

RESEARCH

Open Access



# Altered infective competence of the human gut microbiome in COVID-19

Laura de Nies<sup>1†</sup>, Valentina Galata<sup>1†</sup>, Camille Martin-Gallausiaux<sup>1</sup>, Milena Despotovic<sup>1</sup>, Susheel Bhanu Busi<sup>1</sup>, Chantal J. Snoeck<sup>2</sup>, Lea Delacour<sup>3</sup>, Deepthi Poornima Budagavi<sup>1</sup>, Cédric Christian Laczny<sup>1</sup>, Janine Habier<sup>1</sup>, Paula-Cristina Lupu<sup>1</sup>, Rashi Halder<sup>4</sup>, Joëlle V. Fritz<sup>5</sup>, Taina Marques<sup>6</sup>, Estelle Sandt<sup>7</sup>, Marc Paul O'Sullivan<sup>7</sup>, Soumyabrata Ghosh<sup>8</sup>, Venkata Satagopam<sup>8</sup>, CON-VINCE Consortium, Rejko Krüger<sup>5,6</sup>, Guy Fagherazzi<sup>9</sup>, Markus Ollert<sup>10,11</sup>, Feng Q. Hefeng<sup>10</sup>, Patrick May<sup>8</sup> and Paul Wilmes<sup>1,12\*</sup>

## Abstract

**Background** Infections with SARS-CoV-2 have a pronounced impact on the gastrointestinal tract and its resident microbiome. Clear differences between severe cases of infection and healthy individuals have been reported, including the loss of commensal taxa. We aimed to understand if microbiome alterations including functional shifts are unique to severe cases or a common effect of COVID-19. We used high-resolution systematic multi-omic analyses to profile the gut microbiome in asymptomatic-to-moderate COVID-19 individuals compared to a control group.

**Results** We found a striking increase in the overall abundance and expression of both virulence factors and anti-microbial resistance genes in COVID-19. Importantly, these genes are encoded and expressed by commensal taxa from families such as Acidaminococcaceae and Erysipelatoclostridiaceae, which we found to be enriched in COVID-19-positive individuals. We also found an enrichment in the expression of a betaherpesvirus and rotavirus C genes in COVID-19-positive individuals compared to healthy controls.

**Conclusions** Our analyses identified an altered and increased infective competence of the gut microbiome in COVID-19 patients.

**Keywords** COVID-19, Metagenomics, Metatranscriptomics, Gut microbiome, SARS-CoV-2

<sup>†</sup>Laura de Nies and Valentina Galata contributed equally to this work.

\*Correspondence:

Paul Wilmes  
paul.wilmes@uni.lu

<sup>1</sup> Systems Ecology Group, Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg

<sup>2</sup> Clinical and Applied Virology, Department of Infection and Immunity, Luxembourg Institute of Health, Esch-sur-Alzette, Luxembourg

<sup>3</sup> Luxembourg Centre for Systems Biomedicine, LCSB Operations, University of Luxembourg, Esch-sur-Alzette, Luxembourg

<sup>4</sup> Scientific Central Services, Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg

<sup>5</sup> Transversal Translation Medicine, Luxembourg Institute of Health, Strassen, Luxembourg

<sup>6</sup> Translational Neuroscience Group, Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg

<sup>7</sup> Translational Medicine Operations Hub, Luxembourg Institute of Health, Strassen, Luxembourg

<sup>8</sup> Bioinformatics Core, Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg

<sup>9</sup> Deep Digital Phenotyping Research Unit, Department of Precision Health, Luxembourg Institute of Health, Strassen, Luxembourg

<sup>10</sup> Department of Infection and Immunity, Luxembourg Institute of Health, Esch-sur-Alzette, Luxembourg

<sup>11</sup> Department of Dermatology and Allergy Centre, Odense University Hospital, Odense, Denmark

<sup>12</sup> Department of Life Sciences and Medicine, Faculty of Science, Technology and Medicine, University of Luxembourg, 6, Avenue du Swing, L-4367 Belvaux, Luxembourg



## Background

Coronavirus disease 2019 (COVID-19), which is caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), was declared a global pandemic by the World Health Organization (WHO). COVID-19 exhibits a high degree of clinical heterogeneity, ranging from asymptomatic to severe disease, and may be accompanied by a poor outcome and a relatively high mortality rate [1]. As of 17 October 2022, more than 621 million confirmed SARS-CoV-2 infections and 6.5 million COVID-19-related deaths have been reported [2]. Although COVID-19 is primarily considered a respiratory disease, it clinically often presents with general (fever, myalgia, and/or fatigue) and respiratory symptoms (cough and/or dyspnea). Moreover, an emergence of new variants has led to the more frequent presentation of gastrointestinal symptoms (appetite loss, nausea, vomiting, and diarrhea) [3], indicating a potential involvement of the gastrointestinal tract in COVID-19. More specifically, SARS-CoV-2 has been shown to be able to infect and replicate in enterocytes in vitro [4]. In fact, viral RNA can be detected in fecal samples even after resolution of respiratory symptoms [5]. Additionally, SARS-CoV-2 infections are associated with alterations to the gut microbiome composition that persist for at least 6 months after the initial infection [6]. Thus, an imbalance in the gut microbiome can be linked to disease severity and increased concentrations of inflammatory markers, as well as an increased post-COVID-19 risk, understood as a wide range of symptoms persisting four or more weeks after the initial SARS-CoV-2 infection [6, 7].

Stable ecosystems are important for colonization resistance to pathogens [8]. As such, host and SARS-CoV-2-mediated immune dysregulation and dysbiosis may predispose patients to co-infections or secondary infections of the respiratory and gastrointestinal tracts. In addition, co-infecting microorganisms may alter the intensity of the host immune response [9], thus significantly influencing severity and outcome of the disease. For instance, co-infections with viruses (rhinovirus/enterovirus, respiratory syncytial virus, influenza virus, non-SARS-CoV-2 Coronavirus) [10], bacteria (*Mycoplasma pneumoniae*, *Pseudomonas aeruginosa*, *Haemophilus influenzae*, *Klebsiella pneumoniae*, *Streptococcus pneumoniae*, *Staphylococcus aureus*) [11, 12], or fungi (*Candida spp.*, *Aspergillus spp.*) [13] have been described among SARS-CoV-2-positive cases in different study set-ups. In particular, bacterial co-infections in hospitalized and intensive care unit patients with COVID-19 are associated with prolonged ventilation and an increased mortality rate [11, 14]. Furthermore, hospital-acquired infections with multi-drug-resistant (MDR) pathogens are also linked with increased mortality in COVID-19

patients [15]. These reports collectively suggest a clear shift in COVID-19 patients with respect to an increased abundance of pathogens and potential for harm. Moreover, these shifts may further manifest themselves in relation to the *infective competence*, i.e., the propensity for virulence and increased antibiotic resistance, in the gut microbiome as a consequence of an increased capacity to cause infections.

Major factors that contribute to the success of some of the pathogens highlighted above are virulence factors (VFs). Virulence factors including cell-surface structures, adhesins, siderophores, endo-, and exotoxins enable pathogens to undergo quick adaptive shifts, invade and colonize host niches, as well as evade innate and adaptive immune mechanisms of the host, resulting in inflammation and clinical manifestations of the disease. Another factor facilitating colonization of pathogens, through prevention of effective treatment, is antimicrobial resistance (AMR). Even though AMR is an ancient and natural phenomenon [16], it is usually linked to the human influence on the environment and the use of antibiotics. Overuse of antibiotics is hypothesized to also contribute to the broader problem of antimicrobial resistance [17]. Moreover, although not a VF by itself, AMR shares common characteristics with VFs [18]. Specifically, AMR and VFs: (1) are necessary for the survival of pathogens under unfavourable conditions [19]; (2) can be transmitted between species by horizontal gene transfer [20]; and (3) both processes make use of similar systems, e.g., cell wall alterations, efflux pumps, porins, and two-component systems to activate or repress expression of various genes [18, 21]. Thus, in response to host defence mechanisms and environmental challenges, communities of microorganisms, i.e., microbiomes, may alter their "*infective competence*". The *infective competence* is defined as the ability of microorganisms to constantly adapt and evolve, utilizing VFs and AMR mechanisms, resulting in increased survival, invasion, or growth. Importantly, the combination of host-driven factors, i.e., immune system-mediated effects and antimicrobial peptides, and unfavourable gastrointestinal conditions, e.g., low pH, disruption of the mucus layer, niche competition with other taxa, may confer transiently a selective advantage to a pathogenic lifestyle [22, 23]. This may be reflected in the entire gut microbiome, possibly altering the *infective competence* of the endogenous taxa and subsequently giving rise to pathobiont-dominated communities.

Here, we addressed questions pertaining to the effect of SARS-CoV-2 infection on the endogenous gut microbiome in COVID-19 cases compared to healthy controls using systematic, high-resolution multi-omic data, including metagenomics and metatranscriptomics with a particular focus on VFs and antimicrobial resistance

genes (ARGs). We find that mild, i.e. asymptomatic-to-moderate, COVID-19 does not alter the overall composition of the gut microbiome, unlike the drastic microbiota changes reported previously in severe cases. Importantly, we find that a mild progression of COVID-19 affects the *infective competence* of gut microbiota, wherein taxa encode and express genes facilitating their survival and/or growth. We find specific families such as Acidaminococcaceae and Erysipelatoclostridiaceae to be encoding for and expressing VFs and ARGs, significantly more in individuals with COVID-19. Collectively our data also demonstrates a significantly higher *infective competence* of the endogenous microbiome, suggesting that infection with SARS-CoV-2 may mediate co-infections in the longer term.

**Results**

**Taxonomic and functional profiles indicate minimal changes in COVID-19**

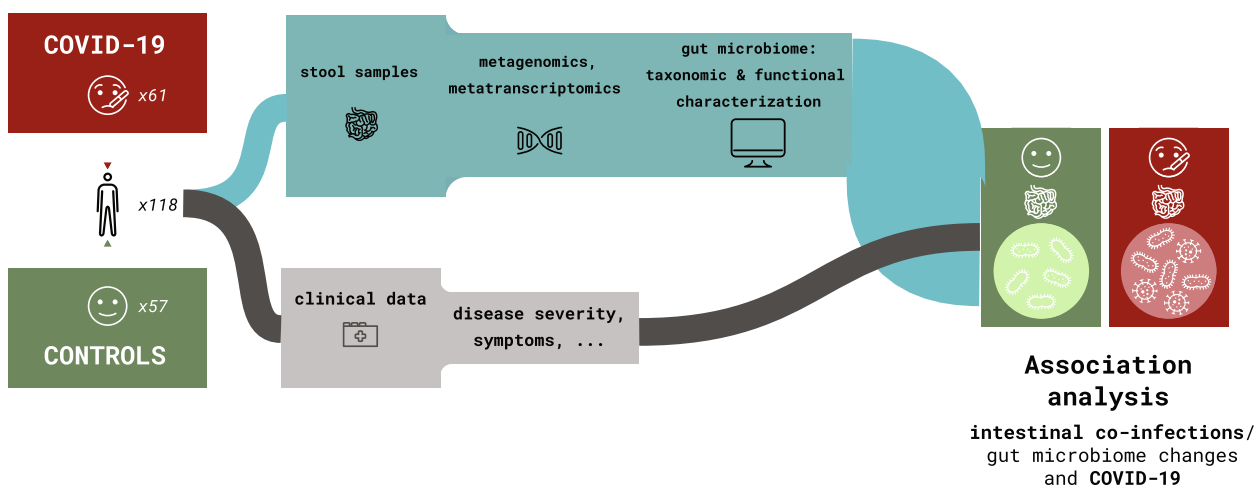
COVID-19 studies have reported an altered gut microbiota composition of hospitalized and critical COVID-19 patients. However, limited attention has been paid to milder forms of COVID-19. Thus, we assessed whether gut microbiota composition was altered in COVID-19 individuals compared to healthy controls. Overall, the gut microbiome compositions, based on the alpha- and beta-diversity metrics, of 61 COVID-19 and 57 individuals from the control group were similar (Fig. 1 and Supplementary Figure S1), with an increased abundance of species belonging to the Lachnospiraceae, Ruminococcaceae, Bacteroidaceae, and Bifidobacteriaceae families in COVID-19 (Fig. 2a). We found specific taxonomic differences within the metagenomes, such as an increase in

the abundance of AM10 47 (Firmicutes phylum), *Prevotella sp.* CAG 520, *Prevotella stercorea* and *Roseburia sp.* CAG 471 in the COVID-19 group (Fig. 2b), along with a decrease in CAG 145 (Firmicutes phylum), *Roseburia faecis* and *Turicibacter sanguinis* (Fig. 2c). Despite these taxonomic differences, we did not observe any significant changes in the overall functional profile of the microbiome between the COVID-19 and control groups. Along similar lines, we did not find a significant correlation between covariates such as age, sex, COVID-19 severity, and other variables in the COVID-19 and control groups in relation to the taxonomic and functional features.

In light of reports, indicating the potential co-infections with viruses along with SARS-CoV-2, we also assessed the virome (Methods) within the COVID-19 patient and the control groups. We did not observe large differences between the groups. However, we found that genes associated with a specific betaherpesvirus and rotavirus were enriched (*adj. p* < 0.05; one-way ANOVA) in the COVID-19 group (Supplementary Table S2).

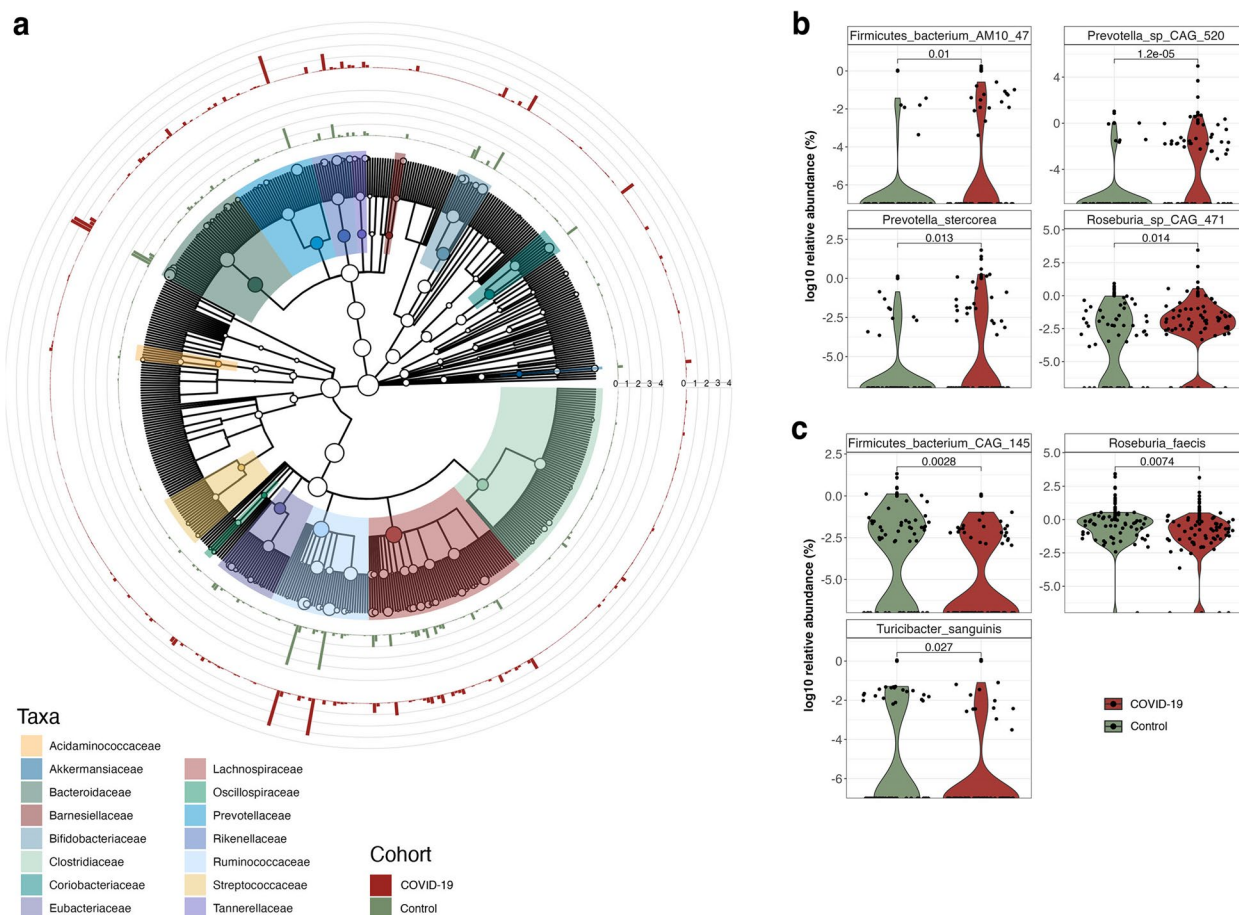
**SARS-CoV-2 is associated with increased abundance and expression of virulence factors**

SARS-CoV-2 infections have been suggested to predispose patients to co-infections or secondary infections of the respiratory and gastrointestinal tracts. Virulence factors in particular enable (pathogenic) microorganisms to colonize host niches and establish infections. We used PathoFact [24] to assess the prevalence of VFs in the co-assembled metagenomic and metatranscriptomic data. PathoFact was designed to contextualize the genomic data and classify VFs and ARGs, allowing to assess the *infective competence* of taxa. To obtain a comprehensive



Icons: OpenMoji

**Fig. 1** Sample collection and study design. Schematic of the project study design, including cohort composition, and data analyses



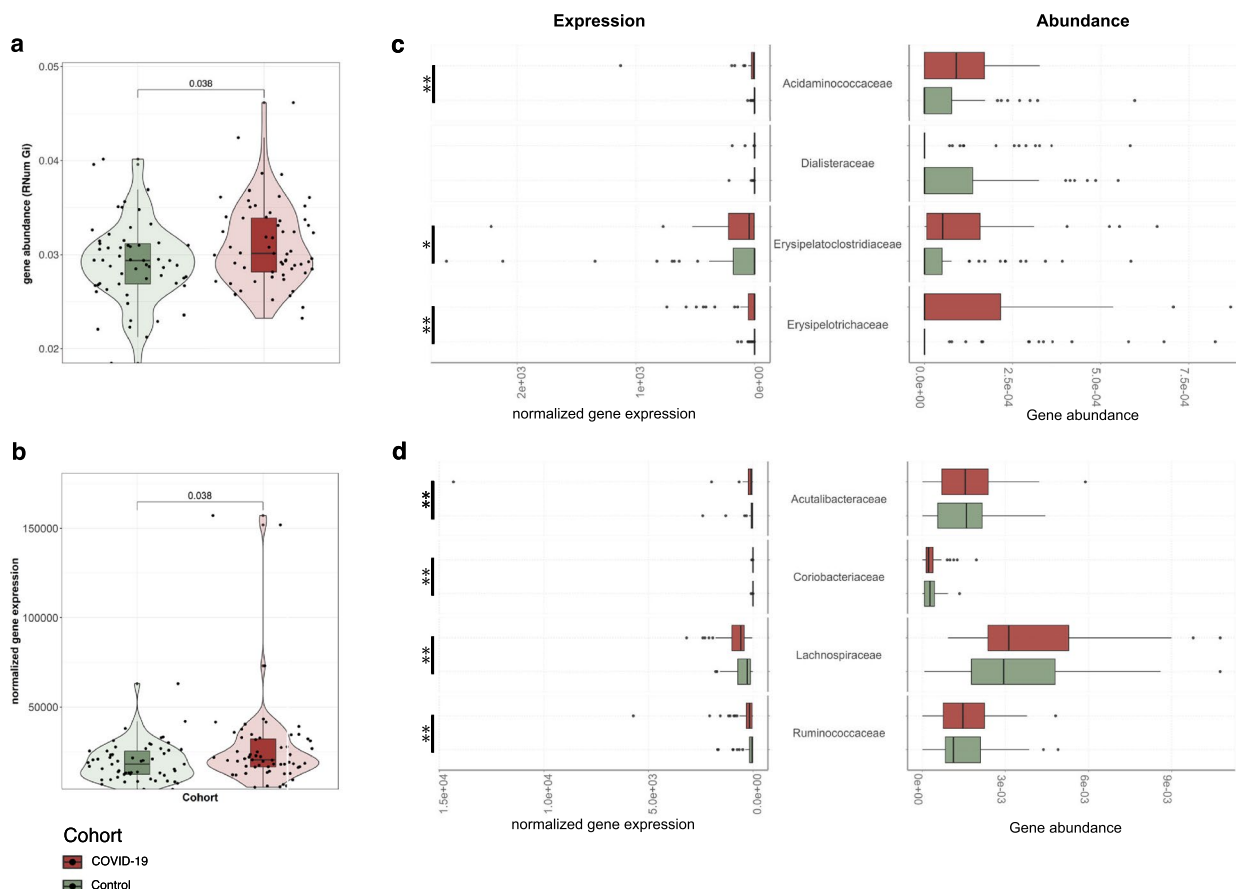
**Fig. 2** Composition of the microbial community. **a** Cladogram representing the microbial community profiles in COVID-19 patients (red) and control group (green). The outer rings represent the relative abundance (%) of the microbial community. **b** Relative abundance of bacterial species significantly enriched in COVID-19 patients compared to the control group [*adj. p* < 0.05; Wilcoxon rank-sum test]. **c** Relative abundance of bacterial species significantly decreased in COVID-19 patients compared to the control group [*adj. p* < 0.05; Wilcoxon rank-sum test]

overview of actual gene expression, we complemented metagenomic analyses with metatranscriptomic information conferring information regarding the transcription levels of identified VFs. Based on the metagenomic data, we found a significant increase (*adj. p* < 0.05; Wilcoxon rank-sum test) in alpha diversity (Supplementary Figure S2) as well as the overall abundance of VFs in the COVID-19 group compared to the control group (Fig. 3a). The metatranscriptomic information further confirmed that these VFs demonstrated significantly increased expression levels (*adj. p* < 0.05; Wilcoxon rank-sum test) in the COVID-19 group compared to the control group (Fig. 3b).

To link the prevalence and expression of the identified VFs to the taxa within the microbial community, we reconstructed metagenome-assembled genomes (MAGs) and further leveraged the iterative workflow of the integrated meta-omic pipeline (IMP) [25].

Overall, we found a significant increase in encoded and expressed VFs between the COVID-19 and control groups (*adj. p* < 0.05; Wilcoxon rank-sum test). Our analyses further linked families such as Acidaminococcaceae, Erysipelatoclostridiaceae, and Erysipelotrichaceae with increased expression of VFs in the COVID-19 group (Fig. 3c). Interestingly, the control group exhibited higher gene abundances and expression of VFs only in the Dialisteraceae family. Furthermore, we found that some families (Acutalibacteraceae, Coriobacteriaceae, Lachnospiraceae, and Ruminococcaceae) demonstrated an increased expression of VFs in the COVID-19 group (Fig. 3d; *adj. p* < 0.05; Wilcoxon rank-sum test), although their respective gene abundances were not different from those found in the control group.





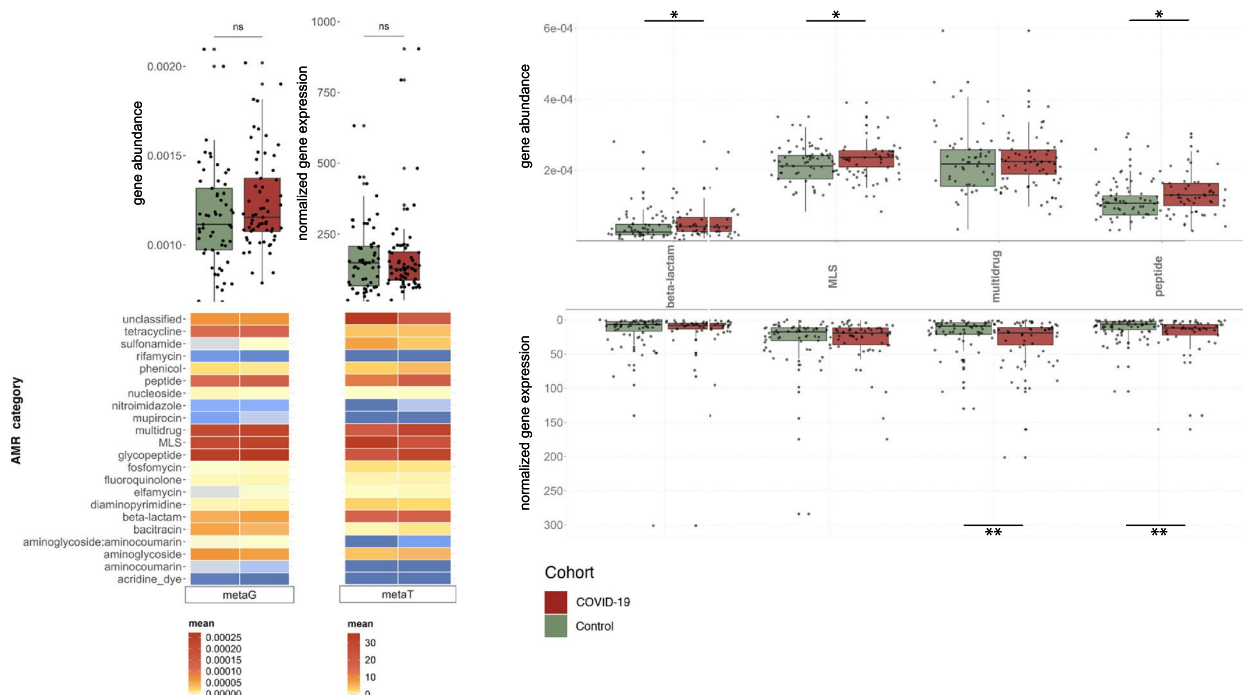
**Fig. 3** Abundance of virulence factors in the microbial community. **a** Overall abundance (metagenome) of virulence factors encoded by the microbiome of COVID-19 patients and control group. The significance of the differential abundance is indicated with the adjusted  $p$  value [ $adj.p < 0.05$ ; Wilcoxon rank-sum test]. **b** Overall expression levels (metatranscriptomics) of virulence factors encoded by the microbiome in COVID-19 patients and the control group [ $adj.p < 0.05$ ; Wilcoxon rank-sum test]. **c** Abundance and expression levels of MAG families where a significant increase in encoded and expressed virulence factors was observed in COVID-19 patients [ $adj.p < 0.05$ ; Wilcoxon rank-sum test, \*  $< 0.05$ , \*\*  $< 0.01$ , \*\*\*  $< 0.001$ ]. **d** Abundance and expression levels of virulence factors in MAGs depicting taxonomic families only demonstrating an increased expression of virulence factors, with no significant difference observed at a metagenomic level [ $adj.p < 0.05$ ; Wilcoxon rank-sum test, \*  $< 0.05$ , \*\*  $< 0.01$ , \*\*\*  $< 0.001$ ]

### Expression of antimicrobial resistance increases together with virulence factors

While co-infections or secondary infections in COVID-19 may exacerbate the disease, the presence of ARGs may limit treatment options. Since the overall abundance and expression of VFs was increased in COVID-19 individuals, we assessed the antimicrobial resistance profile of the microbial community in the COVID-19 and control groups. Specifically, using PathoFact, we characterized the prevalence and relative expression of ARGs (22 categories). While we did not find any significant differences in the overall gene abundances and the normalized expression levels of all ARGs contributing to the resistome, we observed a significant increase ( $adj. p < 0.05$ ; Wilcoxon rank-sum test) in ARG alpha diversity (Supplementary Figure S3) between COVID-19 and the control groups (Fig. 4a). Importantly, when

investigating individual AMR categories, we found that peptide resistance was significantly higher in terms of gene abundance and also more highly expressed within the COVID-19 group (Fig. 4b;  $adj. p < 0.05$ ; Wilcoxon rank-sum test). In addition, we observed that the expression of multi-drug resistance was enriched ( $adj. p < 0.05$ ; Wilcoxon rank-sum test) in the COVID-19 group, while macrolides, lincosamides and streptogramins (MLS) and beta-lactam resistance both exhibited a higher gene abundance in the same group ( $adj. p < 0.05$ ; Wilcoxon rank-sum test).

As described above, we leveraged the MAGs to correlate the differentially abundant and expressed ARGs to the microbial community. In line with our observations with the VFs, we found a significant increase ( $adj. p < 0.05$ ; Wilcoxon rank-sum test) in ARGs encoded and expressed by the Acidaminococcaceae



**Fig. 4** Abundance levels of antimicrobial resistance genes. **a** Overall ARG abundance and expression levels for COVID-19 and control groups (boxplot), coupled with a breakdown of the respective abundance and expression levels to individual AMR categories [ $adj.p < 0.05$ ; Wilcoxon rank-sum test, \*  $< 0.05$ , \*\*  $< 0.01$ , \*\*\*  $< 0.001$ ]. **b** ARG abundance (top) and expression levels (bottom) of individual AMR categories significantly increased in COVID-19 patients compared to the control group [ $adj.p < 0.05$ ; Wilcoxon rank-sum test, \*  $< 0.05$ , \*\*  $< 0.01$ , \*\*\*  $< 0.001$ ]

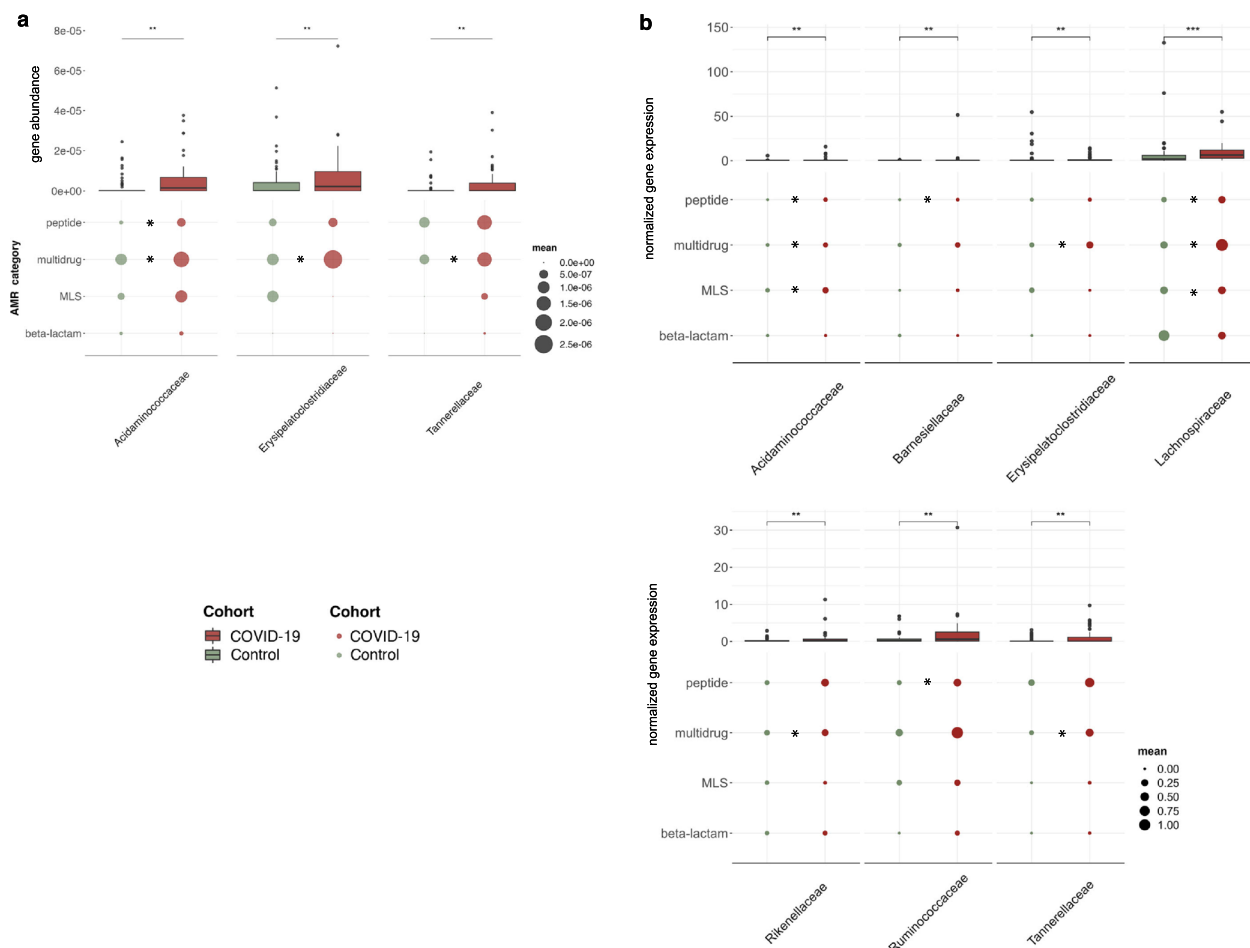
and Erysipelatoclostridiaceae in the COVID-19 group (Fig. 5a, b). Furthermore, an additional family, i.e., Tannerellaceae was also associated with increased abundance and expression of ARGs in the COVID-19 group (Fig. 5a, b). Specifically, in relation to the above-reported AMR categories, we identified a significant increase in multi-drug resistance encoded and expressed by all three of these taxonomic families. In addition, the Acidaminococcaceae also were found to encode a significant increase in ARGs contributing to peptide resistance. Interestingly, we found that several other taxonomic families were also associated with increased ARG expression in the COVID-19 group ( $adj. p < 0.05$ ; Wilcoxon rank-sum test), although their gene abundances did not demonstrate any significant differences (Fig. 5b). These included Barnesiellaceae, Lachnospiraceae, Ruminococcaceae, and Rikenellaceae.

To further validate our findings, especially those linking VFs and ARGs with taxa, we used bias correction-based analyses for microbial compositions (ANCOM-BC). For both abundances and relative expression, across VFs and ARGs, ANCOM-BC revealed similar taxonomic families were enriched in the COVID-19 group, as identified in our initial analyses using MaAsLin2 and subsequent non-parametric tests. Based on the ANCOM-BC analyses, the COVID-19 group had a higher log<sub>2</sub> fold-change

of Erysipelatoclostridiaceae, Acidaminococcaceae, and Erysipelotrichaceae in the VFs compared to the control group (Supplementary Figure S4a-b). Similarly, ARGs in the Acidaminococcaceae and Erysipelatoclostridiaceae families were both abundant and showed higher relative expression in the COVID-19 group (Supplementary Figure S4 c-d).

#### Infective competence of the gut microbiome

Our analyses collectively indicated that both VFs and ARGs were enriched in abundance and expression in the COVID-19 group. Specifically, we found that the abundances of ARGs were correlated with those of the VFs (Fig. 6a,  $R = 0.52$  and  $p < 0.01$ ; Spearman's correlation). Complementing this observation, we found that the expression profiles of ARGs and VFs also correlated with each other (Fig. 6b,  $R = 0.46$  and  $p < 0.01$ ; Spearman's correlation) suggesting a higher propensity for infectious capacity. To further characterize the *infective competence* of the various taxa within the gut microbiome, we estimated the log<sub>2</sub> fold-change of the abundance and expression of VFs and ARGs across taxonomic families found in the COVID-19 group and the control group. We found that ~62% (21/34) of the families had a higher *infective competence* and were enriched in abundance and expression within the COVID-19 group, whereas only ~9%



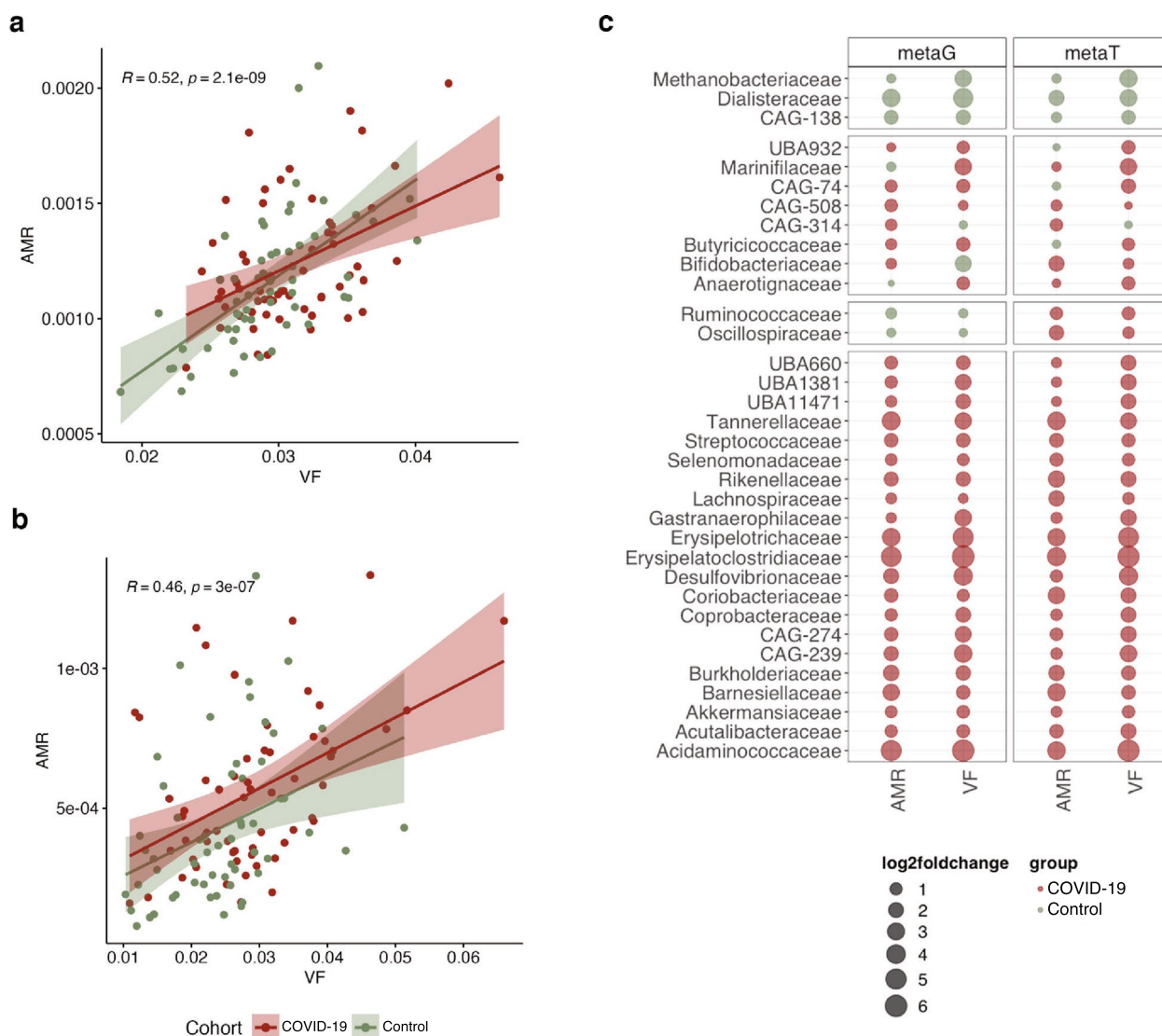
**Fig. 5** Association of AMR with the microbial community. Abundance (**a**) and expression (**b**) levels of ARGs and corresponding to AMR categories linked to MAGs. On top (boxplot) depicting the overall ARG abundance, below the average abundance of selected AMR categories per taxonomic family. The plot depicts taxonomic families in which overall a significant increase in abundance or expression of ARGs was observed [ $adj.p < 0.05$ ; Wilcoxon rank-sum test, \*  $< 0.05$ , \*\*  $< 0.01$ , \*\*\*  $< 0.001$ ]

(3/34) of the families showed increased *infective competence* in the control group (Fig. 6c). In particular, these analyses highlighted the Acidaminococcaceae and Erysipelatoclostridiaceae families, in line with our earlier observations, suggesting a higher *infective competence*, where the abundances and expression levels of the VFs and ARGs were significantly higher in COVID-19 compared to the control group ( $p < 0.05$ ; two-way ANOVA). In the control group, Dialisteraceae, which was also observed earlier, showed increased *infective competence* (Fig. 6c). Collectively, our data suggests that the *infective competence* of taxa found in the COVID-19 group is increased compared to controls.

## Discussion

COVID-19 has become a common condition for which the manifold effects however remain a challenge [26]. Since the onset of the pandemic, the presentation of

gastrointestinal symptoms has indicated the involvement of the gastrointestinal tract in COVID-19 [7]. As we uncover and understand the potential effects of COVID-19 in humans, it is important to also elucidate the concomitant consequences of the disease on the gut microbiome. To this end, several studies have focused on the drastic shifts in the microbiome of COVID-19 patients with severe symptoms. These include changes in diversity including stark enrichments and/or loss of specific taxa [27]. Several Studies have focused on differences in the gut microbiome between patients with severe COVID-19 and controls [7, 28]. Though these findings are essential, the effect on the larger population, wherein the infection is asymptomatic-to-moderate, is not readily represented. To address this particular gap in knowledge, we focused on the effect of COVID-19 in cases with asymptomatic-to-moderate symptoms in comparison to controls. Interestingly, we found that



**Fig. 6** Assessing the *infective competence* of the gut microbiome. **a** Correlation of gene abundances of AMR and virulence factors [ $R = 0.52$  and  $p < 0.01$ ; Spearman's correlation] in COVID-19 patients (red) and the control group (green). **b** Correlation of AMR and virulence factors gene expression levels [ $R = 0.46$  and  $p < 0.01$ ; Spearman's correlation] in COVID-19 patients (red) and negative controls (green). **c** Bubble plot depicting the *infective competence* via the log2 fold change of AMR and virulence factors between COVID-19 patients (red) and control group (green)

the diversity and overall shifts in community composition, highlighted in previous reports between severe and control patients [29], did not manifest themselves when comparing asymptomatic-to-moderate cases to the control group of individuals. However, this was associated with an increased abundances of specific taxa such as *Prevotella* spp., AM10 and CAG145 (Firmicutes phylum), *Roseburia* spp. and a *Turicibacter* spp. in the COVID-19 group. This is in contrast to existing reports [27, 30], suggesting the loss of beneficial taxa such as *Faecalibacterium*, *Bifidobacterium* and *Roseburia* in the context of COVID-19. Since our study did not include patients

with severe COVID-19 or those that were hospitalized, it is likely that the lower disease severity does not lead to significant changes in the abundance of beneficial commensals. Along similar lines, major differences in the virome profile of the COVID-19 group in our study were not observed when compared to the control group. Nevertheless, and importantly, we found that genes associated with rotavirus C were increasingly expressed in the COVID-19 group, despite no differences in overall abundance of this virus between the patient groups. Rotavirus is a known enteric pathogen causing gastroenteritis in the pediatric population; however, their capacity to cause



infections in adults is underappreciated and poorly characterized due to only mild symptoms including nausea, headaches, and diarrhea [31]. Importantly, at the time of writing only one report by Wang et al. [32], indicated the possibility of increased rotavirus A-mediated enteric infections in COVID-19 patients. These findings are intriguing given the propensity for COVID-19 patients to suffer from enteric symptoms [3], including nausea [33] and diarrhea [34]. Whether the rotavirus, especially in adults, is associated with COVID-19 gastrointestinal symptoms, or the enteric effects exacerbate the expression of rotavirus C-associated genes is still unknown and will have to be investigated in dedicated follow-up studies.

In line with the above observations, early in the pandemic, the role of COVID-19 in enhancing co-infections was documented extensively [14, 35]. This is not only limited to mucormycosis [36] which was amplified in certain parts of the world, but also bacterial and viral co-infections that were reported in severe COVID-19 patients [32]. Despite these observations and case studies, the effect of COVID-19 on the *infective competence* of the existing and endogenous microbiota has never been characterized. Our findings, therefore, bridge an important and broad chasm in knowledge, suggesting that COVID-19-mediated shifts may lead to higher microbiome-linked burden with potentially manifold effects. Importantly, we not only found an increased abundance in VFs in the COVID-19 group, but also a concomitant increase in expression of genes associated with virulence. Although it is plausible that a positive correlation between VFs and ARGs exists in de facto pathogens, such infection-linked shifts have not been reported beforehand. Furthermore, this phenomenon has not been reported in commensal organisms. Thereby, *infective competence* may be used to monitor and understand potential future infections in the context of COVID-19-mediated effects. Simultaneously, we found that these VFs were associated with taxa from families such as Acidaminococcaceae, Erysipelotoclostridiaceae, and Erysipelotrichaceae. Though *Acidaminococcus* was recently reported to be associated with a disease-related group in a large-scale meta-analysis [37], the exact role of Acidaminococcaceae in virulence is undocumented. Members of the Erysipelotoclostridiaceae family are typically seen as typical members of the microbiome; however, in specific cases species such as *Erysipelotoclostridium ramosum* have been associated with systemic infection and systemic inflammatory response syndrome [38], while Erysipelotrichaceae have been positively correlated with colorectal cancer [39]. Our observations, especially the increased expression of VF genes associated with these taxa, may pave the way in future explorations to serve as indicators of diseases.

Importantly, it is still unclear whether the enriched *infective competence* is a COVID-19-specific hallmark or one found in all infections. For example, it has previously been hypothesized that selection of pathobionts result from inflammatory responses and/or a dysregulation of the tolerant immune system [40]. Future studies will need to address the extent to which various underlying factors such as a dysbiotic microbiome and an impaired immune system affect the *infective competence* of the gut microbiome. Furthermore, with larger cohorts which would include severe cases of COVID-19, supervised learning analyses may be employed to predict disease status based on the *infective competence* of individuals' microbiome. Further work may also involve the heterologous expression of VF and AMR genes to assess predicted versus realized *infective competence*.

Another important aspect of COVID-19, in particular early on in the pandemic, was the overuse and misuse of antibiotics for treating SARS-CoV-2 [41] which was also associated with the potential increase of AMR [17]. Based on our findings and a recent report from the European Centre for Disease Prevention Control showing a North-to-South as well as a West-to-East AMR gradient in Europe during the COVID-19 pandemic [42], it is imperative to undertake future and detailed analyses accounting for socioeconomic and geographic factors contributing to AMR. In this context, Luxembourg may constitute an important reference population given its geographic location and its diverse demographic composition. Recent studies have reported on the higher incidence of AMR [43] and increased ARGs in COVID-19 patients [44]. However, these reports either refer to patients who were administered antibiotics [44] or include a meta-analysis observing datasets which were generated pre- and post-pandemic, specifically associated with travel [45], or limit characterization of antibiotic-mediated differences at a broad and low-resolution [46]. In contrast to these studies, antibiotic usage was a clear exclusion criterion in our study where individuals included were not administered any antibiotics 3 months prior to sampling. To our knowledge, our findings are the first report to systematically analyze the resistome of COVID-19 and control individuals and importantly to demonstrate that several of these ARGs are indeed expressed significantly higher in the COVID-19 group compared to the control group, regardless of antibiotic treatment. We observe that resistance genes include MLS, multi-drug and peptides, resistance classes where treatments of resistant bacteria are known to be inherently challenging with conventional antibiotics [47]. Strikingly, we found that the increased ARG expression in the COVID-19 group was further associated with the same taxa encoding and expressing VFs. This suggests

that combinatorial effects of VFs and ARGs may exacerbate the *infective competence* of these taxa. This is further supported by our analysis identifying that taxa from the Acidaminococcaceae and Erysipelatoclostridiaceae families demonstrated a predicted higher *infective competence* in the COVID-19 group.

## Conclusions

Our findings suggest that it is imperative to elucidate all the implications of SARS-CoV-2 infection, especially its effect on the gut microbiome community and functions. Although other studies have involved the severe cases of COVID-19 [7, 48], none of these studies include both metagenomic and metatranscriptomic sequencing data. We found that the VFs and ARGs were indeed expressed in higher levels in the COVID-19 group compared to the controls. These key findings would have not been possible by only focussing on metagenomic data. Our collective findings, indicating the enriched abundance and expression of both VFs and ARGs, suggest that COVID-19 may yet have unknown effects that may come to light in the longer term including the shaping of the microbiome across the population. Moreover, we find that none of the commonly reported pathogens (*Salmonella*, *Shigella*, *Klebsiella* etc.) are enriched in the COVID-19 group in our study. In contrast, we find changes in *Prevotella* spp, AM10 and CAG145 (Firmicutes phylum), *Roseburia* spp and a *Turicibacter* spp. Therefore, it will be critically important to evaluate and further validate the effects of COVID-19 on the gut microbiome also in relation to infections by other viral and other pathogens. In particular, it remains unclear at this time, whether infections with other viruses, known to cause respiratory and gastrointestinal distress, e.g., Adenoviruses, respiratory syncytial virus (RSV), influenza viruses, norovirus, would lead to similar community and functional changes within the gut microbiome. Overall, it must be reiterated that pandemic preparedness coupled to the monitoring of VFs in tandem with antibiotic stewardship may be essential components for future strategies to mitigate the longer-term effects of COVID-19 and possibly other viral infections.

## Methods

### Cohort description and patient involvement

Between May and October 2020, stool samples were collected from 61 participants with COVID-19 confirmed by positive SARS-CoV-2 RT-qPCR (Supplementary Table S1) within the framework of the Predi-COVID study [49]. In order to be eligible to participate in the study, an individual must have been residing in Luxembourg and met all the following criteria: (1) signed informed consent form; (2) individuals  $\geq 18$  years old

with confirmed SARS-CoV-2 infection as determined by PCR, performed by one of the certified laboratories in Luxembourg; and (3) hospitalized or at home. In addition to the criteria specific to the Predi-COVID study, samples were excluded if antibiotic treatment was reported. From the individuals, relevant clinical data was collected using a modified version of the International Severe Acute Respiratory and Emerging Infection Consortium (ISARIC) case report form. The participants to be included in the study were classified using an adapted version of the National Institute of Health symptom severity scheme [50]. Briefly, the classification of COVID-19 severity was based on the classification of the National Institute of Health (NIH) in the USA. Disease was classified as moderate if an individual had  $\text{SpO}_2 \geq 94\%$ , and shortness of breath and/or evidence of lower respiratory disease. If  $\text{SpO}_2 < 94\%$ , a ratio of arterial partial pressure of oxygen to fraction of inspired oxygen ( $\text{PaO}_2/\text{FiO}_2$ )  $< 300$  mmHg, a respiratory rate  $> 30$  breaths/min, or lung infiltrates  $> 50\%$ , disease was classified as severe. Subsequently, only asymptomatic to moderate symptoms were reported.

Along with the samples from the COVID-19 confirmed participants, stool samples were collected from a group of 57 individuals who tested negative for SARS-CoV-2 by RT-qPCR, who were participants of the CON-VINCE study, a population-based cohort study which recruited a representative sample of the Luxembourg population, to serve as age-matched controls. Participation in the control group was excluded if matching any of the following criteria: (1) infection of SARS-CoV-2 prior to the study; (2) presence of fever and respiratory distress/cough not attributable to other known chronic disease; (3) usage of antibiotics up to three months prior to enrolment or first SARS-CoV-2 infection. The study design is presented in Fig. 1. Demographic characteristics of the study groups are summarized in Table 1 while additional metadata are included in Supplementary Table S1. Patients were not involved in setting the research questions or the outcome measures of this study.

### Sample collection and processing

Stool samples were collected at home by individuals in Fecal Collection Tubes (Zymo Research). Samples and data were collected at the Integrated BioBank of Luxembourg (IBBL). Around 1 g of stool was sampled, diluted in 9 ml DNA/RNA Shield according to the manufacturer's instructions. Prior to DNA/RNA extraction, stool samples were thawed on ice and aliquoted as follows: 250  $\mu\text{l}$  of sample was aliquoted for DNA extraction, to which 250  $\mu\text{l}$  of lysis solution (ZymoBIOMICS DNA Miniprep Kit; Zymo Research) was added and the sample was subsequently kept frozen at  $-80^\circ\text{C}$  until DNA

**Table 1** Demographic characteristics of the study groups

	COVID-19 (N=61)	Controls (N=57)	P value
Age	43.85 ± 11.92	42.12 ± 3.32	0.297
Sex			
Female	22 (36.07%)	22 (38.6%)	0.776
Male	39 (63.93%)	35 (61.4%)	
COVID-19 severity			
Asymptomatic	4 (6.56%)	N/A	N/A
Mild	45 (73.77%)		
Moderate	12 (19.67%)		
Hospitalization status			
Hospitalized	1 (1.64%)	N/A	N/A
Not hospitalized	60 (98.36%)		
COVID-19 symptoms			
Fever	35 (57.38%)	N/A	N/A
Runny nose	11 (18.03%)		
Sore throat	22 (36.06%)		
Smell and/or taste loss	32 (52.46%)		
Fatigue	40 (65.57%)		
Headache	40 (65.57%)		
Cough	33 (54.1%)		
Shortness of breath	12 (19.67)		
Diarrhea	15 (24.59%)		
Abdominal pain	1 (1.64%)		
Chest pain	9 (14.75%)		
Ear pain	5 (8.2%)		
Joint pain	6 (9.84)		
Muscle pain	30 (49.18%)		
Vomiting/nausea/vertigo	7 (11.47%)		

N Number of participants, N/A Not applicable

extraction was performed. Furthermore, another 700 µl was aliquoted for RNA extraction using ZR BashingBead Lysis Tubes (Zymo Research) and the RNeasy Mini Kit (QIAGEN).

#### DNA and RNA extractions

DNA was extracted using the ZymoBIOMICS DNA Miniprep Kit according to the manufacturer's instructions with the following modifications: samples were inactivated for 7 min at 70 °C prior to homogenization by milling for 3 cycles (5 min of cooling on ice between cycles) for 60 s at 6 m/s in a FastPrep-24 5 G (MP Biomedicals). Prior to DNA purification, a Proteinase K incubation step was performed: 5 µl of 20 mg/ml Proteinase K (New England Biolabs GmbH) was added to each sample and incubated for 30 min at 40 °C. The extraction was performed following the manufacturer's instructions and DNA was eluted in 50 µl DNase/RNase-Free Water (prewarmed to 60 °C). An RNase treatment was performed by adding

2.4 µl of 20 mg/ml Monarch RNase A (New England Biolabs GmbH) to each sample followed by incubation for 10 min at 56 °C. DNA was purified and concentrated using ZR-96 DNA Clean-Up Kit (Zymo Research) following the manufacturer's instructions and DNA was eluted in 50 µl DNase/RNase-Free water (prewarmed to 60 °C). DNA was quantified using Qubit dsDNA BR assay kit (Invitrogen) and purity determined using Nanodrop 2000C (Thermo Scientific). Samples were frozen at – 80 °C until further use.

Samples for RNA extraction were inactivated for 7 min at 70 °C and 600 µl of cold RLT Buffer (containing 10 µl/ml 2-mercaptoethanol) was added to the samples prior to homogenization by milling for 3 cycles (5 min of cooling on ice between cycles) for 60 s at 6 m/s in a FastPrep-24 5 G (MP Biomedicals). Samples were centrifuged for 3 min at full speed and the supernatant was mixed with 1 volume of 70% Ethanol. Lysates were loaded onto a RNeasy Mini Spin Column and centrifuged at 8000 × g for 1 min. This last step was repeated until all supernatants had passed through the filters. Columns were washed according to the manufacturer's instructions whereby 50 µl RNase-free water was added to the centre of the filter and incubated at room temperature for 1 min. RNA was eluted by centrifugation at 8000 × g for 1 min. RNA extracts were filled up to 87.5 µl with RNase-free water, 2.5 µl DNase I stock solution and 10 µl Buffer RDD (both RNase-Free DNase Set, QIAGEN) were added, mixed and incubated for 10 min at room temperature. RNA was purified and concentrated using RNA Clean & Concentrator-5 kit (Zymo Research) following the manufacturer's instructions. RNA was eluted in 15 µl DNase/RNase-Free water. One microliter of obtained RNA was heat-denatured for 2 min at 72 °C and quality-checked using Agilent RNA 6000 Nano kit (Agilent Technologies). RNA was quantified using Qubit RNA HS assay kit (Invitrogen). RNA extracts were frozen at – 80 °C for further use.

#### Metagenomic and metatranscriptomic sequencing

DNA and RNA were extracted from all collected stool samples and sequenced for metagenomic and metatranscriptomic analysis, respectively. One hundred nanograms of DNA was used for metagenomic library preparation using Swift 2S turbo Flexible DNA library kit (cat. no. 45096). The genomic DNA was enzymatically fragmented for 10 min and DNA libraries were prepared without PCR amplification. The average insert size of libraries was 600 bp. Prepared libraries were quantified using Qubit (DNA HS kit, ThermoFischer) and quality-checked with a DNA HS kit on a Bioanalyzer 2100 (Agilent). Sequencing was performed at the LCSB sequencing

platform (RRID: SCR\_021931) on a NextSeq2000 instrument using  $2 \times 150$  bp read lengths.

500 ng of RNA was rRNA depleted using the Illumina Ribo-Zero Plus rRNA Depletion kit (Illumina, 20,037,135). rRNA depleted samples were further processed using the TruSeq Stranded mRNA library preparation kit (Illumina, 20,020,594) which includes the fragmentation and priming steps. The fragmentation time was reduced to 3 min. Prepared libraries were quantified using Qubit (DNA HS kit, ThermoFischer) and quality checked with DNA HS kit on a Bioanalyzer 2100 (Agilent). Sequencing was performed at the LCSB sequencing platform (RRID: SCR\_021931) on a NextSeq500 instrument using  $2 \times 150$  bp read lengths. In total, this resulted in  $\sim 6$  Gbp per sample for the metagenomics and  $\sim 21$  Gbp per sample for the metatranscriptomics.

#### Data processing, including genome reconstruction

The Integrated Meta-omic Pipeline (IMP; v3-commitID #b6f9da0e for preprocessing and #c04edbe for downstream assemblies) [25] was used for the processing and iterative co-assembly of metagenomic and metatranscriptomic reads. The workflow includes pre-processing, assembly, genome reconstruction, and functional and taxonomic annotation based on public and custom databases in a reproducible manner. For the data preprocessing, raw metagenomic reads were first trimmed to the maximal read length of 150 bases using Cutadapt (v3.4) [51]. The preprocessed metagenomic and raw metatranscriptomic reads were further processed using IMP: reads were trimmed using Trimmomatic (v.39) [52], reads mapping to the human genome (hg38 genome) or PhiX genome (gi|9,626,372|ref|NC\_001422.1, Enterobacteria phage phiX174 sensu lato, complete genome) were removed using BWA (v. 0.7.9a) [53], and the metatranscriptomic reads were further filtered using SortMeRNA (v.4.2.0–238-g90cdf6c) [54]. In addition, alpha-diversity was calculated based on metagenomic reads using Nonpareil (v. 3.4.1) [55] as part of the IMP preprocessing step. Quality control was performed on the processed reads by running FastQC (v. 0.11.9) [56] and summarizing the reports using MultiQC (v. 1.10.1) [57]. In addition, Kraken2 (v. 2.1.2) [58] was used with a database containing only the human and PhiX genomes (<https://ndownloader.figshare.com/files/24658262>, from 11.09.2020, provided by Mike Lee) to confirm the successful removal of these contaminants from the processed sequencing data. The tool bbmap (v. 38.90) [59] was used on the preprocessed FASTQ files to extract reads mapping to SARS-CoV-2 reference genomes (same genomes as provided by fastv). Pairwise sample (dis)similarity was calculated using Mash (v. 2.3) [60].

*De novo* co-assembly of the processed metagenomic and metatranscriptomic reads was performed by running Megahit (v2.0) [61] included in IMP, followed by gene calling using an in-house modified Prokka version also allowing for incomplete ORFs [62]. Concurrently, MetaBAT2 [63] and MaxBin2 [64] together with an in-house binning methodology, binny [65], were used to reconstruct metagenome-assembled genomes (MAGs). Subsequently, we obtained a non-redundant set of MAGs using DAS Tool (v1.1.4) [66] with a score threshold of 0.7 for downstream analyses, and those with a minimum completion of 90% and less than 5% contamination as assessed by CheckM (v1.1.3) [67]. Taxonomy was assigned to the MAGs using gtdbtk (v1.7.0) [68]. Finally, MetaQUAST (v. 5.0.2) [69] was run on the created contig FASTA files to compute assembly statistics such as the number and maximal length of contigs, total assembly length, and the N50 and L50 values.

#### Virome analyses

The co-assemblies built using metagenomic and metatranscriptomic data were used for the subsequent identification of viruses and to determine their functional activity. Briefly, the co-assembly was first processed through VIBRANT [70] and CheckV [71]. The CheckV assessment was repeated and any viral contigs with less than 70% completion were removed from further analyses. Subsequently, the complete viral contigs and those passing the 70% completion filter were merged and their respective taxonomies were determined using the IMGVR3 database [72]. To detect other viruses and confirm the status of SARS-CoV-2 infection in the processed reads, we also used fastv (v. 0.8.1, data for SARS-CoV-2 and for other viruses was downloaded on September 11th, 2021) [73]. Taxonomic consensus between the IMGVR3 and the fastv databases were determined to obtain overlapping, robust classification, and subsequently were used for the downstream analyses, where differentially abundant genes were further assessed for differential relative gene expression.

#### Prediction of microbial composition, virulence factors, and antimicrobial resistance

Profiling of the microbial community was performed on the processed reads using MetaPhlAn3.1 (v3.1.0, database “mpa\_v31\_CHOCOPhAn\_201901”) [74]. Simultaneously, profiling of antibiotic resistance factors was done using RGI (v5.2.0, CARD data v3.1.4, prevalence, resistomes and variants data v3.0.9) [75]. To obtain additional in-depth details of ARGs, in addition to the detection of VFs and mobile genetic elements (MGEs), PathoFact (v1.0; modified branch allowing the input of ORFs, #6fa64961) was run [24]. PathoFact is a pipeline



for the prediction of ARGs and VFs, and their localization to MGEs, in metagenomic data. PathoFact was run on the contigs assembled by IMP together with their predicted protein sequences (ORFs) for each sample separately. PathoFact uses DeepARG [76] and RGI [75] for the prediction of ARGs, DeepVirFinder [77] and VirSorter [78] for the prediction of phages and PlasFlow [79] for the prediction of plasmids. Additionally, PathoFact uses its own developed tool, a combination of a HMM database (built on the VFDB [80]) and a random forest model, for the prediction of VFs. To run PathoFact, the input protein sequences were first processed to remove any trailing stop codon symbols (“\*”) and to remove any sequence having an internal stop codon symbol as this is required for the tool RGI for ARG detection. For analyses of the predictions, FeatureCounts (v1.6.4) (Liao et al. 2014) was used to extract the number of reads per functional category. Thereafter, the relative abundance of genes and general expression levels was calculated using the Rnum\_Gi method described by Hu et al. (Hu et al. 2013) which normalizes for both gene length and library size. Subsequently, metatranscriptomic expression levels were further normalized using the respective gene abundances from the metagenomic data (normalized gene expression = gene expression/ gene abundance).

### Statistical testing and data analysis

Statistical analyses of the taxonomic and functional data, as well as further visualizations, were performed using version 4.1.1 of the R statistical software package [81]. The R package MaAsLin2 [82] was used to determine associations between the cohort data and microbial features (e.g., functional and taxonomic profiles). Furthermore, MaAsLin2 identified significant differences were further validated by Wilcoxon rank-sum tests with adjustments using the ‘Benjamini-Hochberg’ method for multiple testing, specifically the ‘p.adjust’ function from the *stats* R package was used. To additionally, validate our findings with respect to the VF and ARGs, we used ANCOM-BC [83]. The *tidyverse*, *microbiomeViz*, *tidytree*, and *ggtree* packages were used to visualize the microbiome data, including using cladogram visualizations. The *tidyverse* package, including *ggplot2*, was used to generate all violin plots, box plots, and bubble plots. Finally, the *hmisc* and *corrplot* packages were used for all correlation plots.

### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40168-023-01472-7>.

**Additional file 1: Supplementary Figure S1.** Dissimilarity of COVID-19 and control groups’ microbiome profiles. Alpha- (Shannon) (a) and

beta-diversity (b) metrics of the gut microbiome compositions of the COVID-19 and control groups.

**Additional file 2: Supplementary Figure S2.** VF diversity between the COVID-19 and control groups. Alpha- (a) and beta-diversity (b) metrics of the VF abundances are depicted along with the VF relative expression alpha- (c) and beta-diversity (d) between the COVID-19 and control groups.

**Additional file 3: Supplementary Figure S3.** AMR diversity between the COVID-19 and control groups. Alpha- (a) and beta-diversity (b) metrics of the AMR abundances. AMR relative expression alpha- (c) and beta-diversity (d) between the COVID-19 and control groups.

**Additional file 4: Supplementary Figure S4.** Analysis of VF and ARG compositionality using ANCOM-BC. Log2 fold change (LFC) of (a) abundance and (b) normalised gene expression of VFs per taxonomic family between the COVID-19 and control groups. ARG abundance (c) and normalised gene expression (d) differences between the COVID-19 and control groups.

**Additional file 5: Supplementary Table 1.** Cohort metadata and description.

**Additional file 6: Supplementary Table 2.** Identified and significantly enriched viral genes in the COVID-19 group compared to control individuals.

### Acknowledgements

We acknowledge the active involvement of all the participants of the Predi-COVID and CON-VINCE cohorts. We are also thankful for the support of the cohort recruitment and processing teams. We also acknowledge Prof. Laetitia Huiart for her efforts in initiating the Predi-COVID cohort and Aurélie Sausy, Sophie Mériaux, and Jean-Yves Servais for their technical help. We would like to thank Susana Martinez Arbas for the initial support during data preprocessing. The computational analyses were performed at the HPC facilities at the University of Luxembourg (<https://hpc.uni.lu>).

### The CON-VINCE Consortium

Geeta Acharya, Luxembourg Institute of Health, Strassen, Luxembourg; Gloria Aguayo, Luxembourg Institute of Health, Strassen, Luxembourg; Wim Ammerlaan, Luxembourg Institute of Health, Strassen, Luxembourg; Ariane Assele-Kama, Luxembourg Institute of Health, Strassen, Luxembourg; Christelle Bahlawane, Luxembourg Institute of Health, Strassen, Luxembourg; Katy Beaumont, Luxembourg Institute of Health, Strassen, Luxembourg; Nadia Beaupain, Luxembourg Institute of Health, Strassen, Luxembourg; Lucrèce Beckers, Luxembourg Institute of Health, Strassen, Luxembourg; Camille Bellora, Luxembourg Institute of Health, Strassen, Luxembourg; Fay Betsou, Luxembourg Institute of Health, Strassen, Luxembourg; Sandie Boly, Luxembourg Institute of Health, Strassen, Luxembourg; Dirk Brenner, Luxembourg Institute of Health, Strassen, Luxembourg; Eleftheria Charalambous, Luxembourg Institute of Health, Strassen, Luxembourg; Emilie Charpentier, Luxembourg Institute of Health, Strassen, Luxembourg; Manuel Counson, Luxembourg Institute of Health, Strassen, Luxembourg; Brian De Witt, Luxembourg Institute of Health, Strassen, Luxembourg; Olivia Domingues, Luxembourg Institute of Health, Strassen, Luxembourg; Claire Dording, Luxembourg Institute of Health, Strassen, Luxembourg; Bianca Dragomir, Luxembourg Institute of Health, Strassen, Luxembourg; Tessa Fautsch, Luxembourg Institute of Health, Strassen, Luxembourg; Jean-Yves Ferrand, Luxembourg Institute of Health, Strassen, Luxembourg; Ana Festas Lopes, Luxembourg Institute of Health, Strassen, Luxembourg; Joëlle Véronique Fritz, Luxembourg Institute of Health, Strassen, Luxembourg; Manon Gantenbein, Luxembourg Institute of Health, Strassen, Luxembourg; Laura Georges, Luxembourg Institute of Health, Strassen, Luxembourg; Jérôme Graas, Luxembourg Institute of Health, Strassen, Luxembourg; Gael Hamot, Luxembourg Institute of Health, Strassen, Luxembourg; Anne-Marie Hanff, Luxembourg Institute of Health, Strassen, Luxembourg; Maxime Hansen, Luxembourg Institute of Health, Strassen, Luxembourg; Lisa Hefele, Luxembourg Institute of Health, Strassen, Luxembourg; Estelle Henry, Luxembourg Institute of Health, Strassen, Luxembourg; Margaux Henry, Luxembourg Institute of Health, Strassen, Luxembourg; Eve Herkenne, Luxembourg Institute of Health, Strassen, Luxembourg; Christiane Hilger, Luxembourg Institute of

Health, Strassen, Luxembourg; Judith Hübschen, Luxembourg Institute of Health, Strassen, Luxembourg; Laetitia Huiart, Luxembourg Institute of Health, Strassen, Luxembourg; Alexander Hundt, Luxembourg Institute of Health, Strassen, Luxembourg; Gilles Iserentant, Luxembourg Institute of Health, Strassen, Luxembourg; Stéphanie Kler, Luxembourg Institute of Health, Strassen, Luxembourg; Rejko Krüger, Luxembourg Institute of Health, Strassen, Luxembourg; Pauline Lambert, Luxembourg Institute of Health, Strassen, Luxembourg; Sabine Lehmann, Luxembourg Institute of Health, Strassen, Luxembourg; Morgane Lemaire, Luxembourg Institute of Health, Strassen, Luxembourg; Andrew Lumley, Luxembourg Institute of Health, Strassen, Luxembourg; Monica Marchese, Luxembourg Institute of Health, Strassen, Luxembourg; Sophie Mériaux, Luxembourg Institute of Health, Strassen, Luxembourg; Maura Minelli, Luxembourg Institute of Health, Strassen, Luxembourg; Alessandra Mousel, Luxembourg Institute of Health, Strassen, Luxembourg; Maeva Munsch, Luxembourg Institute of Health, Strassen, Luxembourg; Mareike Neumann, Luxembourg Institute of Health, Strassen, Luxembourg; Markus Ollert, Luxembourg Institute of Health, Strassen, Luxembourg; Marc Paul O'Sullivan, Luxembourg Institute of Health, Strassen, Luxembourg; Magali Perquin, Luxembourg Institute of Health, Strassen, Luxembourg; Achilleas Pexaras, Luxembourg Institute of Health, Strassen, Luxembourg; Jean-Marc Plesseria, Luxembourg Institute of Health, Strassen, Luxembourg; Lucie Remark, Luxembourg Institute of Health, Strassen, Luxembourg; Estelle Sandt, Luxembourg Institute of Health, Strassen, Luxembourg; Bruno Santos, Luxembourg Institute of Health, Strassen, Luxembourg; Aurélie Sausy, Luxembourg Institute of Health, Strassen, Luxembourg; Margaux Schmitt, Luxembourg Institute of Health, Strassen, Luxembourg; Sneeha Seal, Luxembourg Institute of Health, Strassen, Luxembourg; Jean-Yves Servais, Luxembourg Institute of Health, Strassen, Luxembourg; Florian Simon, Luxembourg Institute of Health, Strassen, Luxembourg; Chantal Snoeck, Luxembourg Institute of Health, Strassen, Luxembourg; Kate Sokolowska, Luxembourg Institute of Health, Strassen, Luxembourg; Hermann Thien, Luxembourg Institute of Health, Strassen, Luxembourg; Johanna Trouet, Luxembourg Institute of Health, Strassen, Luxembourg; Jonathan Turner, Luxembourg Institute of Health, Strassen, Luxembourg; Michel Vaillant, Luxembourg Institute of Health, Strassen, Luxembourg; Daniela Valoura Esteves, Luxembourg Institute of Health, Strassen, Luxembourg; Charlène Verschueren, Luxembourg Institute of Health, Strassen, Luxembourg; Tania Zamboni, Luxembourg Institute of Health, Strassen, Luxembourg; Pinar Alper, Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-Belval, Luxembourg; Piotr Gawron, Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-Belval, Luxembourg; Soumyabrata Ghosh, Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-Belval, Luxembourg; Enrico Glaab, Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-Belval, Luxembourg; Clarissa Gomes, Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-Belval, Luxembourg; Borja Gomez Ramos, Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-Belval, Luxembourg; Vyrion Gorgogietas, Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-Belval, Luxembourg; Valentin Groues, Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-Belval, Luxembourg; Wei Gu, Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-Belval, Luxembourg; Laurent Heirendt, Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-Belval, Luxembourg; Ahmed Hemedan, Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-Belval, Luxembourg; Sascha Herzinger, Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-Belval, Luxembourg; Anne Kaysen, Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-Belval, Luxembourg; Jacek Jaroslaw Lebioda, Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-Belval, Luxembourg; Tainà Marques, Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-Belval, Luxembourg; François Massart, Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-Belval, Luxembourg; Patrick May, Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-Belval, Luxembourg; Christiane Olesky, Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-Belval, Luxembourg; Venkata PSatagopam, Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-Belval, Luxembourg; Claire Pauly, Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-Belval, Luxembourg; Laure Pauly, Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-Belval,

Luxembourg; Lukas Pavelka, Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-Belval, Luxembourg; Guilherme Ramos Meyers, Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-Belval, Luxembourg; Armin Rauschenberger, Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-Belval, Luxembourg; Basile Rommes, Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-Belval, Luxembourg; Kirsten Rump, Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-Belval, Luxembourg; Reinhard Schneider, Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-Belval, Luxembourg; Valerie Schröder, Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-Belval, Luxembourg; Amna Skrozic, Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-Belval, Luxembourg; Lara Stute, Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-Belval, Luxembourg; Noua Toukourou, Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-Belval, Luxembourg; Christophe Trefois, Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-Belval, Luxembourg; Carlos Vega Moreno, Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-Belval, Luxembourg; Maharshi Vyas, Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-Belval, Luxembourg; Xinhui Wang, Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-Belval, Luxembourg; Anja Leist, University of Luxembourg, Esch-sur-Alzette, Luxembourg; Annika Lutz, University of Luxembourg, Esch-sur-Alzette, Luxembourg; Claus Vögele, University of Luxembourg, Esch-sur-Alzette, Luxembourg; Linda Hansen, Centre Hospitalier de Luxembourg, Luxembourg, Luxembourg; João Manuel Loureiro, Centre Hospitalier de Luxembourg, Luxembourg, Luxembourg; Beatrice Nicolai, Centre Hospitalier de Luxembourg, Luxembourg, Luxembourg; Alexandra Schweicher, Centre Hospitalier de Luxembourg, Luxembourg, Luxembourg; Femke Wauters, Centre Hospitalier de Luxembourg, Luxembourg, Luxembourg; Tamir Abdelrahman, Laboratoire National de Santé, Dudelange, Luxembourg; Estelle Coibion, Laboratoire National de Santé, Dudelange, Luxembourg; Guillaume Fournier, Laboratoire National de Santé, Dudelange, Luxembourg; Marie Leick, Laboratoire National de Santé, Dudelange, Luxembourg; Friedrich Mühlshlegel, Laboratoire National de Santé, Dudelange, Luxembourg; Marie France Pirard, Laboratoire National de Santé, Dudelange, Luxembourg; Nguyen Trung, Laboratoire National de Santé, Dudelange, Luxembourg; Philipp Jägi, Laboratoire Réunis, Junglinster, Luxembourg; Henry-Michel Cauchie, Luxembourg Institute of Science and Technology, Luxembourg, Luxembourg; Delphine Collart, Luxembourg Institute of Science and Technology, Luxembourg, Luxembourg; Leslie Ogorzaly, Luxembourg Institute of Science and Technology, Luxembourg, Luxembourg; Christian Penny, Luxembourg Institute of Science and Technology, Luxembourg, Luxembourg; Cécile Walczak, Luxembourg Institute of Science and Technology, Luxembourg, Luxembourg.

#### Authors' contributions

The authors are solely responsible for all aspects of this article including the research, the interpretation, and the writing thereof. PW, LD, CMG, SBB, and CCL conceived the study, while LD, VG, and SBB performed data analysis, data interpretation, and in conjunction with MD wrote the manuscript. CS, FH, and PM contributed to critical experiments and analyses. RH, JH, and PCL were instrumental in sequencing the samples. GF, RK, MO, JVF, TM, ES, MPOS, SG, and VS were responsible for the data and sample collection of the Predi-COVID or CON-VINCE cohorts. All authors were involved in writing and approval of the manuscript.

#### Funding

The Predi-COVID study is supported by the Luxembourg National Research Fund (FNR: 14716273), the Fondation André Losch, and the European Regional Development Fund (FEDEP, no. 2018–04-026–21). F.Q.H. was partially supported by FNR AFR-RIKEN bilateral program (TregBAR, 11228353, F.Q.H. and M.O.) and PRIDE program (11012546/NEXTIMMUNE and 10907093/CRITICS). The CON-VINCE study is supported by the Fonds National de la Recherche (FNR: 14716281/CON-VINCE/Kruger) and the André Losch Foundation. PW acknowledges funding from the European Research Council under the European Union's Horizon 2020 research and innovation program (no. 863664). LD, N, PM, and PW are supported by the Luxembourg National Research Fund (FNR; PRIDE17/11823097) awarded to PW. We would also like to acknowledge the following funding instruments: FNR: COVID-19/2020–1/14719552/FunBiome. The work was further supported by the Luxembourg Government

through the CoVaLux program. This research has also been supported by the Swiss National Science Foundation (CRSII5\_180241) supporting SBB.

#### Availability of data and materials

All code used for metagenomic and metatranscriptomic data analysis can be found in the following repository: <https://gitlab.lcsb.uni.lu/ESB/co-infectomics>. The repository at [https://gitlab.lcsb.uni.lu/laura.denies/co-infectomics\\_r\\_analysis](https://gitlab.lcsb.uni.lu/laura.denies/co-infectomics_r_analysis) provides the code for the subsequent analyses in R. Individual workflows (sequencing data processing, QC, profiling, assembly, and analysis) were implemented using Snakemake [84]. Preprocessed and filtered reads, i.e., those filtered against the human genome (*hg38*), were submitted to the Sequence Read Archive hosted by NCBI and can be found under the accession PRJNA890008. Associated metadata is included in the code repository listed above. Additional data including assemblies, taxonomic, virome, functional pathway profiles along with MultiQC reports, can be found on Zenodo under the following link: <https://doi.org/10.5281/zenodo.7192682>.

#### Declarations

##### Ethics approval and consent to participate

The study was performed in accordance with the Good Clinical Practice and the Declaration of Helsinki. All participants in this study provided their written informed consent. The study was approved by the National Research Ethics Committee of Luxembourg (study number 202006/03). All biological samples and data were collected in the frame of Predi-COVID [49] and CON-VINCE studies. The studies were approved by the National Research Ethics Committee of Luxembourg (Predi-COVID study number 202003/07; CON-VINCE study number 202004/01).

##### Consent for publication

Not applicable.

##### Competing interests

The authors declare no competing interests.

Received: 31 October 2022 Accepted: 24 January 2023

Published online: 09 March 2023

#### References

- Merad M, Blish CA, Sallusto F, Iwasaki A. The immunology and immunopathology of COVID-19. *Science*. 2022;375:1122–7.
- WHO Coronavirus (COVID-19) dashboard. <https://covid19.who.int/>. Accessed 10 Sept 2022.
- Fischer A, et al. Long COVID symptomatology after 12 months and its impact on quality of life according to initial coronavirus disease 2019 disease severity. *Open Forum Infect Dis*. 2022;9:ofac397.
- Lamers MM, et al. SARS-CoV-2 productively infects human gut enterocytes. *Science*. 2020;369:50–4.
- Wu Y, et al. Prolonged presence of SARS-CoV-2 viral RNA in faecal samples. *Lancet Gastroenterol Hepatol*. 2020;5:434–5.
- Chen Y, et al. Six-month follow-up of gut microbiota richness in patients with COVID-19. *Gut*. 2022;71:222–5.
- Yeoh YK, et al. Gut microbiota composition reflects disease severity and dysfunctional immune responses in patients with COVID-19. *Gut*. 2021;70:698–706.
- Sorbara MT, Pamer EG. Interbacterial mechanisms of colonization resistance and the strategies pathogens use to overcome them. *Mucosal Immunol*. 2019;12:1–9.
- Devi P, et al. Co-infections as Modulators of Disease Outcome: Minor Players or Major Players? *Front Microbiol*. 2021;12:664386.
- Kim D, Quinn J, Pinsky B, Shah NH, Brown I. Rates of co-infection between SARS-CoV-2 and other respiratory pathogens. *JAMA*. 2020;323:2085–6.
- Lansbury L, Lim B, Baskaran V, Lim WS. Co-infections in people with COVID-19: a systematic review and meta-analysis. *J Infect*. 2020;81:266–75.
- Garcia-Vidal C, et al. Incidence of co-infections and superinfections in hospitalized patients with COVID-19: a retrospective cohort study. *Clin Microbiol Infect*. 2021;27:83–8.
- Rutsaert L, et al. COVID-19-associated invasive pulmonary aspergillosis. *Ann Intensive Care*. 2020;10:71.
- Silva DL, et al. Fungal and bacterial coinfections increase mortality of severely ill COVID-19 patients. *J Hosp Infect*. 2021;113:145–54.
- Santoso P, et al. MDR pathogens organisms as risk factor of mortality in secondary pulmonary bacterial infections among COVID-19 patients: observational studies in two referral hospitals in West Java. *Indonesia Int J Gen Med*. 2022;15:4741–51.
- D'Costa VM, et al. Antibiotic resistance is ancient. *Nature*. 2011;477:457–61.
- Pelfrene E, Botgros R, Cavaleri M. Antimicrobial multidrug resistance in the era of COVID-19: a forgotten plight? *Antimicrob Resist Infect Control*. 2021;10:21.
- Beceiro A, Tomás M, Bou G. Antimicrobial resistance and virulence: a successful or deleterious association in the bacterial world? *Clin Microbiol Rev*. 2013;26:185–230.
- Martínez JL, Baquero F. Interactions among strategies associated with bacterial infection: pathogenicity, epidemicity, and antibiotic resistance. *Clin Microbiol Rev*. 2002;15:647–79.
- Burrus V, Waldor MK. Shaping bacterial genomes with integrative and conjugative elements. *Res Microbiol*. 2004;155:376–86.
- Rybak B, et al. Antibiotic resistance, virulence, and phylogenetic analysis of *Escherichia coli* strains isolated from free-living birds in human habitats. *PLoS ONE*. 2022;17:e0262236.
- Masri L, et al. Host-pathogen coevolution: the selective advantage of *Bacillus thuringiensis* virulence and its cry toxin genes. *PLoS Biol*. 2015;13:e1002169.
- Wu Y, et al. Resident bacteria contribute to opportunistic infections of the respiratory tract. *PLoS Pathog*. 2021;17:e1009436.
- de Nies L, et al. PathoFact: a pipeline for the prediction of virulence factors and antimicrobial resistance genes in metagenomic data. *Microbiome*. 2021;9:49.
- Narayanasamy S, et al. IMP: a pipeline for reproducible reference-independent integrated metagenomic and metatranscriptomic analyses. *Genome Biol*. 2016;17:260.
- Sanchez-Ramirez DC, Normand K, Zhaoyun Y, Torres-Castro R. Long-term impact of COVID-19: a systematic review of the literature and meta-analysis. *Biomedicines*. 2021;9(8):900.
- Hazan S, et al. Lost microbes of COVID-19: Bifidobacterium, Faecalibacterium depletion and decreased microbiome diversity associated with SARS-CoV-2 infection severity. *BMJ Open Gastroenterol*. 2022;9(1):e000871.
- Lymberopoulos E, Gentili GI, Budhdeo S, Sharma N. COVID-19 severity is associated with population-level gut microbiome variations. *Front Cell Infect Microbiol*. 2022;12:963338.
- Yamamoto S, et al. The human microbiome and COVID-19: A systematic review. *PLoS ONE*. 2021;16:e0253293.
- Zuo T, et al. Alterations in gut microbiota of patients with COVID-19 during time of hospitalization. *Gastroenterology*. 2020;159:944–955.e8.
- Anderson EJ, Weber SG. Rotavirus infection in adults. *Lancet Infect Dis*. 2004;4:91–9.
- Wang L-P, et al. The changing pattern of enteric pathogen infections in China during the COVID-19 pandemic: a nation-wide observational study. *Lancet Reg Health West Pac*. 2021;16:100268.
- Wong SH, Lui RN, Sung JJ. Covid-19 and the digestive system. *J Gastroenterol Hepatol*. 2020;35:744–8.
- D'Amico F, Baumgart DC, Danese S, Peyrin-Biroulet L. Diarrhea during COVID-19 infection: pathogenesis, epidemiology, prevention, and management. *Clin Gastroenterol Hepatol*. 2020;18:1663–72.
- Teklemariam AD, Hashem AM, Saber SH, Almuhayawi MS, Haque S, Abujamel TS, et al. Bacterial coinfections and antimicrobial resistance associated with the Coronavirus Disease 2019 infection. *Biotechnol Genet Eng Rev*. 2022;19:1–22. <https://doi.org/10.1080/02648725.2022.2122297>. Epub ahead of print. PMID: 36123822.
- Islam MR, Rahman MM, Ahasan MT, Sarkar N, Akash S, Islam M, et al. The impact of mucormycosis (black fungus) on SARS-CoV-2-infected patients: at a glance. *Environ Sci Pollut Res Int*. 2022;29(46):69341–66. <https://doi.org/10.1007/s11356-022-22204-8>.

37. Manor O, et al. Health and disease markers correlate with gut microbiome composition across thousands of people. *Nat Commun.* 2020;11:5206.
38. Milosavljevic MN, et al. Antimicrobial treatment of *Erysipelatoclostridium ramosum* invasive infections: a systematic review. *Rev Inst Med Trop Sao Paulo.* 2021;63:e30.
39. Candela M, et al. Inflammation and colorectal cancer, when microbiota-host mutualism breaks. *World J Gastroenterol.* 2014;20:908–22.
40. Chandra H, Sharma KK, Tuovinen OH, Sun X, Shukla P. Pathobionts: mechanisms of survival, expansion, and interaction with host with a focus on *Clostridioides difficile*. *Gut Microbes.* 2021;13:1979882.
41. Chedid M, et al. Antibiotics in treatment of COVID-19 complications: a review of frequency, indications, and efficacy. *J Infect Public Health.* 2021;14:570–6.
42. WHO Regional Office for Europe/European Centre for Disease Prevention and Control. Antimicrobial resistance surveillance in Europe 2022 – 2020 data. Copenhagen: WHO Regional Office for Europe; 2022. <https://www.ecdc.europa.eu/en/publications-data/antimicrobial-resistance-surveillance-europe-2022-2020-data>. Accessed 10 Sept 2022.
43. Kariyawasam RM, et al. Antimicrobial resistance (AMR) in COVID-19 patients: a systematic review and meta-analysis (November 2019–June 2021). *Antimicrob Resist Infect Control.* 2022;11:45.
44. Kang Y, et al. Alterations of fecal antibiotic resistome in COVID-19 patients after empirical antibiotic exposure. *Int J Hyg Environ Health.* 2022;240:113882.
45. Peng Y, et al. Gut microbiome and resistome changes during the first wave of the COVID-19 pandemic in comparison with pre-pandemic travel-related changes. *J Travel Med.* 2021;28(7):067.
46. Cao J, et al. Integrated gut virome and bacteriome dynamics in COVID-19 patients. *Gut Microbes.* 2021;13:1–21.
47. Bassetti M, Righi E, Carnelutti A, Graziano E, Russo A. Multidrug-resistant *Klebsiella pneumoniae*: challenges for treatment, prevention and infection control. *Expert Rev Anti Infect Ther.* 2018;16:749–61.
48. Ke S, Weiss ST, Liu Y-Y. Dissecting the role of the human microbiome in COVID-19 via metagenome-assembled genomes. *Nat Commun.* 2022;13:5235.
49. Fagherazzi G, et al. Protocol for a prospective, longitudinal cohort of people with COVID-19 and their household members to study factors associated with disease severity: the Predi-COVID study. *BMJ Open.* 2020;10:e041834.
50. "Clinical Spectrum." n.d. COVID-19 Treatment Guidelines. <https://www.covid19treatmentguidelines.nih.gov/overview/clinical-spectrum/>. Accessed 7 Sept 2022.
51. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* 2011;17:10–2.
52. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 2014;30:2114–20.
53. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2009;25:1754–60.
54. Kopylova E, Noé L, Touzet H. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics.* 2012;28:3211–7.
55. Rodriguez-R LM, Konstantinidis KT. Nonpareil: a redundancy-based approach to assess the level of coverage in metagenomic datasets. *Bioinformatics.* 2014;30:629–35.
56. Andrews S. FastQC: A Quality Control Tool for High Throughput Sequence Data [Online]. 2010. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>, <https://qubeshub.org/resources/fastqc>.
57. Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics.* 2016;32:3047–8.
58. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. *Genome Biol.* 2019;20:257.
59. Bushnell B. BBMap: A fast, accurate, splice-aware aligner. 2014. <https://www.osti.gov/biblio/1241166>.
60. Ondov BD, et al. Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biol.* 2016;17:132.
61. 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;31:1674–6.
62. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics.* 2014;30:2068–9.
63. Kang DD, et al. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ.* 2019;7:e7359.
64. Wu Y-W, Simmons BA, Singer SW. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics.* 2016;32:605–7.
65. Heintz-Buschart A, et al. Integrated multi-omics of the human gut microbiome in a case study of familial type 1 diabetes. *Nat Microbiol.* 2016;2:16180.
66. Sieber CMK, et al. Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. *Nat Microbiol.* 2018;3:836–43.
67. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* 2015;25:1043–55.
68. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics.* 2019;36:1925–7.
69. Mikhchenko A, Saveliev V, Gurevich A. MetaQUAST: evaluation of metagenome assemblies. *Bioinformatics.* 2016;32:1088–90.
70. Kieft K, Zhou Z, Anantharaman K. VIBRANT: automated recovery, annotation and curation of microbial viruses, and evaluation of viral community function from genomic sequences. *Microbiome.* 2020;8:90.
71. Johansen J, et al. Genome binning of viral entities from bulk metagenomics data. *Nat Commun.* 2022;13:965.
72. Roux S, et al. IMG/VR v3: an integrated ecological and evolutionary framework for interrogating genomes of uncultivated viruses. *Nucleic Acids Res.* 2021;49:D764–75.
73. Chen S, He C, Li Y, Li Z, Melançon CE. A computational toolset for rapid identification of SARS-CoV-2, other viruses and microorganisms from sequencing data. *Brief Bioinform.* 2021;22:924–35.
74. Beghini F, et al. Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. *Elife.* 2021;10:e65088.
75. Alcock BP, et al. CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Res.* 2020;48:D517–25.
76. Arango-Argoty G, et al. DeepARG: a deep learning approach for predicting antibiotic resistance genes from metagenomic data. *Microbiome.* 2018;6:23.
77. Ren J, et al. Identifying viruses from metagenomic data by deep learning. *arXiv [q-bio.GN]*. 2018.
78. Roux S, Enault F, Hurwitz BL, Sullivan MB. VirSorter: mining viral signal from microbial genomic data. *PeerJ.* 2015;3:e985.
79. Krawczyk PS, Lipinski L, Dziembowski A. PlasFlow: predicting plasmid sequences in metagenomic data using genome signatures. *Nucleic Acids Res.* 2018;46:e35.
80. Chen L, et al. VFDB: a reference database for bacterial virulence factors. *Nucleic Acids Res.* 2005;33:D325–8.
81. R Core Team (2020). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>, <https://www.organizingcreativity.com/2020/08/citing-r-and-rstudio/>.
82. Mallick H, et al. Multivariable association discovery in population-scale meta-omics studies. *PLoS Comput Biol.* 2021;17:e1009442.
83. Lin H, Peddada SD. Analysis of compositions of microbiomes with bias correction. *Nat Commun.* 2020;11:3514.
84. Köster J, Rahmann S. Snakemake—a scalable bioinformatics workflow engine. *Bioinformatics.* 2018;34:3600.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.