


# Distinct gut microbiota and health outcomes in asymptomatic infection, viral nucleic acid test re-positive, and convalescent COVID-19 cases

Ruqin Lin<sup>1,6,#</sup>, Mingzhong Xiao<sup>2,7,#</sup>, Shanshan Cao<sup>3,9,#</sup>, Yu Sun<sup>1,6,#</sup>, Linhua Zhao<sup>5</sup>, Xiaoxiao Mao<sup>1,6</sup>, Peng Chen<sup>2,7</sup>, Xiaolin Tong<sup>5</sup>, Zheyuan Ou<sup>1,6</sup>, Hui Zhu<sup>2,7</sup>, Dong Men<sup>3,9,\*</sup>, Xiaodong Li<sup>2,7,\*</sup>, Yiqun Deng<sup>1,6,\*</sup>, Xian-En Zhang<sup>4,8,\*</sup>, and Jikai Wen<sup>1,6,\*</sup> 

Edited by Liping Zhao, Rutgers University, USA; Received November 29, 2021; Accepted May 4, 2022; Published online June 15, 2022

## Abstract

Gut microbiota composition is suggested to associate with coronavirus disease 2019 (COVID-19) severity, but the impact of gut microbiota on health outcomes is largely unclear. We recruited 81 individuals from Wuhan, China, including 13 asymptomatic infection cases (Group A), 24 COVID-19 convalescents with adverse outcomes (Group C), 31 severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) re-positive cases (Group D), and 13 non-COVID-19 healthy controls (Group H). The microbial features of Groups A and D were similar and exhibited higher gut microbial diversity and more abundant short-chain fatty acid (SCFA)-producing species than Group C. Group C was enriched with opportunistic pathogens and virulence factors related to adhesion and toxin production. The abundance of SCFA-producing species was negatively correlated, while *Escherichia coli* was positively correlated with adverse outcomes. All three groups (A, C, and D) were enriched with the mucus-degrading species *Akkermansia muciniphila*, but decreased with *Bacteroides*-encoded carbohydrate-active enzymes. The pathways of vitamin B6 metabolic and folate biosynthesis were decreased, while selenocompound metabolism was increased in the three groups. Specifically, the secondary bile acid (BA) metabolic pathway was enriched in Group A. Antibiotic resistance genes were common among the three groups. Conclusively, the gut microbiota was related to the health outcomes of COVID-19. Dietary supplementations (SCFAs, BA, selenium, folate, vitamin B6) may be beneficial to COVID-19 patients.

**Keywords:** COVID-19; gut microbiota; health outcomes; re-positive; SARS-CoV-2

## Impact statement

This is the first study that identifies the compositional and functional differences in the gut microbiome among the asymptomatic, the severe acute respiratory syndrome coronavirus 2 re-positive and the convalescent coronavirus disease 2019 (COVID-19) cases. These microbiome profiles may serve as a reference for future studies to highlight the importance of maintaining a healthy gut amongst COVID-19 patients. Gut microbiota was associated with the outcomes of COVID-19 patients, and thus maintaining a healthy gut is important in the recovery of COVID-19. The widespread resistance genes in the gut microbiota of COVID-19-related patients indicate that antibiotics should be used with caution in disease treatment. Probiotics could be used as an alternative. Dietary supplements (short-chain fatty acid, bile acid, selenium, folate, vitamin B6) may alleviate COVID-19 symptoms and promote prognosis.

<sup>1</sup>Guangdong Provincial Key Laboratory of Protein Function and Regulation in Agricultural Organisms, College of Life Sciences, South China Agricultural University, Guangzhou, China. <sup>2</sup>Hepatic Disease Institute, Hubei Key Laboratory of Theoretical and Applied Research of Liver and Kidney in Traditional Chinese Medicine, Hubei Provincial Hospital of Traditional Chinese Medicine, Wuhan, China. <sup>3</sup>State Key Laboratory of Virology, Wuhan Institute of Virology, Center for Biosafety Mega-Science, Chinese Academy of Sciences, Wuhan, China. <sup>4</sup>National Laboratory of Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, Beijing, China. <sup>5</sup>Department of Endocrinology, Guang'anmen Hospital, China Academy of Chinese Medical Sciences, Beijing, China. <sup>6</sup>Guangdong Laboratory for Lingnan Modern Agriculture, South China Agricultural University, Guangzhou, China. <sup>7</sup>Hubei Province Academy of Traditional Chinese Medicine, Wuhan, China. <sup>8</sup>Faculty of Synthetic Biology, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China. <sup>9</sup>University of Chinese Academy of Sciences, Beijing, China.

\* **Correspondence:** Dong Men, [d.men@wh.iov.cn](mailto:d.men@wh.iov.cn); Xiaodong Li, [lixiaodong555@126.com](mailto:lixiaodong555@126.com); Yiqun Deng, [yqdeng@scau.edu.cn](mailto:yqdeng@scau.edu.cn); Xian-En Zhang, [zhangxe@ibp.ac.cn](mailto:zhangxe@ibp.ac.cn); Jikai Wen, [jkw@scau.edu.cn](mailto:jkw@scau.edu.cn)

#Ruqin Lin, Mingzhong Xiao, Shanshan Cao, and Yu Sun contributed equally to this study.

DOI: [10.1002/mlf2.12022](https://doi.org/10.1002/mlf2.12022)

This is an open access article under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

## INTRODUCTION

The coronavirus disease 2019 (COVID-19) caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has spread rapidly worldwide. The symptoms of COVID-19 are fever, cough, and gastrointestinal symptoms, including nausea, vomiting, and diarrhea<sup>1</sup>. A meta-analysis of 4243 COVID-19 patients in 60 studies shows that the prevalence of gastrointestinal symptoms is 17.6%, and the positive rate of virus RNA in stool samples is 48.1%<sup>2</sup>. Live viruses are also isolated from the fecal specimens of COVID-19 patients<sup>3,4</sup>. SARS-CoV-2 recognizes the receptor angiotensin-converting enzyme 2 (ACE2) for entry and the serine protease TMPRSS2 for S-protein priming<sup>5</sup>. The ACE2 protein is abundantly expressed in the gastrointestinal epithelium<sup>6</sup>, and several colonic cell types are also permissive to SARS-CoV-2 infection<sup>7</sup>. Accordingly, acute SARS-CoV-2 infection induces gastrointestinal inflammation and systemic inflammatory response<sup>8</sup>. Accumulating evidence unveils that SARS-CoV-2 infection affects the compositions and diversities of fecal microbiota<sup>9–12</sup>, and that the gut microbiota richness is associated with the recovery process of COVID-19<sup>13</sup>, suggesting that gut microbiota may affect the health outcomes of COVID-19 patients. However, the effects of microbial functional compositions, such as metabolic pathways, virulence, and resistance genes, on the health outcomes of COVID-19 patients, are still to be characterized.

SARS-CoV-2 infection shows complex clinical symptoms and adverse outcomes, ranging from asymptomatic to severe and even fatal. Some discharged COVID-19 patients have persistent symptoms beyond 3 or 4 weeks from the onset of acute symptoms and are defined as “postacute COVID-19 syndrome”<sup>14</sup>. The proportion of COVID-19 convalescent patients with adverse outcomes is 68% at 6 months and 49% at 12 months<sup>15</sup>. COVID-19 survivors who are severely ill and have impaired pulmonary diffusion capacities during hospitalization are the main targeted population for the intervention of long-term recovery<sup>16</sup>. A recent study found that the gut microbiome might affect the susceptibility to postacute COVID-19 syndrome<sup>17</sup>. Also, SARS-CoV-2 nucleic acid test re-positive is a common clinical phenomenon. An epidemiology study reported that 14% of discharged cases reported re-positive, but none of them was caused by active reinfection<sup>18</sup>. Overall, the factors that affect different clinical outcomes of SARS-CoV-2 infection still need investigation.

In this study, we characterized the compositional and functional differences in gut microbiome among asymptomatic, SARS-CoV-2 re-positive and convalescent COVID-19 cases. These microbiome profiles revealed that the characteristic gut bacteria and microbial metabolic pathways were related to clinical symptoms and outcomes of COVID-19 patients. Practical applications such as dietary supplements might alleviate the adverse outcomes amongst COVID-19 patients.

## RESULTS

### Characteristics of the participants

We enrolled 81 participants, including 13 asymptomatic infection cases (Group A), 24 convalescent cases with

adverse outcomes (Group C), 31 discharged patients with SARS-CoV-2 re-positive (Group D), and 13 non-COVID-19 healthy controls (Group H) in this study (Figure S1). The clinical characteristics of participants are shown in Table 1 and the schematic diagram of sample collection is shown in Figure 1. Sex and age were matched among groups. The median ages of Groups A, C, D, and H were 53, 58, 52, and 50 years, and 53.8%, 58.3%, and 45.2% of Group A, C, and D cases had comorbidities, respectively. Hypertension was the most common comorbidity in COVID-19-related groups (A, C, and D) and was also matched with Group H. Group A showed no clinical symptoms during the infection period and no sequelae at follow-up. The count of lymphocytes (LYM) was normal, and only one case (1/13) showed a slightly higher white blood cell (WBC) count in blood at admission (Figure S2A,B). Group C cases had symptoms during hospitalization, including fever (79.2%), respiratory symptoms (66.7%), and gastrointestinal symptoms (50%). Of the cases, 33.3%, 50%, and 16.7% were classified as mild, moderate, and severe COVID-19 patients at admission, respectively. Group C participants had persistent chest computed tomography (CT) abnormalities (100%) or/and symptoms (79.2%) at the time of fecal sample collection (median discharged time was 92.5 days). Symptoms included respiratory symptoms (41.7%), fatigue (45.8%), muscle soreness (16.7%), sleep disturbances (8.3%), idrosis (8.3%), ageusia (4.2%), and hand paralysis (4.2%). Of the participants, 22.7% showed low levels of LYM or WBC in the blood (Figure S2A,B) and 37.5% had persistent symptoms at the 6-month follow-up after being discharged, and the symptoms included fatigue (29.2%), sleep disturbances (8.3%), and ageusia (4.2%). Group D patients that were diagnosed as COVID-19 cases with moderate symptoms during hospitalization had respiratory symptoms (100%), fever (48.4%) and gastrointestinal symptoms (3.2%) at first admission. These patients were approved to be discharged from the hospital with no obvious clinical symptoms and were later tested negative for SARS-CoV-2. However, Group D patients were subsequently re-positive for SARS-CoV-2, and 29% of them had mild clinical symptoms, including breathlessness (19.4%), fatigue (6.5%), and cough (3.2%). Of the patients, 23.3% showed abnormal levels of LYM or WBC in the blood (Figure S2A,B). All Group D patients were positive in the SARS-CoV-2 IgG antibody test, and 16.1% of patients were positive in the SARS-CoV-2 IgM antibody test in serum at re-admission. The median SARS-CoV-2 re-positive duration was 18 days. None of them showed clinical symptoms at follow-up after SARS-CoV-2 was negative.

### Distinct fecal microbial diversity and composition in COVID-19-related groups

A total of 2483 gut microbial species were characterized, including 2434 bacteria, 2 archaea, 14 eukaryotes (13 fungi

**Table 1.** Clinical information of the study subjects.

| Variable                               | Group A (asymptomatic infection) | Group C (postacute COVID-19 syndrome)                        | Group D (SARS-CoV-2 test re-positive) | Group H (non-COVID-19 healthy) |
|--|----------------------------------|--|---------------------------------------|--------------------------------|
| Number of cases                        | 13                               | 24   | 31                                    | 13                             |
| Male (%)                               | 6 (46.2%) <sup>a</sup>           | 10 (41.7%) <sup>a</sup>                                      | 16 (51.6%) <sup>a</sup>               | 5 (38.5%) <sup>a</sup>         |
| Age (years)                            | 53 ± 11.5 <sup>a</sup>           | 58 ± 12.5 <sup>a</sup>                                       | 52 ± 24 <sup>a</sup>                  | 50 ± 22.5 <sup>a</sup>         |
| Comorbidities (%)                      | 7 (53.8%) <sup>a</sup>           | 14 (58.3%) <sup>a</sup>                                      | 14 (45.2%) <sup>a</sup>               | 2 (15.4%) <sup>a</sup>         |
| Hypertension                           | 4 (30.8%) <sup>a</sup>           | 11 (45.8%) <sup>a</sup>                                      | 9 (29.0%) <sup>a</sup>                | 2 (15.4%) <sup>a</sup>         |
| Coronary heart disease                 | 3                                | 0  | 3                                     | 0                              |
| Diabetes                               | 1                                | 4  | 4                                     | 0                              |
| Fatty liver disease                    | 0                                | 2  | 1                                     | 0                              |
| Hyperlipidemia                         | 0                                | 6  | 0                                     | 0                              |
| Hyperuricemia                          | 0                                | 3  | 0                                     | 0                              |
| Chronic kidney disease                 | 0                                | 1  | 1                                     | 0                              |
| Chronic hepatitis B carrier            | 2                                | 0  | 0                                     | 0                              |
| Others                                 | 1                                | 1  | 4                                     | 0                              |
| Diseases categorized at admission      | Asymptomatic infection           | Mild: 8 (33.3%);<br>moderate: 12 (50%);<br>severe: 4 (16.7%) | Moderate: 31(100%)                    | -                              |
| Infection period (days)                | 16 ± 2 <sup>a</sup>              | 25 ± 14.75 <sup>b</sup>                                      | 27 ± 22 <sup>b</sup>                  | -                              |
| Symptoms during first hospitalization  |                                  |  |                                       |                                |
| Fever                                  | 0                                | 19 (79.2%)   | 15 (48.4%)                            | -                              |
| Gastrointestinal symptoms              | 0                                | 12 (50%)   | 1 (3.2%)                              | -                              |
| Respiratory symptoms                   | 0                                | 16 (66.7%)   | 31 (100%)                             | -                              |
| Symptoms/chest CT at sample collection |                                  |  |                                       |                                |
| Chest CT abnormalities                 | 0                                | 22/22 (100%)   | 8 (25.8%)                             | -                              |
| Respiratory symptoms                   | 0                                | 10 (41.7%)   | 9 (22.6%)                             | -                              |
| Fatigue                                | 0                                | 11 (45.8%)   | 1 (3.2%)                              | -                              |
| Muscle soreness                        | 0                                | 4 (16.7%)  | 0                                     | -                              |
| Sleep disturbances                     | 0                                | 2 (8.3%)   | 0                                     | -                              |
| Idrosis                                | 0                                | 2 (8.3%)   | 0                                     | -                              |
| Ageusia                                | 0                                | 1 (4.2%)   | 0                                     | -                              |
| Hand paralysis                         | 0                                | 1 (4.2%)   | 0                                     | -                              |
| Symptoms after discharge for 6 months  |                                  |  |                                       |                                |
| Fatigue                                | 0                                | 7 (29.2%)  | 0                                     | -                              |
| Sleep disturbances                     | 0                                | 2 (8.3%)   | 0                                     | -                              |
| Ageusia                                | 0                                | 1 (4.2%)   | 0                                     | -                              |

Continuous variables are presented as median ± interquartile range. Categorical variables are presented as percentages. Differences between groups were compared by using Kruskal–Wallis for nonnormal continuous variables and the  $\chi^2$  test for categorical variables. Labeled means without a common letter differ,  $p < 0.05$ .

and 1 oomycete), and 33 DNA viruses. According to the species accumulation curves, the sampling size and sequencing depth were sufficient to represent the overall microbial community (Figure S3A). Group variations in microbial  $\alpha$ - and  $\beta$ -diversity were characterized. Groups A and D had a higher Chao1 and Shannon index than Group C (analysis of variance [ANOVA],  $p < 0.05$ ; Figure 2A,B). The principal coordinate analysis (PCoA) plot showed that Groups A and H were mainly located on the left side of axis 1, and Group C was mainly located on the right side of axis 1 (Figure 2C). The partial least-squares discriminant analysis (PLS-DA) gave a more apparent separation pattern where Groups A, C, and D had unique microbial compositions that differed from Group H (Figure 2D). Quantitative calculations showed that microbial community differed significantly between Groups A–H, C–H, D–H, A–C, and C–D (permutational multivariate analysis of variance [PERMANOVA] and analysis of similarities [ANOSIM],  $p < 0.05$ ), whereas Group A did not significantly differ from Group D (PERMANOVA,  $p = 0.06$ ; ANOSIM,  $p = 0.30$ ) (Figure 3A). Firmicutes/Bacteroidetes ratio is a widely used marker to assess gut dysbiosis and pathological

conditions<sup>19</sup>. Herein, the average Firmicutes/Bacteroidetes ratio for Group H was 0.38 (Figure 3A), similar to the human microbiome project (HMP, 0.30)<sup>20</sup>, whereas the Firmicutes/Bacteroidetes ratios for Groups A, C, and D were 1.80, 2.19, and 1.32, respectively, all higher than Group H (Mann–Whitney test,  $p < 0.05$ ). The abundance of Proteobacteria, a hallmark of gut dysbiosis<sup>21</sup>, was higher in Group C than in other groups (all  $p < 0.05$ ).

Eight hundred and thirty species were common to all four groups, while 74, 207, 242, and 50 species were unique in Groups A, C, D, and H, respectively (Figure S3B). Linear discriminant analysis effect size (LEfSe) analysis was performed to show feature microbial taxa in Groups A, C, and D (Figures 3B–E and Table S1). Overall, *Bacteroides* sp. and *Clostridium* sp. were enriched in Group H, while *Akkermansia muciniphila* and *Streptococcus thermophilus* were enriched in COVID-19-related groups (Figure 3E). More specifically, short-chain fatty acid (SCFA)-producing and probiotic species were enriched in Group A, including *Faecalibacterium prausnitzii*, *Bifidobacterium longum*, *Blautia obeum*, *Roseburia hominis*, *Gemmiger formicilis*, and *Ruminococcus* sp.

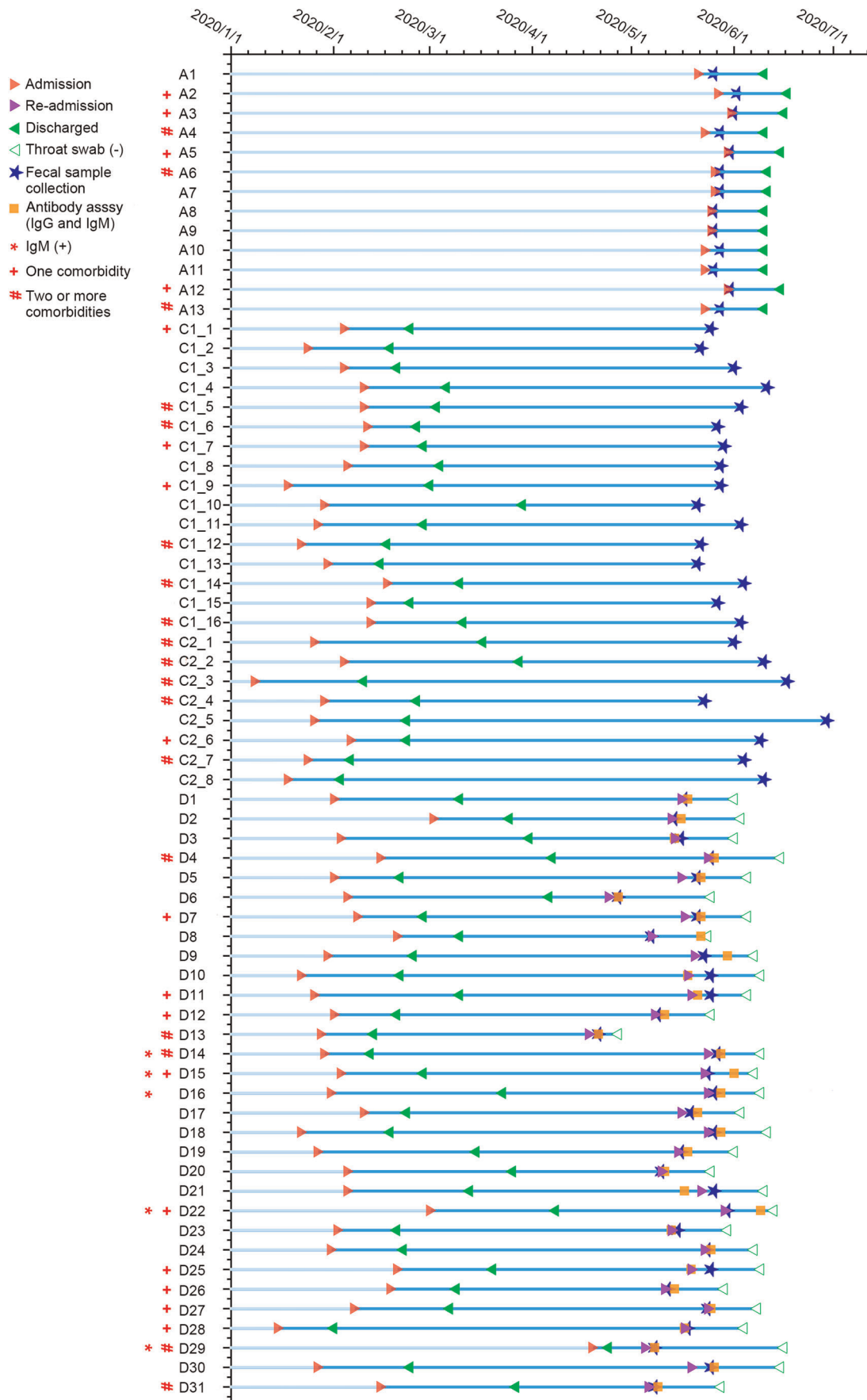
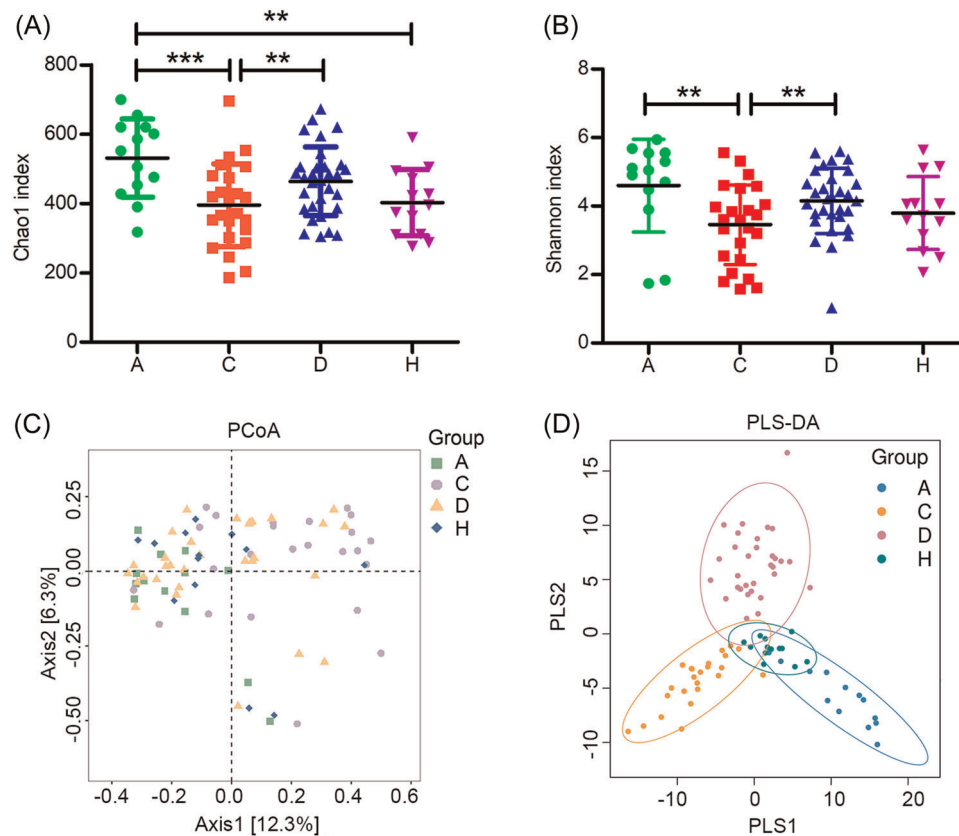


FIGURE 1 Continued.



**Figure 2.** Variation of microbial  $\alpha$ - and  $\beta$ -diversity in coronavirus disease 2019 (COVID-19)-related and control groups. Microbial  $\alpha$ -diversity was represented by Chao1 index (A) and Shannon index (B) in Group A ( $N = 13$ ), Group C ( $N = 24$ ), Group D ( $N = 31$ ), and Group H ( $N = 13$ ). One-way analysis of variance with the least significant difference (LSD) test was used to calculate the variation in microbial  $\alpha$ -diversity. Data are shown as mean  $\pm$  SD. \*\* $p < 0.01$ ; \*\*\* $p < 0.001$ . Microbial  $\beta$ -diversity variation was visualized by (C) principal coordinate analysis (PCoA) and (D) partial least-squares discriminant analysis (PLS-DA) ordination in four groups. PCoA ordination was calculated based on the Bray–Curtis distance matrix.

TF11-2AC. Interestingly, a positive correlation was observed between Chao1 index and the abundance of *Faecalibacterium prausnitzii* ( $\rho = 0.579$ ), *Blautia obeum* ( $\rho = 0.723$ ), *Roseburia hominis* ( $\rho = 0.489$ ), and *Gemmiger formicilis* ( $\rho = 0.61$ ), respectively (Spearman's correlation, all  $p < 0.001$ ), suggesting the SCFA-producing species might promote higher microbial diversity. Other SCFA-producing gut commensals were also enriched in Group D, including *Phascolarctobacterium faecium*, *Lactobacillus amylovorus*, and *Butyricoccus pullicaecorum*. The abundance of SCFA-producing species of these nine species was highest in Group A, followed by Group D and was least in Group C (15.4%, 8.9%, 4.6%, Kruskal–Wallis test,  $p = 0.003$ ) and was negatively correlated with adverse outcomes (logistic regression, odds ratio [OR]: 0.27; 95% confidence interval [CI]: 0.14–0.55;  $p < 0.001$ ). By

contrast, opportunistic pathogens were enriched in Group C, including *Escherichia coli*, *Erysipelatoclostridium ramosum*, *Raoultella ornithinolytica*, and *Hungatella hathewayi*. Notably, the abundance of *Escherichia coli* was negatively correlated with Chao1 index (Spearman's correlation,  $\rho = -0.354$ ) and was positively correlated with adverse outcomes (OR: 3.10; 95% CI: 1.43–6.68;  $p = 0.004$ ). Opportunistic pathogen *Fusobacterium ulcerans* was enriched in Group D.

Growth Rate InDex (GRiD) was used to identify the fast-growing species (Table S2). *Bacteroides ovatus*, *Bacteroides vulgatus*, and *Clostridium* sp. AM32-2 grew fast in Group H ( $t$  test,  $p < 0.05$ ). *Faecalibacterium prausnitzii*, *Bifidobacterium longum*, *Blautia obeum*, *Roseburia hominis*, *Faecalibacterium* sp. AF10-46, and *Streptococcus thermophilus* grew fast in Group A ( $p < 0.05$ ). Similarly, SCFA-producing

**Figure 1.** Schematic diagram of fecal sample collection and patient information. Three categories of coronavirus disease 2019 (COVID-19)-related cases were recruited, including 13 asymptomatic infection (Group A), 24 COVID-19 convalescents (Group C), 31 discharged patients with recurrent severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) positivity (Group D). The infection period in Group A was 16–21 days (median 16 days), and the fecal samples were collected within 1–6 days (median 2 days). The acute infection period in Group C was 12–60 days (median 25 days), and the fecal samples were collected after the patients' discharge for 53–128 days (median 92.5 days). The acute infection period in Group D was 5–61 days (median 27 days) and recurred SARS-CoV-2 positive on Days 11–106 (median 70) after the first discharge. The fecal samples of Group D were collected on Days 0–7 (median 2) after recurred SARS-CoV-2 positive (re-admission). The re-positive patients of Group D showed SARS-CoV-2 negative on Days 9–42 (median 18) after re-admission.

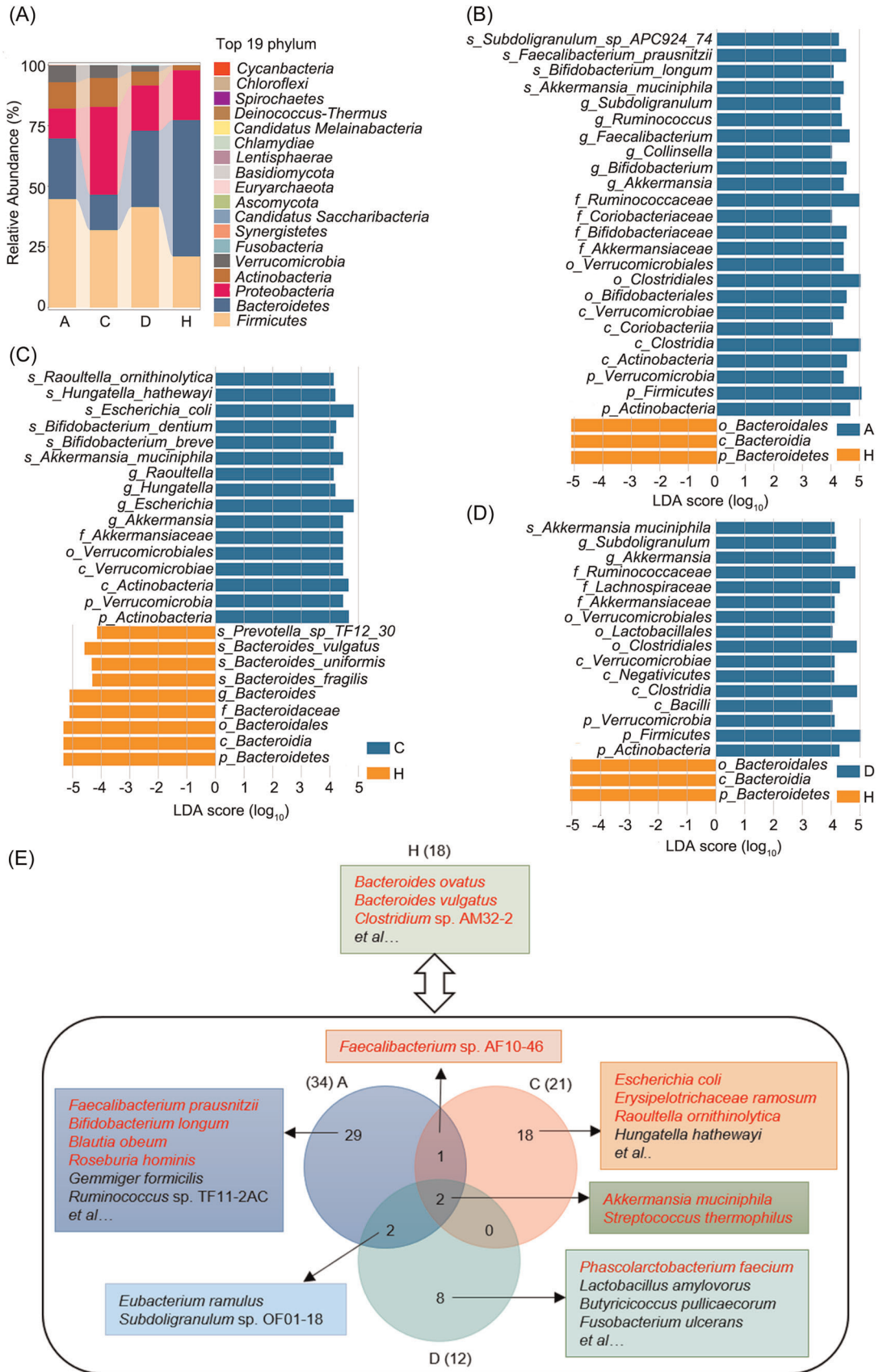


FIGURE 3 Continued.

species *Phascolarctobacterium faecium* and the shared feature species, *Akkermansia muciniphila* and *Streptococcus thermophilus* grew fast in Group D ( $p < 0.05$ ). Opportunistic pathogens (*Escherichia coli*, *Erysipelotrichaceae ramosum*, and *Raoultella ornithinolytica*) and *A. muciniphila* grew fast in Group C ( $p < 0.05$ ), consistent with the LEfSe and regression analysis. Overall, these results indicated that the SCFA-producing species were protective factors, while *Escherichia coli* and other opportunistic pathogens were risk factors for adverse outcomes in COVID-19 patients.

### Characterizing microbial functional pathways in COVID-19-related groups

Next, we profiled the functional pathways in the four groups (Figures S4 and S5). PLS-DA plot showed that Groups A, C, D, and H were clustered separately (ANOSIM,  $p = 0.001$ ), indicating large compositional variation in functional genes (Figure 4A). According to eggNOG annotation, Groups A and D had more abundant genes involved in defense mechanisms, whereas Group C had more genes associated with amino acid and carbohydrate transport and metabolism and energy production and conversion (LDA score  $> 2$ ,  $p < 0.05$ ; Figure 4B and Table S3). According to KO annotations, 11 metabolism pathways were enriched in Group H, such as fatty acid biosynthesis, folate biosynthesis, and vitamin B6 metabolism, indicating that these metabolic activities were decreased in COVID-19-related groups (LDA score  $> 2$ ,  $p < 0.05$ ; Table S4). Four genes (*folB*, *folK*, *queD*, and *queF*) were enriched in the folate biosynthesis pathway, and three genes (*pdxB*, *pdxA*, and *pdxJ*) were enriched in the vitamin B6 metabolism pathway in Group H. Compared with Group H, Groups A, C, and D had 25, 32, and 15 characteristic pathways, respectively (Figure 4C). Selenocompound metabolism pathway was enriched in COVID-19-related groups and the gene *trxB* was enriched in this pathway. The immune disease pathway was enriched in Groups C and D, indicating potential immune damage by COVID-19. Nutrient metabolism pathways were enriched in Group A, including secondary bile acid biosynthesis, amino acid metabolism and starch, sucrose metabolism, and lysine biosynthesis. The gene *baiN* was enriched in the secondary bile acid biosynthesis pathway. The mineral absorption pathway was enriched in Group D. Cell motility, flagellar assembly, and bacterial chemotaxis pathways were enriched in Group C, all of which were related to virulence factors, consistent with highly abundant opportunistic pathogens.

### Enrichment of virulence and resistance genes in COVID-19-related groups

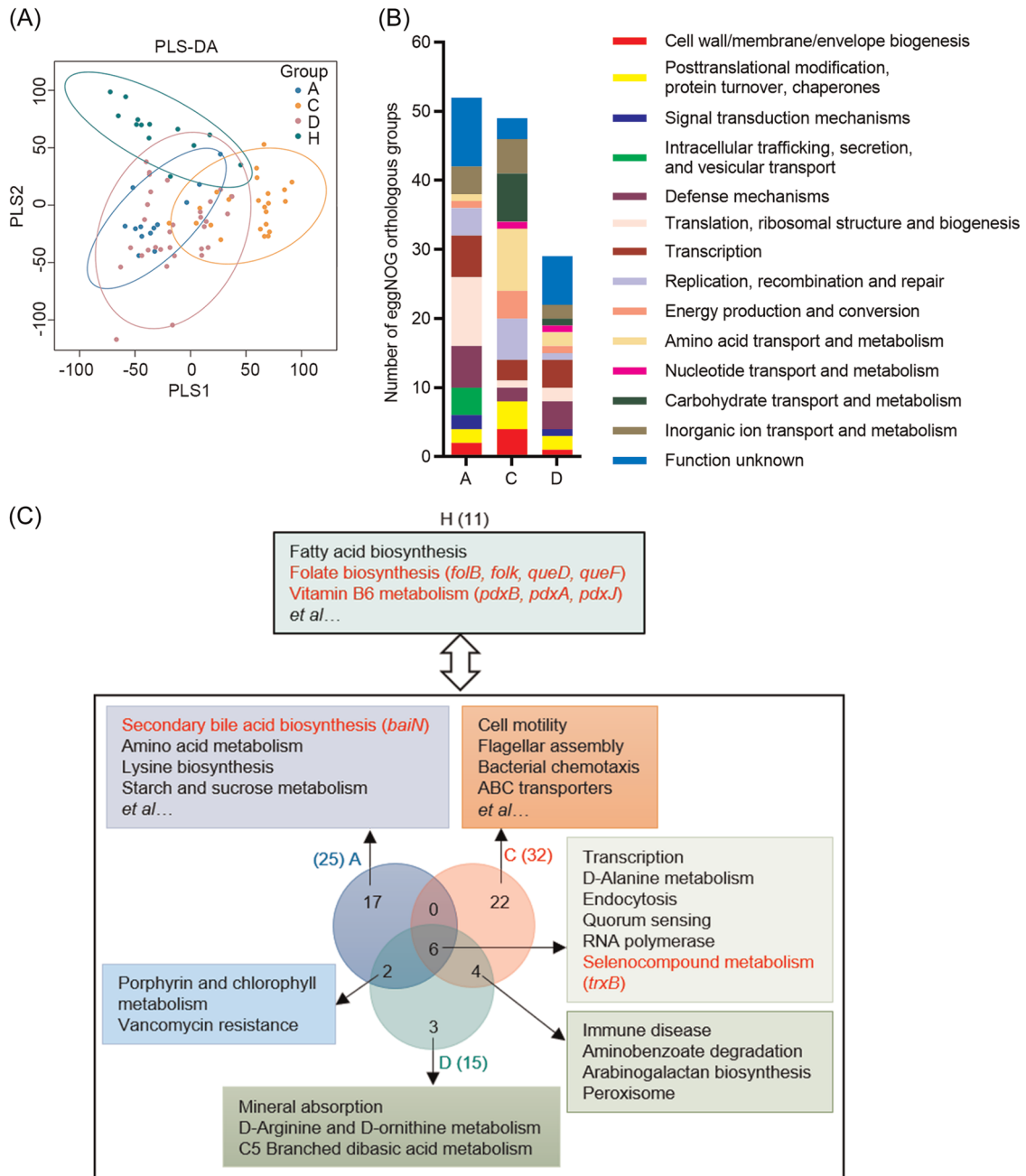
According to virulence factor database (VFDB) annotations, virulence factors of gut microbiota are divided into five categories: (i) colonization, adhesion, and invasion; (ii) anti-phagocytosis and immune escape; (iii) nutrition uptake, growth, and spread; (iv) toxin and endotoxin; and (v) others. Compared with Group H, 94, 140, and 52 virulence factors were enriched in Groups A, C, and D, respectively (LDA score  $> 2$ ,  $p < 0.05$ ; Figure 5A and Table S5). The nutrition uptake genes enriched in three COVID-19-related groups were mainly iron intake genes (*fbpC* and *hitC*). More virulence genes were enriched in Group C, specifically colonization, adhesion, and invasion-associated genes, including *bopD*, *fimD*, and *bcfD*. Blast searching of these virulence sequences indicated that a large proportion of these genes were derived from *Escherichia coli*, *Hungatella hathewayi*, *Citrobacter freundii*, and other opportunistic pathogens (Data not shown). *ecpE*, *fdeC*, and *entE*, genes related to adhesion and enterobactin, were also enriched in Group C and explicitly derived from *Escherichia coli* (LDA score  $> 2$ ,  $p < 0.05$ ).

The antibiotic resistome of gut microbiota was analyzed by searching against the comprehensive antibiotic resistance database (CARD). Compared with Group H, 67, 66, and 46 genes were enriched in Groups A, C, and D, respectively. There were also 37 resistance genes enriched in Group H. Most of these genes are involved in antibiotic target alteration or antibiotic efflux (LDA score  $> 2$ ,  $p < 0.05$ ; Figures S6 and S7A), including macrolide, cephalosporin, lincosamide, and carbapenem resistance in COVID-19-related groups (Figures S6 and S7B). Overall, COVID-19-related groups had more antibiotic target protection genes than Group H. The widespread resistance genes in the gut microbiota of COVID-19-related groups indicate that antibiotics should be cautiously used in the course of treatment. Especially for Group C participants, the enriched resistance genes (*otrC*, *oleB*, *msrE*, *salA*, *vmfR*, *vgaE*, *vgaB*, *optrA*, and *poxxA*) were derived from *Escherichia coli*, suggesting that the opportunistic pathogen has the potential to resist antibiotic treatment.

### Carbohydrate-active enzyme (CAZy) families are altered in COVID-19-related groups

Human gut microbiota encodes hundreds of CAZy genes to degrade various dietary and carbohydrates, which is

**Figure 3.** Microbial taxonomic composition and enriched taxa in COVID-19-related and control groups. (A) Average relative abundance of microbial phyla in four groups. The Firmicutes/Bacteroidetes ratio was calculated in each group, and a *t* test was performed to compare the differences between the control group (H) and each COVID-19-related group (A/C/D). Enriched microbial taxa in Group A (B), Group C (C), and Group D (D) compared with Group H (linear discriminant analysis effect size analysis, least-squares discriminant analysis (LDA) score  $> 4$ ,  $p < 0.05$ ). (E) Representative enriched microbial species in Groups A, C, D, and H (LDA score  $> 3$ ,  $p < 0.05$ ). There were 34, 21, 12, and 18 species enriched in Groups A, C, D, and H, respectively. Two enriched species, *Akkermansia muciniphila* and *Streptococcus thermophilus*, were shared by three COVID-19-related groups. Two enriched species, *Eubacterium ramulus* and *Subdoligranulum* sp. OF01-18, were shared by Groups A and D. The bacterial Growth Rate InDex (GRiD) was calculated based on read coverage differences in bacterial replication origin and terminal regions. A one-sample *t* test was used to assess the GRiD values. The fast-growing species are formatted in red font.

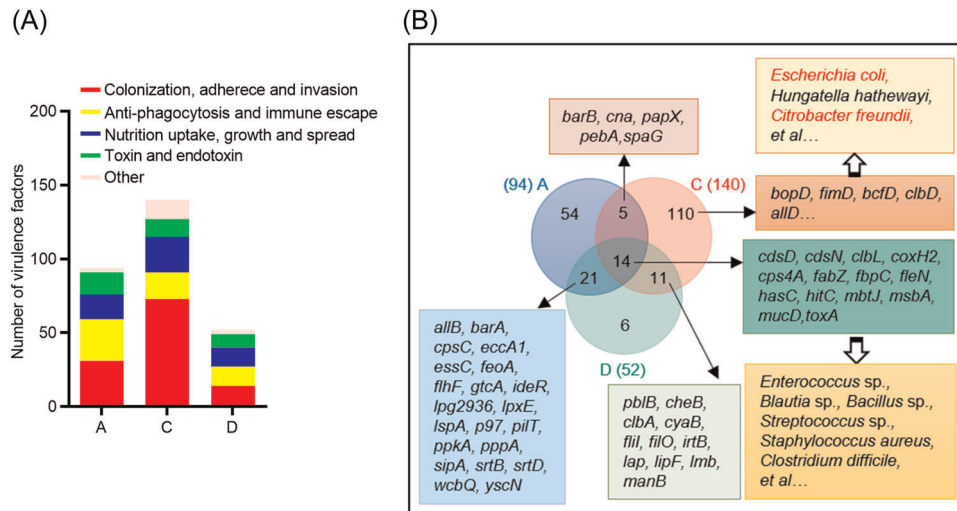


**Figure 4.** Microbial functional pathways enriched in COVID-19-related and control groups, The functional profiling of gut microbiota in COVID-19-related and healthy groups based on eggNOG and KO annotation. (A) Partial least-squares discriminant analysis (PLS-DA) ordination in Groups A, C, D, and H according to eggNOG annotations. (B) Number of enriched eggNOG orthologous groups in Groups A, C, and D compared with Group H (linear discriminant analysis effect size (LEfSe) analysis, LDA score  $> 2$ ,  $p < 0.05$ ). There were 53, 49, and 29 enriched eggNOG orthologous groups in Groups A, C, and D, respectively. (C) Enriched microbial functional pathways in Groups A, C, D, and H based on KO annotations (LEfSe analysis, LDA score  $> 2$ ,  $p < 0.05$ ). There were 25, 32, 15, and 11 enriched functional pathways in Groups A, C, D, and H, respectively. Six enriched pathways were shared by three COVID-19 groups, two were shared by Groups A and D, and four were shared by Groups C and D. The enriched genes and pathways are formatted in red font.

important for human health. PLS-DA and statistical analysis showed overall CAZy variation among groups (Figure 6A, ANOSIM,  $p = 0.001$ ). The relative abundance of glycosyltransferases (GTs), glycoside hydrolases (GHs), carbohydrate-binding modules (CBMs), carbohydrate esterases (CEs), polysaccharide lyases (PLs), and auxiliary activities (AAs) families are shown in Figure 6B. LEfSe

analysis found 33 CAZy families, including 14 GHs, 8 CBMs, 6 GTs, 3 CEs, and 2 PLs were enriched in group H (LDA score  $> 2$ ,  $p < 0.05$ ; Figure 6C and Table S6), indicating that these CAZy families were decreased in COVID-19 groups. *Bacteroidetes* encode more CAZy families than other phyla, suggesting that it can use a larger range of carbohydrate substrates. The abundance of CAZy families was decreased





**Figure 5.** Virulence genes enriched in COVID-19-related groups. (A) Number of enriched virulence genes in COVID-19-related groups compared with the healthy group (LEfSe analysis, LDA score > 2,  $p < 0.05$ ). The virulent genes were annotated based on the virulence factor database (VFDB). Group A had 31 enriched genes related to colonization, 28 related to anti-phagocytosis and immune escape, 17 related to nutrition uptake, growth, and spread and 15 related to toxin and endotoxin production. Group C had 73 enriched genes related to colonization, 18 related to anti-phagocytosis, and immune escape, 24 related to nutrition uptake, growth, and spread, and 12 related to toxin and endotoxin production. Group D had 14 enriched genes related to colonization, 13 related to anti-phagocytosis and immune escape, 13 related to nutrition uptake, growth, and spread, and 12 related to toxin and endotoxin production. (B) Enriched virulence genes in COVID-19-related groups compared with the control group. The COVID-19-related groups had 14 shared virulence genes, mainly derived from *Enterococcus* sp., *Blautia* sp., *Bacillus* sp., *Streptococcus* sp., *Staphylococcus aureus*, *Clostridium difficile*, and other opportunistic pathogens, based on sequence similarity analysis. Group C had 110 specific virulence genes, mainly derived from *Escherichia coli*, *Hungatella hathewayi*, and *Citrobacter freundii*.

in COVID-19-related groups, consistent with reduced *Bacteroidetes* in COVID-19-related groups. The gut microbiome confers the metabolic abilities to compensate for the paucity of GHs and PLs encoded in the human genome. PL family was not enriched in Groups A, C, and D, suggesting inadequate degradation of polysaccharides by gut microbiota in COVID-19 patients.

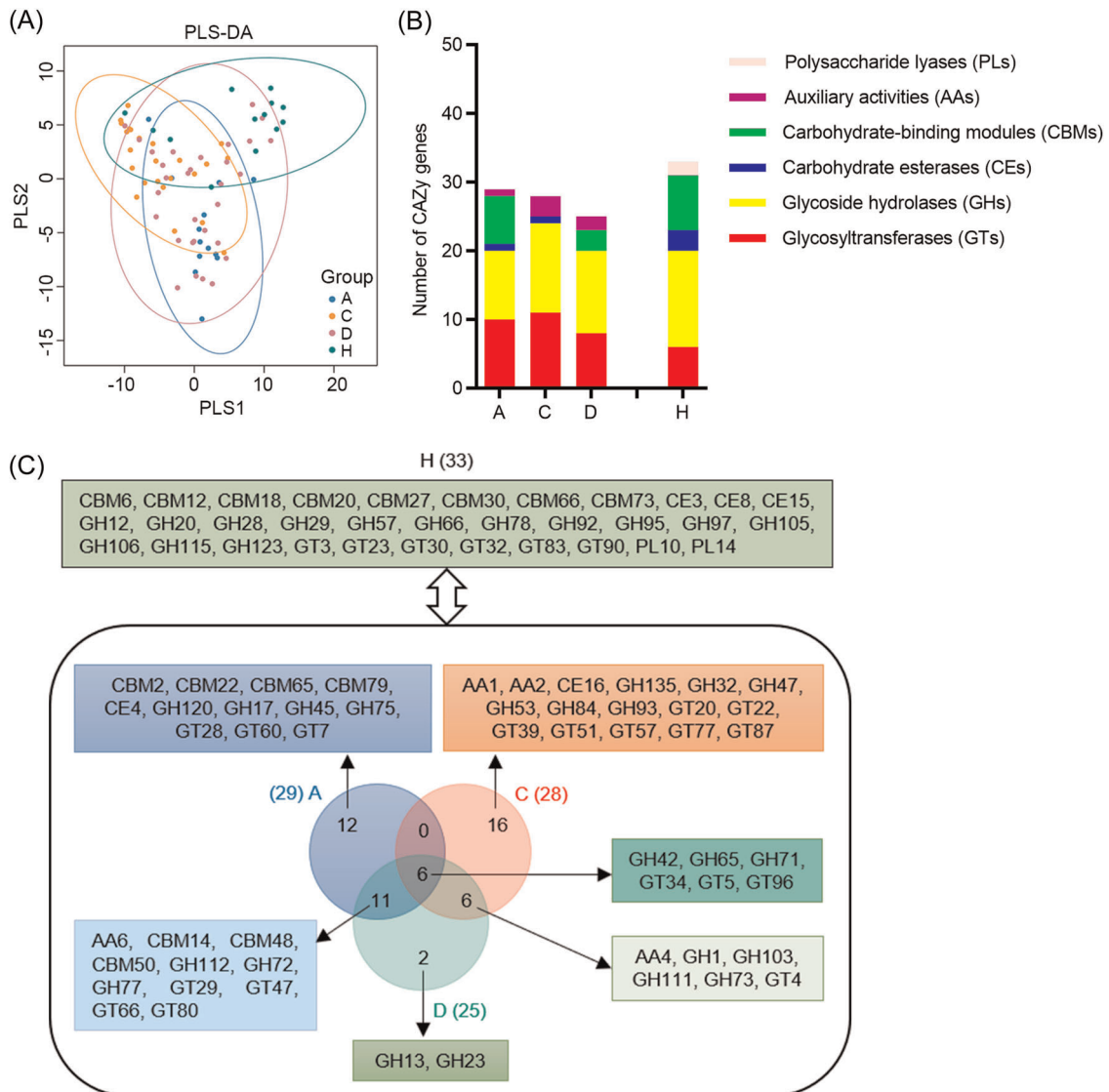
Compared with Group H, 29, 28, and 25 CAZy families were enriched in Groups A, C, and D, respectively (LDA score > 2,  $p < 0.05$ ). GH42, GT34, and GT96 families, enriched in three COVID-19-related groups, are related to galactose metabolism. CBMs play a key role in the hydrolysis of cellulose, xylan, and chitin. Seven CBMs families were enriched in Group A, and three of them were common in Group D. CBM14, an enriched CAZy family in Groups A and D, was first identified as an antimicrobial protein with the chitin-binding ability<sup>22</sup>. There was no characteristic CBM enrichment in Group C, which might lead to poor enzymatic activity in the hydrolysis of cellulose and xylan. GH73 was enriched in Groups C and D, and the enzyme is used during the host-cell invasion, such as the virulence-associated peptidoglycan hydrolase<sup>23</sup>, suggesting that the CAZy family might facilitate potential pathogenesis.

## DISCUSSION

The “biodiversity hypothesis” suggests that maintaining microbial diversity (richness) in the gastrointestinal tract is crucial for human health; losing microbial diversity can lead to gut dysbiosis and many chronic, metabolic and immune diseases<sup>24,25</sup>. Two recent studies reported a lower gut

microbial diversity in COVID-19 patients when compared to healthy controls<sup>9,26</sup>, but the pattern was not found in another COVID-19 gut microbiota survey<sup>27</sup>. In this study, we found that Groups A and D (no symptom after discharge) subjects had a significantly higher microbial diversity than Group C (with adverse outcomes), consistent with that of COVID-19 patients with lower microbial richness who had worse pulmonary functions<sup>13</sup>. Moreover, Group A had higher  $\alpha$ -diversity than healthy controls. Asymptomatic infections are more common in populations of young and middle-aged individuals than in the elderly<sup>28</sup>. It has been reported that the viral load of SARS-CoV-2 detected in asymptomatic cases was similar to that in symptomatic patients<sup>29</sup>. The asymptomatic and symptomatic cases displayed different gut microbiome, suggesting that gut microbiota may be related to the symptoms or progression of COVID-19<sup>30</sup>. Thus, the higher microbial richness partially explained the asymptomatic clinical characteristics of participants. COVID-19 vaccine protects against COVID-19 by shifting toward more asymptomatic infections<sup>31</sup>. All the cases (including asymptomatic infection cases) enrolled in this study were unvaccinated, indicating the symptom and microbiota variation is not driven by vaccination.

Besides the microbial richness, the  $\beta$ -diversity indicated the distinct fecal microbial composition in COVID-19-related groups. The SCFA-producing species *Faecalibacterium prausnitzii*, *Blautia obeum*, *Roseburia hominis*, and *Gemmiger formicilis* were enriched in Group A, and their abundance was positively correlated with microbial richness, suggesting that SCFA-producing species might promote



**Figure 6.** Carbohydrate-active enzyme (CAZY) families enriched in COVID-19-related groups. (A) Partial least-squares discriminant analysis (PLS-DA) ordination in Groups A, C, D, and H according to CAZY annotations. (B) Number of enriched CAZY families in COVID-19-related groups compared with the control group (LEfSe analysis, LDA score > 2,  $p < 0.05$ ). Thirty-three CAZY families were enriched in Group H, including 14 glycoside hydrolases (GHs), 8 carbohydrate-binding modules (CBMs), 6 glycosyltransferases (GTs), 3 carbohydrate esterases (CEs), and 2 polysaccharide lyases (PLs). Twenty-nine CAZY families were enriched in Group A, including 10 GTs, 10 GHs, 1 CE, 7 CBMs, and 1 auxiliary activities (AAs). Twenty-eight CAZY enriched in Group C, including 11 GTs, 13 GHs, 1 CE, and 3 AAs. Twenty-five CAZY enriched in Group D, including 8 GTs, 12 GHs, 3 CBMs, and 2 AAs. (C) Enriched CAZY families in COVID-19-related groups compared with the control group. A total of 6 enriched CAZY families were shared by 3 COVID-19-related groups, 11 shared by Groups A and D, and 6 shared by Groups C and D.

higher microbial diversity and contribute to the positive health consequences. Several other SCFA-producing species, *Bifidobacterium longum*, *Eubacterium ramulus*, *Pascolarctobacterium faecium*, *Butyrivococcus pullicaecorum*, and *Lactobacillus amylovorus* were also enriched in Groups A and D, and all these nine SCFA-producing species were negatively correlated with poor prognosis in COVID-19-related cases. In contrast, the butyrate-producing species (including *Faecalibacterium prausnitzii* and *Blautia obeum*) were significantly depleted in patients with postacute COVID-19 syndrome<sup>17</sup>. SCFAs have antiinflammatory effects by modulating the overproduction of inflammatory cytokines and

chemokines during viral infection<sup>32</sup>. The overproduction of inflammatory cytokines, known as “cytokine storm”, can cause tissue damage in COVID-19 patients and is one of the major reasons for the disease's mortality and morbidity. SCFAs might play an important role in antagonizing the inflammatory response induced by SARS-CoV-2. As Groups A and D were asymptomatic or manifested with only mild symptoms despite being tested positive for SARS-CoV-2, and since none of them developed sequelae, these results indicate that maintaining balanced and diverse gut microbiota and SCFAs are important in limiting sequelae. In addition, based on the low positive rate of IgM antibody (16.1%)

in Group D, and that all of them were home quarantined, the re-positivity of SARS-CoV-2 was unlikely due to reinfection. In a large epidemiologic study, the reinfection incidence in recovered COVID-19 patients is only 0.02%<sup>33</sup>, consistent with our results. Thus, the persistent viral presence is not associated with adverse health outcomes, suggesting that the gut microbiota is likely a better indicator to predict prognosis than persistent SARS-CoV-2 presence. Group C was enriched with opportunistic pathogens, such as *Erysipelatoclostridium ramosum*, *Raoultella ornithinolytica*, *Hungatella hathewayi* and *Escherichia coli*. *Erysipelatoclostridium ramosum* (previously called *Clostridium ramosum*) and *Hungatella hathewayi* (previously called *Clostridium hathewayi*) were the top bacteria that were positively associated with COVID-19 disease severity<sup>10</sup>. Furthermore, *Erysipelatoclostridium ramosum* was also enriched in patients with postacute COVID-19 syndrome<sup>17</sup>. These results suggested that *Erysipelatoclostridium ramosum* and *Hungatella hathewayi* were positively associated with postacute COVID-19 syndrome. *Escherichia coli* was a risk factor for the adverse outcomes in postacute COVID-19 individuals and was overrepresented with virulence genes related to adhesion and toxins production. The colibactin-producing *Escherichia coli* strains could cause inflammatory bowel disease (IBD) and colorectal cancer<sup>34</sup>. Indeed, a previous study found that adherent *Escherichia coli* strains in ileal mucosa could cause intestinal inflammatory symptoms, including abdominal pain, diarrhea, and fatigue<sup>35</sup>. Notably, fatigue was the most common symptom in Group C. These results suggest that the opportunistic pathogens *Escherichia coli* may be the risk factor for postacute COVID-19 syndrome, such as fatigue. According to a previous study<sup>17</sup> and this study, opportunistic pathogens are the important risk factors in shaping postacute COVID-19 syndrome. However, it might be inappropriate to use antibiotics in COVID-19 patients, as the antibiotics resistance genes of gut microbiota were enriched in the COVID-19-related cases. Antibiotic therapy might exacerbate gut dysbiosis. A probiotic supplement is one of the possible replacement options for antibiotics.

Compared with the control group, *Bacteroides ovatus* and *Bacteroides vulgatus* were decreased in COVID-19-related groups. However, it also reported that the patients with postacute COVID-19 syndrome had higher levels of *Bacteroides vulgatus*<sup>17</sup>. This inconsistency might result from the different symptoms of postacute COVID-19 syndrome and the geographic variation of participants in the two studies. The associated risk factors of the post-COVID-19 syndrome may include female sex, more than five early symptoms, early dyspnea, and prior psychiatric disorders<sup>36</sup>. Several studies have demonstrated that *Bacteroides ovatus* and *Bacteroides vulgatus* could reduce intestinal inflammation<sup>37–39</sup>. Previous studies reported an inverse correlation of *Bacteroides* with fecal SARS-CoV-2 load and ACE2 expression<sup>10,40</sup>. These results suggest that the viral load and entry in COVID-19 patients may affect *Bacteroides* colonization. *Bacteroidetes* encode many CAZy families,

specifically GH and PL, to digest dietary polysaccharides<sup>41</sup>. Low abundant *Bacteroidetes* might limit dietary polysaccharides utilizations in COVID-19 patients. The role of *Bacteroides*, including *Bacteroides vulgatus* and *Bacteroides ovatus*, in the postacute COVID-19 syndrome needs to be further clarified. Besides, the mucin-degrading bacteria, *Akkermansia muciniphila*, were enriched in three COVID-19-related groups. It has been reported that virus infection promoted the growth of *Akkermansia muciniphila* in mice<sup>42</sup>. The overrepresentation of *Akkermansia muciniphila* and reduced Firmicute/Bacteroidetes ratio occurred downstream of CD8 T-cell responses during the early phase of infection with a fast-spreading and persistent virus<sup>43</sup>. These results indicated that virus infection elicited host immune responses and altered the ecology of the gut microbiota, such as the blooming of *Akkermansia muciniphila*. *Akkermansia muciniphila* degrades the mucus layer to disrupt intestinal barrier function and induces inflammation<sup>44</sup>, and its abundance was positively associated with infection and immune responses in COVID-19 patients, including expression of interleukin-1 $\beta$  (IL-1 $\beta$ ) and IL-6<sup>27</sup>, suggesting adverse effects in COVID-19 cases. *Akkermansia muciniphila* grew faster in Groups C and D than in Group A, indicating that active *A. muciniphila* may degrade more mucus and induce higher levels of inflammatory cytokines in Groups C and D. SCFAs such as butyrate stimulate mucin production in intestinal epithelial cells, possibly counteracting mucin degradation by *Akkermansia muciniphila* to maintain mucus integrity and microbiota homeostasis, and consequently reducing inflammation in Groups A and D. Targeted quantitative metabolomics showed a drop in SCFAs and changes in several bile acids in the SARS-CoV-2-infected nonhuman primates<sup>45</sup>. The secondary bile acid biosynthesis associated pathways were also enriched in Group A. Secondary bile acids exhibit anti-inflammatory effects and restrict alpha virus dissemination by promoting immune response<sup>46,47</sup>. The differences in immune response determine the COVID-19 disease presentation and severity<sup>48</sup>. It is likely that the enriched SCFAs and secondary bile acid suppress inflammation and improve immunity during SARS-CoV-2 infection.

A recent study found that COVID-19 patients with severe/critical illness showed impaired SCFAs and L-isoleucine biosynthesis in the gut microbiome<sup>49</sup>. However, a previous study also showed that fecal samples with high SARS-CoV-2 infectivity had a higher functional capacity for amino acid biosynthesis<sup>50</sup>. Herein, we found that the amino acid metabolism and lysine biosynthesis associated pathways were enriched in group A. Most amino acids were depleted in the SARS-CoV-2-infected cells, while the aspartate and asparagine were upregulated, likely caused by activation of the cellular integrated stress response<sup>51</sup>. These results suggest that amino acid was essential for the metabolic activity of both SARS-CoV-2 and the host cell. Due to the important role of amino acids in maintaining body health, the possible supplementation of amino acids could be considered in COVID-19 patients. Compared with the control group, vitamin B6 metabolism

and folate biosynthesis were decreased in the COVID-19-related groups. Vitamin B6 in the body is mainly derived from diet and gut bacteria synthesis via intestinal absorption. Dysregulated gut microbiota and vitamin B6 deficiency led to autism-like behavior in mice<sup>52</sup>. Vitamin B6 may be useful for COVID-19 patients in ameliorating the severity of COVID-19 and its complications<sup>53</sup>. Low levels of serum folate were common among COVID-19 patients, but were not associated with the clinical outcomes<sup>54</sup>. Folate biosynthesis is a major biochemical feature of the human microbiome, and its deficiency might lead to human health disorders<sup>55</sup>. Furthermore, both folate and vitamin B6 have been demonstrated to have critical roles in supporting the human immune system and reducing the risk of infections<sup>56</sup>. Thus, folate and vitamin B6 supplementations may also help combat COVID-19. According to the microbial metabolic pathways, the seleno-compound metabolism pathway was enriched in COVID-19-related groups. A previous study found that 42% of the COVID-19 patients were selenium deficient<sup>57</sup>. Due to the antioxidant and antiinflammatory properties of selenium, it plays a crucial role in human physiology, such as maintaining the immune system and balance of gut microbiota<sup>58</sup>. Gut microbiota could remove selenium from the host under selenium-limiting conditions, and dietary selenium could also affect the composition and colonization of gut microbiota<sup>59</sup>. Selenium supplementation should be considered in the dietary supplement of COVID-19 patients. Overall, increasing gut microbial diversity, either by supplementing with probiotics, prebiotics or dietary supplementations (SCFA, secondary bile acid, selenium, folate, vitamin B6), could be a potential strategy to reduce clinical symptoms and improve prognosis in COVID-19 patients.

Although many new insights have been proposed, there are several limitations in this study. This study has a small sample size, especially in healthy controls. The Firmicutes/Bacteroidetes ratio for Group H was similar to the ratio described in the HMP project, a large-scale international collaboration project in characterizing the human microbiome. The ratios for Groups A, C, and D were significantly higher than Group H and HMP reference, indicating the COVID-19-related groups deviated from the normal standard. Although the participants were matched for sex, age, and region, gut microbiota can be influenced by other variables, such as host milieu, diet, and lifestyle. Our findings should be further confirmed in the larger cohorts with different populations. Herein, only one fecal sample was collected for each subject, and thus the longitudinal variation of the fecal microbiome cannot be characterized. Moreover, the functional potentials of the gut microbiome were inferred from the shotgun metagenomics. The approach characterizes the relative abundance of the DNA sequences and infers the functional potentials based on genomic data. To validate the abundance analysis results, we applied growth rate analysis to show that most characteristic species were actively replicating, suggesting the active metabolism. Future metabolomics should be a more optimal approach to validate the functional potentials of gut microbial metabolites between COVID-19 and healthy individuals.

## MATERIALS AND METHODS

### Study design and participants

This study was approved by the Ethics Committee of Hubei Provincial Hospital of Traditional Chinese Medicine (Permit number: HUYZ2020-C21-01). Participants were recruited from April 17, 2020 to June 30, 2020 in Hubei Provincial Hospital of Traditional Chinese Medicine (Wuhan, China) and were followed up until December 31, 2020. All study participants provided informed consent, and the records were kept at the hospital. The participants were diagnosed, categorized, and received standard treatment according to “diagnosis and treatment guidelines” issued by the National Health Commission of China. We finally enrolled 81 participants, including 13 asymptomatic infections (Group A), 24 convalescent cases with adverse outcomes<sup>14</sup> (Group C), 31 discharged patients with SARS-CoV-2 re-positive (Group D), and 13 non-COVID-19 healthy controls with age- and sex-matched (Group H) (excluding individuals subsequently found to be ineligible for enrollment; Figure S1). Group A participants were tested positive for SARS-CoV-2 without clinical symptoms. The fecal/blood samples were collected during the infection period before any drug usage. The fecal/blood samples of Group C were collected in COVID-19 convalescent cases that were discharged from the hospital and had persisting chest CT abnormalities or symptoms for more than 8 weeks. Group D patients were discharged upon being tested to be negative for SARS-CoV-2 and no obvious clinical symptoms, but were later found re-positive during home quarantine by two consecutive reverse transcription-polymerase chain reaction tests of throat swabs. The fecal/blood samples were collected when SARS-CoV-2 re-positivity was confirmed. The time of sample collection in participants is shown in Figure 1. No antibiotic was used in the last 3 months before fecal collection. Medical records of patients, including the epidemiological, clinical, laboratory, radiological characteristics, treatment, and outcome data, were reviewed and extracted by experienced clinicians. Clinical outcomes were followed up after patients were discharged for about 6 months by phone. Healthy controls were tested negative for SARS-CoV-2 during the COVID-19 pandemic. The inclusion criteria of non-COVID-19 healthy controls are (i) no mental illness, (ii) no underlying infectious or acute disease, and (iii) the residents who lived in the Wuhan area. The exclusion criteria are (i) antibiotic or prebiotic use in the last 3 months, (ii) during pregnancy or lactation, (iii) bowel surgery in the last 6 months and history of gastrointestinal disease, and (iv) known history of severe organ failure.

Blood and fecal samples from patients were collected by hospital staff, and non-COVID-19 healthy controls provided fecal samples at home by self-sampling. Samples were stored at  $-80^{\circ}\text{C}$  until processing.

### DNA extraction, library construction, and metagenomic sequencing

Total genomic DNA was extracted from the fecal samples using Stool DNA Kit (Omega Bio-tek) according to the

manufacturer's instructions. The fecal samples were inactivated at 56°C for 30 min before DNA extraction to inactivate SARS-CoV-2 viruses. The quantity and quality of the extracted DNA were assessed by Fluorometer (Promega) and agarose gel electrophoresis. For the qualified samples, the shotgun metagenomic libraries were constructed by the TruSeq Nano DNA High Throughput Library Prep Kit (Illumina). Read lengths were 2 × 150 bp with an insert size of ~400 bp. Dual indexed barcodes were applied for sample multiplexing. The prepared libraries were stored in a freezer at -20°C and sequenced by the Illumina HiSeq X Ten platform (Illumina).

### Metagenomic sequencing, assembly, and statistical analysis

Raw reads were demultiplexed according to the barcode information. Adapter sequences were removed by Cutadapt (v1.2.1)<sup>60</sup>. A 5-bp sliding window algorithm was applied to remove low-quality regions (average base quality <20) in the raw reads. Trimmed reads with length <50 bp were discarded. Human reads were removed by BMTagger (v.3.101) and KneadData (v.0.9.0). The processed clean reads were deposited in the Genome Sequence Archive (<https://bigd.big.ac.cn/gsa>) with the accession number CRA003945<sup>61,62</sup>. The clean reads were *de novo* assembled by MEGAHIT (v.1.2.9)<sup>63</sup>. The CDS (coding sequences >300 bp) regions were predicted by MetaGeneMark (v.3.25)<sup>64</sup> and clustered by CD-HIT (v.4.8.1)<sup>65</sup> to produce a set of nonredundant genes with an amino acid sequence identity cutoff of 90%. The abundance of the nonredundant genes was calculated by coverage function from SOAPdenovo2 (v.1.0)<sup>66</sup>. Microbial taxonomy was annotated by MEGAN (v.6.0)<sup>67</sup> using a lowest common ancestor algorithm based on searching genes against the NCBI-NT database (BLASTN, *e* value <0.001). Microbial functional genes were annotated by searching genes against the KEGG database (release 97.0) and eggNOG (v4.0) by DIAMOND (v.2.0.4) with *e* value <0.001 and coverage ratio >40%<sup>68-70</sup>. Microbial nonredundant genes were also searched against CAZy database to describe the structurally related catalytic and carbohydrate-binding enzymes that degrade, modify, or create glycosidic bonds<sup>71</sup>. Antimicrobial resistance genes were annotated by resistance gene identifier (v.5.2.0) against CARD (v3.1.3)<sup>72</sup>. Microbial virulent factors were annotated by DAMOND (v.2.0.4) against the VFDB (v.2019)<sup>73</sup>.

The characteristic microbial taxa and functional genes/pathways in sample groups were analyzed by LEfSe analysis<sup>74</sup> on the Galaxy website (<http://huttenhower.sph.harvard.edu/galaxy/>, v.1.0). Microbial community variation ( $\beta$ -diversity) was calculated by Bray-Curtis ordination, unconstrained ordination (PCoA), and constrained ordination (PLS-DA) at the species level among groups<sup>75-77</sup>. PERMANOVA/Adonis (10,000 permutations) and ANOSIM were conducted to quantify the microbial taxonomic and functional composition variation between sample groups. The growth rates of characteristic species (LDA >2 in LEfSe

analysis) were calculated by GRiD (v.1.3)<sup>78</sup>. Only species with taxa annotation resolved at the species level were included. Reference genomes were downloaded from the prokaryote database in NCBI. If multiple reference genomes were available, the genome with the smallest scaffold number was chosen. R (v.3.6.1) and Quantitative Insights Into Microbial Ecology (v.1.8.0) were used to analyze and visualize data throughout the study<sup>79</sup>.

Demographic characteristics data were presented as median ± interquartile range and as percentages for categorical variables. Differences between the two groups were compared by using Student's *t* test for normal continuous variables, the Kruskal-Wallis test for nonnormal continuous variables and the  $\chi^2$  test or Fisher's exact test for categorical variables. Binary logistic regression was performed to analyze the associations between the abundance of microbial taxa ( $\log_{10}$  transformed) and health outcomes. Correlations between enriched species and microbial richness (Chao1 index) were tested with Spearman's correlation. Statistical analyses were performed using SPSS V.21 (SPSS). Differences with a *p* < 0.05 (two-sided) were considered statistically significant.

### ACKNOWLEDGMENTS

We thank Dr. Ding Gao in the core facility and the technical support of the Wuhan Institute of Virology and Prof. Fei Deng in the National Virus Resource Center for her technical support. We thank all the volunteer subjects who enrolled in the study. This study was supported by the National Key Research and Development Program (Grant Nos. 2020YFC0861100 and 2020YFC0845000), the Strategic Priority Research Program of the Chinese Academy of Sciences (Grant No. XDB29050100), the Scientific Research Fund of Traditional Chinese Medicine of Hubei Health Commission (Grant No. ZY2021M035), and the Youth Innovation Promotion Association of CAS.

### AUTHOR CONTRIBUTIONS

Ruqin Lin, Shanshan Cao, and Yu Sun performed the experiments, data analyses, and drafted the manuscript. Mingzhong Xiao, Linhua Zhao, Peng Chen, Xiaolin Tong, Hui Zhu, and Xiaodong Li recruited subjects and collected clinical samples and data. Xiaoxiao Mao and Zheyuan Ou performed the statistical analysis. Dong Men, Xiaodong Li, Yiqun Deng, Xian-En Zhang, and Jikai Wen designed and supervised the study. All authors approved the manuscript.

### ETHICS STATEMENT

This study was approved by the Ethics Committee of Hubei Provincial Hospital of Traditional Chinese Medicine (Permit number: HUYZ2020-C21-01). The study was performed in accordance with the Declaration of Helsinki and the Rules of Good Clinical Practice.

### CONFLICT OF INTERESTS

The authors declare no conflict of interests.

## DATA AVAILABILITY

Data are available in a public, open access repository. The processed clean reads were deposited in the Genome Sequence Archive (<https://bigd.big.ac.cn/gsa>) with the accession number CRA003945.

## SUPPORTING INFORMATION

Additional Supporting Information for this article can be found online at <https://doi.org/10.1002/mlf2.12022>.

## ORCID

Jikai Wen  <http://orcid.org/0000-0002-4010-4416>

## REFERENCES

- Guan WJ, Ni ZY, Hu Y, Liang WH, Ou CQ, He JX, et al. Clinical characteristics of coronavirus disease 2019 in China. *N Engl J Med*. 2020;382:1708–20.
- Cheung KS, Hung IFN, Chan PPY, Lung KC, Tso E, Liu R, et al. Gastrointestinal manifestations of SARS-CoV-2 infection and virus load in fecal samples from a Hong Kong Cohort: systematic review and meta-analysis. *Gastroenterology*. 2020;159:81–95.
- Wang W, Xu Y, Gao R, Lu R, Han K, Wu G, et al. Detection of SARS-CoV-2 in different types of clinical specimens. *JAMA*. 2020;323:1843–4.
- Xiao F, Sun J, Xu Y, Li F, Huang X, Li H, et al. Infectious SARS-CoV-2 in feces of patient with severe COVID-19. *Emerging Infect Dis*. 2020;26:1920–2.
- Hoffmann M, Kleine-Weber H, Schroeder S, Krüger N, Herrler T, Erichsen S, et al. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell*. 2020;181:271–80.e8.
- Xiao F, Tang M, Zheng X, Liu Y, Li X, Shan H. Evidence for gastrointestinal infection of SARS-CoV-2. *Gastroenterology*. 2020;158:1831–3.e3.
- Han Y, Duan X, Yang L, Nilsson-Payant BE, Wang P, Duan F, et al. Identification of SARS-CoV-2 inhibitors using lung and colonic organoids. *Nature*. 2021;589:270–5.
- Effenberger M, Grabherr F, Mayr L, Schwaerzler J, Nairz M, Seifert M, et al. Faecal calprotectin indicates intestinal inflammation in COVID-19. *Gut*. 2020;69:1543–4.
- Ren Z, Wang H, Cui G, Lu H, Wang L, Luo H, et al. Alterations in the human oral and gut microbiomes and lipidomics in COVID-19. *Gut*. 2021;70:1253–65.
- Zuo T, Zhang F, Lui GCY, Yeoh YK, Li AYL, Zhan H, et al. Alterations in gut microbiota of patients with COVID-19 during time of hospitalization. *Gastroenterology*. 2020;159:944–55.e8.
- Yeoh YK, Zuo T, Lui GC, Zhang F, Liu Q, Li AY, et al. Gut microbiota composition reflects disease severity and dysfunctional immune responses in patients with COVID-19. *Gut*. 2021;70:698–706.
- Zuo T, Wu X, Wen W, Lan P. Gut microbiome alterations in COVID-19. *Genomics Proteomics Bioinformatics*. 2021;S1672–0229:00206–0.
- Chen Y, Gu S, Chen Y, Lu H, Shi D, Guo J, et al. Six-month follow-up of gut microbiota richness in patients with COVID-19. *Gut*. 2021;71:222–5.
- Nalbandian A, Sehgal K, Gupta A, Madhavan MV, McGroder C, Stevens JS, et al. Post-acute COVID-19 syndrome. *Nat Med*. 2021;27:601–15.
- Huang L, Yao Q, Gu X, Wang Q, Ren L, Wang Y, et al. 1-year outcomes in hospital survivors with COVID-19: a longitudinal cohort study. *Lancet*. 2021;398:747–58.
- Huang C, Huang L, Wang Y, Li X, Ren L, Gu X, et al. 6-month consequences of COVID-19 in patients discharged from hospital: a cohort study. *Lancet*. 2021;397:220–32.
- Liu Q, Mak JWY, Su Q, Yeoh YK, Lui GC, Ng SSS, et al. Gut microbiota dynamics in a prospective cohort of patients with post-acute COVID-19 syndrome. *Gut*. 2022;71:544–52.
- Lu J, Peng J, Xiong Q, Liu Z, Lin H, Tan X, et al. Clinical, immunological and virological characterization of COVID-19 patients that test re-positive for SARS-CoV-2 by RT-PCR. *EBioMedicine*. 2020;59:102960.
- Magne F, Gotteland M, Gauthier L, Zazueta A, Poeso S, Navarrete P, et al. The Firmicutes/Bacteroidetes ratio: a relevant marker of gut dysbiosis in obese patients? *Nutrients*. 2020;12:1474.
- Lloyd-Price J, Mahurkar A, Rahnavard G, Crabtree J, Orvis J, Hall AB, et al. Strains, functions and dynamics in the expanded Human Microbiome Project. *Nature*. 2017;550:61–6.
- Dang AT, Marsland BJ. Microbes, metabolites, and the gut–lung axis. *Mucosal Immunol*. 2019;12:843–50.
- Kawabata S, Nagayama R, Hirata M, Shigenaga T, Agarwala KL, Saito T, et al. Tachycitin, a small granular component in horseshoe crab hemocytes, is an antimicrobial protein with chitin-binding activity. *J Biochem*. 1996;120:1253–60.
- Bublitz M, Polle L, Holland C, Heinz DW, Nimitz M, Schubert WD. Structural basis for autoinhibition and activation of Auto, a virulence-associated peptidoglycan hydrolase of *Listeria monocytogenes*. *Mol Microbiol*. 2009;71:1509–22.
- Haahela T, Holgate S, Pawankar R, Akdis CA, Benjaponpitak S, Caraballo L, et al. The biodiversity hypothesis and allergic disease: world allergy organization position statement. *World Allergy Organ J*. 2013;6:3.
- Bello MGD, Knight R, Gilbert JA, Blaser MJ. Preserving microbial diversity. *Science*. 2018;362:33–4.
- Chen Y, Gu S, Chen Y, Lu H, Shi D, Guo J, et al. Six-month follow-up of gut microbiota richness in patients with COVID-19. *Gut*. 2021;7:222–5.
- Yeoh YK, Zuo T, Lui GC-Y, Zhang F, Liu Q, Li AY, et al. Gut microbiota composition reflects disease severity and dysfunctional immune responses in patients with COVID-19. *Gut*. 2021;70:698–706.
- Gao Z, Xu Y, Sun C, Wang X, Guo Y, Qiu S, et al. A systematic review of asymptomatic infections with COVID-19. *J Microbiol Immunol Infect*. 2021;54:12–6.
- Zou L, Ruan F, Huang M, Liang L, Huang H, Hong Z, et al. SARS-CoV-2 viral load in upper respiratory specimens of infected patients. *N Engl J Med*. 2020;382:1177–9.
- Liu Y, Zhang H, Tang X, Jiang X, Yan X, Liu X, et al. Distinct meta-genomic signatures in the SARS-CoV-2 infection. *Front Cell Infect Microbiol*. 2021;11:706970.
- Mehrotra DV, Janes HE, Fleming TR, Annunziato PW, Neuzil KM, Carpp LN, et al. Clinical endpoints for evaluating efficacy in COVID-19 vaccine trials. *Ann Intern Med*. 2021;174:221–8.
- Vinolo MAR, Rodrigues HG, Nachbar RT, Curi R. Regulation of inflammation by short chain fatty acids. *Nutrients*. 2011;3:858–76.
- Lumley SF, O'Donnell D, Stoesser NE, Matthews PC, Howarth A, Hatch SB, et al. Antibody status and incidence of SARS-CoV-2 infection in health care workers. *N Engl J Med*. 2021;384:533–40.
- Prorok-Hamon M, Friswell MK, Alswied A, Roberts CL, Song F, Flanagan PK, et al. Colonic mucosa-associated diffusely adherent *afaC+* *Escherichia coli* expressing *lpfA* and *pks* are increased in inflammatory bowel disease and colon cancer. *Gut*. 2014;63:761–70.
- Darfeuille-Michaud A, Neut C, Barnich N, Lederman E, Di Martino P, Desreumaux P, et al. Presence of adherent *Escherichia coli* strains in ileal mucosa of patients with Crohn's disease. *Gastroenterology*. 1998;115:1405–13.
- Yong SJ. Long COVID or post-COVID-19 syndrome: putative pathophysiology, risk factors, and treatments. *Infect Dis*. 2021;53:737–54.
- Ihekweazu FD, Engevik MA, Ruan W, Shi Z, Fultz R, Engevik KA, et al. *Bacteroides ovatus* promotes IL-22 production and reduces trinitrobenzene sulfonic acid-driven colonic inflammation. *Am J Pathol*. 2021;191:704–19.
- Li S, Wang C, Zhang C, Luo Y, Cheng Q, Yu L, et al. Evaluation of the effects of different *Bacteroides vulgatus* strains against DSS-induced colitis. *J Immunol Res*. 2021;2021:9117805.
- Wang C, Xiao Y, Yu L, Tian F, Zhao J, Zhang H, et al. Protective effects of different *Bacteroides vulgatus* strains against lipopolysaccharide-induced acute intestinal injury, and their underlying functional genes. *J Adv Res*. 2021;36:27–37.

- 40 Geva-Zatorsky N, Sefik E, Kua L, Pasman L, Tan TG, Ortiz-Lopez A, et al. Mining the human gut microbiota for immunomodulatory organisms. *Cell*. 2017;168:928–43.e11.
- 41 El Kaoutari A, Armougom F, Gordon JI, Raoult D, Henrissat B. The abundance and variety of carbohydrate-active enzymes in the human gut microbiota. *Nat Rev Microbiol*. 2013;11:497–504.
- 42 Hu X, Zhao Y, Yang Y, Gong W, Sun X, Yang L, et al. *Akkermansia muciniphila* improves host defense against influenza virus infection. *Front Microbiol*. 2020;11:586476.
- 43 Labarta-Bajo L, Gramalla-Schmitz A, Gerner RR, Kazane KR, Humphrey G, Schwartz T, et al. CD8 T cells drive anorexia, dysbiosis, and blooms of a commensal with immunosuppressive potential after viral infection. *Proc Natl Acad Sci USA*. 2020;117:24998–5007.
- 44 Sencio V, Machado MG, Trottein F. The lung–gut axis during viral respiratory infections: the impact of gut dysbiosis on secondary disease outcomes. *Mucosal Immunol*. 2021;14:296–304.
- 45 Sokol H, Contreras V, Maisonnasse P, Desmons A, Delache B, Sencio V, et al. SARS-CoV-2 infection in nonhuman primates alters the composition and functional activity of the gut microbiota. *Gut Microbes*. 2021;13:1–19.
- 46 Campbell C, McKenney PT, Konstantinovskiy D, Isaeva OI, Schizas M, Verter J, et al. Bacterial metabolism of bile acids promotes generation of peripheral regulatory T cells. *Nature*. 2020;581:475–9.
- 47 Winkler ES, Shrihari S, Hykes BL, Jr., Handley SA, Andhey PS, Huang YS, et al. The intestinal microbiome restricts alphavirus infection and dissemination through a bile acid-type I IFN signaling axis. *Cell*. 2020;182:901–18.
- 48 Brodin P. Immune determinants of COVID-19 disease presentation and severity. *Nat Med*. 2021;27:28–33.
- 49 Zhang F, Wan Y, Zuo T, Yeoh YK, Liu Q, Zhang L, et al. Prolonged Impairment of short-chain fatty acid and L-isoleucine biosynthesis in gut microbiome in patients with COVID-19. *Gastroenterology*. 2022;162:548–61.e4.
- 50 Zuo T, Liu Q, Zhang F, Lui GC, Tso EY, Yeoh YK, et al. Depicting SARS-CoV-2 faecal viral activity in association with gut microbiota composition in patients with COVID-19. *Gut*. 2021;70:276–84.
- 51 Zhang Y, Guo R, Kim SH, Shah H, Zhang S, Liang JH, et al. SARS-CoV-2 hijacks folate and one-carbon metabolism for viral replication. *Nat Commun*. 2021;12:1676.
- 52 Li Y, Luo ZY, Hu YY, Bi YW, Yang JM, Zou WJ, et al. The gut microbiota regulates autism-like behavior by mediating vitamin B6 homeostasis in EphB6-deficient mice. *Microbiome*. 2020;8:120.
- 53 Kumrungsee T, Zhang P, Chartkul M, Yanaka N, Kato N. Potential role of vitamin B6 in ameliorating the severity of COVID-19 and its complications. *Front Nutr*. 2020;7:562051.
- 54 Meisel E, Efron O, Bleier J, Beit Halevi T, Segal G, Rahav G, et al. Folate levels in patients hospitalized with coronavirus disease 2019. *Nutrients*. 2021;13:812.
- 55 Engevik MA, Morra CN, Röth D, Engevik K, Spinler JK, Devaraj S, et al. Microbial metabolic capacity for intestinal folate production and modulation of host folate receptors. *Front Microbiol*. 2019;10:2305.
- 56 Calder PC. Nutrition, immunity and COVID-19. *BMJ Nutr Prev Health*. 2020;3:74–92.
- 57 Im JH, Je YS, Baek J, Chung MH, Kwon HY, Lee JS. Nutritional status of patients with COVID-19. *Int J Infect Dis*. 2020;100:390–3.
- 58 Ferreira RLU, Sena-Evangelista KCM, de Azevedo EP, Pinheiro FI, Cobucci RN, Pedrosa LFC. Selenium in human health and gut microbiota: bioavailability of selenocompounds and relationship with diseases. *Front Nutr*. 2021;8:685317.
- 59 Kasaikina MV, Kravtsova MA, Lee BC, Seravalli J, Peterson DA, Walter J, et al. Dietary selenium affects host selenoproteome expression by influencing the gut microbiota. *FASEB J*. 2011;25:2492–9.
- 60 Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J*. 2011;17:10–2.
- 61 Partners NGDCMa. Database resources of the National Genomics Data Center in 2020. *Nucleic Acids Res*. 2020;48:D24–33.
- 62 Wang Y, Song F, Zhu J, Zhang S, Yang Y, Chen T, et al. GSA: Genome Sequence Archive. *Genomics Proteomics Bioinformatics*. 2017;15:14–8.
- 63 Li D, Liu CM, Luo R, Sadakane K, Lam TW. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics*. 2015;31:1674–6.
- 64 Zhu W, Lomsadze A, Borodovsky M. Ab initio gene identification in metagenomic sequences. *Nucleic Acids Res*. 2010;38:e132.
- 65 Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*. 2012;28:3150–2.
- 66 Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, et al. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience*. 2012;1:1–18.
- 67 Huson DH, Auch AF, Qi J, Schuster SC. MEGAN analysis of metagenomic data. *Genome Res*. 2007;17:377–86.
- 68 Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nat Methods*. 2015;12:59–60.
- 69 Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, et al. KEGG for linking genomes to life and the environment. *Nucleic Acids Res*. 2008;36:D480–4.
- 70 Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, et al. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res*. 2019;47:D309–14.
- 71 Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res*. 2014;42:D490–5.
- 72 Alcock BP, Raphenya AR, Lau TTY, Tsang KK, Bouchard M, Edalatmand A, et al. CARD 2020: antibiotic resistance surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Res*. 2020;48:D517–25.
- 73 Liu B, Zheng D, Jin Q, Chen L, Yang J. VFDB 2019: a comparative pathogenomic platform with an interactive web interface. *Nucleic Acids Res*. 2019;47:D687–92.
- 74 Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS, et al. Metagenomic biomarker discovery and explanation. *Genome Biol*. 2011;12:R60.
- 75 Ramette A. Multivariate analyses in microbial ecology. *FEMS Microbiol Ecol*. 2007;62:142–60.
- 76 Bray JR, Curtis JT. An ordination of upland forest communities of southern Wisconsin. *Ecol Monograph*. 1957;27:325–49.
- 77 Barker M, Rayens W. Partial least squares for discrimination. *J Chemom*. 2003;17:166–73.
- 78 Emiola A, Oh J. High throughput in situ metagenomic measurement of bacterial replication at ultra-low sequencing coverage. *Nat Commun*. 2018;9:4956.
- 79 Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods*. 2010;7:335–6.

**How to cite this article:** Lin R, Xiao M, Cao S, Sun Y, Zhao L, Mao X, et al. Distinct gut microbiota and health outcomes in asymptomatic infection, viral nucleic acid test re-positive, and convalescent COVID-19 cases. *mLife*. 2022;1:183–197. doi:10.1002/mlf2.12022