


ORIGINAL ARTICLE

Meta-analysis of the microbial biomarkers in the gut–lung crosstalk in COVID-19, community-acquired pneumonia and *Clostridium difficile* infections

S. Aishwarya^{1,2}  and K. Gunasekaran²¹ Department of Bioinformatics, Stella Maris College (Autonomous), Chennai, India² Centre for Advanced Study in Crystallography and Biophysics, University of Madras, Chennai, India

Significance and Impact of the study: The human microbiota is emerging as an influential key player engaged in the onset and progression of pathogenic diseases. The post-genomic era evidently unravels the little known indispensable microbiota profiles of the lung infections and their influence on the gut. Integration of machine learning concepts with the genomics complements successful and crucial evaluation of serious infectious disease treatment and prevention options. The current study of microbial profiles of gut–lung axis when extended to the brain, skin, oral, liver, faecal, vaginal and secretion fluids would offer incredible information of well-established microbial interplay involved in health and wellness of humans.

Keywords

diversity, gut–lung axis, interplay, microbiota, random forest classifier.

Correspondence

S. Aishwarya, Department of Bioinformatics, Stella Maris College (Autonomous), Chennai 600086, India.

E-mail: aishwarya@stellamariscollege.edu.in

2022/LAMICRO-2022-0246.R1: received 2 May 2022, revised 3 July 2022 and accepted 26 July 2022

doi:10.1111/lam.13798

Abstract

Respiratory infections are the leading causes of mortality and the current pandemic COVID-19 is one such trauma that imposed catastrophic devastation to the health and economy of the world. Unravelling the correlations and interplay of the human microbiota in the gut–lung axis would offer incredible solutions to the underlying mystery of the disease progression. The study compared the microbiota profiles of six samples namely healthy gut, healthy lung, COVID-19 infected gut, COVID-19 infected lungs, *Clostridium difficile* infected gut and community-acquired pneumonia infected lungs. The metagenome data sets were processed, normalized, classified and the rarefaction curves were plotted. The microbial biomarkers for COVID-19 infections were identified as the abundance of *Candida* and *Escherichia* in lungs with *Ruminococcus* in the gut. *Candida* and *Staphylococcus* could play a vital role as putative prognostic biomarkers of community-acquired pneumonia whereas abundance of *Faecalibacterium* and *Clostridium* is associated with the *C. difficile* infections in gut. A machine learning random forest classifier applied to the data sets efficiently classified the biomarkers. The study offers an extensive and incredible understanding of the existence of gut–lung axis during dysbiosis of two anatomically different organs.

Introduction

Microbiome research in the current scenario has explored the understanding of the structure, functional and immunological benefits of the microbial communities in human health and disease (Enaud and Prevel 2020). The microbiome composed of bacteria, fungi, virus, archaea and phages is found to be involved in the human

pathologies and disruption of homeostasis (Ning and Yang 2020). Under normal conditions of healthy individuals, a balanced homeostasis, crosstalk and cross-regulation of the microbiome is established. They contribute to various health benefits, offer a commensal relationship to the host and avoid the overgrowth of potentially harmful microbes. A disturbance in the structure and function of the microbiome over prolonged

years would result in dysbiosis (Kim and Womble 2021). Multitude of factors including the diet, intake of antibiotics, chronic inflammations, stress, environmental factors, immune suppression etc. influences dysbiosis which in turn can pose a threat to the distant organs as well Li *et al.* 2021.

Respiratory disorders that affect the lungs and airways throw a global burden resulting in one-sixth of all mortality across the world. The “Big five” respiratory disorders constituting chronic obstructive pulmonary disorders (COPD), asthma, tuberculosis, acute respiratory infections and lung cancer are more lethal (Kogan *et al.* 2019). Lungs have been a home to abundant noncultivable bacteria that are disrupted during the respiratory disorders that have been reported in COPD (Wang and Bafadhel 2016), asthma (Loverdos and Bellos 2019), lung cancer (Liu and Ma 2020), respiratory viral infections (Hanada and Pirzadeh 2018), bronchiectasis (Richardson and Dicker 2019), pneumonia (Hong *et al.* 2021) and cystic fibrosis (Faner and Sibila 2017). The lungs were believed to be sterile for a long time until recently emerged culture independent metagenomics studies reported the coexistence of microbes in the lung (Hilty and Burke 2010). Off late, the little available knowledge of the lung microbiome is being enhanced and evidence suggests the presence of Firmicutes, Bacteroidetes and Proteobacteria in the lower respiratory tract (Faner and Sibila 2017). Several pieces of evidence in recent years have confirmed the existence of diverse, both known and novel microbial communities in various human organs including skin, mouth, gastrointestinal tract, lung, faeces, urine, vagina and nasopharyngeal tracts. Although lung microbiome and their functions are being tapped recently, there is still a significant lag in comparison to the gut microbiome studies.

The two anatomically distinct organs—the gut and the lung exhibit complex and potential communication through their microbiome that reinforces the evidence of the gut–lung axis (GLA) crosstalk. There has been proven evidence of gut dysbiosis in case of respiratory infections and reverse is possible too where the dysbiosis of the gut possibly influences the respiratory infections. The dysbiosis of gut and the role of gut microbes in various ailments have been extensively studied but the inter organ crosstalk such as GLA, gut–brain–lung axis (GBLA) and gut–skin axis (GSA) of the microbiome have not been studied well and should not be overlooked (DeGruttola *et al.* 2016). There is an intelligible interplay of gut–lung microbes that plays a vital role in maintaining homeostasis and immunomodulatory regulations. Hence, there is a pressing need to understand the effect of pathologies in both gut and lung simultaneously in cases of respiratory infections. In the current study, we aimed at inferring the

microbial biomarkers for both lung and gut dysbiosis due to the pathologies—COVID-19, community-acquired pneumonia and *Clostridium difficile* infections. The study is first of its kind where the microbiome of lung and gut has been extensively investigated to unravel the gut–lung intermodulation during lung and gut ailments both individually and collectively.

Results and discussion

Taxonomic classification of the microbiota

Meta-analysis of metagenomic data sets was performed from multiple projects with integration of samples under six categories CG (Control gut), CL (Control Lung), CVIG (COVID-19 infected gut), CVIL (COVID-19 infected lungs), CDIG (*C. difficile* infected gut) and CAPL (community-acquired pneumonia infected lung). The data sets individually were processed, quality checked and the adapters were trimmed. The data sets of the same sample category were pooled after a second quality check and the total number of classified reads was identified using MG-RAST and QIIME2. The data sets were pooled with respect to sample categories due to two reasons: (1) some of the data sets were quite smaller with few samples and (2) very few data sets were available under some categories in the public access database. The pooled in data of sample categories CG, CL, CVIG, CVIL, CDIG and CAPL resulted in an average classification of 74·076, 98·11, 100, 99·1, 49·95 and 69·12% reads respectively.

Phylum level classification of the gut and lung microbiome samples

Phylum level classification in the three gut sample categories—CG, CVIG and CDIG as shown in Table S1 revealed two phylum Bacteroidetes and Firmicutes as predominant and the other phylum were negligible in all three samples. The percentage of Bacteroidetes and Firmicutes in CG was 55·37 and 34·58% which showed significant difference in CVIG and CDIG. In comparison with CG, the Bacteroidetes were higher with 72·33% and Firmicutes were lesser with 23·60% in CVIG, whereas lesser occurrence of Bacteroidetes with 37·41% and higher occurrence of Firmicutes with 56·35%. Increase in the occurrence of Fusobacteria was observed in CDIG which was not seen in the other two.

Phylum classification of microbiota in the lung categories such as CL, CVIL and CDIL as shown in Table S2, the occurrence of Actinobacteria, Firmicutes, Fusobacteria and Bacteroidetes was seen. Higher percentages of Firmicutes (41·59%) were seen in CL along with 26·13% of Fusobacteria and 16·01% of Actinobacteria. In contrast,

CVIL had a significantly higher percentage of Proteobacteria (22.31%) which was very meagre (0.029%) in CL and 11.34% in CAPL. Similar significant enrichment of Ascomycota was seen in CVIL and 35.97% in CAPL which was not at all observed in the control lung. A drastic reduction of 15.22% of Firmicutes was seen in CVIL in contrast to CL and CAPL which showed 41.59 and 37.91% respectively. A very significant deterioration of Fusobacteria was seen in CVIL and CAPL while a dominant percentage of 26.13% was seen in CL. In lung infections like COVID-19 and community-acquired pneumonia, significant enrichment of Ascomycota and Proteobacteria was seen which were not present in the control lung while reduced occurrence of Fusobacteria and Firmicutes was observed. Hence, overexpression of Ascomycota and Proteobacteria, and reduction of Fusobacteria and Firmicutes can be putative prognostic microbial biomarkers in cases of COVID-19 and pneumonia.

Relative abundance of gut and lung microbiome at the genus level

Relative abundance of CG, CVIG and CDIG at the genus levels had a total number of 474215, 1440701 and 309107 classified reads respectively. The *Prevotella* and *Bacteroides* were predominant in CG with 36.80 and 35.80% respectively. A drastic reduction of *Prevotella* was seen in both CVIG and CDIG with 1.1 and 5.1%, respectively, while the occurrence of *Bacteroides* was reduced in CDIG with 21.31% with a more or less similar percentage in CVIG with 29.65%. A prominent increase in the presence of *Ruminococcus* (12.79%) was seen in CVIG against 3.10% in CG and 1.13% in CDIG. A rise in the growth of *Faecalibacterium* (9.83%) and *Enterococcus* (9.20%) was also seen in CVIG in contrast to respective 2.68 and 1.15% in CG. A significant rise in the presence of *Faecalibacterium* with 23.90% and *Clostridium* 29.79% was seen in CDIG in contrast to respective 6.7 and 2.60% in CG. Hence, *Prevotella* and *Bacteroides* are characteristic microbes of control gut with their reduction in COVID-19 infected gut with abundance of *Ruminococcus*. Characteristic microbes in CDI was increased levels of *Faecalibacterium* and *Clostridium* with the reduction in *Prevotella* and *Bacteroides*. The relative abundance of the gut microbiota CG, CVIG and CDIG at the genus level classification can be seen in Fig. 1a.

A total of 687 738, 426 970 and 413 309 reads were classified at the genus level for the lung microbiota of CL, CVIL and CAPL. The relative abundance of lung microbiota at the genus level for the three sample categories CL, CVIL and CAPL was shown in Fig. 1b. One predominant control lung microbiota was *Leptotrichia* with 62.2% or classified reads followed by two genus *Megasphaera*

with 12.84% and with 16.26% abundance. In a strange contrast, the CVIL had three abundant *Candida*, *Acinetobacter* and *Corynebacterium* with 35.37, 14.10 and 12.10%, respectively, against their absence in the healthy lung with the absence of *Leptotrichia*. CAPL exhibited abundance of two genera *Candida* and *Staphylococcus* with 38.78 and 28.50%, respectively, along with noted absence of *Leptotrichia*. Hence, *Leptotrichia* was found to be characteristic microbiota of control lung with *Candida*, *Acinetobacter* and *Corynebacterium* of COVID-19 infected lungs while *Candida* and *Staphylococcus* in case of pneumonia infected lungs. The absence of *Leptotrichia* was a striking feature of the infected lungs.

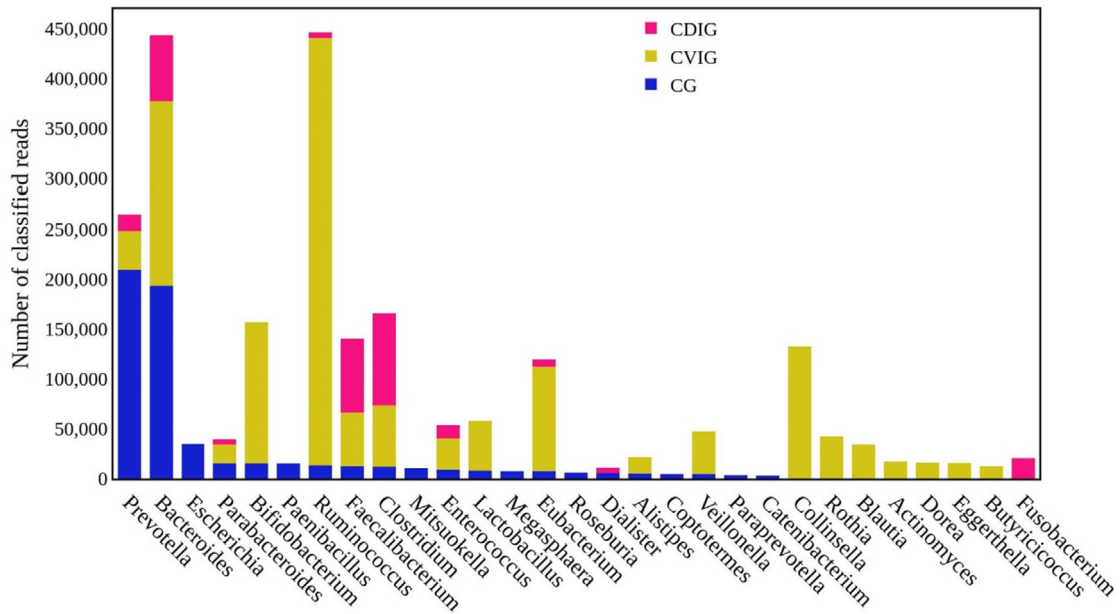
Relative abundance of species in gut–lung crosstalk

Species level identification of microbiome is restricted to 16S rRNA amplicon sequencing, yet a small percentage of predominant species were inferred from the classified reads. Total normalized species count accounted for 79, 80, 65, 63, 42 and 43% in CG, CVIG, CDIG, CL, CVIL and CAPL respectively. The abundant species that were observed in the CG were *Prevotella copri* (28%), *Bacteroides stercoris* (16%), *Chryseobacterium oncorhynchi* (5.2%), *Bacteroides vulgatus* (14.1%) and *Paraprevotella capra* (3.2%), while CL had predominantly *Leptotrichia goodfellowii* (21%), *Streptococcus pneumoniae* (12%), *Leptotrichia* sp. Oral taxon 212 (15%), *Klebsiella pneumoniae* (7%), *Acinetobacter baumannii* (6%) and *Lautropia mirabilis* (5.3%). The relatively abundant species of CVIG were found to be *Chryseobacterium viscerum* (9.2%), *S. pneumoniae* (5.2%), *Chryseobacterium* sp. (6%), *Prevotellamassillia timonensis* (5.1%), *Escherichia coli* (18.1%), *Faecalibacterium prausnitzii* (5.4%), *Staphylococcus aureus* (9.2%) and *Bacteroides stercoris* (4.3%) whereas CVIL exhibited abundance of *E. coli* (24%) also called an intestinal microbe. *Streptococcus pneumoniae* (8.2%), *K. pneumoniae* (4.7%), *Salmonella enterica* (5.6%), *Neisseria meningitidis* (3.3%) and *Dialister succinatiphillus* (2.1%) were the successively abundant microbiota of COVID-19 infected lung. CDIG samples showed abundance in *Ruthenibacterium lactatiformans* (18%), *Parabacteroides distasonis* (13%), *E. coli* (5.2%), *Bacteroides dorei* (8.2%), *C. difficile* (9.9%) and *Clostridium symbiosum* (3.9%) whereas in cases of CAPL, *E. coli* (14%), *S. aureus* (12%), *S. pneumoniae* (5.7%) and *K. pneumoniae* (5.8%) were seen. Figure 2a shows the comparison of normalized species across all the six samples.

Gut–lung axis microbial crosstalk

In the identification of microbial biomarkers across the categories of samples, top 41, 40, 45, 34, 45 and 46

(a) Relative abundance at genus level - gut microbiome



(a) Relative Abundance at genus level - Lung microbiome

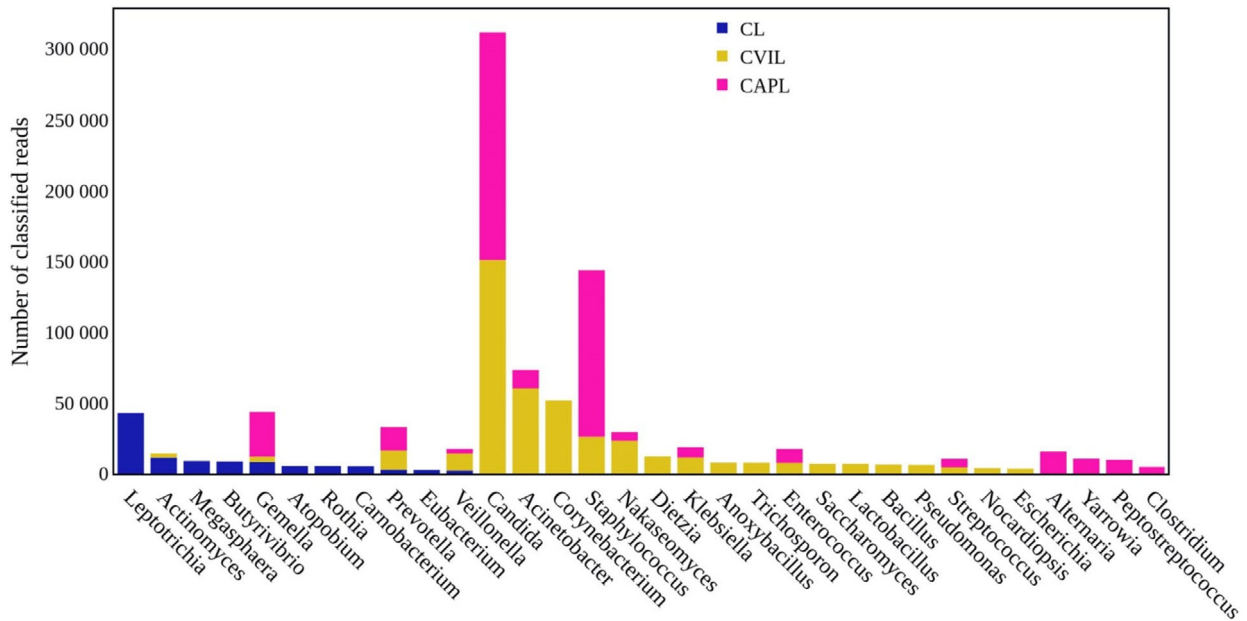
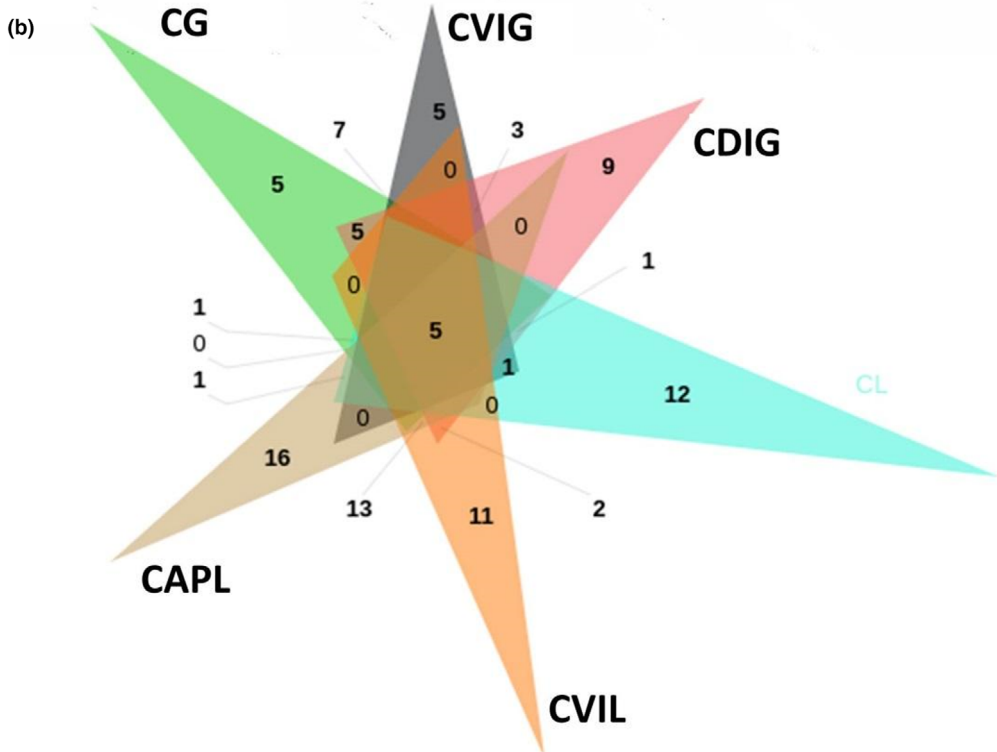
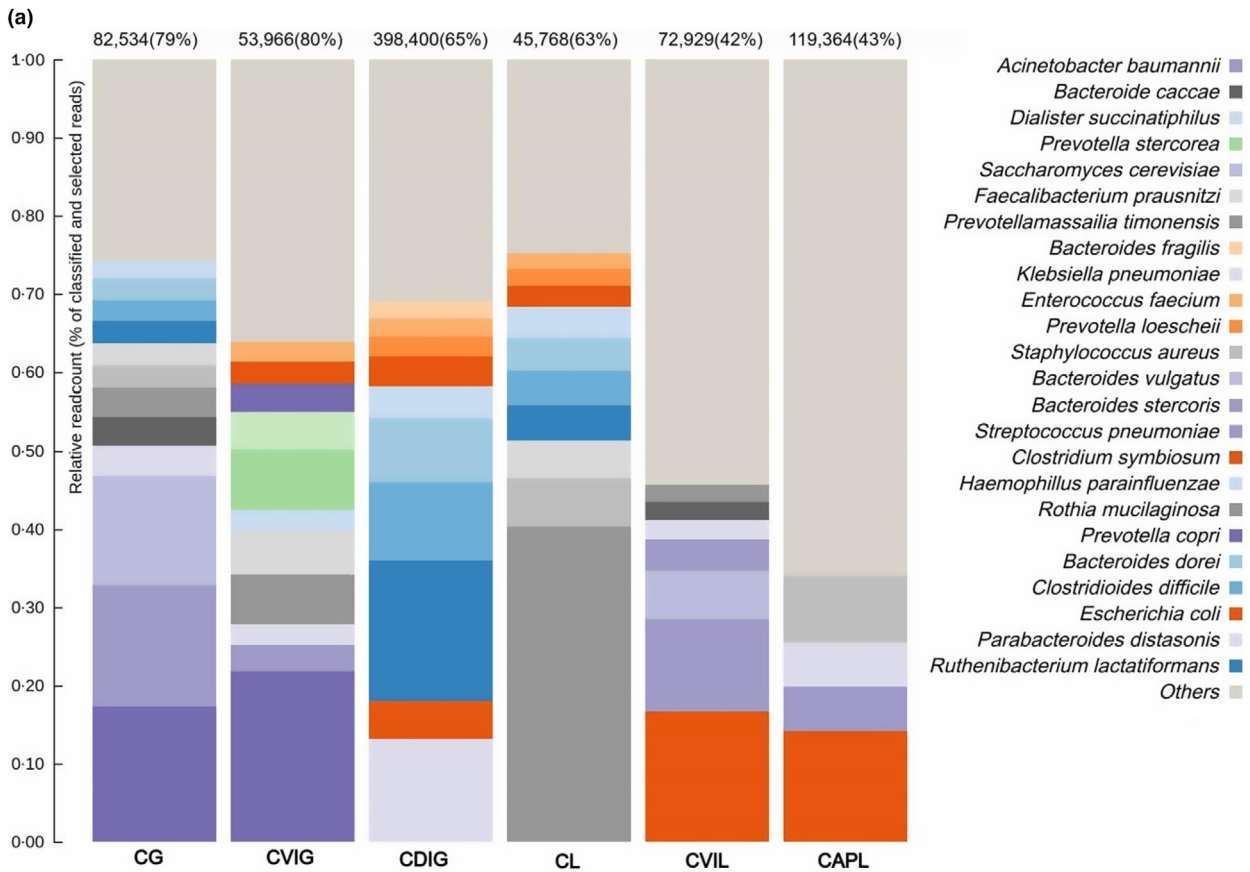


Figure 1 (a) Relative abundance of genus in the gut microbiome in three categories—control gut (CG in blue), COVID-19 infected gut (CVIG in yellow) and *Clostridium difficile* infected gut (CDIG in pink). (b) Relative abundance of genus in the lung microbiome in three categories—control lung (CL in blue), COVID-19 infected lung (CVIL in yellow) and community-associated pneumonia infected lung (CAPL in pink).

Figure 2 (a) Relative abundance of species in all six sample categories—CG, CVIG, CDIG, CL, CVIL and CAPL. (b) Five mutual microbiota occurring in all six categories of samples. A total of 251 microbial genus were compared that included 41 from CG, 40 from CVIG, 45 from CDIG, 34 from CL, 45 from CVIL and 46 from CAPL.



abundant microbes in categories CG, CVIG, CDIG, CL, CVIL and CAPL, respectively, were filtered (>20%) and a comparison of mutual microbes as shown in Fig. 2b, was performed to infer the presence of *Prevotella*, *Enterococcus*, *Veillonella*, *Corynebacterium*, *Actinomyces* in all six samples. In a healthy gut and lung axis, the organisms such as *Prevotella*, *Bacteroides*, *Rothia*, *Enterococcus*, *Clostridium*, *Actinomyces*, *Alkaliphilus*, *Corynebacterium*, *Eubacterium*, *Bacillus*, *Megasphaera*, *Bifidobacterium* and *Veillonella* were possibly common. Microbes that were present in CVIG and CVIL were found to be *Enterococcus*, *Streptococcus*, *Oribacterium*, *Veillonella*, *Gemella*, *Actinomyces*, *Corynebacterium*, *Bacteroides*, *Ruminococcus*, *Atopobium*, *Fusobacterium*, *Eubacterium*, *Lactobacillus*, *Prevotella*. Although *Prevotella* and *Bacteroides* were less abundant, they still occur during the COVID-19 infections in both gut and lung. *Enterococcus*, *Streptococcus*, *Oribacterium*, *Gemella*, *Ruminococcus*, *Atopobium*, *Fusobacterium*, *Eubacterium* and *Lactobacillus* occurrence were noted only during the COVID-19 infections and can be considered putative microbial biomarkers. In cases of lung infections both in COVID-19 and pneumonia, *Candida* and *Escherichia* were predominant followed by *Pseudomonas*, *Staphylococcus*, *Acinetobacter*, *Stenotrophomonas*, *Enterobacter*, *Klebsiella* and *Nakaseomyces*. Significant occurrences of *Candida* and *Escherichia* might be probable diagnostic microbial biomarkers of lung infections. Occurrence of *Bifidobacterium*, *Veillonella*, *Gemella*, *Actinomyces*, *Corynebacterium*, *Enterococcus*, *Prevotella*, *Clostridium*, *Escherichia*, *Atopobium*, *Klebsiella*, *Porphyromonas* and *Lactobacillus* was possible in cases of CDI infections in gut and CAP infections in lung which elicits the role of gut–lung crosstalk of microbiome. COVID-19 and CDI infections together can elicit the occurrence of *Gemella*, *Actinomyces*, *Barnesiella*, *Corynebacterium*, *Fusobacterium*, *Eggerthella* and *Butyricoccus* in the gut. The list of overlapping microbes between the sample categories is listed in Table S3.

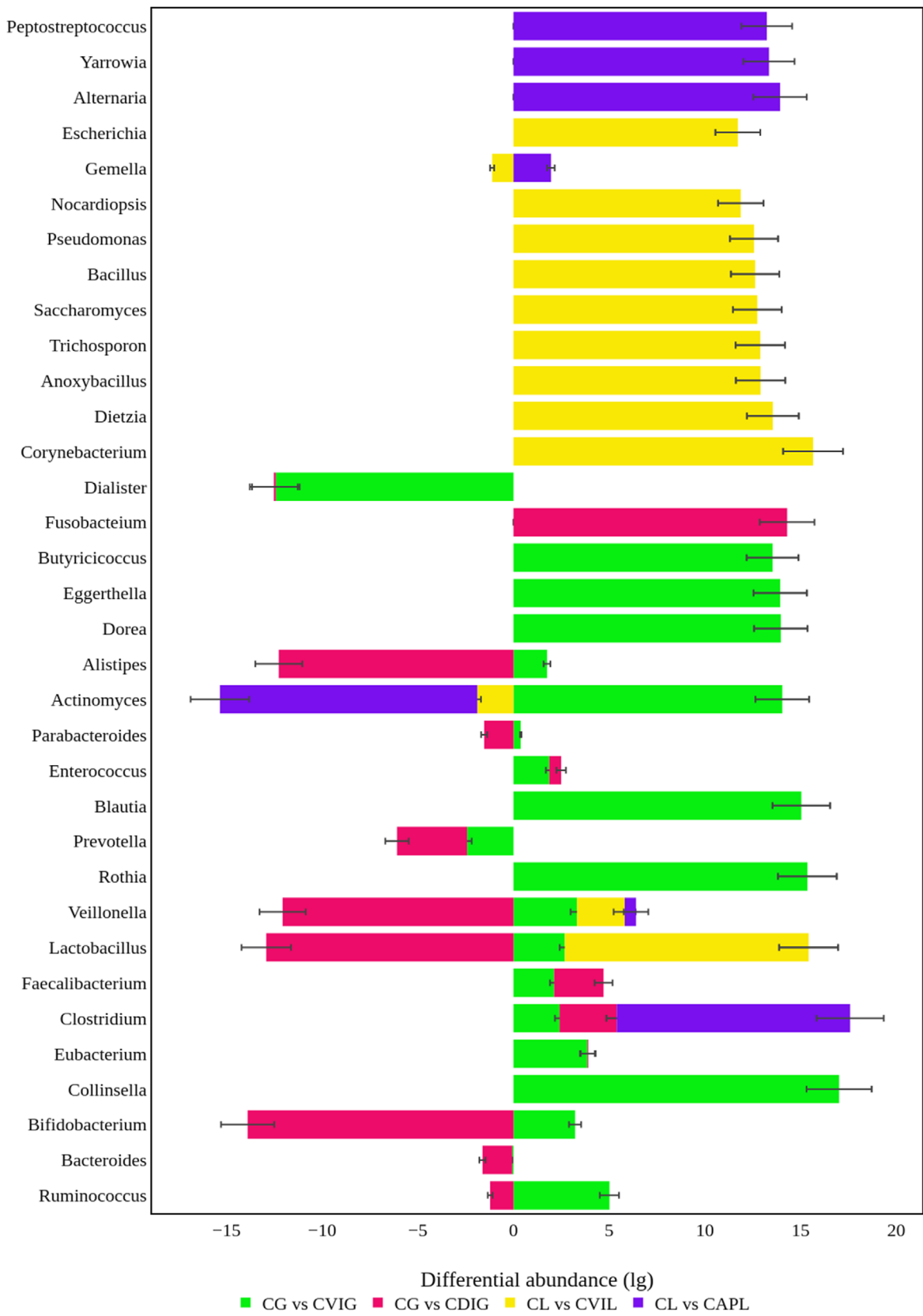
Differential abundance of microbiota in the gut–lung axis

From the mutually existing microbiota, differential abundances of microbiota at the genus level were identified from their log₂FC and $P < 0.05$ using the DESeq2 with logCPM normalization. There were 3, 5 and 3 significant differentially overlapping microbiota filtered for CL versus CG, CVIL versus CVIG and CAPL versus CDIG. The

differentially expressed overlapping microbiota of CL versus CG was *Prevotella*, *Veillonella* and *Eubacterium* with their respective log₂FC values of -6.28 , -0.99 and -1.54 . Occurrence of *Prevotella* and *Eubacterium* was abundant in the gut whereas they were reduced in lungs while occurrence of *Veillonella* is not very different. Comparison of CVIL versus CVIG revealed the overlapping occurrences of *Prevotella*, *Veillonella*, *Actinomyces*, *Enterococcus* and *Lactobacillus* with the log₂FC values of -1.51 , -1.82 , -2.49 , -2.06 and -2.86 respectively. *Actinomyces*, *Enterococcus* and *Lactobacillus* originally seen in the gut were now seen in COVID-19 infected lungs while the occurrence of *Prevotella* and *Veillonella* also increased in lungs. This presumes the movement of microbiota to lungs during infections. Comparison of CAPL and CDIG showed the abundance of *Prevotella*, *Clostridium* and *Enterococcus* with log₂FC ratio being 0.03 , -4.29 and -0.43 respectively. During CAP conditions, the lung acquires more *Prevotella* which is almost equal in abundance with the gut. *Enterococcus*, pathogenic bacteria of the gut, also shows an increase in CAPL. The overlapping differentially significant microbiota of the gut and lungs is shown in Fig. S1.

Corynebacterium, *Dietzia*, *Anoxybacillus*, *Trichosporon*, *Lactobacillus*, *Bacillus*, *Pseudomonas*, *Nocardiopsis* and *Escherichia* were significantly upregulated only during COVID-19 with significant fold changes of 15.65, 13.23, 12.91, 12.89, 12.74, 12.62, 12.51, 11.87 and 11.7 respectively. They are not seen in CL and CAPL. *Lactobacillus* and *Veillonella* were seen upregulated in CVIG than CG while in contrast, *Escherichia* a gut bacterium was not seen in CVIG but was upregulated in CVIL. *Paenibacillus*, *Mitsuokella*, *Paraprevotella* and *Catenibacterium* were completely lost in both CDIG and CVIG but were abundant only in CG. *Actinomyces*, *Alistipes*, *Dorea*, *Eggerthella*, *Collinsella* and *Butyricoccus* were abundant only in the CVIG but not in CDIG and CG. *Actinomyces*, *Eubacterium*, *Carnobacterium*, *Rothia*, *Atopobium*, *Butyrivibrio*, *Megasphaera* and *Leptotrichia* were abundant in CL and absent in CVIL and CAPL. *Candida*, *Acinetobacter* and *Enterococcus* were abundant in CVIL and CAPL. Differentially abundant microbiota of gut and lungs has been depicted in Fig. 3. Thus, upregulated microbial biomarker for COVID-19 infection in lungs are *Corynebacterium*, *Dietzia*, *Anoxybacillus*, *Trichosporon*, *Lactobacillus*, *Bacillus*, *Pseudomonas*, *Nocardiopsis* and *Escherichia* while in the gut were *Collinsella*, *Rothia*, *Blautia*, *Actinomyces*, *Dorea*,

Figure 3 Differentially abundant genus with $P < 0.05$ identified between lungs and gut. Green bars represent differential abundance of CVIG to CG, pink bars represent the differential microbial abundance of CDIG to CG, yellow bars represent differential microbial abundance of CVIL to CL and blue bars represent differential microbial abundance of CAPL to CL.



Eggerthella and *Butyricoccus*. The downregulated biomarkers of COVID-19 in gut were *Catenibacterium*, *Paraprevotella*, *Dialister*, *Mitsuokella*, *Paenibacillus* and *Escherichia* with significant fold changes of -11.53 , -11.79 , -12.42 , -13.29 , -13.85 and -15.06 respectively. Predominant dysbiosis of lungs and gut was observed during COVID-19 and there was drastic abundance of gut microbes in the lungs and vice versa seen both in lungs and gut. During *C. difficile* infections, the *Fusobacterium* ($\log_2FC = 13.22$) is the significantly upregulated microbial marker while *Bifidobacterium* ($\log_2FC = -13.29$), *Lactobacillus* ($\log_2FC = -12.06$) and *Veillonella* ($\log_2FC = -12.92$) were the most promising downregulated biomarkers. Finally during CAPL, significantly upregulated biomarkers were *Alternaria*, *Yarrowia*, *Peptostreptococcus*, *Clostridium* and *Gemella* with \log_2FC of 13.93 , 13.35 , 13.23 , 12.19 and 1.95 , respectively, and one downregulated biomarker was *Actinomyces* ($\log_2FC = -13.44$). The core microbiome of all the six categories was clustered based on average clustering and Euclidean distance matrix as seen in Fig. S2.

Diversity of microbiota—alpha diversity

The biodiversity of the microbiota across various sample categories was inferred from the rarefaction curves plot. The plot represents a varying number of individual OTUs identified with a depth of 2000 rarified reads. Alpha diversity is a metric that measures the rarefied samples specific to the richness and evenness of the species. Simpson index that ranges from 0 to 1 signifies species diversity, higher the range, diverse are the species. Accordingly as shown in Table 1, the values of all six samples were in the range of 0.79–0.87 that showed greater diversity of species. The Shannon index that measures evenness of the species was in the range of 1.5–3.5. For the current categories, the Shannon index ranged between 2.0 and 2.6 in all the six samples confirming the evenness of the sample. Observed species metric deciphers the number of unique OTUs in the sample. It inferred that a wide range of unique OTUs—69 for CG, 369 for CVIG, 64 for CDIG, 70 for CL, 839 for CVIL and 113 for CAPL. The total

number of abundant species occurring in the sample was 107 for CG, 242 for CVIG, 45 for CDIG, 49 for CL, 118 for CVIL and 47 for CAPL. Thus, based on the abundant species per the given sample size, a rarefaction curve for the six samples was plotted and shown in Fig. S3a. From the inferred values, there is a greater diversity, evenness and richness of species in all six sample categories.

Beta diversity

Beta diversity is the degree of difference in the microbial community in relation to the environment gradients or pattern of environments. The range of beta diversity metric lies between 0 and 1 that corresponds, respectively, to similarity and diversity among the samples. Vegan R package deduced the beta diversity by Bray–Curtis from the genera of samples and the values are shown in Table S4. A Euclidean space matrix was created that measured the pairwise distance between the samples' abundant OTUs. A pairwise dissimilarity matrix was calculated for the six samples and a nonlinear dimensionality reduction ordination was plotted with a principal coordinate analysis (PCoA) as shown in Fig. S3b. The ordination of the diseased lung samples is farther to the control lung, and hence, the microbiota is more diverse between each other in gut and lungs.

Machine learning and ROC curve

An additional machine learning strategy was applied to validate the OTUs using a supervised random forest classifier approach. The ROC curve is a measure of specificity and sensitivity of the microbial diversity. From the OTU table, the ROC curve was plotted for true positives (specificity) against false positives (sensitivity) on X and Y axis respectively. As seen in the Fig. S4, AUC was 0.9028 and 0.9381 for CG versus CVIG and CG versus CDIG respectively. The AUC scores for CL versus CVIL and CL versus CAPL, respectively, were 0.9653 and 0.9382. *Bacteroides* and *Prevotella* were associated with a healthy gut and *Leptotrichia* with a healthy lung while signifying that the predicted model was reliable and accurate. *Ruminococcus*, *Collinsella*,

Table 1 Alpha diversity of the microbiota

Categories	Simpson Index (0–1)	Shannon Index (1.5–3.5)	Observed unique species	Number of abundant species
CG	0.875119543891432	2.60045768532565	69	107
CVIG	0.876555953648081	2.50674071849006	369	242
CDIG	0.808983186590101	2.00752693166077	64	45
CL	0.812674010654916	2.24456436175542	70	49
CVIL	0.855636744533066	2.69800679243509	839	118
CAPL	0.791916215737391	2.22549046479948	113	47

Rothia and *Blautia* were associated with COVID-19 gut while *Corynebacterium*, *Lactobacillus* and *Escherichia* with COVID-19 lungs. *Fusobacterium* and *Candida*, *Clostridium* were associated with *Clostridium* infections and community-acquired pneumonia respectively.

Discussion

The synergies of host–microbiota interactions have been widely explored and understanding of the symbiotic relationships has brought various underlying pathological mechanisms to the limelight. Microbiome, be it in the gut, skin, lung or any other organ has a tremendous role to play in the maintenance of homeostasis, immune modulations, gene expressions and regulation of pathological imbalance in the host (Anand and Mande 2018). The current study probed a meta-analysis approach to understand the interplay of gut–lung microbiota in healthy and pathological states by analysing the microbiome of healthy gut, healthy lung, COVID-19 infected lung, COVID-19 infected gut, community-acquired pneumonia and *C. difficile* infected gut. The publicly available amplicon metagenome data sets from NCBI-SRA were retrieved and processed to infer the abundant OTUs in both normal and diseased states under six categories as CG, CVIG, CDIG, CL, CVIL and CAPL.

The abundant phylum classified at CG was Bacteroidetes and Firmicutes. In contrast, CVIG had a higher percentage of Bacteroidetes and lesser abundant Firmicutes while CDIG showed lesser percentage of Bacteroidetes and more abundant Firmicutes as compared to CG. This is an iconic feature that reveals the state of gut dysbiosis during respiratory infections like COVID-19. Fusobacteria abundance was only observed in CDI gut and not in other two states. A simultaneous investigation of lung dysbiosis revealed a significant abundance of Proteobacteria in COVID-19 infections which was strikingly lower in the control lung and can be a probable microbial biomarker phylum candidate of COVID-19. Similarly, highly abundant Ascomycota was also observed in both CVIL and CAPL which proves a characteristic feature of lung dysbiosis due to respiratory infections. A drastic reduction of Fusobacteria and Firmicutes was observed in both infections when compared to the healthy lung. Scarcity of Firmicutes in COVID-19 infections in both gut and lung might possibly indicate the existence of GLA crosstalk. Firmicutes were found to be generally abundant in both the healthy gut and lungs. The genus level classification of the samples resulted in more meaningful interpretations. Significant paucity of *Prevotella* was seen in both CVIG and CDIG when compared to the healthy gut while *Ruminococcus* was on a rise in CVIG along with *Faecalibacterium* and *Enterococcus* with possibilities of being

microbial biomarkers of COVID-19 infection in the gut. *Faecalibacterium* and *Clostridium* were enriched in CDI than the normal gut. Reduction of the *Prevotella* and *Bacteroides* in the gut dysbiosis was seen. *Leptotrichia* was identified as the characteristic microbe of healthy lungs which drastically vanished during COVID-19 and pneumonia infections. Abundance in *Candida*, *Acinetobacter* and *Corynebacterium* and occurrence of *Escherichia* in the CVIL were noted while abundant *Candida* and *Staphylococcus* in CAPL indicated the lung dysbiosis due to lung infections. Abundance of *Candida* and paucity of *Leptotrichia* seem to have significant chances of being considered a diagnostic microbial marker in cases of lung infections. The normal gut and lung exhibit a rather distinct microbial population which is greatly altered in the respiratory and gut infections. Species-level alteration was seen in CVIG where a noticeable percentage of *S. pneumoniae* was seen, which is responsible for causing bacterial pneumonia in lungs and leads to fatality.

CVIL had abundance of gut pathogens such as *Salmonella enterica*, known to cause typhoid salmonellosis (Andino and Hanning 2015), *K. pneumoniae*, known to cause liver abscess (Fung and Lin 2012) and *E. coli* known to cause gastroenteritis (Kittana and Gomes-Neto 2018). CVIL also showed the presence of *A. baumannii* (Kokkonouzis and Christou 2009) and *S. pneumoniae* that are known to cause pneumonia of the lung. *Escherichia coli* abundance was seen in CAPL samples as well. During the COVID-19 infections, the presence of gut pathogens in the lung and vice versa could possibly suggest the translocation or interplay of the GLA microbiome. *Prevotella*, *Enterococcus*, *Veillonella*, *Corynebacterium* and *Actinomyces* were found to exist overlapping in all the six sample categories.

In an attempt to investigate the overlapping abundant genus of the samples, top abundant microbes of each sample were compared. In the healthy gut and lung, the organisms such as *Prevotella*, *Bacteroides*, *Rothia*, *Enterococcus*, *Clostridium*, *Actinomyces*, *Alkaliphilus*, *Corynebacterium*, *Eubacterium*, *Bacillus*, *Megasphaera*, *Bifidobacterium* and *Veillonella* were common. During COVID-19 infections, *Enterococcus*, *Streptococcus*, *Oribacterium*, *Veillonella*, *Gemella*, *Actinomyces*, *Corynebacterium*, *Bacteroides*, *Ruminococcus*, *Atopobium*, *Fusobacterium*, *Eubacterium*, *Lactobacillus* and *Prevotella* were found common in both the gut and lungs. This is a clear indication of the dysbiosis happening in both organs and suggestive of the possible GLA axis communications. *Candida*, a common microbial fungus present in intestines in traces and *Escherichia*, intestinal bacteria are relatively abundant in the lungs during both COVID-19 and pneumonia infections thereby can strongly be considered as putative diagnostic microbial biomarkers of lung infections. Strikingly overlapping

occurrence of *Bifidobacterium*, *Veillonella*, *Gemella*, *Actinomyces*, *Corynebacterium*, *Enterococcus*, *Prevotella*, *Clostridium*, *Escherichia*, *Atopobium*, *Klebsiella*, *Porphyromonas* and *Lactobacillus* in CD infections and CAPL indicate the subsistence of GLA crosstalk. The presence of *Gemella*, *Actinomyces*, *Barnesiella*, *Corynebacterium*, *Fusobacterium*, *Eggerthella* and *Butyrivibrio* was observed in the gut during COVID-19 and CDI from which it is evident that COVID-19 induces gut dysbiosis thus proving the existence of GLA intermodulation.

The microbiota diversity, species richness and evenness metrics are significant in the microbiota comparison studies, and thus, alpha diversity within each sample and beta diversity between the samples were calculated. Alpha diversity of the individual samples was calculated based on the SHE (Species richness, Shannon diversity and Evenness) index (Anonye 2018). Simpson metrics for the six samples were in the range of 0.79–0.87 that lies within the accepted range of 0–1, indicating the greater diversity of microbial taxa in each sample. Shannon diversity index was measured between 2.0 and 2.6 that is suggestive of the evenness of the samples. Richness of the species in the samples were evident as measured by the number of observed unique OTUs were 69 for CG, 364 for CVIG, 64 for CDIG, 70 for CL, 839 for CVIL and 113 for CAPL. Thus, the microbiota of CG, CL, CVIG, CVIL, CDIG and CAPL was all significantly diverse, rich and even. The species diversity between the sample categories was revealed by the mathematical Euclidean similarity matrix and a successive Bray–Curtis algorithm that deciphered the presence of greater diversity of microbes in all four samples except the COVID-19 and CAP infected lung. Statistical significance of the OTU classification was performed with Wilcoxon paired *t*-test, before assigning them to the respective genus. A random classifier was generated using Python modules (Ning and Yang 2020) that resulted in an AUCs of 0.9028, 0.9381, 0.9653 and 0.9382 for CG versus CVIG, CG versus CDIG, CL versus CVIL and CL versus CAPL respectively. *Faecalibacterium*, *Lactobacillus* and *Bifidobacterium* being deterministic of healthy gut while *Leptotrichia*, *Megasphaera* and *Gemella*, of healthy lung. The ROC curve obtained indicated that the classified genus was more accurate and valid (Chen 2021). COVID-19 infection in lungs is associated with the modified microbial abundance in *Corynebacterium*, *Dietzia*, *Anoxybacillus*, *Trichosporon*, *Lactobacillus*, *Bacillus*, *Pseudomonas*, *Nocardiosis* and *Escherichia* while *Collinsella*, *Rothia*, *Blautia*, *Actinomyces*, *Dorea*, *Eggerthella* and *Butyrivibrio* in the gut. The downregulated biomarkers of COVID-19 were *Catenibacterium*, *Paraprevotella*, *Dialister*, *Mitsuokella*, *Paenibacillus* and *Escherichia* with significant fold changes of –11.53, –11.79, –12.42, –13.29, –13.85 and –15.06 respectively. Predominant dysbiosis

of lungs and gut was observed during COVID-19 and there were contrasting abundance and reduction of microbiota seen both in lungs and gut than their normal counterparts. During *C. difficile* infections, the *Fusobacterium* is the significantly enhanced microbial marker while *Bifidobacterium*, *Lactobacillus* and *Veillonella* were significantly diminished biomarkers. Promising abundance of *Alternaria*, *Yarrowia*, *Peptostreptococcus*, *Clostridium* and *Gemella* was associated with community-acquired pneumonia and *Actinomyces* was diminished than the control lung.

There are very few previous research that has confirmed the microbial biomarkers such as *Pseudomonas* spp., Enterobacteriaceae and *Acinetobacter* spp. in critically ill COVID-19 patients, whereas *Burkholderia*, *Chryseobacterium* and Enterobacteriaceae were biomarkers in deceased patients with diminished *Prevotella* spp., *Neisseria* spp., *Veillonella* spp. and *Streptococcus* spp. (Gaibani and Viciani 2021). In a comparative metabolome, microbiome analysis of COVID-19, performed by Liu and Liu (2021) *Megasphaera micronuciformis*, *Prevotella histicola*, *Streptococcus sanguinis*, *Veillonella dispar* and *Lautropia mirabilis* were found in abundance. In a study that analysed gut microbiome of COVID-19 patients revealed the abundance of *Bifidobacterium adolescentis*, *Eubacterium rectale* and *F. prausnitzii* (Patel and Roper 2021). Significantly abundant bacteria such as *Ruminococcus gnavus*, *Ruminococcus torques* and *B. dorei* were observed during antibiotic usage in COVID-19 patients and reduction of *B. adolescentis*, *F. prausnitzii* and *E. rectale* was noted during antibiotic intake (Yeoh *et al.* 2021). The current study identified the higher rates of *E. coli* (24%), an intestinal microbe in the lungs of COVID-19 samples and presume that might be nosocomially acquired either by aspiration or haematogenous dissemination from infected GI tract to cause *E.coli* pneumonia (Jonas 1982). An increase in *S. pneumoniae* in the lungs can elevate acute lung infection, a scariest reason for 1 million deaths every year and the increase is 12 lipoxigenase dependent. The third abundant *K. pneumoniae* was also observed to cause a nosocomial outbreak of pneumonia in lungs when there is gastrointestinal disturbance (Ashurst and Dawson 2022). *Salmonella enterica* the enteropathogen has been associated with gastroenteritis, bacteraemia, cholecystitis, endocarditis, meningitis etc. Its nosocomial presence in the lung causes pneumonia which is very rare and has been attributed to prior lung infection and impaired cellular immunity (Samonis and Maraki 2003). Abundance of *E. coli*, *S. pneumoniae* and *K. pneumoniae* was also seen in CAPL. Hence, it is possible that a GLA exists and whenever there is a lung infection, the possibilities of a gastrointestinal infection rates are higher. *Chryseobacterium indologenes* was identified in COVID-19 associated

bacteraemia with devastating co-harboured antimicrobial resistance genes blaND2, blaCIA and blaCcrA (Yeh and Li 2022). In CDIG, apart from *C. difficile* and *E. coli*, three other abundant bacteria—*R. lactatiformans*, *Parabacteroides distasonis* and *F. prausnitzii* were frequently associated with gut dysbiosis and severity of gastrointestinal pathologies (Schult and Reitmeier 2022).

The current study is one of the first few studies to evaluate the microbial biomarkers of lung and gut dysbiosis simultaneously and unravelled novel microbial biomarkers along with those already identified. The study throws light on the intermodulation of GLA and signifies that a respiratory infection influences the gut alteration and vice versa to disrupt the normal microbial balance both in lungs and gut. The study confirms the microbial biomarkers as *Candida* and *Escherichia* for respiratory infections in lung while *Ruminococcus* for gut dysbiosis due to COVID-19. Although the study is extensive and sheds crucial light on the GLA intermodulation, suffers few limitations such as (1) the absence of lung microbiome data during gut infections, limited the understanding of lung dysbiosis; (2) since the study is computational, the alterations of microbiome could not ascertain whether it was due to translocation; (3) 16s rRNA amplicon data, restricted the precise classification of abundant species and statistically significant functional annotation could not be achieved; and (4) the study warrants further clinical research to validate the results. Nevertheless, the choice of samples and the normalization of read counts before assigning the taxonomy to avoid sample bias had achieved an extensive insight on GLA interplay. This study strongly envisages the higher possibilities of GLA interplay existence during lung infections.

Materials and methods

Selection of data set

For the study, to compare the abundant microbiome in the gut–lung axis, we chose publicly available data sets from SRA that included—(1) gut metagenome data set of healthy control group; (2) gut metagenome of COVID-19-positive patients; (3) gut metagenome of *C. difficile* infected patients (Zhang *et al.* 2015); (4) lung metagenome data set of healthy control group (Ekanayake and Madegedara 2020); (5) lung metagenome of COVID-19 infected patients; and (6) lung metagenome of community-acquired pneumonia patients. The only criteria applied to include data sets were amplicon metagenome sequencing strategy. Since the data were very limited, we ended up taking all the available data sets with different platforms and layouts. The source for gut samples was stool and for lungs was nasopharyngeal swab/bronchoalveolar lavage fluid or sputum. The raw reads of metagenome data sets as specified in Table S5 were retrieved

from the NCBI-SRA database using sra-toolkit. The data sets that were chosen for the study differed in the number of samples; read lengths and number were inferred to be relatively negligible to impact the technical variation in contrast to the biological variation (Nayfach and Pollard 2016).

Data processing

All the data sets were processed individually during the initial validation. Quality check of the raw data was performed by trimming out the low quality (<30 phred score) and short read adapter sequences (<50 bp) using FASTQC package and TRIMMOMATIC tools respectively. Paired-end data sets were assembled into single-end reads using FLASH (Magoc and Salzberg 2011) and later integrated into six different classification such as control gut (CG), COVID-19 infected gut (CVIG), *C. difficile* infected gut (CDIG), control lung (CL), COVID-19 infected lung (CVIL) and community-acquired pneumonia infected lung (CAPL). Burrows-wheeler alignment tool depleted the human genome by aligning the sequence reads with human reference genome hg38 (Menzel *et al.* 2016). The unassembled data were processed using the Meta-Genome Rapid Annotation using Subsystems Technology server (MG-RAST) v3.0. Gene calling of the quality filtered reads was processed with 90% identity to ribosomal sequences and was clustered at 97% identity. The contigs with an average length of 350 bp were assembled using Megahit software (Li and Luo 2016).

Relative abundance and taxonomic classification

The scaffolds of each sample were further classified by discarding operational taxonomic units (OTU) <5 classified reads (Johnson and Spakowicz 2019). Taxonomic classification was obtained by clustering the OTU using BLAST with >90% identity and an *E*-value of 1e-5 and compared with Silva database (www.arb-silva.de). The reads were normalized for all the six samples with read count normalization. Relative abundance of the microbiota from all six samples computed from the least common ancestor algorithm (LCA) was plotted as stacked bar charts (Aishwarya and Gunasekaran 2021) at both genus and species level. Species-level classification was performed only for the OTUs with greater than 80% identity to single species mapped to the NCBI BLAST specific for 16srRNA.

Comparative analysis

Comparison of the microbial samples was conducted with Vegan R package version 3.6.1 (Dixon 2003). Species richness and evenness were calculated with alpha diversity indices such as Shannon, Simpson index and observed

taxa (Sato and Kakuta 2020). Alpha diversity measures were calculated based on a rarefied abundant data equivalent to 2000 reads/sample. Statistical *t*-tests were performed to ensure the significant differences between alpha diversity of the six categories. Beta diversity analysis estimates the difference of species between samples and PCoA was executed to estimate the dissimilarity of taxa between the samples based on Bray–Curtis distance which is a multidimensional scaling metric. It calculates the dissimilarity matrix between the groups of samples based on Euclidean distance and chi-squared distances (D’Argenio *et al.* 2014; Sato and Kakuta 2020). The significant difference was confined to a $P < 0.05$. Rarefaction curves represent the species richness between the samples and were plotted using ggplot2 of R package. Microbiota common to the gut–lung axis was identified from Venn plots using the Jvonn server (Bardou and Mariette 2014).

Statistical analysis and machine learning-based accuracy prediction

The differential abundance at the genus level was performed with the DESeq2 R package to infer log2FC and the significance of difference was limited to $p < 0.05$ corrected with multiple testing Benjamini Hochberg correction. The differences between the samples types were calculated with Wilcoxon two-tailed *t*-test (Older and Diesel 2017). The microbiota for the control versus diseased was classified with a random forest (RF) classifier of Scikit package of Python and a fivefold cross-validation was performed with the decision tree classifier module. OTUs that were rare and <20% of the samples were removed and the remaining OTUs were transformed with standardized techniques for each sample and together included in the feature selection. The RF model was constructed with training sets from 70% of the OTUs and remaining 30% was assigned as test sets. The training set was applied successfully to the test sets and a confusion matrix was created. After a fivefold cross-validation, a successful area under curve (AUC) for the receiver operating characteristic (ROC) curve comparable between gut metagenome and lung metagenomes was computed (Li and Chang 2020).

Acknowledgements

Authors express their sincere gratitude to the management of Stella Maris College, Chennai, for their continued support and the seed fund.

Conflict of Interest

Authors declare that there were no potential conflicts of interest in publishing the manuscript.

Author Contributions

S. Aishwarya involved in idea, conceptualization, writing and analysis. K. Gunasekaran involved in proof reading and supervision.

DATA AVAILABILITY STATEMENT

Data availability The data that were used in the study were retrieved from the public database NCBI-SRA.

References

- Aishwarya, S. and Gunasekaran, K. (2021) Structural, functional, resistome and pathogenicity profiling of the Cooum river. *Microb Pathog* **158**, 105048.
- Anand, S. and Mande, S.S. (2018) Diet, microbiota and gut-lung connection. *Front Microbiol* **9**, 2147.
- Andino, A. and Hanning, I. (2015) Salmonella enterica: survival, colonization, and virulence differences among serovars. *The Sci World J* **2015**, 520179.
- Anonye, B.O. (2018) Commentary: bacteriophage transfer during faecal microbiota transplantation in *Clostridium difficile* infection is associated with treatment outcome. *Front Cell Infect Microbiol* **8**, 104.
- Ashurst, J.V. and Dawson, A. (2022) *Klebsiella pneumoniae*. In *StatPearls*. Treasure Island, FL, StatPearls Publishing.
- Bardou, P. and Mariette, J. (2014) jvonn: an interactive Venn diagram viewer. *BMC Bioinform* **15**, 293.
- Chen, P. (2021) *Gut microbiota and pathogenesis of organ injury*. Singapore: Springer.
- D’Argenio, V., Casaburi, G., Precone, V. and Salvatore, F. (2014) Comparative metagenomic analysis of human gut microbiome composition using two different bioinformatic pipelines. *Biomed Res Int* **2014**, 1–10. <https://doi.org/10.1155/2014/325340>.
- DeGruttola, A.K., Low, D., Mizoguchi, A. and Mizoguchi, E. (2016) Current understanding of dysbiosis in disease in human and animal models. *Inflammatory Bowel Diseases* **22**, 1137–1150.
- Dixon, P. (2003) VEGAN, a package of R functions for community ecology. *J Veg Sci* **14**, 927–930.
- Ekanayake, A. and Madegedara, D. (2020) Respiratory bacterial microbiota and individual bacterial variability in lung cancer and bronchiectasis patients. *Indian J Microbiol* **60**, 196–205.
- Enaud, R. and Prevel, R. (2020) The gut-lung axis in health and respiratory diseases: a place for inter-organ and inter-kingdom crosstalks. *Front Cell Infect Microbiol* **10**, 9.
- Faner, R., Sibila O. (2017) The microbiome in respiratory medicine: current challenges and future perspectives. *Eur Clin Respir J*, **49**, 1602086.
- Fung, C.-P., Lin Y.T. (2012) *Klebsiella pneumoniae* in gastrointestinal tract and pyogenic liver abscess. *Emerg Infect Dis*, **18**, 1322–1325.

- Gaibani, P., Viciani E. (2021) The lower respiratory tract microbiome of critically ill patients with COVID-19. *Sci Rep*, **11**, 10103.
- Hanada, S. and Pirzadeh, M. (2018) Respiratory viral infection-induced microbiome alterations and secondary bacterial pneumonia. *Front Immunol* **9**, 2640.
- Hilty, M., Burke C. (2010) Disordered microbial communities in asthmatic airways. *PLoS One*, **5**, e8578.
- Hong, L., Chen, Y. and Ye, L. (2021) Characteristics of the lung microbiota in lower respiratory tract infections with and without history of pneumonia. *Bioengineered* **12**, 10480–10490.
- Johnson, J.S., Spakowicz D.J. (2019) Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. *Nat Commun*, **10**, 5029.
- Jonas, M. (1982) Bacteremic *Escherichia coli* pneumonia. *Arch Intern Med* **142**, 2157–2159.
- Kim, Y.J., Womble J.T. (2021) The gut/lung microbiome axis in obesity, asthma, and bariatric surgery: a literature review. *J Obes*, **29**, 636–644.
- Kittana, H., Gomes-Neto J.C. (2018) Commensal *Escherichia coli* strains can promote intestinal inflammation via differential interleukin-6 production. *Front Immunol* **9**, 2318.
- Kogan, G., Vidal-Ragout, M., Trzaska, D., Crisafulli, A., Berkouk, K., Hogan, S., Hanrahan, F., Laang, H., et al. (2019) European Union funding of research on airway diseases. In: *Implementing precision medicine in best practices of chronic airway diseases*, ed. Agache, I. and Hellings, P. pp. 167–176. Cambridge, MA: Academic press Science direct.
- Kokkonouzis, I. and Christou, I. (2009) Multiple lung abscesses due to acinetobacter infection: a case report. *Cases J* **2**, 9347.
- Li, Q., Chang Y. (2020) Implication of the gut microbiome composition of type 2 diabetic patients from northern China. *Sci Rep*, **10**, 5450.
- Li, D. and Luo, R. (2016) MEGAHIT v1.0: a fast and scalable metagenome assembler driven by advanced methodologies and community practices. *Methods* **102**, 3–11.
- Li, X. et al. (2021) The intestinal dysbiosis of mothers with gestational diabetes mellitus (GDM) and its impact on the gut microbiota of their newborns. *Can J Infect Dis Med Microbiol AMMI Canada* **2021**, 3044534.
- Liu, J., Liu S. (2021) Association between the nasopharyngeal microbiome and metabolome in patients with COVID-19. *Synth Syst Biotechnol* **6**, 135–143.
- Liu, N.-N., Ma Q. (2020) Microbiome dysbiosis in lung cancer: from composition to therapy. *NPJ Precis Oncol*, **4**, 33.
- Loverdos, K. and Bellos, G. (2019) Lung microbiome in asthma: current perspectives. *J Clin Med Res* **8**, 1967.
- Magoc, T. and Salzberg, S.L. (2011) FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **27**, 2957–2963.
- Menzel, P., Ng, K.L. and Krogh, A. (2016) Fast and sensitive taxonomic classification for metagenomics with kaiju. *Nat Commun* **7**, 11257.
- Nayfach, S. and Pollard, K.S. (2016) Toward accurate and quantitative comparative metagenomics. *Cell*, **166**, 1103–1116.
- Ning, Y. and Yang, G. (2020) Characteristics of the urinary microbiome from patients with gout: a prospective study. *Front Endocrinol* **11**, 272.
- Older, C.E., Diesel A. (2017) The feline skin microbiota: the bacteria inhabiting the skin of healthy and allergic cats. *PLoS One*, **12**, e0178555.
- Patel, P. and Roper, J. (2021) Gut microbiome composition is associated with COVID-19 disease severity. *Gastroenterology* **161**, 722–724.
- Richardson, H., Dicker A.J. (2019) The microbiome in bronchiectasis. *Eur Respir J Eur Res* **28**, 190048.
- Samonis, G. and Maraki, S. (2003) *Salmonella enterica* pneumonia in a patient with lung cancer. *J Clin Microbiol* **41**, 5820–5822.
- Sato, N., Kakuta M. (2020) Metagenomic analysis of bacterial species in tongue microbiome of current and never smokers. *NPJ Biofilms Microbiomes*, **6**, 11.
- Schult, D., Reitmeier S. (2022) Gut bacterial dysbiosis and instability is associated with the onset of complications and mortality in COVID-19. *Gut Microb*, **14**, 2031840.
- Wang, Z., Bafadhel M. (2016) Lung microbiome dynamics in COPD exacerbations. *Eur Resp J*, **47**, 1082–1092.
- Yeh, T.-K. and Li, Z.H. (2022) COVID-19 associated bacteremia with chryseobacterium indologenes co-harboring blaIND-2, blaCIA and blaCcrA. *Infect Drug Resist* **15**, 167–170.
- Yeoh, Y.K., Zuo T., Lui G.C.Y., Zhang F., Liu Q., Li A.Y.L., Chung A.C.K., Cheung C.P., et al. (2021) Gut microbiota composition reflects disease severity and dysfunctional immune responses in patients with COVID-19. *Gut*, **70**, 698–706.
- Zhang, L., Dong, D., Jiang, C., Li, Z., Wang, X. and Peng, Y. (2015) Insight into alteration of gut microbiota in *Clostridium difficile* infection and asymptomatic *C. difficile* colonization. *Anaerobe* **34**, 1–7.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1 Taxonomic classification at phylum level of the gut microbiome.

Table S2. Taxonomic classification at phylum level of the lung microbiome

Table S3. Presence of mutual microbes across the different categories of samples used in the study.

Table S4. Beta diversity calculation by Bray–Curtis algorithm

Table S5. Data sets retrieved from the SRA database for comparison

Figure S1. The significant overlapping microbiota seen in lungs when compared to the gut. They are significant with a $P < 0.05$.

Figure S2. Heat map of the significant core microbiome clustered according to their normalized abundance.

Figure S3. (A) Rarefaction plot depicting the alpha diversity of samples in CG (site 1), CVIG (site 2), CDIG (site 3), CL (site 4), CVIL (site 5) and CAPL (site 6). S 3B. Principal coordinate analysis plot that indicates the beta diversity of the samples.

Figure S4. Receiver operating characteristic curve plot obtained from random forest classifier of operational taxonomic unit classification of microbiota from six sample categories.