Taxonomical and functional changes in COVID-19 faecal microbiome are related to SARS-CoV-2 faecal load

Lucia Grenga
Université Paris-Saclay, CEA, INRAE, Département Médicaments et Technologies pour la Santé (DMTS), SPI
Olivier Pible
CEA
Guylaine Miotello
Université Paris-Saclay, CEA, INRAE, Département Médicaments et Technologies pour la Santé (DMTS), SPI
Karen Culotta
CEA
Sylvie Ruat
Université Paris-Saclay, CEA, INRAE, Département Médicaments et Technologies pour la Santé (DMTS), SPI
Marie-Anne Roncato
Université Paris-Saclay, CEA, INRAE, Département Médicaments et Technologies pour la Santé (DMTS), SPI
Fabienne Gas
Université Paris-Saclay, CEA, INRAE, Département Médicaments et Technologies pour la Santé (DMTS), SPI
Laurent Bellanger
Université Paris-Saclay, CEA, INRAE, Département Médicaments et Technologies pour la Santé (DMTS), SPI
Pierre-Géraud Claret
CHU de Nîmes
Catherine Dunyach-Remy
CHU de Nîmes
Didier Laureillard
CHU de Nîmes
Albert Sotto
CHU de Nîmes
Jean-Philippe Lavigne

CHU de Nîmes

Jean Armengaud (jean.armengaud@cea.fr)

Université Paris-Saclay, CEA, INRAE, Département Médicaments et Technologies pour la Santé (DMTS), SPI https://orcid.org/0000-0003-1589-445X

Article

Keywords: Metaproteomics, fecal microbiota, COVID-19, SARS-CoV-2, coronavirus, infection, dysbiosis

Posted Date: April 26th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-414136/v1

Abstract

Since the beginning of the pandemic caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) the gastro-intestinal (GI) tract has emerged as an important organ influencing the propensity to and potentially severity of the related COVID-19 disease. However, the contribution of the SARS-CoV-2 intestinal infection on COVID-19 pathogenesis remains to be clarified. In this exploratory study, we evidenced that alterations in the composition of the gut microbiota depends on the levels of SARS-CoV-2 RNA in the gastrointestinal tract but not on the presence of SARS-CoV-2 in the respiratory tract, COVID-19 severity and GI symptoms. Altered molecular functions in the microbiota profiles of high SARS-CoV-2 RNA level faeces as established by metaproteomics highlight mechanisms that may contribute to vicious cycles. Uncovering the role of this gut microbiota dysbiosis could drive the investigation of alternative therapeutic strategies to favour the clearance of the virus and potentially mitigate the effect of SARS-CoV-2 infection.

Background

Novel coronavirus pneumonia, COVID-19, is caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). The most common pathognomonic symptoms include cough, fever, and dyspnoea, but the heterogeneity of clinical manifestations is one of several intriguing facets of SARS-CoV-2 infection [1]. Thus, even though COVID-19 is mainly considered to be a respiratory disease, gastrointestinal symptoms have been described in COVID-19 patients, the most recurrent of which are diarrhoea, vomiting, nausea, and abdominal pain. These symptoms suggest that SARS-CoV-2 infection also has a direct impact on the gastrointestinal system [2].

Early in the pandemic, digestive symptoms were reported to be an early symptom of COVID-19 [3]. Various aetiopathogenetic hypotheses were subsequently advanced to explain the occurrence of diarrhoea in COVID-19 patients, including loss of absorption capability by enterocytes [4], microscopic inflammatory damage to gastrointestinal mucosal, and impaired ACE2 function. The role of ACE2 functions in maintaining gut homeostasis is recognized. However, SARS-CoV-2 viral particles were also reported to be detectable in stool samples from patients without diarrhoea [5]. Other coronavirus diseases, such as SARS in 2002 and Middle East Respiratory Syndrome (MERS) in 2012 – with which SARS-CoV-2 shares 70% and 40% genomic sequence similarities, respectively – were also reported to show some enteric involvement [6]. Although the human intestinal tract was demonstrated to serve as an alternative infection route for MERS-CoV [7], the role of intestinal SARS-CoV-2 infection in the pathogenesis of COVID-19 remains to be clarified.

A faecal-oral route of transmission for SARS-COV-2 was proposed based on the evidence that viral RNA is detected in up to 50% of faecal samples from patients diagnosed with COVID-19 despite a negative nasopharyngeal (NP) swab [8,9]. The persistent shedding of SARS-CoV-2 in stools from infected patients between 1 and 12 days after a negative NP test result [10–12] led to the hypothesis that infectious virions

are secreted from infected gastrointestinal cells. Furthermore, re-emergence of coronavirus disease in several regions was recently associated with processing of frozen food products [13], and initial contamination via ingestion cannot be excluded. Evidence obtained using organoid models revealed SARS-CoV-2 to exclusively target the apical surface of mature villous enterocytes expressing high levels of three proteins enhancing virus entry into enterocytes: angiotensin conversion enzyme 2 (ACE2) and two related membrane-bound serine proteases, TMPRSS2 and TMPRSS4 [14,15]. Although infected enterocytes form syncytia and viral particles are shed into the lumen, experiments examining the effect of simulated human colonic fluid on the virus suggest that shed virus is rapidly inactivated as it passes through the colon [16].

Currently, little is known about the impact of the gastrointestinal presence of SARS-CoV-2 on the course of COVID-19. Evidence that the gut microbiome influences ACE2 expression has led to several groups hypothesizing a contribution of intestinal microbes to COVID-19, but only very limited data are available on the profile of the gut microbiome in SARS-CoV-2-infected patients [17,18]. Notably, alterations to the microbiota have mainly been discussed in relation to the detection of SARS-CoV-2 in the respiratory tract and COVID-19 severity. Thus, Zuo et al. [19] investigated transcriptional activity of SARS-CoV-2 and temporal microbiome alterations in faecal samples from patients with COVID-19. They observed that faecal samples showing high SARS-CoV-2 infectivity contained higher abundances of bacterial species *Collinsella aerofaciens, Collinsella tanakaei, Streptococcus infantis*, and *Morganella morganii*. Metatranscriptomics analysis revealed higher functional capacity for de novo nucleotide biosynthesis, amino acid biosynthesis, and glycolysis in these samples. In contrast, faecal samples with low-to-no SARS-CoV-2 infectivity were associated with higher abundances of short-chain fatty acid-producing bacteria.

Nevertheless, transcript presence does not necessarily indicate protein synthesis and thus a functional impact. To obtain this type of information, more precise functional information could be obtained by metaproteomics characterization. Profiling of the human gut microbiome via metaproteomics has proven its value in pathogenesis research in the context of several diseases [20,21]. The potential of these techniques to help guide future clinical diagnosis has also been highlighted [22].

In this study, we used a mass spectrometry-based approach to profile the gut microbiota in terms of bacterial, archaeal, yeast, and fungal content, and analysed the associated metaproteomic functions in patients with intestinal COVID-19 infection. Altered microbiota compositions were found to be independent of the presence of SARS-CoV-2 in the respiratory tract, disease severity, and gastro-intestinal (GI) symptoms, but correlated with GI levels of SARS-CoV-2 RNA. Examination of the functional composition of the metaproteome provided a shortlist of both microbial and human biomarker candidates indicative of intestinal SARS-CoV-2 infection. These biomarkers could be used to monitor infection. Information on how intestinal SARS-CoV-2 affects the microbiota and the host could be useful in the search for alternative therapies promoting viral clearance, with a view to mitigating the impact of SARS-CoV-2 infection.

Results

Levels of SARS-CoV-2 RNA in the gut

A total of 39 faecal samples collected from 39 patients has been studied (Table 1). Among them, 32 were included in the COVID group due to a RT-PCR and/or a positive CT-scan (Fig. 1A). The remaining 7 patients were without COVID-19 diagnosis (RT-PCR and CT-scan negative). The median ages and sex ratios of patients with COVID-19 and non-COVID-19 were 76.5 years (34–96) versus 79 (61–99) and 0.89 versus 1.33, respectively. To investigate the effect of intestinal SARS-CoV-2 infection on the composition of the gut microbiota, faecal samples were characterized by applying the pipeline presented in Fig. 1B. The presence of SARS-CoV-2 in the gut was analysed by subjecting stool samples to RT-qPCR. Of the 39 patients, 10 (25.6%) presented a positive RT-qPCR result. Among them, one (C01P001) belonged to the non-COVID group (RT-PCR and CT-scan negative). The median faecal viral load was 3.7 $\log_{10}$ copies per mg [IQR 4.5–2.7], as estimated by RT-qPCR performed in parallel with standards. A viral load of 1.7 $\log_{10}$ copies per mg (50 copies/mg) was arbitrarily chosen as the cut off value to group samples into high SARS-CoV-2 RNA levels, low SARS-CoV-2 RNA loads, and negative samples (Fig. 2). No obvious association (Kruskal-Wallis test, $p < 0.05$) was found between plasma C-reactive protein (CRP) levels and GI tract SARS-CoV-2 levels ($p = 0.21$). No correlation (Fisher exact test, $p < 0.05$) between the levels of SARS-CoV-2 RNA in the faecal samples and the positivity of nasopharyngeal swab tests on the first hand ($p = 0.40$), and with severity of COVID-19 on the other ($p = 1.00$), were detected.

Table 1
Clinical characteristics of subjects involved in this study.

Patient	Sex	Age	RT-PCR NP swab status	CT-scan signs of COVID-19	SARS-CoV-2 in stool (copies/mg)	Δ RT-PCR NP swab and stool (days)	PlasmaCRP (mg/L)	Severity of pulmonary signs
C01P001	M	99	(-)	No	3.6E + 04	34	68.9	No sign
C01P002	M	66	(-)	Yes	(-)	5	64.5	Mild
C01P003	M	79	(+)	Yes	1.4E + 03	8	32.4	Moderate
C01P004	F	67	(+)	Yes	(-)	17	3.2	Moderate
C01P005	F	91	(+)	No	(-)	1	7.4	No sign
C01P006	F	54	(+)	Yes	(-)	14	30.1	Moderate
C01P007	F	88	(+)	No	1.4E + 01	1	2.4	No sign
C01P008	M	60	(+)	Yes	(-)	20	1.2	Severe
C01P009	F	83	(+)	No	(-)	4	12.3	No sign
C01P010	F	92	(-)	No	(-)	-	103.1	No sign
C01P011	M	85	(+)	No	(-)	7	4.2	No sign
C01P012	F	83	(+)	Yes	(-)	9	3.9	Mild
C01P013	F	47	(+)	Yes	(-)	1	15.6	Moderate
C01P014	M	63	(+)	Yes	(-)	21	1.9	Severe
C01P015	F	54	(-)	Yes	(-)	8	7.6	severe
C01P016	M	82	(+)	No	6.8	5	48.9	No sign
C01P017	M	61	(-)	No	(-)	-	23.3	No sign
C01P018	M	48	(+)	Yes	(-)	11	65.6	Mild
C01P019	F	77	(+)	No	(-)	15	2.2	Mild
C01P020	M	96	(+)	No	(-)	3	37.9	No sign
C01P021	F	56	(+)	Yes	(-)	6	30	Moderate
C01P022	M	76	(+)	Yes	2.5E + 01	10	251.3	Severe
C01P023	F	82	(-)	No	(-)	3	70.9	Mild
C01P024	M	60	(-)	Yes	(-)	30	3.9	Severe

Patient	Sex	Age	RT-PCR NP swab status	CT-scan signs of COVID-19	SARS-CoV-2 in stool (copies/mg)	Δ RT-PCR NP swab and stool (days)	PlasmaCRP (mg/L)	Severity of pulmonary signs
C01P025	M	75	(-)	No	(-)	-	37.9	No sign
C01P027	M	79	(-)	No	(-)	-	13	No sign
C01P028	F	40	(-)	Yes	(-)	-	25.8	Mild
C01P029	F	84	(+)	No	3.6E + 04	4	133.7	No sign
C01P030	F	78	(-)	Yes	(-)	-	2.3	No sign
C01P031	F	76	(-)	No	(-)	-	32.4	No sign
C01P032	M	69	(-)	Yes	(-)	4	8	Severe
C01P033	F	95	(+)	Yes	9.6E-01	2	14.1	Mild
C01P035	M	34	(+)	No	1.5	3	1.6	No sign
C01P036	F	58	(+)	Yes	1.0E + 01	-	66.5	Moderate
C01P037	M	87	(+)	No	(-)	6	38.5	No sign
C01P038	M	96	(+)	No	(-)	1	5	No sign
C01P039	F	68	(+)	Yes	(-)	8	21.2	Moderate
C01P040	M	79	(+)	No	(-)	9	20.8	Moderate
C01P041	M	91	(+)	No	6.9E + 01	3	85	No sign

Significantly altered microbiota profile in patients with a high level of SARS-CoV-2 RNA in the GI tract

To gain insight into the relationship between SARS-CoV-2 infection, microbiota, and host, we differentiated samples based on the levels of SARS-CoV-2 RNA detected in the GI tract. The total metaproteomic dataset from the 117 nanoLC-MS/MS runs comprised 7,761,229 MS/MS spectra. With the adjustment procedure of the peptide quantities to inject based on a pre-screen by mass spectrometry, an overall average of 66,335 ± 3,248 MS/MS was obtained per sample with low variation between samples. Peptides were identified by searching tandem mass spectra in a two-step cascaded search against a sample-specific database. This strategy allowed 31 ± 6% of spectra to be assigned per sample. For each sample, the distribution of the assigned TSMs as a function of their origin – microbial, host, or food-related – (**Supplementary Table S1**) revealed a higher percentage of host signal in samples positive for SARS-CoV-2 RNA (Fig. 3A). The microbial component of the metaproteomes expressed as a proportion of the average protein biomass at the phylum level was dominated by bacteria (61 ± 19%; mean ± SD), followed by fungi and archaea, which represented 5 ± 2% and 3 ± 1%, respectively. *Firmicutes*,

Proteobacteria, *Actinobacteria*, and *Bacteroidetes* were the predominant bacterial phyla, which combined accounted for an average of 48% of the biomass. Among the archaeal signatures, the *Euryarchaeota* phylum was the most highly represented (3 ± 1%). A total of 201 genera were identified in microbiota profiles, with *Clostridium*, unclassified Lachnospiraceae, *Bacteroides*, and *Lachnoclostridium* as dominant taxa in addition to the host. The relative biomass contributions for these groups ranged from less than 1 to 6% for bacterial genera, and up to 58% for *Homo sapiens*. Among the 13 archaeal genera identified, *Methanobrevibacter* (28,792 TSMs) was the most abundant. Even though it remained a minor component of the microbiota, sequence coverage for this genus was high (6392 taxon-specific peptides). *Ascomycota* and S*treptophyta* tended to be the most abundant Eukaryota phyla, with the notable exception of the *Chordata* host (Fig. 3B). After filtering out food-related and host signals, dimension reduction by principal component analysis (PCA) revealed distinct microbiota profiles for patients with high intestinal levels of SARS-CoV-2 RNA (viral load > 50 copies per mg of faecal material) compared to profiles for patients with low or no viral RNA (Fig. 3C). Two outliers were observed: sample C01P003 (faecal viral load 1.4 x 10^3 copies/mg), which clustered among negative samples; and sample C01P033 (1 copy/ mg) which had a microbiota composition resembling high SARS-CoV-2 samples.

Because age-related changes to the microbiota have recently been reported [23], we further examined the alpha diversity of our samples after matching positive (viral load > 50 copies per mg of faecal material) and negative patients by age (88.2 (± 8.7) and 80.5 (± 7.8) years old, respectively). Both Inverse Simpson and Shannon indices indicated that microbial diversity was significantly decreased (Tukey's Honest Significant Difference) in samples containing SARS-CoV-2 RNA compared to negative samples (Fig. 3D). Overall comparison between microbiota determined for patients with a high SARS-CoV-2 load and negative samples (Wilcoxon Rank Sum test, p-value < 0.05) showed that patients with SARS-CoV-2 RNA in the intestine had a higher relative abundance (log2 FC > 1.6) of certain fungal genera, such as *Candida*, *Fusarium*, *Penicillium*, *Aspergillus*, and *Saccharomyces* (Fig. 3E). For archaea, *Methanosphaera* and unclassified *Halobacteriales* were among the most enriched genera in samples containing SARS-CoV-2 (log2 FC = 1.4) (Fig. 3F). Among bacteria, several genera belonging to the Actinobacteria class (e.g., *Streptomyces*, *Actinomadura*, *Amycolatopsis*, *Nocardia*, *Mycobacterium*, and *Arthrobacter*), as well as *Paenibacillus*, *Chitinophaga*, and *Sphingomonas* were significantly enriched in SARS-CoV-2-positive samples compared to controls (log2 FC > 1) (Fig. 3G). In contrast, significantly lower relative abundance was observed for genera belonging to the *Ruminococcaceae* and *Lachnospiraceae* families (log2 FC > -1), and more broadly to the Firmicutes phylum, with the exception of members of the Bacilli class. Several genera from *Odoribacteraceae*, *Prevotellaceae*, *Eggerthellaceae*, and *Coriobacteriaceae* families (log2 FC = -1), as well as *Akkermansia* (log2 FC = -1.7) were also detected at significantly lower abundance in virus-positive samples (Fig. 3G).

Functional composition of the metaproteome reveals potential biomarkers of SARS-CoV-2 infection

To retrieve functional information from the metaproteome for different sample groups, microbiota and host proteins identified with an FDR of 1% were annotated. A total of 88,135 proteins and 60,179 protein

groups were listed in the dataset, functions of which were assigned to 887 KO (KEGG Orthology) entries (**Supplementary Table S2**). The relative abundance of each functional term was calculated at phylum level to reduce loss of information when peptides were unambiguously assigned to the different taxa at a finer resolution. To investigate dissimilarities in microbiota-derived KO functions (n = 664), unsupervised PCA was performed on age-matched SARS-CoV-2-positive (faecal viral load > 50 copies/mg) and negative samples (Fig. 4A**)**. This PCA analysis revealed distinct clustering based on the presence of SARS-CoV-2, with only SARS-CoV-2-positive C01P003 sitting closer to the negative samples. Comparative analysis of the functional profiles (Wilcoxon test, FDR-adjusted $p < 0.05$) revealed 341 KOs to be significantly differentially abundant between the two groups of samples. Of these, 21 were increased in SARS-CoV-2-positive samples (Fig. 4B**/C**). These KOs are included in 67 KEGG pathways, the most populated of which were metabolic pathways (10), biosynthesis of secondary metabolites (5), glycolysis / gluconeogenesis (3), and microbial metabolism in diverse environments (3). Interestingly, the identification of molecular functions related to citrulline flux (KO names OTC and arcA) suggested that Firmicutes species were still adapting to cope with stress and gain an energetic advantage. Evidence for this ongoing adaptation was also provided by the increase in abundance of polypeptides belonging to two-component systems (mcp and yesN) and cobalamin production (cobS - cobV, Firmicutes). Similarly, KOs for drug exporter pump (K06994), NADPH:quinone reductase (qor, CRYZ), and NTE family proteins associated with Actinobacteria suggesting the implementation of mechanisms conferring a competitive advantage in stressful environments, such as the SARS-CoV-2-infected gut. Modules involved in sulfur (soxD, Bacteroidetes) and glutathione metabolism (pepN, Actinobacteria) were also significantly increased. Another marker of the host response, K06856 (IGH, immunoglobulin heavy chain), was increased in association with both Actinobacteria and Firmicutes. Some fungi-associated molecular functions were also altered (PDC, AdhP, SET2). Among these markers, AdhP is known to be involved in retinol metabolism, and the histone modification protein SET2 plays a key role in mucosal immune responses, and could be critically involved in integrating a variety of external signals driving fungal expansion. This fungal expansion would in turn influence the host immune response. Among the 21 altered KOs identified, 9 (K00134, K00344, K01256, K07001, K01568, K22622, K07720, K11686, K16703) were linked to 14 pathways. Levels of these pathways were also significantly increased relative to age-matched samples in which lower levels of SARS-CoV-2 RNA were detected (viral load < 50 copies/mg) (Fig. 4B**/C**).

To investigate alterations in host molecular functions, the relative abundance of host-associated KOs was also analysed. The abundance of a total of 72 out of 187 KOs was altered in samples containing high levels of SARS-CoV-2-RNA (viral load > 50 copies/mg) compared to negative samples (**Supplementary Table S2**). Of these KOs, 42, represented in 117 pathways, were increased in SARS-CoV-2-positive samples with a high viral load (Fig. 4D). These KOs included host molecular functions involved in the ACE2 signalling network (Renin-angiotensin system) such as peptidyl-dipeptidase A (ACE), aminopeptidases (ANPEP/CD13), glutamyl aminopeptidase (ENPEP), and neprilysin (MME). In parallel, functions described by KOs such as latexin (LXN), sphingomyelinase-related protein SMPDL3, dihydrolipoamide dehydrogenase (DLD), SOD2, Cu/Zn superoxide dismutase (SOD1), ferritin heavy chain (FTH1), tissue-nonspecific alkaline phosphatase (K01077), antileukoproteinase (SLPI), bleomycin

hydrolase (BLMH), aminopeptidase (NAALADL1), glutamate carboxypeptidase II (GCPII), and transmembrane serine protease 15 (TMPRSS15), proteins from the enterocyte brush-border membrane – including sucrase-isomaltase, calmodulin, cadherin-17, and glutamate carboxypeptidase II, trefoil factor family peptides (TFF2 and TFF3) were also significantly increased in faecal samples with high levels of SARS-CoV-2 RNA. Interestingly, faeces with high levels of SARS-CoV-2 contained a statistically higher abundance of proteins with CRP-mediated functions: immunomodulatory Gal-9, bone marrow proteoglycan (PRG2), glutathione S transferase (GST), deleted in malignant brain tumour 1 (DMBT1), gastric intrinsic factor (GIF), members of the lysosome pathway (DNASE2, CTSH, CD63 antigen, GAA and NPC2), L-Tryptophan (SLC3A1), and choline (SLC44A2) transporters. These proteins are linked to functional pathways, and as such provide further insight into the host-microbiota crosstalk in SARS-CoV-2-positive GI tracts. In parallel, the increase in MHC class I antigen (MHC1), ADP-Ribosyl Cyclase 2 (BST1), nicastrin (NCSTN), pre-B lymphocyte gene (VPREB), and ferritin heavy chain (FTH1) suggest activation of immune processes. Sixteen of the altered KOs identified (K00382, K01203, K01316, K01346, K01389, K04565, K06497, K11140, K11141, K12316, K13912, K14210, K17286, K18152, K23879, K24332) were also specifically increased in samples containing < 50 copies/mg viral RNA following comparison with their relative levels in SARS-COV-2 positive samples with high SARS-CoV-2 load.

Discussion

The gut microbiota plays multiple critical roles not only in nutrition through food processing, but also by maintaining human health as a result of both local and systemic effects, such as limiting pathogen colonization, helping maintain the intestinal barrier function, and training the immune system. Importantly, in the context of SARS-CoV-2 infection, the GI tract has emerged as an organ significantly influencing the propensity to develop, the ensuing disease, COVID-19, and potentially predict its severity. Studies based on taxonomical molecular biology approaches have demonstrated that respiratory infections are associated with changes in the composition of the gut microbiota [4], but the correlation between respiratory disease and the amount of virions present in the gut has received less attention [19]. Here, we provide a broad profile of gut microbiota obtained by a mass spectrometry-based detecting bacteria, archaea, yeasts, and fungi. Organisms were proteotyped and their respective biomass contributions directly compared. We also investigated whether the presence of SARS-CoV-2 in the gastrointestinal tract was associated with changes to the composition of the microbiota.

In this article, we report that the presence of SARS-CoV-2 RNA in the GI tract is not directly related to the detection of the virus in the respiratory tract or to COVID-19 severity at the time of detection. This result is in agreement with the detection of SARS-CoV-2 in tissues throughout the GI tract, and virus shedding in stools in a significant proportion of patients. GI shedding often continues for prolonged periods following virus clearance from the respiratory tract [17]. In addition, we provide evidence of changes in the composition of the gut microbiota as a function of the abundance of SARS-CoV-2 RNA in the intestine. In particular, our results show that the microbiota was significantly different in patients with a faecal viral load greater than 50 copies per mg, whereas microbiota from patients with a lower or negative viral load

tended to be more similar. Closer examination revealed a relatively high level of SARS-CoV-2 RNA to be associated with a significantly higher abundance of several genera belonging to the Euryarchaeota and Ascomycota phyla, as well as bacterial genera from the Actinobacteria order, *Chitinophaga*, *Paenibacillus*, *Sphingomonas*, and *Bacillus*. The overgrowth of fungi is in itself indicative of a disruption of commensal communities. In addition, fungal overgrowth in the gut can cause macrophage polarization which has been linked to increased infiltration of inflammatory cells in allergic airways [24]. Interestingly, bacterial genera from the Bacteroidetes phylum – known to be associated with suppression of colonic expression of ACE2 in the murine gut [25] – were significantly more abundant in SARS-CoV-2-positive samples. Similarly, the relative abundance of bacterial genera belonging to the Firmicutes phylum, such as *Faecalibacterium*, *Mitsuokella*, and *Frisingicoccus*, correlated well with the absence of SARS-CoV-2 in the gut. Interestingly, this class of bacteria are the main producers of butyrate in the intestinal lumen, which is extensively used by colonic epithelial cells as an energy source [26]. An in-depth shotgun metagenomics analysis of samples from 15 patients hospitalized in Hong Kong [27] revealed a similar shift in the composition of microbiota, particularly an increased proportion of opportunistic fungal pathogens (*Candida albicans*, *Candida auris*, and *Aspergillus flavus*) in COVID-19 patients compared to controls. In this cohort, a negative correlation between the abundance of *Faecalibacterium prausnitzii* (an anti-inflammatory bacterium) and both COVID-19 severity and changes in faecal shedding of SARS-CoV-2 particles was also reported. Other authors have studied connections between amino acid transport based on the association between ACE2 and microbial ecology in the gut during SARS-CoV-2 infection [28], and the possibility that features of gut microbiota could serve as indicators of a predisposition to severe COVID-19 [29].

Our peptide-based functional metaproteome analysis confirmed the differences observed between samples containing high levels of SARS-CoV-2 RNA and negative samples, and the existence of a complex interplay between the gut and SARS-CoV2 infection. In particular, the alterations to host functions observed in faeces containing high levels of SARS-CoV-2 revealed an inflamed GI tract characterized by activation of the immune response, as reflected by the molecular alterations governing the host's antiviral defence system. In addition to serving as a receptor for SARS-CoV-2, ANPEP/CD13 expression is known to be dysregulated in inflammatory diseases. Its detection in wide numbers of gut samples led to the persistent intestinal inflammation hypothesis [30]. In our samples, expression levels for this marker were consistent with increased latexin and SMPDL3, a sphingomyelinase-related protein abundantly expressed on macrophages and dendritic cells [31]. Both of these proteins are upregulated by inflammatory stimuli. In parallel, functions described by KOs such as dihydrolipoamide dehydrogenase (DLD), SOD2, Cu/Zn superoxide dismutase (SOD1), and ferritin heavy chain (FTH1) could reflect mitochondrial dysfunction and interplay between inflammation and oxidative stress. Interestingly, expression of tissue-nonspecific alkaline phosphatase in the colon was also previously reported to be upregulated during inflammatory episodes as a consequence of inflammation-driven tissue-infiltration by neutrophils [32]. The identification of elevated levels of KOs linked to functions located in the enterocyte brush border supports inflammation-induced enterocyte damage and increased intestinal permeability.

This possibility is supported by the over-detection in faecal samples containing SARS-CoV-2 RNA of trefoil factor family peptides. These essential proteins are involved in protection and repair of the gastrointestinal tract [33]. This response suggests that expression of TFF variants could be used to predict prognosis, or to monitor therapeutic efficacy. Similarly, the increased abundance of several leaky-gut-related functions like those associated to antileukoproteinase (SLPI) [34]; bleomycin hydrolase (BLMH), a cytosolic aminopeptidase thought to contribute to MHC class I peptide presentation [35]; aminopeptidase (NAALADL1), glutamate carboxypeptidase II (GCPII), and transmembrane serine protease 15 (TMPRSS15) may contribute to the dysregulation of the protease/antiprotease balance and could potentially lead to epithelial damage and increased intestinal permeability. In parallel, the increase in MHC class I antigen (MHC1), ADP-Ribosyl Cyclase 2 (BST1), nicastrin (NCSTN), pre-B lymphocyte gene (VPREB), and ferritin heavy chain (FTH1) suggests activation of immune processes that could be further explored in future studies to unravel the links between SARS-CoV-2 infection and gastrointestinal perturbations.

In this dysfunctional environment, individual members of the altered microbiota battle to gain an energetic advantage. Given the known regulatory role played by H2S in mucosal inflammation, and the involvement of glutathione in bio-reduction of reactive oxygen species, the observed alterations suggest that the presence of SARS-CoV-2 triggers activation of microbial metabolic pathways that further fuelling inflammation and related immune responses. An Ig-like domain is frequently present in bacterial proteins that affect adhesion to host cells and tissues, as part of invasion, or other steps in the infection process [36].

Our results corroborate the alterations recently reported in the faecal metabolome for COVID-19 patients [37]. However, the question of whether COVID-19-related intestinal lesions are the result of a secondary response following systemic inflammation, of primary intestinal infection, or of the combined consequences of both mechanisms, remains open. The altered molecular functions described here suggest mechanisms feeding into vicious cycles. Furthermore, our results provide markers that could be potentially be valuable in monitoring the progression of SARS-CoV-2 infection and assessing therapeutic strategies to promote viral clearance and restore normal intestinal function. Among the protein functions described, some such as Galectin-9 and CRP are already under investigation as biomarkers of disease severity [38–40]. Interestingly, elevated serum levels of CRP, especially in association with a high concentration of D-dimers, are reported to be indicative of an increased risk of adverse outcomes in COVID-19 [41]. Consequently, faecal CRP levels could represent a useful biomarker to stratify COVID-19 patients as a function of the level of SARS-CoV-2 in the GI tract.

Although the COVID-19 pandemic has led to the launch of a large number of clinical studies [42], to our knowledge, this is the first time the composition and functionality of gut microbiota have been analysed by differential metaproteomics, with samples distinguished based on the presence of SARS-CoV-2 RNA in the intestinal tract. Interestingly, even in stool specimens containing high levels of viral genetic material, no SARS-CoV-2-derived peptides were detected in our discovery metaproteomics approach. This negative

result could be ascribed to the complexity of the samples and a lack of sensitivity of our approach. Indeed, although targeted proteomics can successfully detect SARS-CoV-2 proteins in nasopharyngeal swabs and gargle samples [43–45], viral proteins have never been detected by discovery proteomics approaches applied to complex matrices [42]. Alternatively, these samples may only contain RNA fragments as previously reported for faeces collected at later time points following disease onset [10]. Whatever the case, and despite the relatively modest number of samples analysed and the fact that extra-pulmonary detection of viral RNA does not constitute proof that infectious virus particles are or were present, our results confirm the clinical relevance of testing for viral RNA in faeces because of its direct correlation with altered gut microbiota.

The association between COVID-19 and the presence of SARS-CoV-2 in the intestine remains to be further explored. SARS-CoV-2 is the third highly pathogenic coronavirus (after SARS-CoV and MERS-CoV) to cross to humans within less than 20 years, suggesting that new zoonotic coronavirus spill-overs are likely to occur in the future. The potential role of the gut in the diseases induced by this family of viruses should be further explored in multiple cohorts and settings, longitudinally over the course of different stages of infection. “Long Covid” describes the situation when the effects of Covid-19 continue for weeks or months beyond the initial illness, and has now been reported for a significant number of patients [46]. Some of these cases could be linked to continuous infection of the gut epithelium or severe alterations of the gut microbiota. Ultimately, mechanistic studies will be required to examine how composition of the microbiota affects how SARS-CoV-2 infects the GI tract, and to advance in the search for novel therapies to reduce the severity of COVID-19.

Until now, the impact of SARS-CoV-2 infection on the gut microbiota has been scarcely studied. Here, we assessed on a cohort of 39 patients the gastrointestinal SARS-CoV-2 viral load and correlated it with their full-range microbiota, including Bacteria, Archaea, Fungi and Moulds, accessible through metaproteomics. Our data shows three key results that could be of major importance in the battle against COVID-19: i) faecal SARS-CoV-2 viral load is not correlated to symptoms, ii) an important change in the microbial structure is observed for patients with high faecal SARS-CoV-2 viral load, and iii) a list of microbiome and human markers can be drawn from this study as possible candidates for diagnostic. These results should be incentive for an extensive multi-centric metaproteomics analysis of the gut microbiota of COVID-19 patients.

Methods

Ethics approval, consent to participate, and study population

This research has been performed in accordance with the Declaration of Helsinki. This retrospective study was approved by the local Institutional Review Board (IRB number: 20.05.01), “Comité d'éthique du CHU de Nîmes”. An informed letter was sent to patients to describe the study. Participants (all volunteers)

did not have to provide written informed consent to take part in the study. Stool samples received in the Department of Microbiology from the Department of Infectious Diseases (University Hospital Nîmes, France) from March 17, 2020 to May 11, 2020 were included. During this period, a nationwide lockdown was applied with an emergency state due to the context of the COVID-19 health crisis. The hospital admitted exclusively patients with acute health problems.

The flowchart of the study is presented in Fig. 1A. Among the 41 stools received during the indicated period, two were excluded from the analysis due to the low quantity. Data including demographics, laboratory results and imaging results were extracted from the electronic medical records of the University Hospital management system (Clinicom®, Intersystems SAS, France).

Collecting and processing faecal specimens

All the faecal samples received in the Department of Microbiology were routinely conserved at -80°C, according to the French microbiology standard [47]. The selected stools were transported in dry ice to the analytical facility following international guidelines (number UN3373, B category). Nucleic acid extraction and sample lysis performed on live SARS-CoV-2 samples prior protein extraction were conducted in a level 3 biosafety facility.

Quantification of viral RNA by RT-qPCR in faecal samples

SARS-CoV-2 RNA from stool samples was isolated using the NucleoSpin RNA Virus Mini kit (Macherey-Nagel SAS, Hoerdt, France). Stools (200 mg) was suspended at 10% (v/w) in PBS, homogenized vigorously with a vortex and then centrifuged for 3 min at 500 x g to remove large particles. Proteinase K (20 µL at 20 mg/mL) was added to the resulting supernatant (200 µL), incubated for 5 min at room temperature. Lysis and nucleic acid extraction were performed as recommend in the manufacturer's protocol. Viral RNA was eluted in 50 µL RNase-free H_2O preheated to 70°C. RNA was quantified by RT-qPCR using the SuperScript III Platinum One-Step RT-qPCR Kit (Thermo Fisher, Waltham, USA) and a CFX96 Touch Real-Time PCR Detection System Thermal Cycler (Bio-Rad, Marnes La Coquette, France). Primers targeting IP2 and IP4 (RdRp gene), 0.4 µM per reaction, were those recommended by Grenga et al. [48]: nCoV_IP2-12669Fw (ATGAGCTTAGTCCTGTTG), nCoV_IP2-12759Rv (CTCCCTTTGTTGTGTTGT), nCoV_IP2-12696bProbe(+) (AGATGTCTTGTGCTGCCGGTA [5']Hex [3']BHQ-1), nCoV_IP4-14059Fw (GGTAACTGGTATGATTTCG), nCoV_IP4-14146Rv (CTGGTCAAGGTTAATATAGG) and nCoV_IP4-14084Probe(+) (TCATACAAACCACGCCAGG[5']Fam [3']BHQ-1). Primers targeting E gene were: E_Sarbeco_Fw (ACAGGTACGTTAATAGTTAATAGCGT), E_Sarbeco_Rv (ATATTGCAGCAGTACGCACACA), and E_Sarbeco_Probe(+) (ACACTAGCCATCCTTACTGCGCTTCG[5']Fam [3']BHQ-1). Standard curves were created using *in vitro*-transcribed RNA derived from the BetaCoV_Wuhan_WIV04_2019 strain (EPI_ISL_402124), which contains the amplification regions of the RdRp and E gene as positive strand. Standard samples were prepared with 1011 copies of target sequences diluted in yeast tRNAs to facilitate recovery, and lyophilized.

Proteome extraction and digestion

Following sample homogenization, three 50 mg aliquots of each faecal sample were transferred to fresh tubes to create three technical replicates for each sample. Aliquots were lysed in 100 µL LDS 1X (Lithium dodecyl sulfate) sample buffer (Invitrogen™, Thermo Fisher) supplemented with 5% beta-mercaptoethanol (vol/vol). Samples were sonicated for 5 min in an ultrasonic water bath (VWR ultrasonic cleaner) and incubated at 99°C for 5 min before transfer to 2-mL Screw Cap microtubes (Sarstedt, Marnay, France) containing 200 mg ceramic beads, as previously described [49]. Cell disruption was performed on a Precellys Evolution instrument (Bertin Technologies, Aix en Provence, France) operated at 10,000 rpm for ten 30-s cycles, with 30 s rest between cycles. After lysis, samples were centrifuged at 16,000 × g for 3 min. For each sample, the resulting supernatant (25 µL) was loaded onto a NuPAGE 4–12% Bis-Tris gel, and proteins were subjected to short (5-min) SDS-PAGE migration. Proteins were stained for 5 min with Coomassie SimplyBlue SafeStain (Thermo Fisher) prior to in-gel trypsin proteolysis with Trypsin Gold (Promega) using 0.011% ProteaseMAX surfactant (Promega, Madison, WI, USA), as described in Hartmann et al. [50].

Preliminary quantitation of peptides extracted from faecal samples by high-resolution mass spectrometry survey

For each faecal sample, 1 µL of the extracted peptide mixture was injected for analysis on an LTQ-Orbitrap XL (Thermo Fisher Scientific, Waltham, USA) tandem mass spectrometer coupled to an Ultimate 3000 nano LC system (Thermo Fisher Scientific). The proteolyzed products were desalted online on a reverse-phase PepMap 100 C18 µ-precolumn (5 µm, 100 Å, 300 µm id × 5 mm, ThermoFisher) and resolved on a nanoscale PepMap 100 C18 nanoLC column (3 µm, 100 Å, 75 µm id × 50 cm, Thermo Fisher) at a flow rate of 0.3 µL/min prior to injection into the mass spectrometer. A linear chromatographic gradient of mobile phase A (0.1% HCOOH/100% H2O) and phase B (0.1% HCOOH/80% CH3CN) was applied from 5 to 40% B in 30 min. Full-scan mass spectra were measured from *m/z* 350 to 1500 in data-dependent mode using a Top5 strategy. Briefly, a scan cycle was initiated by a full high mass-accuracy scan in the Orbitrap analyser, operated at 30,000 resolution, followed by MS/MS scans in the linear ion trap on the five most abundant precursor ions. A 10-s dynamic-exclusion window was applied to previously selected ions. Precursor ions were isolated using a 3-*m/z* isolation window and activated with 35% normalized collision energy.

NanoLC-MS/MS characterization of peptides extracted from faecal samples

To normalize the peptide amount injected for each faecal sample, the total ion current (TIC) chromatogram obtained with the LTQ-Orbitrap XL instrument was used to calculate the exact volume (µL) to be injected on a Q-Exactive HF mass spectrometer (Thermo) by dividing 16.1E7 (the target optimal TIC previously established) by the TIC value obtained with the LTQ-Orbitrap XL. The high-field Orbitrap instrument was used in combination with an UltiMate 3000 LC system (Dionex-LC) and operated in data-dependent mode, as previously described [51]. The appropriate volumes of peptides were injected for each aliquot (technical triplicate for each faecal sample), desalted on an Acclaim PepMap100 C18 precolumn

(5 µm, 100 Å, 300 µm id × 5 mm), and then resolved on a nanoscale Acclaim PepMap 100 C18 column (3 µm, 100 Å, 75 µm id × 50 cm) with a 120-min gradient at a flow rate of 0.2 µL/min. The gradient was developed from 4 to 25% of [CH3CN, 0.1% formic acid] over 100 min, and then from 25 to 40% over 20 min. Peptides were analysed during scan cycles initiated by a full scan of peptide ions in the ultra-high-field Orbitrap analyser, followed by high-energy collisional dissociation and MS/MS scans on the 20 most abundant precursor ions (Top20 method). Full-scan mass spectra were acquired from *m/z* 350 to 1500 at a resolution of 60,000 with internal calibration activated on the *m/z* 445.12002 signal. During ion selection for MS/MS fragmentation and measurement, a 10-s dynamic-exclusion window was applied with an intensity threshold of 1.7×10^4. Only ions with positive charges 2 + and 3 + were considered.

Assigning peptides and analysing metaproteomics data

The Mascot Daemon 2.6.1 search engine (Matrix Science) was used to match MS/MS spectra to peptides in a multi-round search process. The search parameters were as follows: full trypsin specificity, maximum of two missed cleavages, mass tolerances of 5 ppm on the parent ion and 0.02 Da on the MS/MS, carbamidomethylated cysteine (+ 57.0215) as fixed modification, and oxidized methionine (+ 15.9949) and deamidation of asparagine and glutamine (+ 0.9848) as variable modifications. The NCBInr database (National Center for Biotechnology Information, NIH, Bethesda) was downloaded as ftp://ftp.ncbi.nlm.nih.gov/blast/db/FASTA/nr.gz (55 GB, downloaded on 2nd January, 2020). Protein accession numbers were mapped to taxids using files from ftp://ftp.ncbi.nlm.nih.gov (downloaded in January 2020). Files included: assembly_summary_refseq.txt and assembly_summary_genbank.txt downloaded from ftp://ftp.ncbi.nlm.nih.gov/genomes/ASSEMBLY_REPORTS/ to map taxids to RefSeq (GCF) and GenBank assemblies (GCA), GCF/GCA *_assembly_report.txt files downloaded from ftp://ftp.ncbi.nlm.nih.gov/genomes/all/ to map GCF/GCA to nucleotides and the GCF/GCA *_genomic.gff.gz files downloaded from the same directory to map GCF/GCA to protein accessions. Python SQLite databases were created to perform fast protein-to-taxid matching. The file prot.accession2taxid.gz, downloaded from ftp://ftp.ncbi.nlm.nih.gov/pub/taxonomy/accession2taxid/ was used to supplement databases, especially for Eukaryota. Files containing the NCBI taxonomy, names.dmp and nodes.dmp, were downloaded as ftp://ftp.ncbi.nlm.nih.gov/pub/taxonomy/taxdmp.zip to directly assign Peptide-spectrum matches (PSMs) to the “canonical” taxonomical levels (species, genus, family, order, class, phylum, and superkingdom) of the NCBI taxonomic tree.

A first-round Mascot search was performed on a reduced NCBInr-based database, termed NCBInrS, containing 45,925 selected taxa. This database comprises 76,112,114 sequences and a total of 31,417,968,972 amino acids. This first round was performed with a selection of the 50,000 best MS/MS spectra, as determined using the open-source algorithm ScanRanker. DirectTag version 1.4.66 [52] was used with the following parameters to apply the algorithm: PrecursorMzTolerance 0.1, FragmentMzTolerance 0.5, StaticMods "C 57.0215", UseChargeStateFromMS true, NumChargeStates 4. Molecular ion peak lists were extracted with Proteome Discoverer daemon v1.4 software (Thermo scientific), applying the following parameters: 400 (minimum mass), 5000 (maximum mass), 0 (grouping tolerance), 0 (intermediate scans), and 1000 (threshold). Peptide-to-MS/MS spectrum assignments

complied with the following constraints: full trypsin specificity, maximum of one missed cleavage, 2 + or 3 + peptide charges, mass tolerances on the parent ion of 3 ppm, and 0.02 Da on the MS/MS, static modification of carboxyamidomethylated cysteine (+ 57.0215), and oxidized methionine (+ 15.9949) as dynamic modification.

The Python version of Matrix Science's msparser (version 2.6.1) was used to parse the Mascot dat file, applying the ms_peptidesummary function. PSMs were validated when they had a Mascot expectation value below 0.3 according to their Mascot homology threshold (MHT); multiple PSMs were allowed per MS/MS spectrum. Each PSM associated with peptide sequences by Mascot was mapped to taxa using in-house SQLite databases built from NCBI data, as described in Pible et al. [53].

The phylopeptidomics procedure [53] was applied to identify taxonomies. The raw PSMs for each taxon (hereafter TSMs, for Taxon-Spectrum Matches), the numbers of matching peptide sequences, specific peptide sequences, and corresponding specific PSMs were determined for the species, genus, family, order, class, phylum, and superkingdom "canonical" taxonomical levels. Acceptance criteria for genera selection were adjusted for each individual dataset. Genera validated in the first-round search, and all their descendants, were extracted from the full NCBInr database for a second Mascot search. All acquired spectra were used at this stage, and the Mascot settings were the same as for the first-round search, except that the maximum number of missed cleavages was set to 2, and 5 ppm mass tolerance was allowed on the parent ion. Proteins validated with a Mascot p-value of 0.05 during the second-round search were extracted from NCBInr, and added to a selection of proteins specific to SARS-CoV-2, as detailed below, for a third-round search. SARS-CoV-2 viral genomic sequences were downloaded (2020/05/11) from the GISAID database [54], and proteins were extracted after genome synchronization against the reference genome from the corresponding GFF file GCA_009858895.3_ASM985889v3_genomic.gff. Both these resources were downloaded from ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/009/858/895/GCA_009858895.3_ASM985889v3/. The 186 non-redundant sequences from the N protein were added to the FASTA file used for the third-round search. This final Mascot search was performed using the same spectra and settings as those used in the second-round search. Peptides and proteins were identified with an FDR of 0.01 calculated from the relevant decoy database search. PSM validated with a Mascot p-value of 0.05 during the second-round search were filtered using an FDR < 1% and subsequently used to infer peptide and protein identifications. Proteins were grouped if they shared at least one peptide. Label-free quantification was performed based on PSM counts for each protein, applying the principle of parsimony. For each phylum, spectral counts for all the proteins in a group were summed to assign abundance values to each protein group. Proteins were KEGG-annotated using the GhostKoala web service (https://www.kegg.jp/ghostkoala/) to match proteins to KEGG Orthology (KO) and KEGG pathway maps. The spectrum count values for peptides mapped to a KEGG through protein mapping were summed to attribute an abundance value to each functional term. Peptide-to-taxon mapping was also performed to allow taxon-resolved functional quantification.

Statistical analyses

Count values from both taxonomic (number of TSMs) and functional data (spectrum counts) were scaled relative to their sum total in the sample. The categorical and continuous variables were compared among the patients using the Fisher exact test and the Kruskal–Wallis test, respectively. Principal component analysis was performed as previously described [55]. The R package metacoder [56] was used to represent taxonomic abundance as a differential heat tree. Univariate differential analysis of taxon and KO abundances between conditions was performed by applying non-parametric Wilcoxon tests corrected for multiple comparisons (Benjamini–Hochberg adjustment). Pairwise alpha-diversity indices were calculated using the vegdist function from the vegan package in R.

Declarations

Data availability

The mass spectrometry and proteomics datasets acquired on faecal samples are available through the ProteomeXchange Consortium via the PRIDE partner repository (https://www.ebi.ac.uk/pride/), under dataset identifiers PXD024990 and 10.6019/PXD024990. [The reviewers may access the currently private dataset using reviewer_pxd024990@ebi.ac.uk as Username and cnATF1JU as Password].

Ethics approval and consent to participate

This research has been performed in accordance with the Declaration of Helsinki. This retrospective study was approved by the local Institutional Review Board (IRB number: 20.05.01), “Comité d'éthique du CHU de Nîmes”. An informed letter was sent to patients to describe the study. Participants (all volunteers) did not have to provide written informed consent to take part in the study. Stool samples received in the Department of Microbiology from the Department of Infectious Diseases (University Hospital Nîmes, France) from March 17, 2020 to May 11, 2020 were included. During this period, a nationwide lockdown was applied with an emergency state due to the context of the COVID-19 health crisis. The hospital admitted exclusively patients with acute health problems.

Availability of mass spectrometry and proteomics data

The mass spectrometry and proteomics datasets acquired on faecal samples are available through the ProteomeXchange Consortium via the PRIDE partner repository (https://www.ebi.ac.uk/pride/), under dataset identifiers PXD024990 and 10.6019/PXD024990. [The reviewers may access the currently private dataset using reviewer_pxd024990@ebi.ac.uk as Username and cnATF1JU as Password].

Competing interests

The authors declare that they have no competing interests.

Funding

J.P.L. was supported by Montpellier University d’Excellence (MUSE) (MicroCOVID grant 2020).

Authors' contributions

LG, OP, CDR, AS, JPL, and JA conceived the project. LG and JA designed the overall experimental approach and analysed the data with OP. LG drafted the manuscript with inputs from JA and other authors. GM performed the mass spectrometry measurements under the supervision of LG and JA. SR and FG performed the RT-qPCR. MAR and LG performed sample inactivation and extraction. OP and KC contributed to data interpretation scripts. PGC, CDR, DL, AS and JPL were in charge of patient recruitment and diagnostics. All authors read and approved the final manuscript.

Acknowledgements

We thank the Nîmes University hospital for its structural, human, and financial support through the award obtained by our team during the internal call for tenders “Thématiques phares”. C.D.R, D.L., A.S. and J.P.L. belong to the FHU InCh (Federation Hospitalo Universitaire Infections Chroniques, Aviesan). We also express our gratitude to CEA for its long-term support for developing the phylopeptidomics methodology, as well as GIS IBISA (“Infrastructures en Biologie Santé et Agronomie”) and Région Occitanie for their support to the “ProGénoMix” mass spectrometry facility.

References

1. Guan, W. J. *et al.* Clinical Characteristics of Coronavirus Disease 2019 in China. *N Engl J Med* **382**, 1708–1720, doi:10.1056/NEJMoa2002032 (2020).
2. Wan, Y. *et al.* Enteric involvement in hospitalised patients with COVID-19 outside Wuhan. *Lancet Gastroenterol Hepatol* **5**, 534–535, doi:10.1016/S2468-1253(20)30118-7 (2020).
3. Song, Y. *et al.* SARS-CoV-2 induced diarrhoea as onset symptom in patient with COVID-19. *Gut* **69**, 1143–1144, doi:10.1136/gutjnl-2020-320891 (2020).
4. Guo, M., Tao, W., Flavell, R. A. & Zhu, S. Potential intestinal infection and faecal-oral transmission of SARS-CoV-2. *Nat Rev Gastroenterol Hepatol*, doi:10.1038/s41575-021-00416-6 (2021).
5. Wang, D. *et al.* Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China. *JAMA* **323**, 1061–1069, doi:10.1001/jama.2020.1585 (2020).
6. Chu, D. K. W. *et al.* Molecular Diagnosis of a Novel Coronavirus (2019-nCoV) Causing an Outbreak of Pneumonia. *Clin Chem* **66**, 549–555, doi:10.1093/clinchem/hvaa029 (2020).
7. Zhou, J. *et al.* Human intestinal tract serves as an alternative infection route for Middle East respiratory syndrome coronavirus. *Sci Adv* **3**, eaao4966, doi:10.1126/sciadv.aao4966 (2017).
8. Scaldaferri, F. *et al.* The Thrilling Journey of SARS-CoV-2 into the Intestine: From Pathogenesis to Future Clinical Implications. *Inflamm Bowel Dis* **26**, 1306–1314, doi:10.1093/ibd/izaa181 (2020).
9. Wang, W. *et al.* Detection of SARS-CoV-2 in Different Types of Clinical Specimens. *JAMA* **323**, 1843–1844, doi:10.1001/jama.2020.3786 (2020).
10. Xiao, F. *et al.* Infectious SARS-CoV-2 in Feces of Patient with Severe COVID-19. *Emerg Infect Dis* **26**, 1920–1922, doi:10.3201/eid2608.200681 (2020).

11. Xing, Y. H. *et al.* Prolonged viral shedding in feces of pediatric patients with coronavirus disease 2019. *J Microbiol Immunol Infect* **53**, 473–480, doi:10.1016/j.jmii.2020.03.021 (2020).
12. Xu, Y. *et al.* Characteristics of pediatric SARS-CoV-2 infection and potential evidence for persistent fecal viral shedding. *Nat Med* **26**, 502–505, doi:10.1038/s41591-020-0817-4 (2020).
13. Xie, Y., Chen, Y., Ma, M., He, D. & Yi, H. Re-emergence of coronavirus disease in Chinese cities associated with chilled and frozen food products. *J Infect*, doi:10.1016/j.jinf.2020.12.028 (2021).
14. Xiao, F. *et al.* Evidence for Gastrointestinal Infection of SARS-CoV-2. *Gastroenterology* **158**, 1831–1833 e1833, doi:10.1053/j.gastro.2020.02.055 (2020).
15. Zhang, H. *et al.* Specific ACE2 expression in small intestinal enterocytes may cause gastrointestinal symptoms and injury after 2019-nCoV infection. *Int J Infect Dis* **96**, 19–24, doi:10.1016/j.ijid.2020.04.027 (2020).
16. Zang, R. *et al.* TMPRSS2 and TMPRSS4 promote SARS-CoV-2 infection of human small intestinal enterocytes. *Sci Immunol* **5**, doi:10.1126/sciimmunol.abc3582 (2020).
17. Segal, J. P. *et al.* The gut microbiome: an under-recognised contributor to the COVID-19 pandemic? *Therap Adv Gastroenterol* **13**, 1756284820974914, doi:10.1177/1756284820974914 (2020).
18. Villapol, S. Gastrointestinal symptoms associated with COVID-19: impact on the gut microbiome. *Transl Res* **226**, 57–69, doi:10.1016/j.trsl.2020.08.004 (2020).
19. Zuo, T. *et al.* Depicting SARS-CoV-2 faecal viral activity in association with gut microbiota composition in patients with COVID-19. *Gut* **70**, 276–284, doi:10.1136/gutjnl-2020-322294 (2021).
20. Lai, L. A., Tong, Z., Chen, R. & Pan, S. Metaproteomics Study of the Gut Microbiome. *Methods Mol Biol* **1871**, 123–132, doi:10.1007/978-1-4939-8814-3_8 (2019).
21. Wang, Y., Zhou, Y., Xiao, X., Zheng, J. & Zhou, H. Metaproteomics: A strategy to study the taxonomy and functionality of the gut microbiota. *J Proteomics* **219**, 103737, doi:10.1016/j.jprot.2020.103737 (2020).
22. Long, S. *et al.* Metaproteomics characterizes human gut microbiome function in colorectal cancer. *NPJ Biofilms Microbiomes* **6**, 14, doi:10.1038/s41522-020-0123-4 (2020).
23. Nagpal, R. *et al.* Gut microbiome and aging: Physiological and mechanistic insights. *Nutr Healthy Aging* **4**, 267–285, doi:10.3233/NHA-170030 (2018).
24. Kim, Y. G. *et al.* Gut dysbiosis promotes M2 macrophage polarization and allergic airway inflammation via fungi-induced PGE(2). *Cell Host Microbe* **15**, 95–102, doi:10.1016/j.chom.2013.12.010 (2014).
25. Geva-Zatorsky, N. *et al.* In vivo imaging and tracking of host-microbiota interactions via metabolic labeling of gut anaerobic bacteria. *Nat Med* **21**, 1091–1100, doi:10.1038/nm.3929 (2015).
26. Kalantar-Zadeh, K., Ward, S. A., Kalantar-Zadeh, K. & El-Omar, E. M. Considering the Effects of Microbiome and Diet on SARS-CoV-2 Infection: Nanotechnology Roles. *ACS Nano* **14**, 5179–5182, doi:10.1021/acsnano.0c03402 (2020).

27. Zuo, T. *et al.* Alterations in Fecal Fungal Microbiome of Patients With COVID-19 During Time of Hospitalization until Discharge. *Gastroenterology* **159**, 1302–1310 e1305, doi:10.1053/j.gastro.2020.06.048 (2020).
28. Chen, L., Li, X., Chen, M., Feng, Y. & Xiong, C. The ACE2 expression in human heart indicates new potential mechanism of heart injury among patients infected with SARS-CoV-2. *Cardiovasc Res* **116**, 1097–1100, doi:10.1093/cvr/cvaa078 (2020).
29. Wanglong Gou, Y. F., Liang Yue, Geng-dong Chen, Xue Cai, Menglei Shuai, Fengzhe Xu, Xiao Yi, Hao Chen, Yi Zhu, Mian-li Xiao, Zengliang Jiang, Zelei Miao, Congmei Xiao, Bo Shen, Xiaomai Wu, Haihong Zhao, Wenhua Ling, Jun Wang, Yu-ming Chen, Tiannan Guo, Ju-Sheng Zheng. Gut microbiota may underlie the predisposition of healthy individuals to COVID-19. *medRxiv 2020.04.22.20076091*, doi:doi: https://doi.org/10.1101/2020.04.22.20076091 (2020).
30. Bauvois, B. & Dauzonne, D. Aminopeptidase-N/CD13 (EC 3.4.11.2) inhibitors: chemistry, biological evaluations, and therapeutic prospects. *Med Res Rev* **26**, 88–130, doi:10.1002/med.20044 (2006).
31. Heinz, L. X. *et al.* The Lipid-Modifying Enzyme SMPDL3B Negatively Regulates Innate Immunity. *Cell Rep* **11**, 1919–1928, doi:10.1016/j.celrep.2015.05.006 (2015).
32. Lalles, J. P. Intestinal alkaline phosphatase: novel functions and protective effects. *Nutr Rev* **72**, 82–94, doi:10.1111/nure.12082 (2014).
33. Ghanemi, A., Yoshioka, M. & St-Amand, J. Trefoil Factor Family Member 2 (TFF2) as an Inflammatory-Induced and Anti-Inflammatory Tissue Repair Factor. *Animals (Basel)* **10**, doi:10.3390/ani10091646 (2020).
34. Van Spaendonk, H. *et al.* Regulation of intestinal permeability: The role of proteases. *World J Gastroenterol* **23**, 2106–2123, doi:10.3748/wjg.v23.i12.2106 (2017).
35. Towne, C. F., York, I. A., Watkin, L. B., Lazo, J. S. & Rock, K. L. Analysis of the role of bleomycin hydrolase in antigen presentation and the generation of CD8 T cell responses. *J Immunol* **178**, 6923–6930, doi:10.4049/jimmunol.178.11.6923 (2007).
36. Bodelon, G., Palomino, C. & Fernandez, L. A. Immunoglobulin domains in Escherichia coli and other enterobacteria: from pathogenesis to applications in antibody technologies. *FEMS Microbiol Rev* **37**, 204–250, doi:10.1111/j.1574-6976.2012.00347.x (2013).
37. Lv, L. *et al.* The faecal metabolome in COVID-19 patients is altered and associated with clinical features and gut microbes. *Anal Chim Acta* **1152**, 338267, doi:10.1016/j.aca.2021.338267 (2021).
38. Ahnach, M., Zbiri, S., Nejjari, S., Ousti, F. & Elkettani, C. C-reactive protein as an early predictor of COVID-19 severity. *J Med Biochem* **39**, 500–507, doi:10.5937/jomb0-27554 (2020).
39. Gong, J. *et al.* Correlation analysis between disease severity and inflammation-related parameters in patients with COVID-19: a retrospective study. *BMC Infect Dis* **20**, 963, doi:10.1186/s12879-020-05681-5 (2020).
40. Moar, P. & Tandon, R. Galectin-9 as a biomarker of disease severity. *Cell Immunol* **361**, 104287, doi:10.1016/j.cellimm.2021.104287 (2021).

41. Smilowitz, N. R. *et al.* C-reactive protein and clinical outcomes in patients with COVID-19. *Eur Heart J*, doi:10.1093/eurheartj/ehaa1103 (2021).
42. Grenga, L. & Armengaud, J. Proteomics in the COVID-19 Battlefield: First Semester Check-Up. *Proteomics* **21**, e2000198, doi:10.1002/pmic.202000198 (2021).
43. Gouveia, D. *et al.* Proteotyping SARS-CoV-2 Virus from Nasopharyngeal Swabs: A Proof-of-Concept Focused on a 3 Min Mass Spectrometry Window. *J Proteome Res* **19**, 4407–4416, doi:10.1021/acs.jproteome.0c00535 (2020).
44. Ihling, C. *et al.* Mass Spectrometric Identification of SARS-CoV-2 Proteins from Gargle Solution Samples of COVID-19 Patients. *J Proteome Res* **19**, 4389–4392, doi:10.1021/acs.jproteome.0c00280 (2020).
45. Saadi, J. *et al.* Quantitative Assessment of SARS-CoV-2 Virus in Nasopharyngeal Swabs Stored in Transport Medium by a Straightforward LC-MS/MS Assay Targeting Nucleocapsid, Membrane, and Spike Proteins. *J Proteome Res* **20**, 1434–1443, doi:10.1021/acs.jproteome.0c00887 (2021).
46. Norton, A. *et al.* Long COVID: tackling a multifaceted condition requires a multidisciplinary approach. *Lancet Infect Dis*, doi:10.1016/S1473-3099(21)00043-8 (2021).
47. Rémic. *Rémic: référentiel en microbiologie médicale 6.2.* 2018 edn, Vol. 1 (Société française de microbiologie (SFM), 2018).
48. Grenga, L. *et al.* Shotgun proteomics analysis of SARS-CoV-2-infected cells and how it can optimize whole viral particle antigen production for vaccines. *Emerg Microbes Infect* **9**, 1712–1721, doi:10.1080/22221751.2020.1791737 (2020).
49. Hayoun, K. *et al.* Evaluation of Sample Preparation Methods for Fast Proteotyping of Microorganisms by Tandem Mass Spectrometry. *Front Microbiol* **10**, 1985, doi:10.3389/fmicb.2019.01985 (2019).
50. Hartmann, E. M., Allain, F., Gaillard, J. C., Pible, O. & Armengaud, J. Taking the shortcut for high-throughput shotgun proteomic analysis of bacteria. *Methods Mol Biol* **1197**, 275–285, doi:10.1007/978-1-4939-1261-2_16 (2014).
51. Klein, G. *et al.* RNA-binding proteins are a major target of silica nanoparticles in cell extracts. *Nanotoxicology* **10**, 1555–1564, doi:10.1080/17435390.2016.1244299 (2016).
52. Ma, Z. Q. *et al.* ScanRanker: Quality assessment of tandem mass spectra via sequence tagging. *J Proteome Res* **10**, 2896–2904, doi:10.1021/pr200118r (2011).
53. Pible, O. *et al.* Estimating relative biomasses of organisms in microbiota using "phylopeptidomics". *Microbiome* **8**, 30, doi:10.1186/s40168-020-00797-x (2020).
54. Elbe, S. & Buckland-Merrett, G. Data, disease and diplomacy: GISAID's innovative contribution to global health. *Glob Chall* **1**, 33–46, doi:10.1002/gch2.1018 (2017).
55. Gouveia, D., Grenga, L., Pible, O. & Armengaud, J. Quick microbial molecular phenotyping by differential shotgun proteomics. *Environ Microbiol* **22**, 2996–3004, doi:10.1111/1462-2920.14975 (2020).

56. Foster, Z. S., Sharpton, T. J. & Grunwald, N. J. Metacoder: An R package for visualization and manipulation of community taxonomic diversity data. *PLoS Comput Biol* **13**, e1005404, doi:10.1371/journal.pcbi.1005404 (2017).

Figures

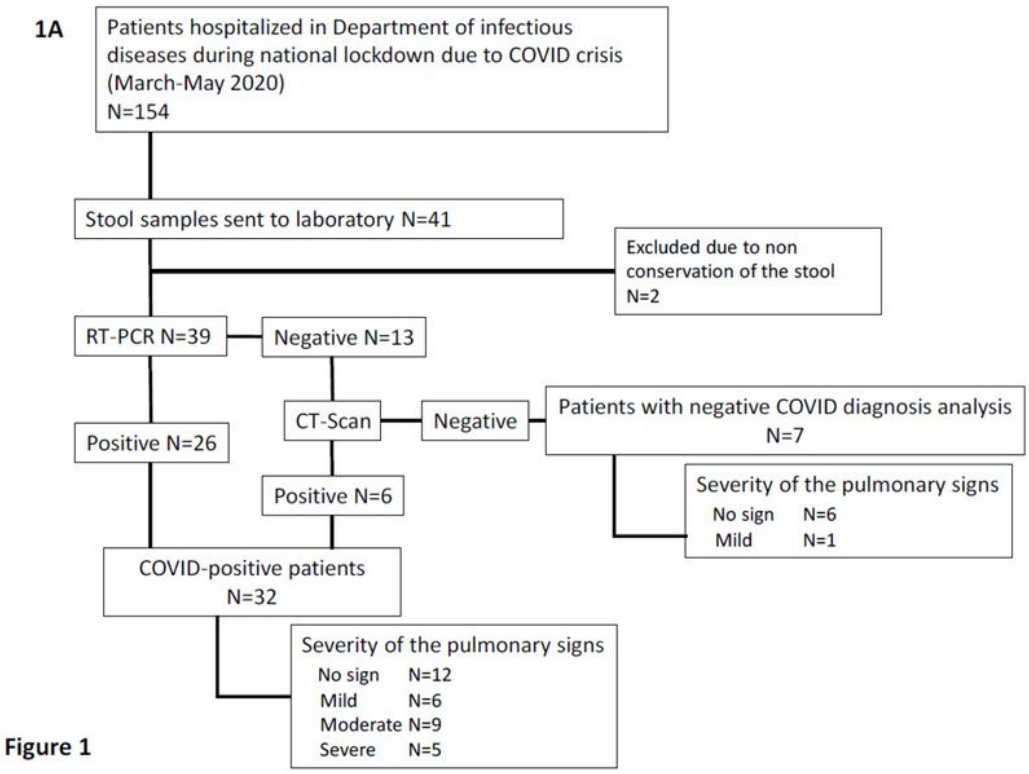

Figure 1

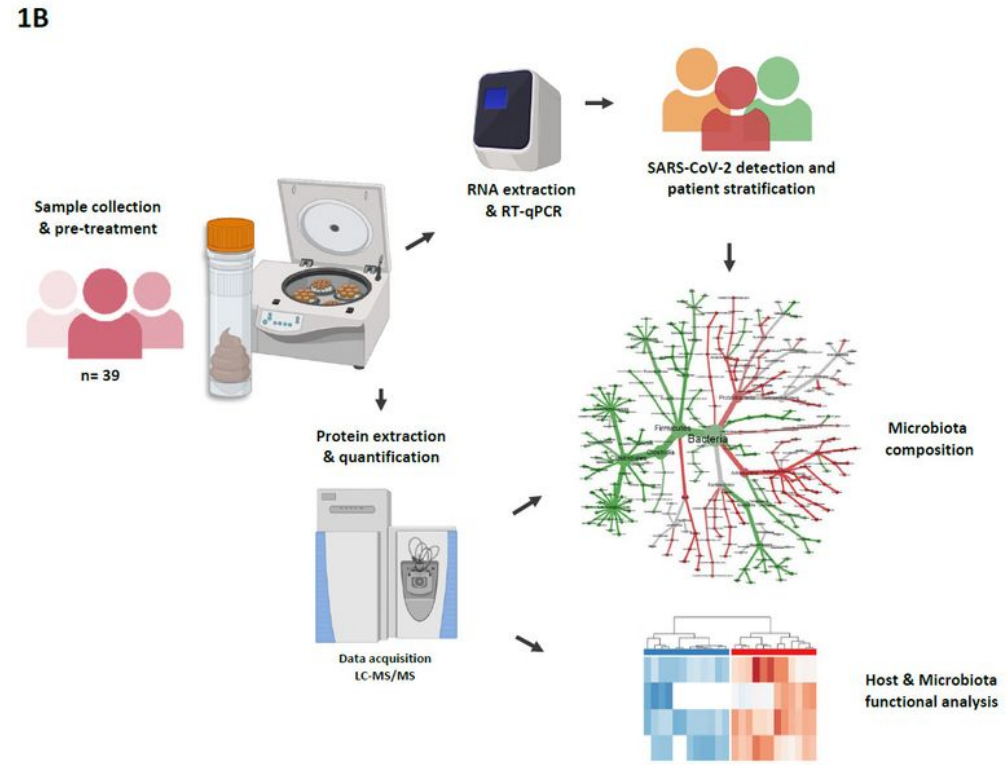

Figure 1

Microbiota profiling of SARS-CoV-2 infected patients. 1A) Flowchart of the study. 1B) Schematic representation of the experimental analytical workflow. Stool samples were analysed in parallel by RT-qPCR to detect the presence of SARS-CoV-2 RNA in the gut, and by shotgun tandem mass spectrometry to investigate the taxonomical and functional composition of the microbiota.

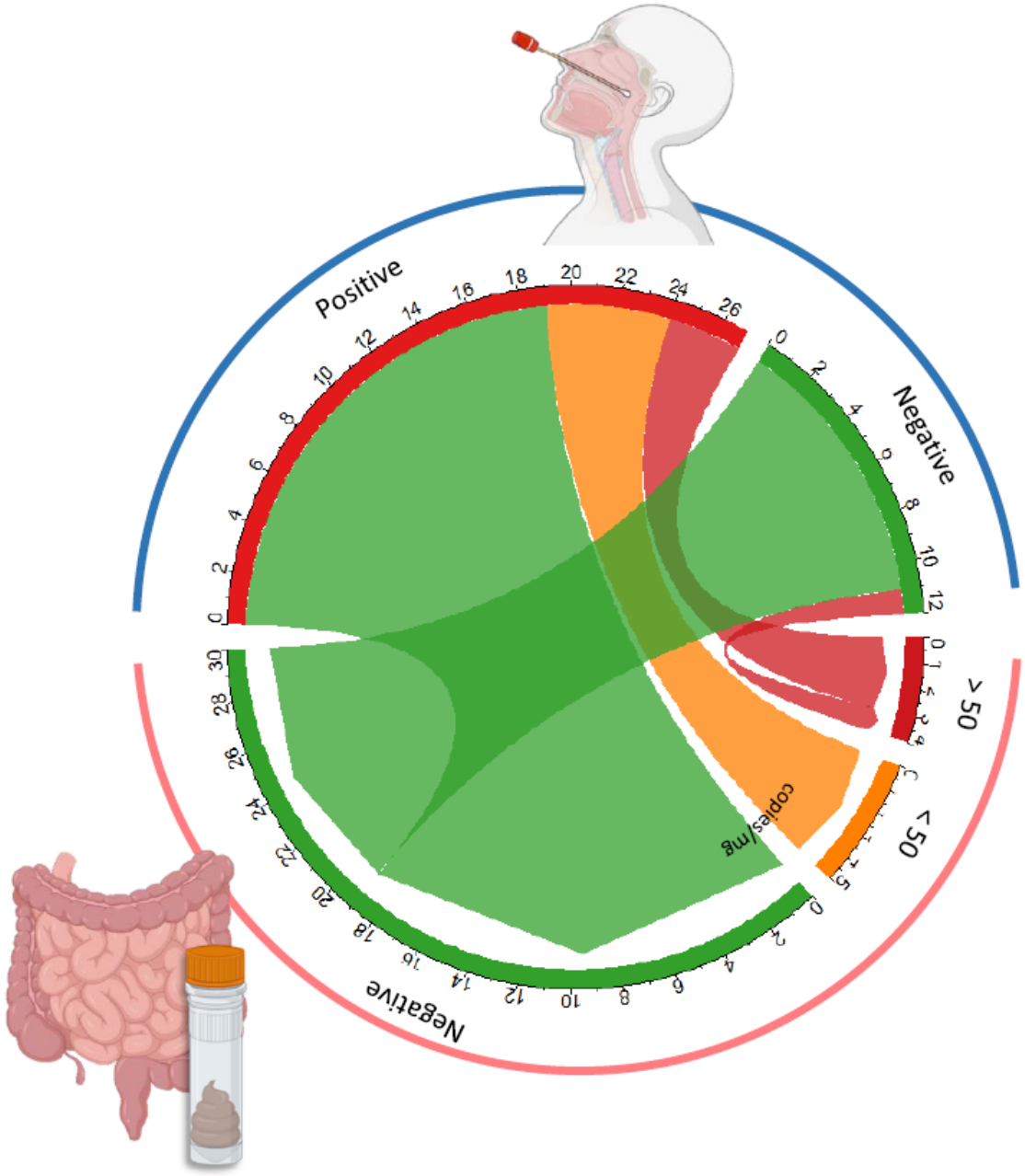

Figure 2

Presence of SARS-CoV-2 in the gut. Circos plot showing RT-qPCR results for stool samples by RT-qPCR and nasopharyngeal swab tests. The size of each arc section reflects the number of samples falling into each group. Red and orange indicate positive samples with high and low levels of SARS-COV-2, respectively; green indicates negative samples.

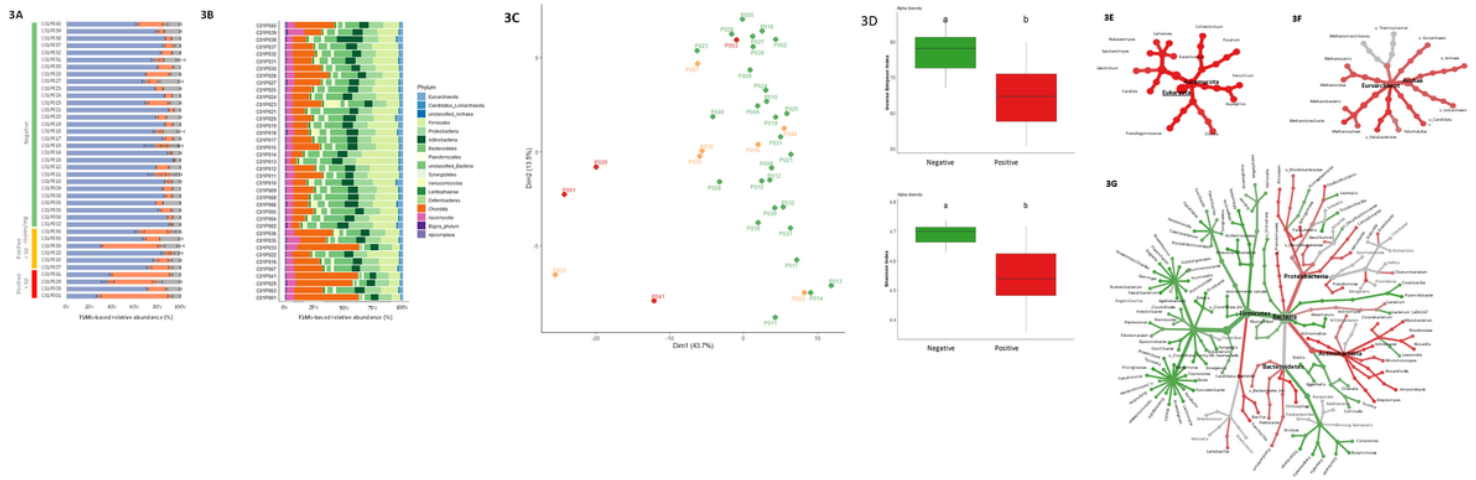

Figure 3

Metaproteomic profiling of gut microbiota. 3A) Distribution of the assigned TSMs as a function of their origin, microbial (blue), host (orange) or food-related (grey). Samples are grouped based on their SARS-CoV-2 RNA content. 3B) Relative biomass contributions of the phyla identified for each sample. Each sample was analysed in three replicates, bars represent average relative abundances for each analysis. 3C) PCA plot of the microbiota profiles. Red and orange indicate positive samples with high and low level of SARS-COV-2, respectively; green indicates negative samples. Each point corresponds to one sample, represented as the sum of the contribution of its three corresponding replicates. 3D) Box plots showing variation in alpha diversity across age-matched samples from different groups based on the Inverse Simpson and Shannon indices. Red bars indicate positive samples with high levels of SARS-COV-2, green bars indicate negative samples. a, b: significant difference between groups (Tukey's Honest Significant Difference (HSD) test). 3E) Differential heat tree for Eukaryota. 3F) Differential heat tree for Archaea. 3G) Differential heat tree for Bacteria. The trees show comparisons of microbiota compositions between samples containing high levels of SARS-CoV-2 RNA and negative samples. Phyla and genera are indicated. Taxa coloured red are more abundant in positive samples, green taxa are more abundant in negative samples.

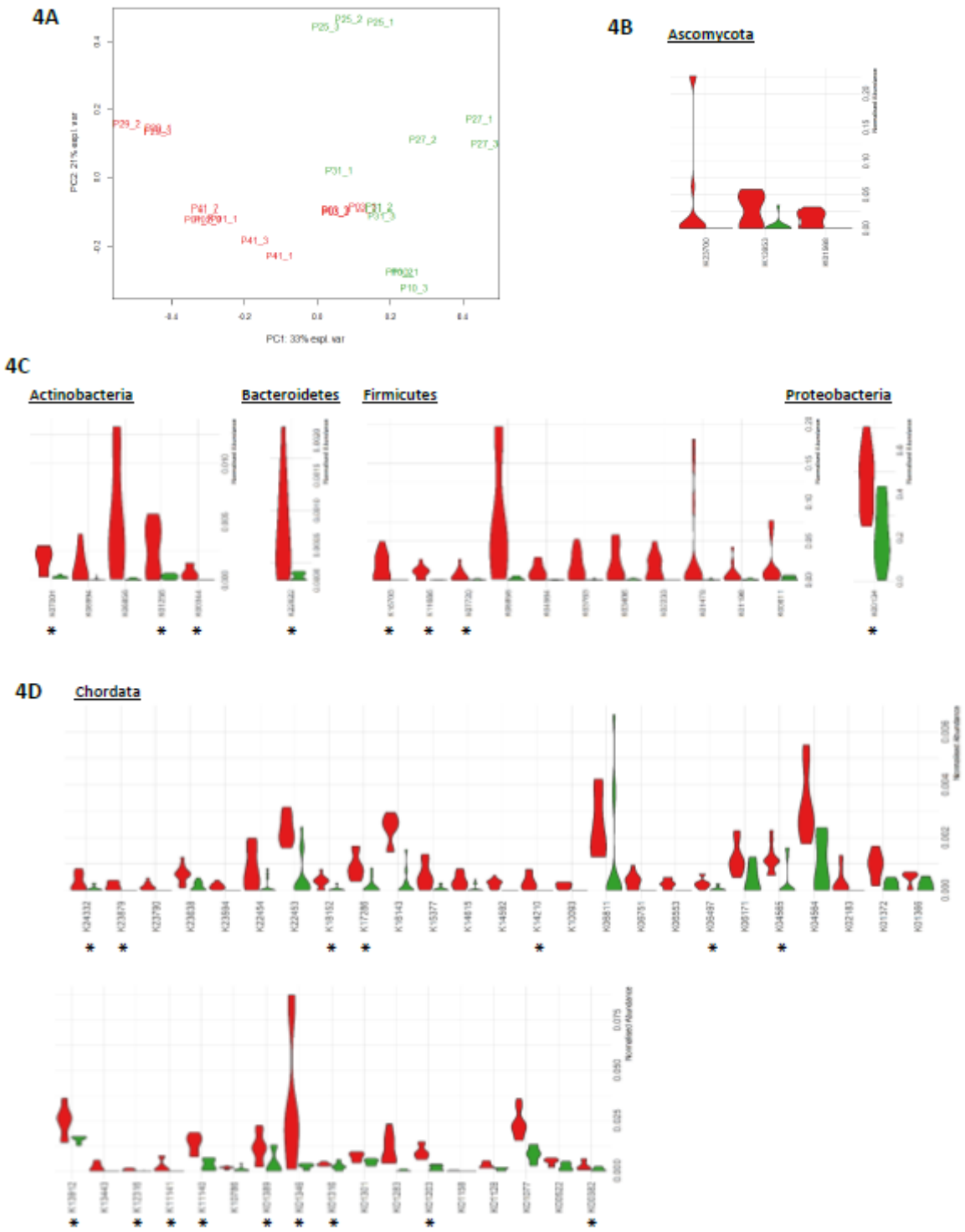

Figure 4

Comparative analysis of the functional composition of the metaproteome. 4A) PCA analysis of functional microbiota profiles from aged-matched samples containing high levels of SARS-CoV-2-RNA, or negative for viral RNA. 4B) Violin plots showing the relative abundance of the Ascomycota-related KOs increased in SARS-CoV-2-positive samples (red) or negative samples (green). Asterisks denote KOs that were also significantly increased when compared to levels detected in age-matched samples containing

lower levels of SARS-CoV-2 RNA. 4C) Violin plots showing the relative abundance of the bacterial-related KOs increased in SARS-CoV-2-positive samples (red) or negative samples (green). Asterisks denote KOs that were also significantly increased when compared to levels detected in age-matched samples containing lower levels of SARS-CoV-2 RNA. 4D) Violin plots showing the relative abundance of host-related KOs increased in SARS-CoV-2-positive samples (red) compared to negative samples (green). Asterisks denote KOs that were also significantly increased when compared to levels detected in age-matched samples containing lower levels of SARS-CoV-2 RNA.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- COVID19microbiomeTableS120210411.xlsx
- COVID19microbiomeTableS220200411.xlsx