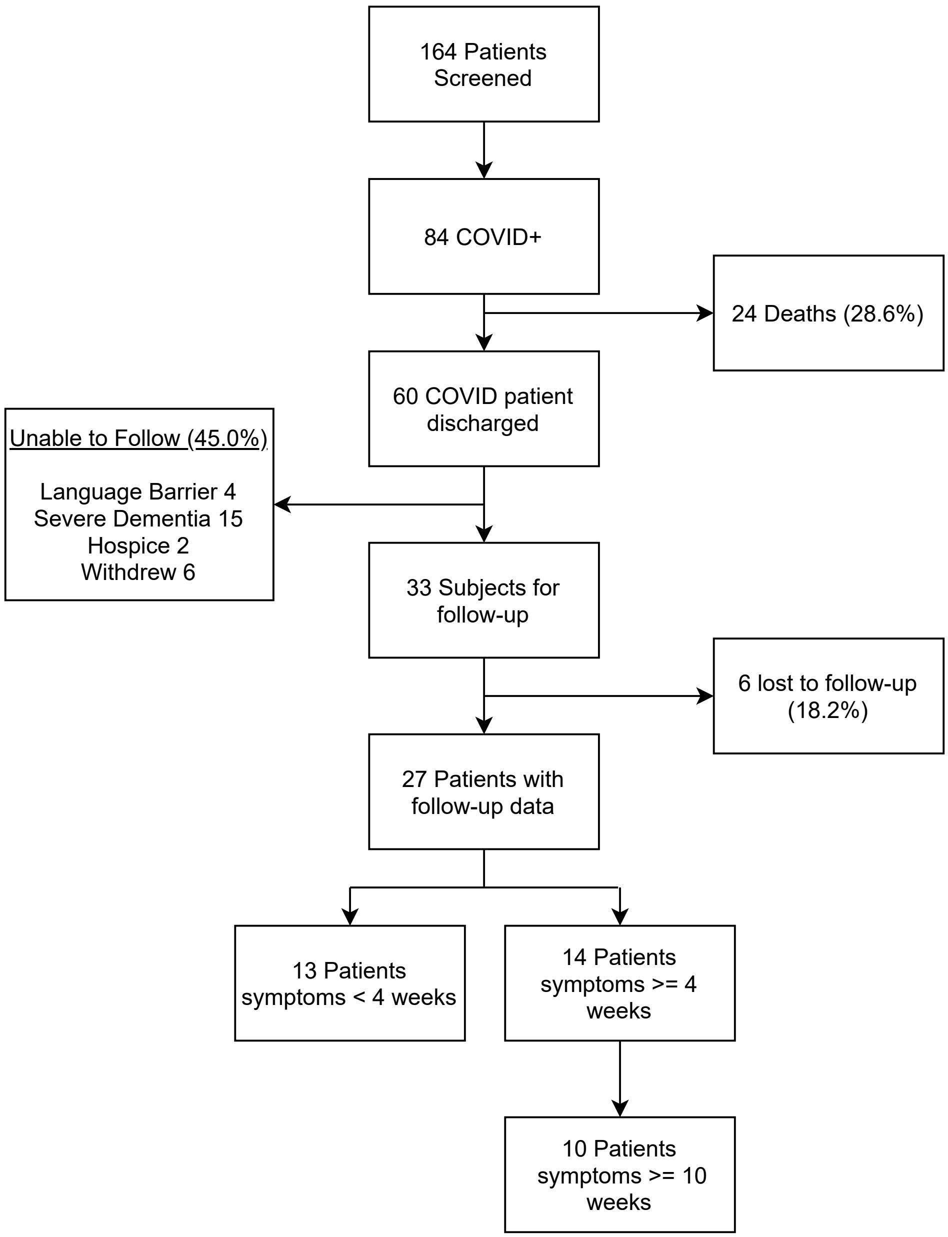
96. Langmead B, and Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods.* 2012;9(4):1-3.

97. Haran JP, Bucci V, Dutta P, Ward D, and McCormick B. The nursing home elder microbiome stability and associations with age, frailty, nutrition, and physical location. *Journal of medical microbiology.* 2018;67(1):40-51.

98. Beghini F, McIver LJ, Blanco-Míguez A, Dubois L, Asnicar F, Maharjan S, et al. Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. *bioRxiv.* 2020:2020.11.19.388223.

99. CDC's Diagnostic Test for COVID-19 Only and Supplies. *https://wwwcdcgov/coronavirus/2019-ncov/lab/virus-requestshtml CENTERS FOR DISEASE CONTROL AND PREVENTION (CDC) Accessed 01/20/21.*

**Figure 1: Study Enrollment Flow Chart**



664

| Table 1: Demographics, hospital treatments, and symptoms by outcome category | | | | |
|---|---|---|---|---|
| Patient Characteristic[a] | Early Symptom Resolution (n=13) | Ongoing Symptomatic COVID-19 (n=4) | Long-COVID (n=10) | p-Value |
| Demographics and Medical | | | | |
| Age (mean [SD]) (yr) | 62.3 (14.3) | 63.8 (13.5) | 62.5 (10.9) | 0.98 |
| Male | 11 (84.6) | 3 (75.0) | 5 (50.0) | 0.19 |
| White | 9 (69.2) | 2 (50.0) | 7 (70.0) | 0.75 |
| African American | 1 (7.7) | 1 (25.0) | 0 (0.0) | 0.27 |
| Hispanic | 3 (23.1) | 2 (25.0) | 3 (30.0) | 0.93 |
| Smoker | 4 (30.8) | 2 (50.0) | 3 (30.0) | 0.75 |
| CCI (mean [SD]) | 4.1 (3.1) | 1.75 (1.5) | 3.2 (2.2) | 0.31 |
| Hypertension | 9 (69.2) | 1 (25.0) | 6 (60.0) | 0.29 |
| Diabetes | 6 (46.2) | 0 (0.0) | 2 (20.0) | 0.15 |
| Chronic Obstructive Lung Disease | 1 (7.7) | 1 (25.0) | 3 (30.0) | 0.37 |
| BMI (mean [SD]) | 30.2 (6.4) | 39.3 (5.3) | 31.5 (4.8) | 0.77 |
| ICU Admission | 2 (15.4) | 1 (25.0) | 1 (10.0) | 0.77 |
| Remdesivir | 5 (38.5) | 4 (100.0) | 6 (60.0) | 0.09 |
| Clinical Trial | 4 (30.8) | 1 (25.0) | 1 (10.0) | 0.49 |
| Longest Lasting Symptoms | | | | |
| Fatigue | 6 (46.2) | 1 (25.0) | 8 (80.0) | 0.11 |
| Respiratory | 10 (76.9) | 3 (75.0) | 9 (90.0) | 0.68 |

| | | | | |
|---|---|---|---|---|
| GI Symptoms | 3 (23.1) | 0 (0.0) | 1 (10.0) | 0.45 |
| Fever | 2 (15.4) | 0 (0.0) | 0 (0.0) | 0.31 |
| Ageusia / Anosmia | 3 (23.1) | 0 (0.0) | 1 (10.0) | 0.45 |
| Confusion / "Brain fog" | 0 (0.0) | 1 (25.0) | 5 (50.0) | 0.017 |
| Duration of Symptoms Days (mean [SD]) | 18.8 (11.5) | 47.8 (5.4) | 80.1 (10.7) | <0.001 |

[a]Data are presented as the number (%), unless otherwise specified.

CCI, Charlson Comorbidity Index; BMI, body mass index; ICU, intensive care unit; Advanced O2, if patients received oxygen beyond nasal canula (i.e. high flow, continuous positive airway pressure); Clinical Trial, if patient received therapy as part of a clinical trial; GI, gastrointestinal.

$\chi^2$ test was used to compare categoric variables and analysis of variance for continuous variables

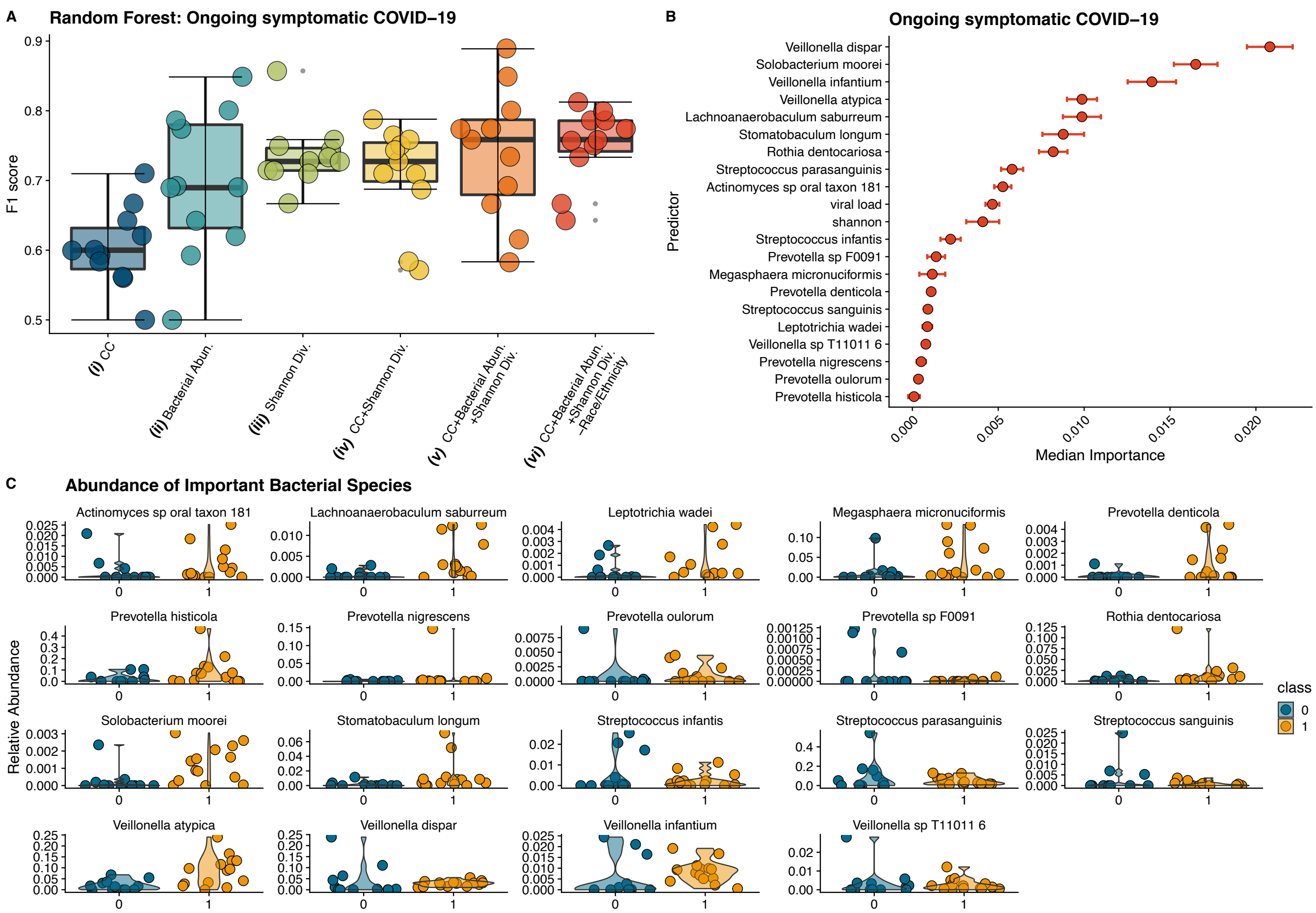



**Figure 2: Bacterial abundances predict ongoing symptomatic COVID-19 disease.** Random forest classification modeling to identify predictors of ongoing symptomatic COVID-19 disease using six different combinations of data modalities. A) F1-scores for the different RFC models trained on different sets of covariates. Boxplot represents the median and interquartile range. B) Ranking of forest predictors based on median permutated variable importance for the top performing model. C) Relative abundances for each bacteria found to be important in predicting ongoing symptomatic COVID-19 disease from the top performing random forest classification model (vi). Violin plots showing the distribution of relative abundance for microbes in each patient with symptoms <4 weeks and >= 4 weeks. 0 indicates No, 1 indicate Yes ongoing symptomatic COVID-19 disease. CC, clinical covariates; Abn., abundances; Div., diversity.

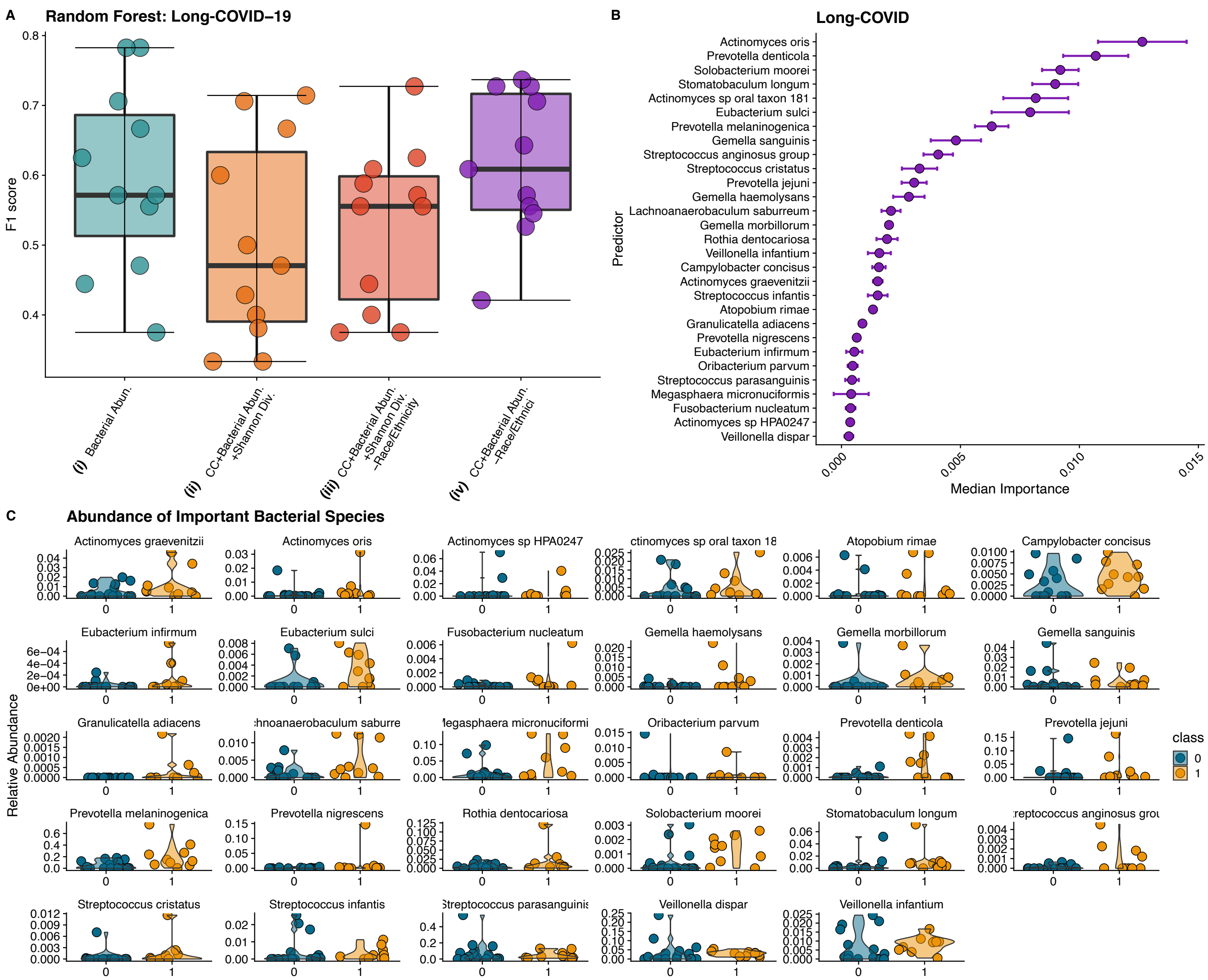



**Figure 3: Bacterial abundances can predict long-COVID-19 disease.** Random forest classification modeling to predict long-COVID-19 disease. A) F1 scores for all subsets of trainable RFC models. B) Ranking of top 29 predictors associated with long-COVID based on median permutated variable importance from the top performing model (iv). C) Relative abundances for each bacteria identified by model (iv) as important for predicting long-COVID-19 disease are presented as violin plots. Long-COVID (orange plots). CC, clinical covariates; Abn., abundances; Div., diversity.

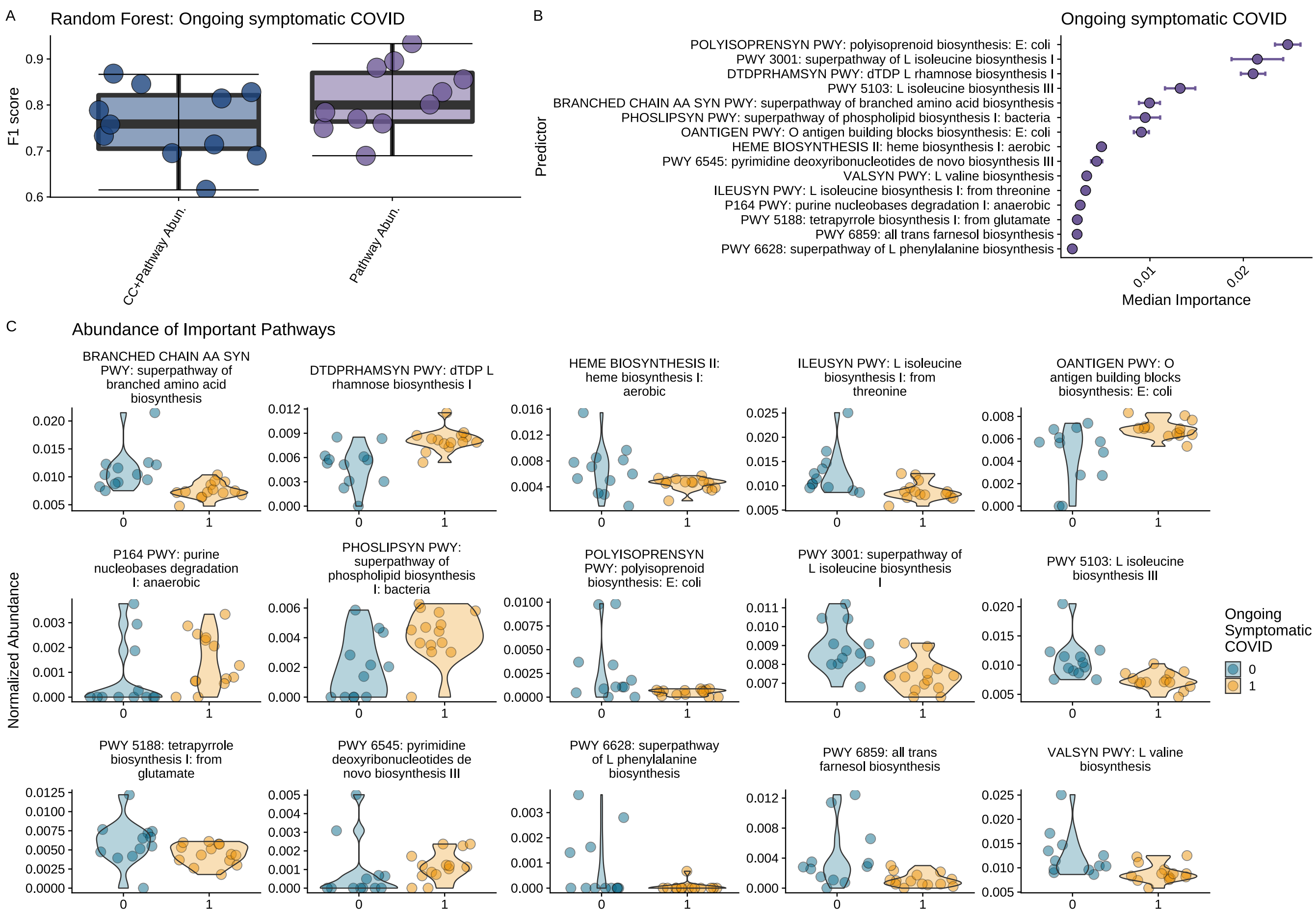



**Figure 4: Bacterial metabolic pathways involving inflammation are significantly associated with ongoing symptomatic disease.** Results from random forest classification modeling using to predict ongoing symptomatic and long-COVID-19 disease from HUMAnN3 pathway abundances. A) F1 scores for (i) demographics + clinical covariates + pathway abundances and, (ii) only on pathway abundances. B) Ranking of forest predictors based on median permutated variable importance from the top performing model, (ii) pathways only, for each outcome C) Relative pathway abundances for each pathway found to be important in predicting ongoing symptomatic and long-COVID-19 disease, respectively, by random forest classification modeling using (ii) only pathway abundances. We report violin plots showing the distribution of the relative abundance of pathways in patients with symptoms with <4 weeks (blue) and > 4 weeks (yellow) in 4C.

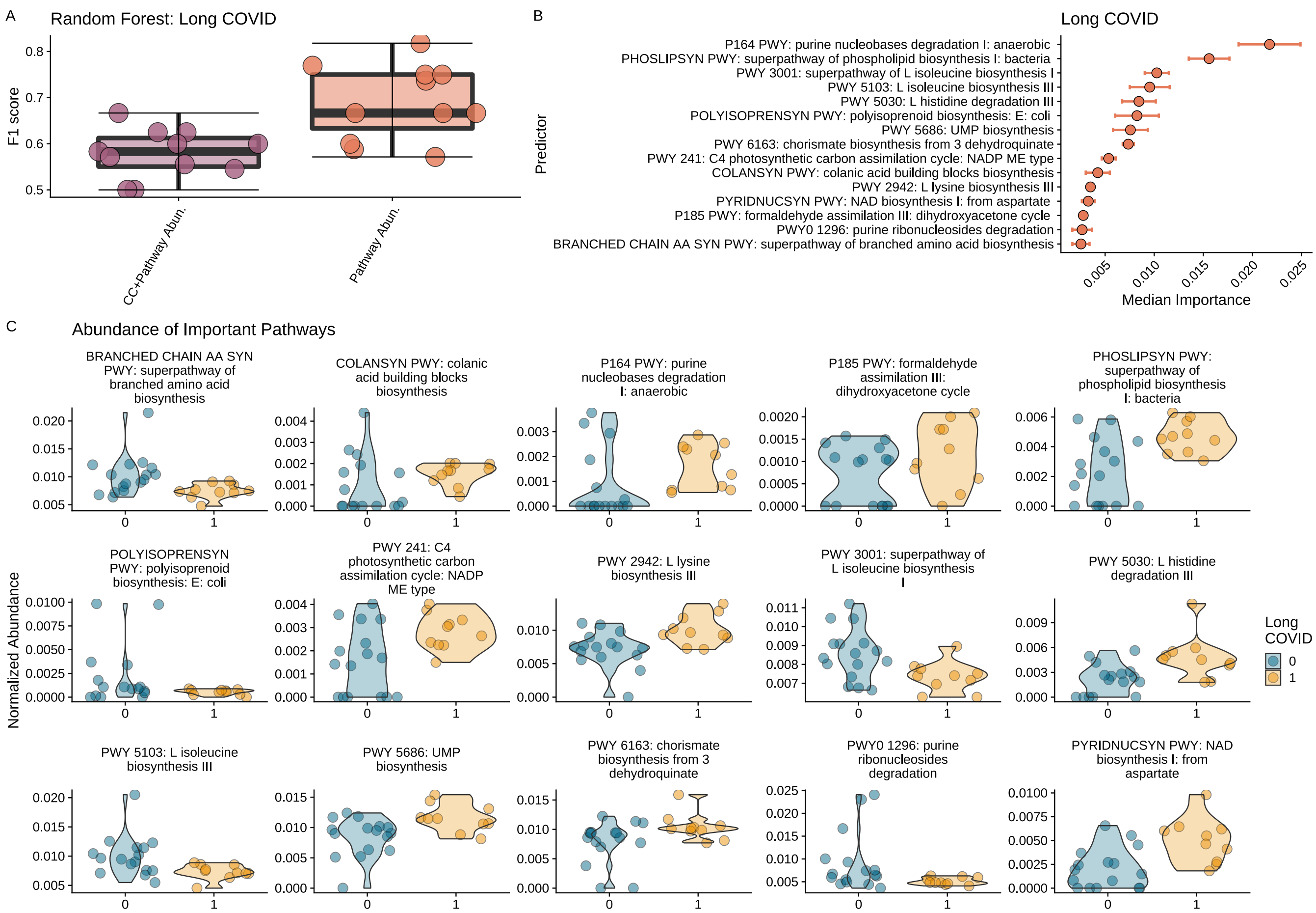



**Figure 5: Bacterial metabolic pathways involving inflammation are significantly associated with long-COVID-19 disease.** Results from random forest classification modeling using to predict ongoing symptomatic and long-COVID-19 disease from HUMAnN3 pathway abundances. A) F1 scores for (i) demographics + clinical covariates + pathway abundances and, (ii) only on pathway abundances. B) Ranking of forest predictors based on median permutated variable importance from the top performing model, (ii) pathways only, for each outcome C) Relative pathway abundances for each pathway found to be important in predicting long-COVID-19 disease, respectively, by random forest classification modeling using (ii) only pathway abundances. We report violin plots showing the distribution of the relative abundance of pathways in patients with symptoms with <10 weeks (blue) and >= 10 weeks (yellow) in 5C.