Metagenomics Approaches to Investigate the Gut Microbiome of COVID-19 Patients

Bioinformatics and Biology Insights
Volume 15: 1–8

DOI: 10.1177/1177932221999428

Sofia Sehli[1,*], Imane Allali[2,*], Rajaa Chahboune[3], Youssef Bakri[2], Najib Al Idrissi[4], Salsabil Hamdi[5], Chakib Nejjari[6,7], Saaïd Amzazi[2] and Hassan Ghazal[1,8]

[1]Department of fundamental sciences, School of Medicine, Mohammed VI University of Health Sciences, Casablanca, Morocco. [2]Laboratory of Human Pathologies Biology, Department of Biology, Faculty of Sciences, and Genomic Center of Human Pathologies, Faculty of Medicine and Pharmacy, University Mohammed V, Rabat, Morocco. [3]School of Medicine and Pharmacy, University Abdelmalek Essaâdi, Tangier, Morocco. [4]Department of Surgery, School of medicine, Mohammed VI University of Health Sciences, Casablanca, Morocco. [5]Environmental Health Laboratory, Department of Research, Institut Pasteur Du Maroc, Casablanca, Morocco. [6]Department of Epidemiology and Biostatistics, International School of Public Health, Mohammed VI University of Health Sciences, Casablanca, Morocco. [7]Department of Epidemiology and Public Health, Faculty of Medicine, University Sidi Mohammed Ben Abdellah, Fez, Morocco. [8]Scientific Department, National Centre for Scientific and Technical Research, Rabat, Morocco. *These authors have contributed equally to this work as co-first authors.

ABSTRACT: Over the last decade, it has become increasingly apparent that the microbiome is a central component in human well-being and illness. However, to establish innovative therapeutic methods, it is crucial to learn more about the microbiota. Thereby, the area of metagenomics and associated bioinformatics methods and tools has become considerable in the study of the human microbiome biodiversity. The application of these metagenomics approaches to studying the gut microbiome in COVID-19 patients could be one of the promising areas of research in the fight against the SARS-CoV-2 infection and disparity. Therefore, understanding how the gut microbiome is affected by or could affect the SARS-CoV-2 is very important. Herein, we present an overview of approaches and methods used in the current published studies on COVID-19 patients and the gut microbiome. The accuracy of these researches depends on the appropriate choice and the optimal use of the metagenomics bioinformatics platforms and tools. Interestingly, most studies reported that COVID-19 patients' microbiota are enriched with opportunistic microorganisms. The choice and use of appropriate computational tools and techniques to accurately investigate the gut microbiota is therefore critical in determining the appropriate microbiome profile for diagnosis and the most reliable antiviral or preventive microbial composition.

KEYWORDS: COVID-19, SARS-CoV-2, human microbiome, bioinformatics, metagenomics, gene marker analysis, 16S rRNA gene, metatranscriptomics, metaproteomics

RECEIVED: January 28, 2021. **ACCEPTED:** February 9, 2021.

TYPE: Review

FUNDING: The author(s) received no financial support for the research, authorship, and/or publication of this article.

DECLARATION OF CONFLICTING INTERESTS: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

CORRESPONDING AUTHOR: Hassan Ghazal, National Centre for Scientific and Technical Research, Corner of FAR and Allal El Fassi avenues, Hay Ryad, B.P. 8027 N.U., Rabat 10102, Morocco. Email: hassan.ghazal@fulbrightmail.org

Introduction

The evolution of affordable next-generation sequencing (NGS) technologies transformed microbial research from culture-based to genome-based methods, named metagenomics.[1] This approach has been used to explore the microbial communities that interfere with the human host cells.[2] Infectious pathologies have generally relied on one disease-equivalent pathogen.[3] Nevertheless, we now realize that microbiota dysbiosis is often correlated with several distinct disease states. In many cases, a decline in the microbial diversity and growth of some particular species may contribute to harmful consequences, such as inflammation or infections.[4] Studies have given evidence of the possible participatory of the microbiome in nearly all forms of complications associated with our health, including COVID-19.

Nowadays, many researchers focus on studying the composition of the gut microbiota in relation to a disease, such as the COVID-19 caused by the SARS-CoV-2 virus.[5] A multitude of different microorganisms lives in our gut. Hence, imbalances in the composition of the intestinal microbes may induce an intestinal microbiota dysbiosis.[5] Studies in mice and humans have revealed the existing correlation between the gut microbiota dysbiosis and disease through a large scope of chronic disorders.[5] For instance, Allali et al[6] found relevant differences between the microbiota composition of healthy individuals and patients with colorectal cancer.[6] This is noteworthy that the angiotensin-converting enzyme (ACE2), known to be the host cell entry molecule of SARS-CoV-2, is found at high-level concentrations in the gastrointestinal epithelial cells and regulates intestinal inflammations. In fact, the gut microbiota is influenced by the ACE2 in an indirect way, which may indirectly induce a cardiopulmonary risk.[7]

Yet, microbiome studies yield large data that require advanced computing methods, and the technologies used are constantly progressing. In addition, researchers also need to maintain vigilance that microbiome classification, data processing, and modelling is just a tiny component of the process of discovery and should be used to supplement standard *in vitro* and *in vivo* modelling approaches. Nevertheless, the use of metagenomics is changing microbiology by specifically quizzing the group

Table 1. List of published Microbiome studies in COVID-19 patients by September 30, 2020.

STUDY APPROACH	COUNTRY	TOOLS	RESULTS	PUBLICATION
WGS metagenomics	Hong-Kong	MetaPhlAn2	Enrichment of opportunistic microorganisms and reduction of helpful commensals in COVID-19 patients	[12]
Gene marker analyses (16S)	China	UPARSE	Increased relative abundance of *Enterococcus* and *Rhodococcus*, and decreased relative abundance of *Faecalibacterium* and *Clostridium* XIVa in COVID-19 patients	[13]
Metatranscriptomics	China	HUMAnN2	Increased transcriptional activities of *Escherichia coli* and *Klebsiella pneumoniae*, virulence factors, and antibiotic resistance genes, and decreased activities of *Faecalibacterium prausnitzii* in patients of COVID-19	[13]
Metaproteomics	China	Machine-learning model	The proinflammatory cytokines were mainly found among elderly individuals compared to younger ones and that they were positively associated with the proteomic risk score	[14]
WGS metagenomics	China	BWA, MEGAHIT	Microbiota of COVID-19 patients was found to contain the most known pathogens	[15]

composition in an impartial course, allowing further species identification and decreasing dependency on cultural approaches[8,9] (p. 4). The future application of these tools to enhance diagnostics and in public health environments has already been acknowledged on a large scale.[7,10] Comprehensive ongoing projects are being conducted to resolve the issues with the therapeutic employment of these techniques.[7,11]

Herein, we present and discuss bioinformatics methods and tools that are available to characterize the influence of the gut microbiome on the evolution of the COVID-19 disease or to evaluate the impact of the disease on that microbiome (Table 1). We will also describe the studies on the human gut microbiota in relationship to the SARS-CoV-2 virus infection and discuss how these microbes would be involved in the fight against harmful microorganisms.

Bioinformatics Tools for Studying the Microbiome

There are a variety of technologies accessible to research the gut microbiome. One of these technologies is gene marker analysis that uses the NGS platforms for sequencing. The common softwares used with it are QIIME2,[16] Mothur,[17] VEGAN,[18] phyloseq,[19] and DADA2.[20] The latter tools are cost-effective, and its analytical pipelines are widely accepted but lack clear functional information and can produce errors in the taxa differentiation.[21] Another widely used technology is the shotgun metagenomics, even though it is expensive and computationally demanding, it is a great approach to capture all microbial genomes within the sample. Mostly used softwares to analyse shotgun metagenomics data include IDBA-UD,[22] SPades,[23] MEGAHIT,[24] MetaPhlAn2,[25] MG-RAST,[26] and HUMAnN2.[27] Metatranscriptomics is another approach for studying the microbiome to assess gene expression level with the SOAPdenovo,[28] the commonly used software. Another approach that allows the identification and quantification of proteins within the sample is metaproteomics, but this technology does not give the protein's origin.[29] Metabolomics is a further technology used to perform profiling of the metabolites produced by the gut microbiota.[29]

Shotgun Metagenomics to Perceive the Microbiome in Patients with COVID-19

One of the first studies on the gut microbiome in COVID-19 patients has been carried out in Hong Kong.[12] The authors performed metagenomic analysis of 15 faecal samples extracted from patients with COVID-19. Bacterial population profiling was done using MetaPhlAn2 (v2.9),[25] which is a computational tool for profiling the composition of microbial communities from metagenomic shotgun sequencing data.[25] The obtained reads have been mapped against clade-specific markers, that are defined as coding sequences highly preserved within the genomes of the clade and that have no local similarity outside the clade with any sequence.[12] As a result, intestinal microbiome profiles correlated with disease severity and higher faecal release of SARS-CoV-2. Zuo et al[12] observed that patients with COVID-19 have substantial changes in faecal microbiome relative to the control, marked by enrichment of opportunistic microorganisms and reduction of helpful commensals, within hospitalization and at all times throughout hospitalization.[12] Depleted symbiotics and intestinal dysbiosis continued well after the elimination of SARS-CoV-2, assessed from the throat swabs, and the relief of respiratory problems.[30] The residual abundance of *Coprobacillus*, *Clostridium hathewayi*, and *Clostridium ramosum* was associated with the infection's severity; the abundance of *Faecalibacterium prausnitzii*, an anti-inflammatory bacterium, was inversely correlated with the disease severity. In the middle of hospitalization, *Bacteroides dorei*, *Bacteroides massiliensis*, *Bacteroides thetaiotaomicron*, and *Bacteroides ovatus*, that reduces the expression of ACE2 in the murine intestine, are inversely linked with SARS-CoV-2 content in faecal samples of patients.[30]

In another study, Zuo et al[31] (p. 19) conducted RNA metagenomic sequencing of repeated faecal viral extractions

taken from 15 hospitalized patients with COVID-19. They analysed faecal microbiome diversity and the gut microbiota functionality, in accordance with faecal SARS-CoV-2 infectivity profiles[31 (p. 19)]. The authors used MetaPhlAn2 (V2.9)[25] to perform taxonomic profiling of the faecal bacterial communities by mapping reads to clade-specific markers[31 (p. 19)]. HUMAnN V2.0.18[27] was used to perform functional profiling of faecal bacterial communities[31 (p. 19)]. Seven out of 15 patients had stool positivity for SARS-CoV-2 by viral RNA metagenomic sequencing. Furthermore, in the absence of gastrointestinal manifestations, the 7 patients displayed significantly increased coverage and density of the 3′ versus 5′ end of the coronavirus genome in their faecal viral metagenome samples[31 (p. 19)]. The faecal viral metagenome of 3 cases persisted to exhibit the signature of active viral infection[31 (p. 19)]. Faecal samples containing elevated SARS-CoV-2 infectiousness had greater concentrations of the bacterial species, *Morganella morganii*, *Collinsella tanakaei*, *Streptococcus infantis*, and *Collinsella aerofaciens*; while faecal samples with reduced-to-none SARS-CoV-2 infectivity had elevated abundance of bacteria that produce short-chain fatty acid, namely *Bacteroides stercoris*, *Parabacteroides merdae*, and *Alistipes onderdonkii*[31 (p. 19)].

In addition, Wu et al[13] examined in another study the intestinal microbiome properties of a group of COVID-19 patients throughout probiotic-assisted therapy. Wu et al[13] used UPARSE[21] to assign each operating taxonomic unit (OTU) representative sequences to a taxonomic level in the RDP database, with aid of the RDP classifier at an 80% confidence level. The authors also conducted a metatranscriptomic analysis of faecal samples using an Illumina Hiseq 4000 platform. The resulting reads were taxonomically profiled using Kraken,[13] and the relative abundance of metabolic pathways in MetaCyc[32] database was calculated using HUMAnN2.[27] ShortBRED[33] was used to quantify the abundance of antibiotic resistance genes and virulence genes against the CARD database and Virulence Factors Database.[13]

Furthermore, to look for any potential agent causing pneumonia besides the coronavirus, a metagenomic study was conducted using samples of the BronchoAlveolar Large Fluid (BALF) from a patient held in the intensive care unit. The aforementioned BALF samples were used for RNA extraction and NGS sequencing using both BGI MGISEQ 2000 and Illumina MiSeq 3000 sequencers, while the metagenomic analysis was carried through the MGmapper[34] bioinformatics platform. The raw NGS reads were primarily analysed by Cutadapt (v.1.18)[35] and alignment performed using BWA (v.0.7.12-r1039)[36] against a local database. To filter reads of the hosts' genomes before aligning them against the virus database, Zhou et al[37] used a local nucleic acid database for humans and mammals. The authors used Geneious (v.11.0.3) and MEGAHIT[24] (v.1.2.9) to assemble the NGS reads. The Clone Manager Professional Suite 8 (Sci-Ed Software) was used to annotate the genomes. Results showed that more than 87% of the sequences retrieved from the BALF matched perfectly with the SARS-CoV-2 genome,[37] while the remaining ones belonged to six other viruses. Therefore, the most potentially harmful agent to be considered was the SARS-CoV-2.[37] According to the same study, a heat map of the lung microbiota composition was done by clustering the microbiota into 3 different groups (Type I, Type II, and Type III). Type I microbiota had the most pathogens, while Type II and III, respectively, contained environmental organisms and commensal species. The microbiota of patients with COVID-19 was clustered more in type I, which is considered as the most pathogenic one.[37]

Another recent study analysed alterations in the mycobiome of 30 patients with COVID-19 by Zuo et al,[38] reporting that these patients compared to the controls had significant alterations in their faecal mycobiome, distinguished by the enrichment of *Candida albicans* and a highly heterogeneous mycobiome configuration, at the time of hospitalization.[38] With that being said, 22 of the 30 patients with COVID-19 did not significantly change from the controls during the hospitalization period.[38] Therefore, more studies are needed to conclude if the mycobiome changes and enrichment of fungal pathogens contribute to the COVID-19 progression or can be used as a predictor in it.[38]

In addition to that Wang et al[39] performed some analysis on 159 Italian patients with pneumonia using a phage-display method to characterize circulating antibodies binding to 93,904 viral peptides encoded by 1,276 strains of human viruses. These researchers developed VirScan's, a tool that predicts SARS-CoV peptides and its clinical severity, an effective tool used to detect SARS-CoV-2 antibodies in the host plasma.[39]

Actually, some researchers are suggesting phages as a means of therapy that could build a protective barrier to eukaryotic virus particles by an increase in phage transcytosis by epithelial cells, given that lung epithelium is also involved in transcytosis of phages.[40] Therefore, such a phenomenon may play an intriguing role in protecting those lung epithelium cells from invasion by coronaviruses. Nevertheless, further studies need to be done to conclude whether phages have the potential to at least be an adjunct treatment of the SARS-CoV-2 infection.[40]

Surprisingly, none of the analysed studies measured immunity in dependence of the gut flora. However, Donati Zeppa et al,[41] stated that the various responses to the virus infection may be explained by an adaptive immune system that is not effective enough, or/and pneumonia can begin before the immune system responds.[6] The first line of defence against SARS-CoV-2 is innate immunity, the response of which, unlike the adaptive response, is activated within a few hours of the infection.[6] The natural history of the disease is determined by this first encounter between the

innate immunity of the host and SARS-CoV-2 virus, and by exposure over the following 2 weeks to the virus.[41]

As in another study by Yeoh et al,[42] the gut microbiota composition of COVID-19 patients was found to be consistent with the infection's severity and that many inflammatory cytokines, chemokines, and blood markers of tissue damage were found with immense concentrations in the host plasma.[42] In addition to that, the authors found that patients with COVID-19 were depleted in gut bacteria with known immunomodulatory potential, namely *Faecalibacterium prausnitzii*, *Eubacterium rectal*, and numerous bifidobacterial species.[42]

However, this aspect of intestinal microbiota alterations in combination with immune dysregulation has shown that intestinal microorganisms are likely to engage in modulating host inflammatory reactions in COVID-19 infection.[42] With proof that intestinal microorganisms are associated with inflammatory diseases inside and outside the intestine, these discoveries underscore an important need to consider the particular involvement of intestinal microorganisms in human immune function and systemic inflammation.[42]

From studies that performed taxonomic profiling, all COVID-19 patients (mild, moderate, critical, and severe) had an increase in pathogenic and opportunistic microorganisms, wherein severe cases had a further depletion of many bacterial species that are commensal microorganisms beneficial to the healthy and effective immunity of the host including *Faecalibacterium prausnitzii* that was very recurrent[12] (see Table 2). Twenty-three bacterial taxa were found to be significantly associated with COVID-19 disease severity, most of which (15 of 23) were from the Firmicutes phylum, among them, Erysipelotrichia and Actinobacteria classes showed positive correlation with the disease severity, while Gammaproteobacteria, Bacteroidia, and Clostridia classes showed negative correlation with disease severity.[12]

Statistical studies (PERMANOVA test) on host's factors that mostly affect the gut microbiota of COVID-19 patients showed that SARS-CoV-2 infection and antibiotics treatment affect the gut microbiota, while age and gender had no significant correlation with microbiome alteration.[12]

A Metatranscriptomic Approach to Assess the Expression of the SARS-CoV-2 Host Receptor Molecule

Furthermore, a metatranscriptomic analysis has been recently performed, to profile the transcriptionally active gut microbiota in patients with different types of pneumonia, which is associated with the immunity response in the lung. This study was conducted including 8 COVID-19 patients. Data from 25 patients with community-acquired pneumonia (CAP) and 20 healthy controls, have been analysed for metatranscriptome comparison.[15] Quality control comprised the trimming of adapters and the elimination of reads with low quality, using fastp software (version 0.20.0).[43] Komplexity software was used to remove low-complexity reads, Shen et al[15] used the bmtagger software to remove the hosts' reads and SortMeRNA software (version 2.1b)[44] to remove ribosomal reads. BLAST + software (version 2.9.0)[44] was used to map the subsequent reads against the NCBI Nucleotide Database. In addition, Shen et al[15] conducted a taxonomic classification using MEGAN programme (version 6.11.0).[45] Following a per-mutational multivariate analysis of variance and general key coordinate analysis, samples, and microorganisms were selected for advanced studies. Samples containing less than 5000 microbial reads have been dismissed. As a result, only samples from the BALF belonging to COVID-19 patients had SARS-CoV-2, yet some mild β-coronaviruses species were detected in the healthy and CAP patients.[15]

In addition, Shen et al[15] used the BWA-MEM package (version 0.7.12)[36] to map clean metatranscriptomics reads against the COVID-19 reference genome.[15] The authors eliminated duplicated reads using Picard tool (version 2.18.22),[46] then created the mpileup file using samtools software (version 1.8)[47] and used VarScan software (version 2.3.9)[48] to define intrahost variants. For the intrahost variants in the genome of SARS-CoV-2, 84 variants were identified with a minor allele frequency (MAF) greater than 5%, and 25 variants were detected with MAF higher than 20%. Therewith, the variant number was proportional to the gene length, but only 2 out of 84 of these variants were identified in multiple patients.[15] The analysis of the lung microbiota revealed a difference between CAP, COVID-19 patients, and the healthy group, implying a dysbiosis in the lung microbiota of the unhealthy CAP and COVID-19 patients.[15]

Gene Marker Analyses Approach to Discern Key Underlying Factors of COVID-19 Disparity

To identify one of the key underlying factors of COVID-19 disparity, in terms of the disease's severity, a very recent study suggests that the composition of the gut microbiota could partially explain the difference in susceptibility.[49] As in a set of 336 individuals, these gut microbiota features were highly correlated with proinflammatory cytokines[14] (p. 19). The authors used UPARSE[21] to cluster sequences that have a 97% similarity into one OTU, and the RDP classifier to assign the OTUs taxonomy and to align sequences[14] (p. 19). They additionally used the QIIME software 1.9.0 to analyse OTUs.[16] Hence, the predisposition of normal individuals to severe COVID-19 may be predicted by the gut microbiome, which brings a completely new aspect of what is currently understood about the virus[14] (p. 19).

In another study, Tao et al[50] used the 16S rRNA amplicon profiling to investigate the possible effect of SARS-CoV-2 infection on the intestinal flora composition. The authors used a customized pipeline that combined USEARCH (v8.1),[49]

Table 2. A comparative evaluation of different metagenomics studies of COVID-19 patients.

STUDY TITLE	AIM	METHODS	DATA SET	RESULTS
Alterations in Gut Microbiota of Patients With COVID-19 During Time of Hospitalization (Zuo et al, 2020)[12]	Study the variation of the faecal microbiome of COVID-19 patients during hospitalization	Isolation of the faecal microbiome of 15 COVID-19 patients during the hospitalization 2 to 3 time per week, compare them with those of 15 healthy subjects and 6 patients with common pneumonia, study the association of the microbiome variation with the clinical status of the patient (mild, moderate, severe, or critical)	WGS metagenomics data Bioproject: PRJNA624223	• Significant alterations of faecal microbiomes of COVID-19 patients compared with controls. • Enrichment of COVID-19 patients' microbiome with opportunistic pathogens and depletion of beneficial commensals from the beginning and during hospitalization. Gut dysbiosis persistence even after clearance of SARS-CoV-2. • Positive correlation between the severity of the disease and the baseline abundance of Coprobacillus, *Clostridium ramosum*, and *Clostridium hathewayi* • Negative correlation between COVID-19 severity and the abundance of *Faecalibacterium prausnitzii*.
The Volatile and Heterogenous Gut Microbiota Shifts of COVID-19 Patients Over the Course of a Probiotics-Assisted Therapy (Wu et al, 2020)[13]	Study the association of transcriptional activity of SARS-CoV-2 with longitudinal faecal microbiome alterations in patients with COVID-19.	Performing serial faecal viral extractions from 15 hospitalized COVID-19 patients and sequencing RNA using shotgun metagenomics, then assessing the faecal microbiome composition and microbiome functionality and its association with signatures of faecal SARS-CoV-2 infectivity	Data are available on request from the authors	• Positivity for viral RNA metagenomics in 46.7% of patients with COVID-19, even without gut-intestinal clinical manifestations. • Persistence of viral metagenomes in 3 patients even after clearance of the virus. • Faecal samples with signature of high SARS-CoV-2 infectivity had higher abundances of bacterial species *Collinsella aerofaciens*, *Collinsella tanakaei*, *Streptococcus infantis*, *Morganella morganii*, and higher functional capacity for nucleotide de novo biosynthesis, amino acid biosynthesis, and glycolysis, whereas faecal samples with signature of low-to-none SARS-CoV-2 infectivity had higher abundances of short-chain fatty acid producing bacteria, *Parabacteroides merdae*, *Bacteroides stercoris*, *Alistipes onderdonkii*, and *Lachnospiraceae bacterium* 1_1_57FAA.
Gut Microbiota May Underlie the Predisposition of Healthy Individuals to COVID-19 (Gou et al, 2020)[14]	Construction of a 'proteomic risk' score based on biomarkers to predict the progression of the severity.	Building a proteome risk score using machine learning on data set of biomarkers from 301 individual	16S rRNA of stool samples of COVID-19 patients and healthy subjects CNGBdb ID: CNP0000829	• Gut microbiota features are highly correlated with proinflammatory cytokines • Potential amino acid-related pathways linking gut microbiota to inflammation
Genomic Diversity of Severe Acute Respiratory Syndrome-Coronavirus 2 in Patients With Coronavirus Disease 2019 (Shen et al, 2020)[15]	Exploring the interaction between the virus and the lung microorganisms	Conducting a metatranscriptome sequencing for the bronchoalveolar lavage fluid of 8 SARS-CoV-2 patients, 25 community-acquired pneumonia patients, and 20 healthy controls	RNA-seq from stools of 8 COVID-19 patients, 25 CAP patients, and 20 healthy subjects Bioproject ID: PRJNA605907	• The median number of intrahost virus variants was 1 to 4 in SARS-CoV-2 infected patients, which ranged between 0 and 51 in different samples. • The distribution of variants on genes was similar to those observed in the Chinese population data. • Very few intrahost variants were observed in the viral genome in the population as polymorphism, implying either a bottleneck or purifying selection involved in the transmission of the virus, or a consequence of the limited diversity represented in the current polymorphism data of the virus. • Current evidence did not support the transmission of intrahost variants in a person-to-person spread; the risk should not be overlooked. • The microbiota in SARS-CoV-2 infected patients was similar to those in CAP, either dominated by the pathogens or with elevated levels of oral and upper respiratory commensal bacteria.

Table 3. A table summarizing the methods of performing metagenomic studies.

STUDY APPROACH	STUDY TITLE	SAMPLE	METHOD	BIOINFORMATICS TOOL USED IN THE STUDIES CITED
WGS metagenomics	Alterations in Gut Microbiota of Patients With COVID-19 During Time of Hospitalization[12]	Metagenomics sequencing data set is available under the following BioProject accession number PRJNA624223.	The authors performed shotgun metagenomic sequencing analyses of faecal samples from 15 patients with COVID-19	The profiling of bacterial communities was done using Metaphlan2 V2.9 which was done by mapping reads to specific clade markers.
Gene marker analysis	The Volatile and Heterogenous Gut Microbiota Shifts of COVID-19 Patients Over the Course of a Probiotics-Assisted Therapy[13]	The cohort study included 13 hospitalized COVID-19 patients, 15 patients with pneumonia and 15 healthy controls.	DNA samples were extracted targeting the V3V4 regions of the 16S rRNA gene	Pandaseq was used to merge paired reads, then Uparse was used to cluster sequence reads into operational taxonomic units (OTUs) and taxonomy profiling against the RDP database
Metatranscriptomics	The Volatile and Heterogenous Gut Microbiota Shifts of COVID-19 Patients Over the Course of a Probiotics-Assisted Therapy[13]	The cohort study included 13 hospitalized COVID-19 patients, 15 patients with pneumonia and 15 healthy controls.	Metatranscriptomic sequencing of samples was conducted on Illumina Hiseq 4000 platform with 150 bp paired-end read length.	• Kraken[13] was used to perform a taxonomy profiling of the metatranscriptomic reads. • Then the relative abundance of metabolic pathways in MetaCyc database was calculated using HUMAnN2. • Antibiotic resistance genes and virulence genes against the CARD database were quantified using ShortBRED. • StrainPhlAn was used to perform a strain-level profiling. • Bowtie2 was used to map reads against the MetaphlAn2 markers. • A multiple sequence alignment was performed of the consensus sequences of references and samples with MUSCLE. • RAxML was used to build a phylogenetic tree.
WGS Metagenomics	Alterations in Faecal Fungal Microbiome of Patients With COVID-19 During Time of Hospitalization until Discharge[38]	Faecal fungal mycobiome samples from 30 COVID-19 patients, 9 subjects with community-acquired pneumonia and 30 healthy individuals	Authors performed shotgun metagenomic sequencing analysis of faecal samples.	• Kneaddata was used to filter human reads contamination. • MiCoP was used to perform fungal taxonomy. • Vegan package in *R* was used to perform a nonmetric multidimensional scaling analysis based on Bray-Curtis dissimilarities.
Metatranscriptomics	Metatranscriptomic Characterization of Coronavirus Disease 2019 Identified a Host Transcriptional Classifier Associated With Immune Signaling[53]	RNA-seq isolated from the nasopharyngeal swab extracted from a cohort of 113 patients and a validation cohort of 74 patients	• RNA-seq isolated with the QIAamp ViralRNA mini kit, filtered from human rRNA, then reverse transcripted, and sequenced using Illumina NextSeq sequencer. • The alpha diversity of the respiratory microbiome for each patient was assessed using the Shannon diversity index (SDI). • Diversity values were then compared between patients with and without SARS-CoV-2 infection within each group using the Wilcoxon rank sum testSpecies with differential abundance were identified within each group using DESeq2 in R at a false discovery rate (FDR) $\leqslant 0.1$, fold change $\geqslant 2$, and $P \leqslant .05$.	• Vegan package in *R* was used to study the alpha diversity using the Shannon Wiener index. • DESeq 2 was used to study the differential abundance of species.

VSEARCH (v2.13.0),[51] and QIIME (v1.9.1)[16] to analyse data for microbial diversity. Unfortunately, the aforementioned tools do not give clear functional information and may give errors in taxa differentiation. Alpha-diversity analysis demonstrated that the gut microbiota composition was less diverse in COVID-19 patients in contrast with both flu patients and control cases.[50] Assessed by weighted UniFrac,[52] the genus level in each group revealed that the abundance and composition of faecal bacteria in COVID-19 patients varied from those in both control cases and seasonal influenza patients.[50] The increased abundance of *Streptococcus* in COVID-19 patients was indicative of the risk of infection by opportunistic pathogenic bacteria in this group. However, the approach used in this study was not enough to capture all microbial genomes within samples and thus not reaching the species level.[50]

Conclusion

Pithily, we are on the brink of a quickly evolving research field that holds an enormous opportunity to clarify and describe microbial interactions on the human interrelatedness. Early in the COVID-19 pandemic, researchers smartly pointed to the microbiome as a key element in understanding the etiology, infection, and transmissibility processes of the emerging SARS-CoV-2 virus. Three out of 5 studies reported that there was (1) a significant enrichment of opportunistic microorganisms such as *Clostridium hathewayi*, *Actinomyces viscosus*, *Bacteroides nordii*;[29] (2) reduction of helpful commensals in COVID-19 patients, and (3) an interesting decrease of *Faecalibacterium prausnitzii* in patients with COVID-19.[1-3] *Faecalibacterium prausnitzii* plays an important role in promoting gut health.[4] The aforementioned species may be a useful potential biomarker in diagnostics and prognostics for certain diseases, such as Crohn's disease, and ulcerative colitis.[4] *Faecalibacterium prausnitzii* has frequently been identified as one of the major butyrate contributors in the gut.[4] Butyrate plays a significant part in the physiology of the intestines and in the well-being of the host.[4] Butyrate can minimize inflammation in the intestinal mucosa by inhibiting the activation of NF-κB transcription factor, up-regulating PPARγ, and inhibiting interferon gamma (IFN-γ).[4] Moreover, anti-inflammatory properties have been linked to this species due to its capacity to cause a tolerogenic cytokine profile (with quite low secretion of proinflammatory cytokines such as interleukin [IL]-12 and IFN-γ) and enhanced secretion of anti-inflammatory cytokines IL-10.[4]

We reported these few studies to shed further light on this promising route for new diagnostics and therapeutic strategies, yet we acknowledge that it is too early to make strong conclusions. It is noteworthy that most available metagenomics tools do not reach the optimal level of accuracy. After all, most taxonomic classifiers are also encumbered by a large number of false positives at the poor abundance that needs to be discussed. Further than this, significant advances and innovation will be required in so many other ways, namely in managing experimental contamination sources and bias and in managing the rapid growth of reference databases, to establish world-changing improvements in metagenomics classification towards microbial identification and classification (Table 3). Actually, our current scientific information and understanding of the gut microbiota and COVID-19 relationship is less than precise but continues to evolve fast. Thus, understanding how to choose the computational tools and strategies to analyse efficiently the gut microbiota is one important thing to decipher the most pertinent microbiome profile for diagnostics and the precise antiviral or preventive microbial composition.

Author Contributions

S.S: Draft writing and Data acquisition; I.A: Draft writing and Data acquisition; R.C: Critical revision of the article; Y.B: Critical revision of the article; N.A: Critical revision of the article; S.H: Critical revision of the article; C.N: Conceptualisation; S.A: Conceptualisation; H.G: Conceptualisation, Design of the work and writing.

ORCID iD

Hassan Ghazal https://orcid.org/0000-0002-0638-2180

REFERENCES

1. Riesenfeld CS, Schloss PD, Handelsman J. Metagenomics: genomic analysis of microbial communities. *Annu Rev Genet*. 2004;38:525-552.
2. Song E-J, Lee E-S, Nam Y-D. Progress of analytical tools and techniques for human gut microbiome research. *J Microbiol*. 2018;56:693-705.
3. Bantock GG. The modern doctrine of bacteriology, or the germ theory of disease. *Br Med J*. 1899;1:846-848. doi:10.1136/bmj.1.1997.846.
4. Belkaid Y, Hand T. Role of the microbiota in immunity and inflammation. *Cell*. 2014;157:121-141. doi:10.1016/j.cell.2014.03.011.
5. Lynch SV, Pedersen O. The human intestinal microbiome in health and disease. *N Engl J Med*. 2016;375:2369-2379. doi:10.1056/NEJMra1600266.
6. Allali I, Boukhatem N, Bouguenouch L, et al. Gut microbiome of Moroccan colorectal cancer patients. *Med Microbiol Immunol*. 2018;207:211-225.
7. Miller S, Naccache SN, Samayoa E, et al. Laboratory validation of a clinical metagenomic sequencing assay for pathogen detection in cerebrospinal fluid. *Genome Res*. 2019;29:831-842.
8. Knights D, Kuczynski J, Charlson ES, et al. Bayesian community-wide culture-independent microbial source tracking. *Nat Methods*. 2011;8:761-763. doi:10.1038/nmeth.1650.
9. Loman NJ, Constantinidou C, Christner M, et al. A culture-independent sequence-based metagenomics approach to the investigation of an outbreak of Shiga-toxigenic Escherichia coli O104:H4. *JAMA*. 2013;309:1502-1510. doi:10.1001/jama.2013.3231.
10. Miller RR, Montoya V, Gardy JL, Patrick DM, Tang P. Metagenomics for pathogen detection in public health. *Genome Med*. 2013;5:81.
11. Chiu CY, Miller SA. Clinical metagenomics. *Nat Rev Genet*. 2019;20:341-355. doi:10.1038/s41576-019-0113-7.
12. Zuo T, Zhang F, Lui GCY, et al. Alterations in gut microbiota of patients with COVID-19 during time of hospitalization. *Gastroenterology*. 2020;159:944-955.
13. Wu C, Xu Q, Cao Z, et al. The volatile and heterogenous gut microbiota shifts of COVID-19 patients over the course of a probiotics-assisted therapy [Published online ahead of print January 21, 2021]. *Res Sq*. doi:10.21203/rs.3.rs-72753/v1.
14. Gou W, Fu Y, Yue L, et al. Gut microbiota may underlie the predisposition of healthy individuals to COVID-19 [Published online ahead of print April 25, 2020]. *medRxiv*. doi:10.1101/2020.04.22.20076091.
15. Shen Z, Xiao Y, Kang L, et al. Genomic diversity of severe acute respiratory syndrome-coronavirus 2 in patients with coronavirus disease 2019. *Clin Infect Dis off Publ Infect Dis Soc Am*. 2020;71:713-720. doi:10.1093/cid/ciaa203.
16. Caporaso JG, Kuczynski J, Stombaugh J, et al. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods*. 2010;7:335-336.
17. Schloss PD, Westcott SL, Ryabin T, et al. Introducing Mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol*. 2009;75:7537-7541.

18. Oksanen J, Blanchet FG, Friendly M, et al. *Vegan: community ecology package*, https://CRAN.R-project.org/package=vegan (2020, accessed 28 January 2021).
19. McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE*. 2013;8:e61217.
20. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: high-resolution sample inference from Illumina amplicon data. *Nat Methods*. 2016;13:581-583.
21. Edgar RC. UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat Methods*. 2013;10:996-998.
22. Peng Y, Leung HCM, Yiu SM, Chin FYL. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics*. 2012;28:1420-1428. doi:10.1093/bioinformatics/bts174.
23. Bankevich A, Nurk S, Antipov D, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol*. 2012;19:455-477.
24. Li D, Liu C-M, Luo R, Sadakane K, Lam T-W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics*. 2015;31:1674-1676. doi:10.1093/bioinformatics/btv033.
25. Segata N, Waldron L, Ballarini A, Narasimhan V, Jousson O, Huttenhower C. Metagenomic microbial community profiling using unique clade-specific marker genes. *Nat Methods*. 2012;9:811-814. doi:10.1038/nmeth.2066.
26. Meyer F, Paarmann D, D'Souza M, et al. The metagenomics RAST server – a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics*. 2008;9:386. doi:10.1186/1471-2105-9-386.
27. Franzosa EA, McIver LJ, Rahnavard G, et al. Species-level functional profiling of metagenomes and metatranscriptomes. *Nat Methods*. 2018;15:962-968.
28. Luo R, Liu B, Xie Y, et al. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience*. 2015;4:s13742.
29. Galloway-Peña J, Hanson B. Tools for analysis of the microbiome. *Dig Dis Sci*. 2020;65:674-685. doi:10.1007/s10620-020-06091-y.
30. Claesson MJ, Cusack S, O'Sullivan O, et al. Composition, variability, and temporal stability of the intestinal microbiota of the elderly. *Proc Natl Acad Sci*. 2011;108:4586-4591. doi:10.1073/pnas.1000097107.
31. Zuo T, Liu Q, Zhang F, et al. Depicting SARS-CoV-2 faecal viral activity in association with gut microbiota composition in patients with COVID-19. *Gut*. 2021;70:276-284.
32. Caspi R, Billington R, Fulcher CA, et al. The MetaCyc database of metabolic pathways and enzymes. *Nucleic Acids Res*. 2018;46:D633-D639. doi:10.1093/nar/gkx935.
33. Kaminski J, Gibson MK, Franzosa EA, Segata N, Dantas G, Huttenhower C. High-specificity targeted functional profiling in microbial communities with ShortBRED. *PLoS Comput Biol*. 2015;11:e1004557.
34. Petersen TN, Lukjancenko O, Thomsen MCF, et al. MGmapper: reference based mapping and taxonomy annotation of metagenomics sequence reads. *PLoS ONE*. 2017;12:e0176469. doi:10.1371/journal.pone.0176469.
35. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *Embnet.Journal*. 2011;17:10-12. doi:10.14806/ej.17.1.200.
36. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*. 2009;25:1754-1760. doi:10.1093/bioinformatics/btp324.
37. Zhou P, Yang X-L, Wang X-G, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*. 2020;579:270-273. doi:10.1038/s41586-020-2012-7.
38. Zuo T, Zhan H, Zhang F, et al. Alterations in fecal fungal microbiome of patients with COVID-19 during time of hospitalization until discharge. *Gastroenterology*. 2020;159:1302-1310.
39. Wang L, Candia J, Ma L, et al. Serological responses to human virome define clinical outcomes of Italian patients infected with SARS-CoV-2 [Published online ahead of print September 7, 2020]. *medRxiv*. doi:10.1101/2020.09.04.20187088.
40. Górski A, Międzybrodzki R, Żaczek M, Borysowski J. Phages in the fight against COVID-19. *Future Microbiol*. 2020;15:1095-1100.
41. Donati Zeppa S, Agostini D, Piccoli G, Stocchi V, Sestili P. Gut microbiota status in COVID-19: an unrecognized player. *Front Cell Infect Microbiol*. 2020;10:576551.
42. Yeoh YK, Zuo T, Lui GC-Y, et al. Gut microbiota composition reflects disease severity and dysfunctional immune responses in patients with COVID-19 [Published online ahead of print January 11, 2021]. *Gut*. doi:10.1136/gutjnl-2020-323020.
43. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*. 2018;34:i884-i890. doi:10.1093/bioinformatics/bty560.
44. Camacho C, Coulouris G, Avagyan V, et al. BLAST+: architecture and applications. *BMC Bioinformatics*. 2009;10:421.
45. Huson DH, Auch AF, Qi J, Schuster SC. MEGAN analysis of metagenomic data. *Genome Res*. 2007;17:377-386.
46. Brouard J-S, Schenkel F, Marete A, Bissonnette N. The GATK joint genotyping workflow is appropriate for calling variants in RNA-seq experiments. *J Anim Sci Biotechnol*. 2019;10:44.
47. Li H, Handsaker B, Wysoker A, et al. The sequence alignment/map format and SAMtools. *Bioinformatics*. 2009;25:2078-2079. doi:10.1093/bioinformatics/btp352.
48. Koboldt DC, Zhang Q, Larson DE, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res*. 2012;22:568-576.
49. Edgar R. *Usearch*. Berkeley, CA: Lawrence Berkeley National Lab (LBNL), 2010.
50. Tao W, Zhu S, Zhang G, et al. Analysis of the intestinal microbiota in COVID-19 patients and its correlation with the inflammatory factor IL-18 and SARS-CoV-2-specific IgA [Published online ahead of print August 14, 2020]. *medRxiv*. doi:10.1101/2020.08.12.20173781.
51. Rognes T, Flouri T, Nichols B, Quince C, Mahé F. VSEARCH: a versatile open source tool for metagenomics. *PeerJ*. 2016;4:e2584.
52. Lozupone C, Knight R. UniFrac: a new phylogenetic method for comparing microbial communities. *Appl Environ Microbiol*. 2005;71:8228-8235.
53. Zhang H, Ai J-W, Yang W, et al. Metatranscriptomic characterization of coronavirus disease 2019 identified a host transcriptional classifier associated with immune signaling. *Clin Infect Dis*. 2020;2020:ciaa663. doi:10.1093/cid/ciaa663.