frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 20 October 2022
DOI 10.3389/fmicb.2022.979320

Check for updates

OPEN ACCESS

EDITED BY
Qi Zhao,
University of Science and Technology
Liaoning, China

REVIEWED BY
Joao Carlos Gomes-Neto,
University of Nebraska-Lincoln,
United States
Christophe Becavin,
Université Côte d'Azur, France

*CORRESPONDENCE
Anirban Bhar
anirban.bhar@uni-greifswald.de

SPECIALTY SECTION
This article was submitted to
Systems Microbiology,
a section of the journal
Frontiers in Microbiology

RECEIVED 27 June 2022
ACCEPTED 21 September 2022
PUBLISHED 20 October 2022

CITATION
Bhar A, Gierse LC, Meene A, Wang H, Karte C, Schwaiger T, Schröder C, Mettenleiter TC, Urich T, Riedel K and Kaderali L (2022) Application of a maximal-clique based community detection algorithm to gut microbiome data reveals driver microbes during influenza A virus infection. *Front. Microbiol.* 13:979320. doi: 10.3389/fmicb.2022.979320

COPYRIGHT
© 2022 Bhar, Gierse, Meene, Wang, Karte, Schwaiger, Schröder, Mettenleiter, Urich, Riedel and Kaderali. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Application of a maximal-clique based community detection algorithm to gut microbiome data reveals driver microbes during influenza A virus infection

Anirban Bhar[1]*, Laurin Christopher Gierse[2], Alexander Meene[2], Haitao Wang[2], Claudia Karte[3], Theresa Schwaiger[3], Charlotte Schröder[3], Thomas C. Mettenleiter[3], Tim Urich[2], Katharina Riedel[2] and Lars Kaderali[1]

[1]Institute of Bioinformatics, University Medicine Greifswald, Greifswald, Germany, [2]Institute of Microbiology, University of Greifswald, Greifswald, Germany, [3]Friedrich-Loeffler-Institut, Greifswald-Insel Riems, Greifswald, Germany

Influenza A Virus (IAV) infection followed by bacterial pneumonia often leads to hospitalization and death in individuals from high risk groups. Following infection, IAV triggers the process of viral RNA replication which in turn disrupts healthy gut microbial community, while the gut microbiota plays an instrumental role in protecting the host by evolving colonization resistance. Although the underlying mechanisms of IAV infection have been unraveled, the underlying complex mechanisms evolved by gut microbiota in order to induce host immune response following IAV infection remain evasive. In this work, we developed a novel Maximal-Clique based Community Detection algorithm for Weighted undirected Networks (MCCD-WN) and compared its performance with other existing algorithms using three sets of benchmark networks. Moreover, we applied our algorithm to gut microbiome data derived from fecal samples of both healthy and IAV-infected pigs over a sequence of time-points. The results we obtained from the real-life IAV dataset unveil the role of the microbial families *Ruminococcaceae*, *Lachnospiraceae*, *Spirochaetaceae* and *Prevotellaceae* in the gut microbiome of the IAV-infected cohort. Furthermore, the additional integration of metaproteomic data enabled not only the identification of microbial biomarkers, but also the elucidation of their functional roles in protecting the host following IAV infection. Our network analysis reveals a fast recovery of the infected cohort after the second IAV infection and provides insights into crucial roles of *Desulfovibrionaceae* and *Lactobacillaceae* families in combating Influenza A Virus infection. Source code of the community detection algorithm can be downloaded from https://github.com/AniBhar84/MCCD-WN.

KEYWORDS

16S rRNA gene sequencing, microbiome, metaproteome, influenza A virus infection, community detection

1. Introduction

Trillions of microorganisms are living inside of multicellular, living organisms (Elgamal et al., 2021). The bacteria in the gut microbiome play an instrumental role in food digestion, regulation of the immune system and protecting the host against infectious disease caused by pathogenic bacteria. The microbial community is formed by interacting microorganisms which act as either predators or symbionts. Symbionts interact with each other in order to gain benefits from each other whereas, predators grapple with each other for the same source of nutrition (Kuntal et al., 2019). Dysbiosis, caused by changes in the gut microbiome, may not only trigger autoimmune diseases such as obesity, hypertension, diabetes, allergic disorders such as food allergy (Lee et al., 2021), but can also lead to the development of colorectal, hepatocellular and breast cancers (Vernocchi et al., 2020).

Influenza A Virus (IAV) is known to cause acute upper respiratory tract infection in animals such as human, pig etc. Following IAV infection, the host immune system plays an instrumental role to counter the infection and prevent serious disease, whereas IAV tries to provoke viral replication by escaping from host's immune surveillance (Chen et al., 2018). Interaction between commensal microbiota and invading viruses often disrupts the healthy gut microbiota community and such disruptions facilitate increasing the transmission of viral infection (Li et al., 2019). In order to protect the host from viral infection, microbiota adopts complex mechanisms such as induction of host immune response (Khan et al., 2021). Previous studies showed the occurrences of co-infections with viral pathogens in a substantial number of patients with a respiratory tract disease (Stefanska et al., 2013; Hoefnagels et al., 2021; Pacheco et al., 2021; Pettigrew et al., 2021). The presence of certain diseases in mammals may lead to an alteration in microbial community and such alteration may have an impact on host-immune response. Although a number of studies have been carried out to unravel the underlying mechanisms of IAV infection (Yildiz et al., 2018; Kaul et al., 2020), the mechanisms through which microbiota evolves the induction of host immune response still remain unknown. Hence, it is necessary to gain a better understanding of the interaction between microbial groups and provide insights into the driver microbial families and their functions that facilitate the induction of the host immune system in order to mitigate IAV infection.

Due to the advancement of next-generation sequencing technology such as 16s ribosomal RNA (rRNA) sequencing, it has become feasible not only to profile hundreds of microorganisms from a single analysis, but also semi-quantify the relative abundance of microbiome members during the progression of infectious diseases. However, although 16s rRNA sequencing has been widely used to mine the microbiome, while it provides information about abundance of species, it gives only very limited information about the function of identified microbial communities (Cortes et al., 2019). This limitation is dealt with the use of shotgun metagenomics. Since, shotgun metagenomics is expensive, proteomics has turned up as an alternative approach. Metaproteomics facilitates the study of all proteins in microbiomes, facilitating an understanding of the functional consequences of the microbiome (Cortes et al., 2019). Hence, the integration of both 16s rRNA sequencing and metaproteomics data may be advantageous to unveil key microbiome families and their role in the progression of infectious diseases.

Given the complexity of the microbial interactions, gut microbiota can be modeled as a network where each node represents an Amplicon Sequence Variant (ASV) and an edge between each pair of nodes represents a predicted association between the corresponding ASVs. In the context of network analysis, community detection algorithms are often used to partition a set of nodes into a number of communities, where nodes belonging to each community are tightly connected with each other. Nodes in each community are supposed to have many within-community edges, but few between-community edges (Yang et al., 2016). Many community detection algorithms have been proposed over the last two decades (Newman, 2004; Derényi et al., 2005; Ahn et al., 2010; Horvath, 2011; Alvarez et al., 2015; Benson et al., 2016; Yang et al., 2016; Lu et al., 2018). These algorithms are broadly categorized into modularity optimization based approaches (Newman, 2006a; Reichardt and Bornholdt, 2006; Blondel et al., 2008), clique based methods (Derényi et al., 2005; Benson et al., 2016; Lu et al., 2018), minimum-cut based methods (Newman, 2004) and hierarchical clustering based approaches (Ahn et al., 2010; Horvath, 2011; Alvarez et al., 2015).

Clique-based community detection methods have achieved augmented attention over the last decades (Derényi et al., 2005; Benson et al., 2016; Lu et al., 2018). One such algorithm is the k-clique percolation method (Derényi et al., 2005) which starts with finding all the k-cliques in a graph and subsequently merges two k-cliques having $k - 1$ edges in common between them. A single giant clique can be produced if the value of k is small, whereas a larger value of k yields multiple small communities. Hence, the selection of k affects the outcome of the algorithm. Another algorithm (Benson et al., 2016) based on network motifs was proposed to detect communities in a given graph. Moreover, this algorithm uses the size of the network motifs to build the co-occurrence count matrix in which each value is set to the number of motifs containing the corresponding pair of nodes. Thus, this algorithm ignores the edge weights and the effects of partitions on the co-occurrence count matrix. Recently (Lu et al., 2018), proposed another clique based community detection algorithm which constructs a clique-conductance matrix accumulating both the size and number of maximal cliques containing a pair of

vertices. Although this algorithm overcomes the limitations of the algorithm proposed by Benson et al. (2016), it ignores the edge weights. In addition, most of the existing clique-based algorithms (Derényi et al., 2005; Benson et al., 2016; Lu et al., 2018) use only the local properties of nodes to compute the similarity matrix and ignore the global importance of the nodes. In this paper, we proposed a novel Maximal-Clique based Community Detection algorithm for Weighted Network (MCCD-WN) in order to partition an undirected weighted graph by leveraging the information contained in all the maximal cliques in the given graph and taking both the local and global influence of the nodes into consideration. Moreover, maximal-cliques have biological importance in microbial co-occurrence network. Connectivity between the nodes tends to be higher in infected microbial network compared to healthy microbial network due to colonization activity during infection. Such variability in the node connectivity can be observed in maximal-clique structure.

The identified communities can further be used to detect different types hubs such as kinless, provincial and connector hubs. Theoretically the connector and provincial hubs are defined as the nodes which are highly connected between the modules and within a module, respectively. The kinless hubs are highly connected not only between the modules, but also within a module and hence, the kinless hubs are functionally more important to conciliate stabilization in the functional networks in the gut microbiome (Mangangcha et al., 2020; Shi et al., 2020). In this work, we applied our MCCD-WN algorithm to a time-series 16s rRNA sequencing dataset containing the abundance of ASVs in healthy and IAV infected cohorts in order to (i) identify driver ASVs and their corresponding microbial families at each time-point, (ii) find out different types of hub ASVs in the co-abundance network and their corresponding families at each time-point and (iii) provide insights into the functionalities of the microbiome families of the identified driver and hub ASVs at each time-point for both the healthy and infected states by integrating a metaproteome data.

2. Materials and methods

2.1. Materials

2.1.1. Artificial datasets

In order to evaluate the performance of the proposed community detection algorithm MCCD-WN, we used the LFR benchmark (Lancichinetti et al., 2008; Lancichinetti and Fortunato, 2009), which is a generalization of Girvan-Newman benchmark (Girvan and Newman, 2002) and assumes a power-law degree distribution of the nodes, similar to biological networks (Jing et al., 2021). The LFR benchmark uses a parameter called mixing parameter (μ_w) which is delineated as the ratio of the external degree of a node to the total degree of the node. Hence, a higher value of μ_w disrupts the community structure as each node will have a higher external degree than within-community degree. In this work, we generated three different sets of networks, AN_1, AN_2, and AN_3 using the LFR benchmark with 500, 3000 and 5000 nodes, respectively. Each of these three sets of networks contains a set of disjoint communities. For AN_1 and AN_3, we set the parameters k (average degree) to 25, *beta* (exponent for the weight distribution) to 1.1, *minc* (minimum for the community sizes) to 10, and *maxc* (maximum for the community sizes) to 50. In case of AN_2, we set the parameters k (average degree) to 15, *beta* (exponent for the weight distribution) to 1.1, *minc* (minimum for the community sizes) to 20, and *maxc* (maximum for the community sizes) to 50. The values of *maxk* (maximum degree) for AN_1, AN_2 and AN_3 benchmark networks are set to 50, 300, and 500, respectively. The value of the mixing parameter for the edge weights (μ_w) varies from 0.1 to 0.8 for each set of the networks.

2.1.2. Real-life datasets

The real data we use is from an influenza infection experiment carried out in pigs at the Department of Experimental Animal Facilities and Biorisk Management of the Friedrich Loeffler Institute within the H1N1pdm09 animal experiment (Schwaiger et al., 2019; Gierse et al., 2020, 2021). H1N1 infection was carried out in 19 pigs, three additional animals were used as controls. Fecal samples of all animals in the healthy cohort were used subsequently to generate 16s rRNA gene sequences, while samples of 19, 16, 12, and eight animals in the infected cohort were taken at days 0, 7, 21, and 25, respectively. Sequences obtained from 16s rRNA sequencing experiment of the infected cohort are available at European Nucleotide Archive (ENA) with project number PRJEB42450 and accession number ERP126308 whereas, the sequences for the healthy cohort can be downloaded from European Nucleotide Archive (ENA) using project number PRJEB39963 and accession number ERP123542. Fecal samples of all pigs from the healthy cohort and three pigs from the infected cohort were used for metaproteomic analysis. The mass spectrometry proteomics data can be obtained from ProteomeXchange Consortium using the dataset identifier PXD020775. Abundances of the microbial proteins measured across the same set of time-points i.e., days 0, 7, 21, and 25, were used in this work. Days 0 and 21 are reported to be the days of first and second IAV infection, respectively. The purpose of the second infection was to trigger the host immune response and explicate the crosstalk between host immune system and gut microbiome. Moreover, it is reported that the presence of influenza A virus matrix protein was observed after the first infection, whereas it was not detected after the second infection, implying a fast recovery of the infected animals (Gierse et al., 2021).

2.2. Methods

Figure 1 shows a schematic diagram of the overall workflow used in this work. Initially, co-abundance networks were built for both the healthy and infected cohorts. Subsequently, the co-abundance networks were analyzed to extract the driver ASVs. In addition, the co-abundance networks were used to find communities which were thereafter used to provide insights into the role of microbial families in subduing pathogen colonization and identify ASVs acting as different types of hubs. Finally, we leveraged the metaproteomics data in order to provide insights into functional role of microbial families of the identified driver and hub ASVs.

2.2.1. Co-abundance network construction

In order to build a reliable co-abundance network, amplicon sequence variants (ASVs) present in at least 30% of all the samples and having an average relative abundance value $> 0.01\%$ were used. SparCC, a method designed to estimate the correlation values for compositional data (Friedman and Alm, 2012), was then used to compute the correlation coefficient between each pair of ASVs. The significance of the correlation coefficients was established using randomly permuted data and a p-value cutoff < 0.05. In this work, we built one network at each time-point for both the healthy and infected cohorts.

2.2.2. Identification of key-drivers

After constructing the co-abundance network at each time-point, we considered ASVs present in both the healthy and infected networks in order to obtain a comparable set of nodes for performing a comparative analysis on both the networks. The properties of the co-abundance networks were computed using the R package *igraph* (Csardi and Nepusz, 2006). The topological properties of each of the co-abundance networks are shown in Figure 2A. For each of the time-points, a lower Jaccard Edge Index (JEI) score between the healthy and infected co-abundance networks implies the occurrence of rewiring. Moreover, it is of interest to observe that the infected network has a higher edge density and lower average distance as compared to the healthy network at each time-point. This tendency might be an indication of colonization activity in the infection setting (Kuntal et al., 2019). In order to identify driver ASVs at each time-point, we used the neighborhood shift (netshift) score (Kuntal et al., 2019). The netshift score of a node i is computed as

$$nesh(i) = 1 - (T1 - (T2 + T3)), \tag{1}$$

where

$$T1 = \frac{[\Lambda_i]^H \cap [\Lambda_i]^I}{[\Lambda_i]^H \cup [\Lambda_i]^I}, \tag{2}$$

$$T2 = \frac{[\Lambda_i]^I - [\Lambda_i]^H}{max\ degree\ in\ I}, \tag{3}$$

$$T3 = \frac{[\Lambda_i]^I - [\Lambda_i]^H}{[\Lambda_i]^H \cup [\Lambda_i]^I}, \tag{4}$$

and $[\Lambda_i]^I$ (or $[\Lambda_i]^H$) represents the set of direct neighbors of node i in the infected (or healthy) network. A node having a higher *nesh* score is considered to have a higher neighborhood shift in the infected network over a healthy network. Since the netshift score does not account for the edge weights between a node and its interacting partners, we used the betweenness centrality score and a Shannon entropy based node importance metric to estimate the importance of each node in the healthy and infected networks. The betweenness centrality score (Kuntal et al., 2019) of the ith node in the healthy (or infected) network is normalized as

$$BW_{scaled}(i) = \frac{BW(i) - BW_{min}}{BW_{max} - BW_{min}}, \tag{5}$$

where $BW(i)$ represents the betweenness value of node i in the healthy (or infected) network, and BW_{min} and BW_{max} denote the minimum and maximum values of all betweenness centrality scores in the healthy (or infected) network, respectively. Subsequently, the relative importance of the ith node is computed using equation-6

$$\triangle BW(i) = (BW_{scaled_I}(i) - BW_{scaled_H}(i)), \tag{6}$$

where $BW_{scaled_H}(i)$ and $BW_{scaled_I}(i)$ denote the normalized betweenness centrality scores of node i in the healthy and infected networks, respectively. A node having a positive $\triangle BW$ value has more importance in the infected network than in the healthy network. In addition to the betweenness centrality score, we used a Shannon entropy based metric to estimate the node importance based on its strength in the healthy and infected network. The Shannon entropy of a node i (Li et al., 2015) can be computed as

$$\mathbb{H}(i) = -\sum_{j=1}^{n} p_{ij} log(p_{ij}), \tag{7}$$

where $p_{ij} = \frac{w_{ij}}{\sum_{j=1}^{n} w_{ij}}$; w_{ij} denotes the weight between nodes i and j. Subsequently, the entropy of a network is computed as

$$\mathbb{H}(V) = -\sum_{i=1}^{n}\sum_{j=1}^{n} p_{ij} log(p_{ij}), \tag{8}$$

where V represents the set of n number of nodes in the network. In order to compute the importance of a node i in the network, we removed the ith node along with its edges from the network and then computed the entropy ($\mathbb{H}(V-i)$) of the resultant graph (Omar and Plapper, 2020). Subsequently, the normalized node

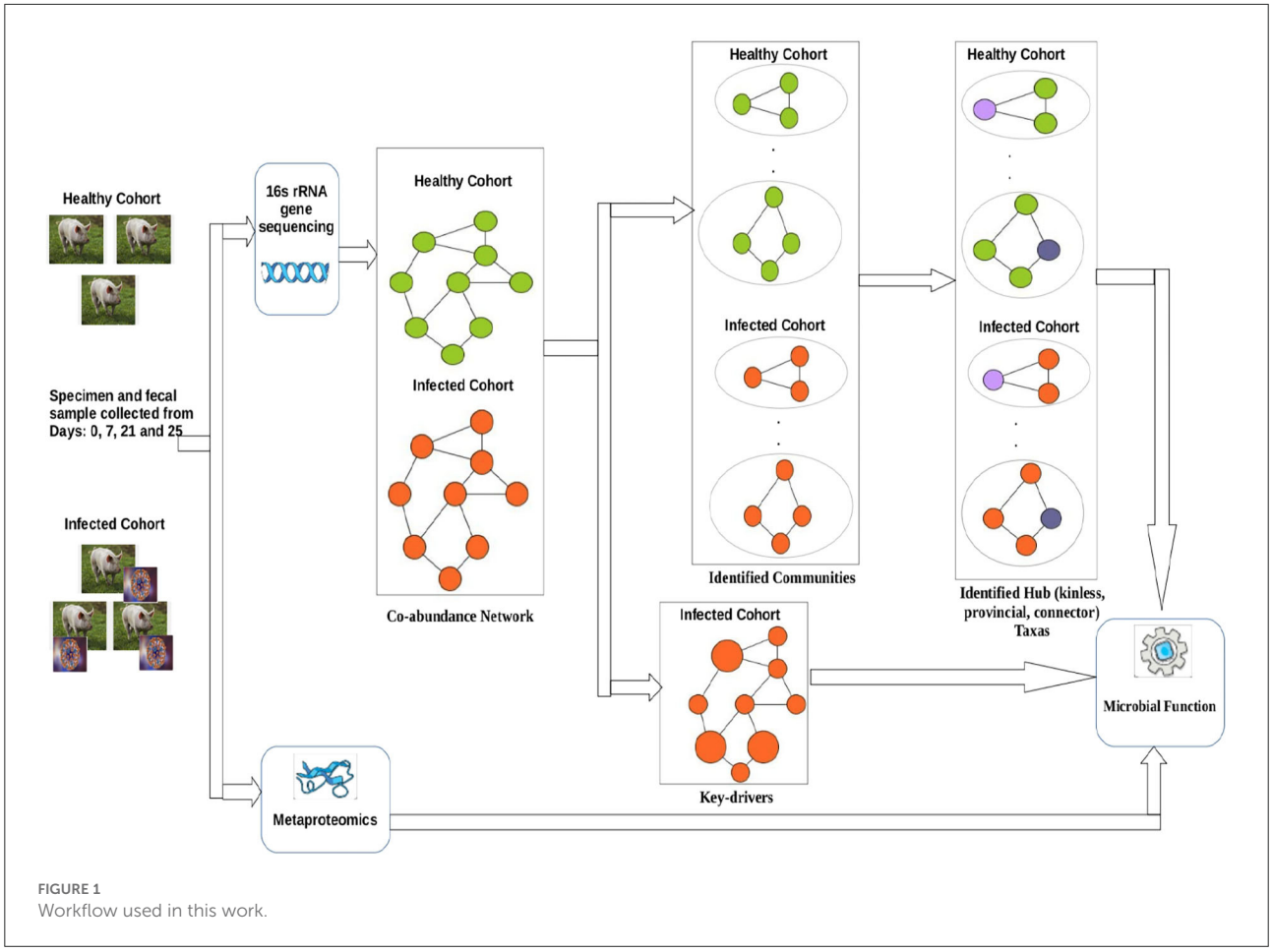

FIGURE 1
Workflow used in this work.

importance score is computed using equation 9 so that $\mathbb{H}(V)$ is always scaled to 1,

$$imp_{scaled}(i) = \frac{\mathbb{H}(V-i) - min_{\forall i \in V}(\mathbb{H}(V-i))}{\mathbb{H}(V) - min_{\forall i \in V}(\mathbb{H}(V-i)}. \quad (9)$$

Finally, the relative importance of the ith node based on the Shannon entropy based metric is computed as

$$\triangle imp(i) = (imp_{scaled_I}(i) - imp_{scaled_H}(i)), \quad (10)$$

where $imp_{scaled_I}(i)$ and $imp_{scaled_H}(i)$ denote the importance of the ith node in the infected and healthy networks, respectively. An ASV i is considered to be a driver ASV if it has a higher neighborhood-shift score and positive $\triangle BW(i)$ and $\triangle imp(i)$ values.

2.2.3. Identification of modules of bacterial ASVs

To extract modules from the co-abundance network, we applied a novel community detection algorithm MCCD-WN which leverages all the maximal cliques of a given network. An undirected weighted graph is defined as $G = \{V, E, W\}$, where V corresponds to a set of vertices, E corresponds to a set of edges and W is a symmetric weight matrix of size $\|V \times V\|$ where each element $W[i, j]$ represents the weight between each pair of connected vertices i and j. In graph theory, a clique is defined as a subgraph where every pair of nodes is connected with each other. A maximal clique is a clique that can not be expanded by adding any other vertex which is not a part of it. A k-clique is a clique having k number of vertices. In this work, we leveraged all the maximal k-cliques to find a set of modules in a given network by ensuring that there are very few cliques having lower density between different modules, whereas the number of cliques having higher density within the same module is relatively high. To achieve our goal, we first constructed a similarity matrix by utilizing both the global importance and local information of the nodes. In order to compute the global importance of the ith node, we used the PageRank algorithm (Page et al., 1999). PageRank method aims at ranking a set of web pages by assigning a weight based on the behavior of a random surfer which is similar to a Markov chain process. The estimated weight of a node represents the probability that a random surfer visits a web page either by following a hyperlink or directly by

A

Day	Stage	No. of Nodes	No. of Edges	Jaccard Edge Index	Edge Density	Average Distance
0	Healthy	89	153	0.03252033	0.07814096	3.447419
0	Infected	89	228		0.1164454	3.072268
7	Healthy	103	154	0.01327434	0.05863316	4.67613
7	Infected	103	304		0.1157434	3.053113
21	Healthy	128	311	0.02146465	0.07652559	3.587895
21	Infected	128	498		0.1225394	2.793799
25	Healthy	160	530	0.02903501	0.08333333	3.239408
25	Infected	160	675		0.1061321	3.110142

B

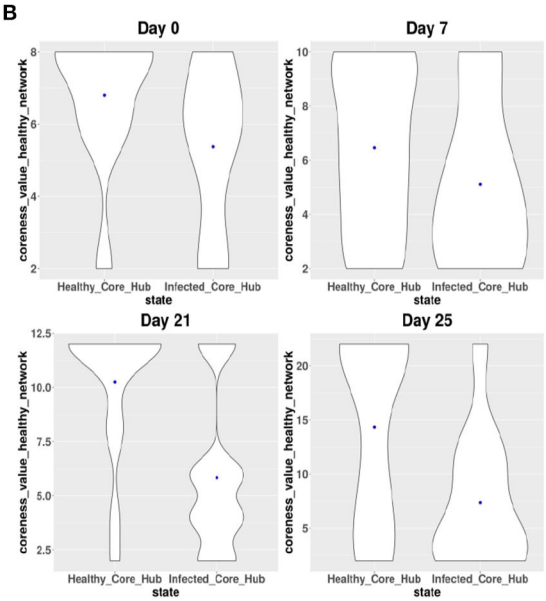

FIGURE 2
(A) Topological properties of healthy and infected co-abundance networks at each of the time-points. **(B)** Coreness values of the ASVs belonging to the infected-core-hub community in the control network and the ASVs belonging to the control-core-hub community.

entering the address of the page in the web browser. Thus a pagerank score captures the influence of a node on every other vertex in a network and can be used as a global importance score of a node in a given network. The pagerank score is computed by equation 11

$$imp_global_i = \beta \sum_{j \in \Lambda_i} \frac{imp_global_j}{deg_j} + \frac{1-\beta}{N}, \tag{11}$$

where Λ_i denotes the direct neighbors of the $i$the node, deg_j corresponds to the degree of node j and N denotes the total number of vertices in the given graph. In the PageRank algorithm, $\beta \in \{0, 1\}$ is used to denote the probability that a random surfer will continue to follow the hyperlink structure and is usually set to 0.85. Subsequently, we used Equation (12) to compute each element of the first structural similarity matrix Sim_1, based on the global importance values of the nodes,

$$Sim_1[i,j] = \frac{\sum_{x \in (\Lambda_i \cap \Lambda_j)} imp_global_x}{\sqrt{\sum_{x \in \Lambda_i} imp_global_x}\sqrt{\sum_{x \in \Lambda_j} imp_global_x}}. \tag{12}$$

The weight matrix W is used as the second similarity matrix Sim_2 as denoted by Equation (13):

$$Sim_2[i,j] = W[i,j]. \tag{13}$$

Finally, Equation (14) was used to compute the final similarity matrix,

$$Sim[i,j] = (\alpha * Sim_1[i,j]) + ((1-\alpha) * Sim_2[i,j]), \tag{14}$$

where the value of α is set experimentally. In the context of spectral clustering, the problem of optimizing the conductance function is considered as computationally intractable and can be converted into a relaxed, tractable eigenvector problem in order to obtain an approximate solutions. In the context of an undirected weighted graph partitioning problem, the conductance function is given by

$$\phi(C_1, C_2, .., C_K) = \frac{cut(C_i, \overline{C_i})}{volume(C_i)}, \tag{15}$$

where $\{C_1, C_2, .., C_K\}$ are the set of communities such that $(C_i \cap C_j) = \emptyset$, $\forall (i,j)_{i \neq j} \in \{1, 2, .., K\}$ and $\bigcup_i C_i = V$. In spectral clustering, we aim at minimizing the ratio $\phi(C_1, C_2, .., C_K)$. Moreover, the normalized cut ensures maximization of the intra-cluster similarity as long as the $volume(C)$ is maximized and the cut with the remaining vertices is minimized.

Now, we define the clique conductance function based on the information obtained from all maximal k-cliques in order to partition the given undirected weighted graph. The weight matrix for a clique-induced graph is computed using Equation (16),

$$W_{clq}[i,j] = \sum_{l \in Clq_k} \frac{\sum_{(i,j) \in l} Sim[i,j]}{\|l\| * (\|l\| - 1)}, \tag{16}$$

where Clq_k denotes the set of all maximal k-cliques, $\|l\|$ is the size of lth maximal clique and $W_{clq}[i,j]$ represents how tightly two vertices i and j are connected in the lth maximal k-clique.

In the context of a clique conductance function optimization problem, $cut(C, \overline{C}) = \sum_{i \in C, j \in \overline{C}} W_{clq}[i,j]$ and $volume(C) = \sum_{i \in C, j \in V} W_{clq}[i,j]$. Although the identification of all maximal cliques is known to be an NP-hard problem, retrieval of all maximal cliques is not problematic because the co-abundance network of bacterial ASVs is sparse. Algorithm 1 demonstrates the steps of our method to partition a weighted undirected graph. For a binary clustering problem, we divided the nodes into two groups based on the sign of values in the eigenvector corresponding to the second smallest eigenvalue of the normalized laplacian matrix L_{clq} of the clique-induced graph. In order to find more than two clusters, we have used the K-median algorithm as it is less sensitive to outliers. The problem of computing cluster centroids is converted into a L_1-median computing problem and solved using the algorithm proposed by Vardi and Zhang (2000). The L_1-median computing algorithm proposed by Vardi et al., an improved version of the Weiszfeld algorithm (Weiszfeld and Plastria, 2009), is capable of dealing with the situation where the median is found to be one of the data points in the cluster. In order to obtain better performance and reproducible clusters, we used Algorithm 2 which is similar to the K-means++ algorithm (Arthur and Vassilvitskii, 2007) for initial centroids selection. Algorithm 2 starts with finding a seed node as the one having the lowest betweenness centrality score. The intuition behind the selection of the seed node is that the seed node is unlikely to act as a bridge node between different communities.

Require: Undirected weighted graph $G = \{V, E, W\}$, where $W \in \mathbb{R}^{n \times n}$ and K, the number of communities.
Ensure: A set of disjoint communities $\{C_1, C_2, .., C_K\}$
  Step I. Extract all maximal cliques using the algorithm proposed by Eppstein et al. (2010)
  Step II. Compute the weight matrix W_{clq}, for the clique induced graph using equation 16.
  Step III. Compute the degree matrix D_{clq} from W_{clq} and normalized laplacian matrix L_{clq} using $D_{clq}^{-\frac{1}{2}}(D_{clq} - W_{clq})D_{clq}^{-\frac{1}{2}}$.
  if $K = 2$ **then**
    Step IVa. Compute the second eigenvector $E_{clq} \in \mathbb{R}^n$ of L_{clq}
    Step IVb. Assign ith object to community C_1 if the ith element of E_{clq} is greater than 0; otherwise, assign ith object to community C_2.
  else
    Step Va. Compute the first K eigenvectors $M_{clq} \in \mathbb{R}^{n \times K}$ of L_{clq}
    Step Vb. Compute a matrix $T_{clq} \in \mathbb{R}^{n \times K}$ by normalizing the rows of matrix M_{clq} to norm 1 by setting $T_{clq}[i,j] = \frac{M_{clq}[i,j]}{(\sum_{k=1}^{K} M_{clq}[i,k]^2)^{\frac{1}{2}}}$
    Step Vc. Select K number of initial centroids using Algorithm 2.
    Step Vd. Apply K-median clustering algorithm to T_{clq} to get K number of disjoint communities $\{C_1, C_2, .., C_K\}$
  end if

Algorithm 1. Proposed weighted graph partitioning algorithm.

2.2.4. Estimating the number of communities

In order to estimate the number of communities, we used an iterative approach shown in Figure 3. We applied our algorithm to extract different number (K) of communities and computed the modularity score (Q) (Newman, 2006b) for every value of K using Equation (17),

$$Q = \frac{1}{2m} \sum_{i,j \in V} (W[i,j] - \frac{deg_i deg_j}{2m}) \delta(C_i, C_j), \qquad (17)$$

where deg_i and deg_j denote the degree of nodes i and j. $\delta(C_i, C_j)$ is set to 1 if both the nodes belong to the same community, otherwise $\delta(C_i, C_j)$ is set to 0. $W[i,j]$ represents the weight of the edge between nodes i and j. m is the sum of all the edge weights in the network. Finally, the number of communities is set to the value of K which results in the maximum modularity (Q) score.

Require: Undirected weighted graph $G = \{V, E, W\}$, where $W \in \mathbb{R}^{n \times n}$, normalized matrix $T_{clq} \in \mathbb{R}^{n \times K}$ and K, the number of communities.
Ensure: K number of initial centroids.
  Step I. $S \leftarrow \emptyset$
  Step II. Compute the betweenness centrality score of each vertex in the graph G and select the vertex (p) having the lowest betweenness centrality score as the initial centroid.
  Step III. $S \leftarrow (S \cup p)$
  while $\|S\| < K$ **do**
    Step IVa. For each data point $q \notin S$, compute the distance ($dist(q)$) between q and the nearest point in S.
    Step IVb. Select a new point $q \notin S$ having the highest probability value $\frac{dist(q)^2}{\sum_{q \notin S} dist(q)^2}$, as the new center and add it to S ($S \leftarrow (S \cup q)$)
  end while

Algorithm 2. Initial centroids selection algorithm.

2.2.5. Identification of ASVs preventing pathogen colonization

In order to provide insights into the commensal bacteria playing a crucial role in conquering pathogen colonization,

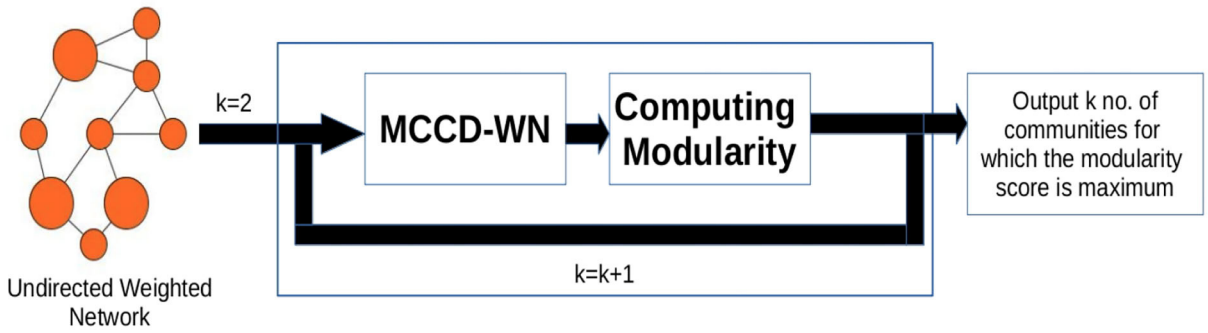

FIGURE 3
Workflow for estimating the number of communities. *k* denotes the number of communities.

we identified the core-hub community in both the healthy and infected networks as the one having the highest coreness value. Subsequently, we computed the coreness values of the ASVs belonging to the infected-core-hub community in the control network and compared with the coreness values of the ASVs belonging to the healthy-core-hub community. Figure 2B shows that the ASVs in the infected-core-hub community have lower coreness values in the healthy network as compared to the healthy-core-hub community members at each time-point. This phenomenon may imply the prevention of pathogen colonization by the corresponding microbial families of the commensal bacterial ASVs.

2.2.6. Identification of hub ASVs

We computed the participation coefficient score (Hall et al., 2019) for each node using Equation (18),

$$PC_i = 1 - \sum_{k \in K} \frac{deg_i(k)}{deg_i}, \tag{18}$$

where K denotes the number of communities, $deg_i(k)$ denotes the degree of node i within community k. A higher value of PC_i represents that the node i has higher inter-module connections, relative to intra-module connections. In addition to the participation coefficient, we calculated the within-community z-score (Hall et al., 2019).

$$z_i = \frac{deg_i(C_i) - \bar{deg}(C_i)}{\sigma^{deg(C_i)}}, \tag{19}$$

where C_i is the community containing node i, $deg_i(C_i)$ denotes the intra-module degree of node i, $\bar{deg}(C_i)$ and $\sigma^{deg(C_i)}$ stand for the mean and standard deviation of intra-module degrees of the nodes belonging to community C_i, respectively. Finally, the nodes having a z_i score greater than 0 are considered to be hubs. Furthermore, a hub node having a PC_i score $\geq$ 0.75 is defined as a kinless hub, a hub node having a PC_i score $\geq$ 0.30 and $<$ 0.75 is considered as a connector hub and a hub node having a PC_i score $<$ 0.30 is defined as a provincial hub.

2.2.7. Prediction of microbial biomarkers and their functions associated with healthy and infected cohorts

Since very limited information about the function of microbial community is provided by 16s rRNA sequencing experiment, we leveraged a metaproteome data obtained from the same healthy and infected cohorts for unveiling the functional potential of microbiome family and the microbial proteins during IAV infection. In order to achieve this, we applied Linear discriminant analysis Effect Size (LEfSe) to the relative abundance of functional categories of the protein groups belonging to the microbial families identified as hubs, key-drivers and members of the infected core-hub community at each time-point (Segata et al., 2011). Functional categories having a $LDA > 2.0$ and $p\ value < 0.05$ are considered to be the ones which explain the difference between the healthy and infected classes. Finally, the LEfSe analysis was carried out using the relative abundance of protein groups belonging to the significantly associated functional categories in order to identify the microbial proteins which differentiate the healthy and infected classes.

3. Results

3.1. Results on the LFR benchmark datasets

We applied our MCCD-WN algorithm to three sets of LFR benchmark networks. Although the number of communities can be provided by the user, we estimated the number of communities along with the parameter α by optimizing the modularity score (Q). We initialized the value of α to 1 and

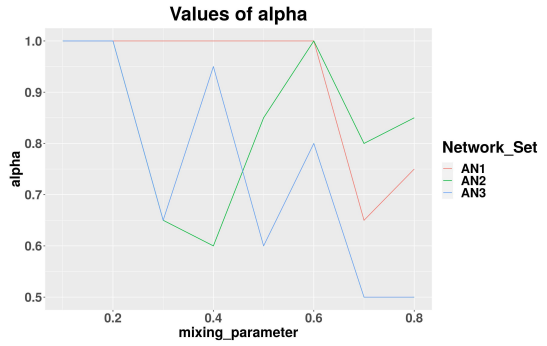

FIGURE 4
Different values of the parameter α for three sets of LFR benchmark networks AN_1, AN_2, and AN_3.

kept reducing its value by 0.05 in every iteration. Figure 4 shows the changes in the values of α for different values of the mixing parameter (μ_w), in case of each set of LFR benchmark networks. It is of interest to observe that in order to achieve the best performance of the MCCD-WN algorithm, the global node importance score is playing a crucial role for a relatively higher value of the mixing parameter μ_w. In order to evaluate the performance of the proposed MCCD-WN algorithm and compare its performance with the other existing algorithms, we have used the F_1 score (Laarhoven and Marchiori, 2016) as an evaluation metric, which can be defined by equation 20,

$$F_1(C, C^*) = 2 * \frac{|C \cap C^*|}{|C| + |C^*|}, \tag{20}$$

where C and C^* indicate the ground truth community and community produced by the community detection algorithm. The F_1 score denotes how well a community detection algorithm finds the ground-truth communities. F_1 ranges from 0 to 1, where a higher value indicates a better performance of the algorithm. Figure 5 demonstrates that the proposed algorithm obtained the highest F_1 score in case of the benchmark datasets AN_1 ($\mu_w = 0.7$ and 0.8), AN_2 ($\mu_w = 0.8$) and AN_3 ($\mu_w = 0.7$) whereas, it produces very similar F_1 scores as the other best performing algorithms for the remaining LFR benchmark networks.

In addition to F_1 score, we also used Modularity score (Q) as delineated in equation 17 for evaluating the performance of the proposed algorithm. Modularity score ranges from −1 to +1 and a higher modularity score indicates a better structure of the communities found in the given network. Figure 6 shows that the proposed algorithm achieved either equal or very close modularity scores to the ones obtained by the best performing algorithms for all the benchmark networks.

Moreover we carried out a study on the computing time for each of the benchmark networks. From Figure 7, it can be well seen that the proposed algorithm has reasonable computation speeds on each of the benchmark networks.

3.1.1. Similarity between real networks and LFR benchmark networks

In order to compare the real co-abundance networks obtained from Section 2.2.1 with the LFR benchmark networks, we used AN_1 *in silico* networks. Supplementary Figures 1, 2 compare AN_1 benchmark networks with the co-abundance networks in terms of the number of nodes, number of edges, edge density value and clustering coefficient. Moreover, we estimated the quality of the community structures present in the real networks by computing the absolute difference between the clustering coefficients of the real networks with that of each benchmark network. From Supplementary Figures 3, 4 it can be well seen that the real networks are comparable to the benchmark networks generated using a mixing parameter (μ_w) value less than or equal to 0.5.

3.2. Results on real-life data

3.2.1. Reproducibility testing of the proposed workflow

In order to test the reproducibility of the proposed workflow, we computed the size of intersection between the microbial families reported to be instrumental in Borey et al. (2021) and Gierse et al. (2021) and the ones found by applying our workflow to these two datasets. The sequences of the first dataset (Borey et al., 2021) were downloaded from the NCBI Sequence Read Archive using accession number PRJNA647267 and processed using SHAMAN tool (Volant et al., 2020) in order to obtain the OTU abundance matrix and annotation table. In case of the second experiment (Gierse et al., 2021), We used the sequence data obtained from the fecal samples of healthy and infected cohorts. Supplementary Figures 5A,B show that we achieved descent Jaccard similarity coefficient scores 0.70 and 0.66 for these two datasets, respectively.

3.2.2. Infected networks indicate a faster recovery after the second infection

The network analysis of gut microbiome data of IAV infected pigs revealed a strong negative correlation between the edge density values and modularity scores of the co-abundance network of infected cohort across all the time-points. Figure 8A exhibits the lowest modularity score and highest edge density value at day 21. This phenomenon indicates the maximum disruption of microbial communities and highest rise in the colonization activity triggered by the second IAV infection, whereas at day 25, a relatively lower edge density value and higher modularity score indicate a reduction in both the

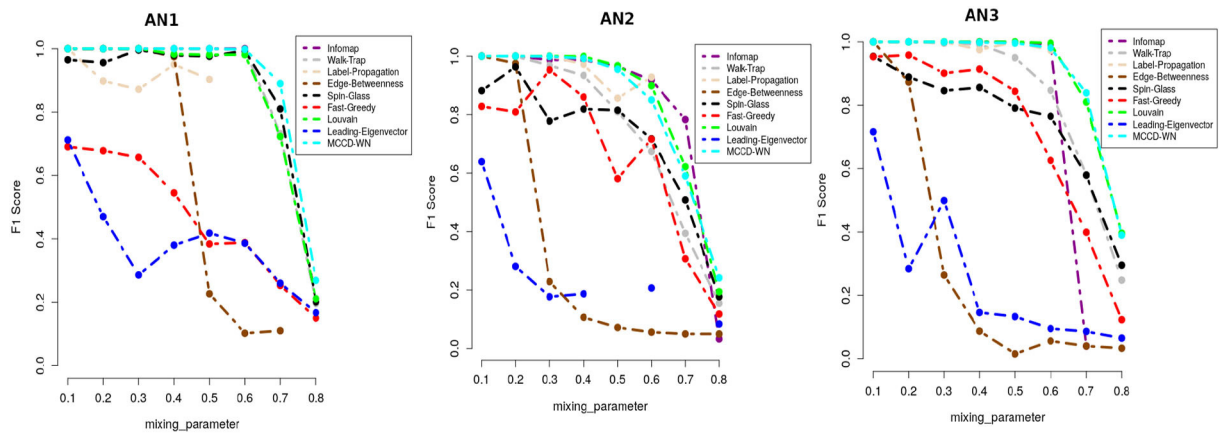

FIGURE 5
Performance comparison in terms of F1-score.

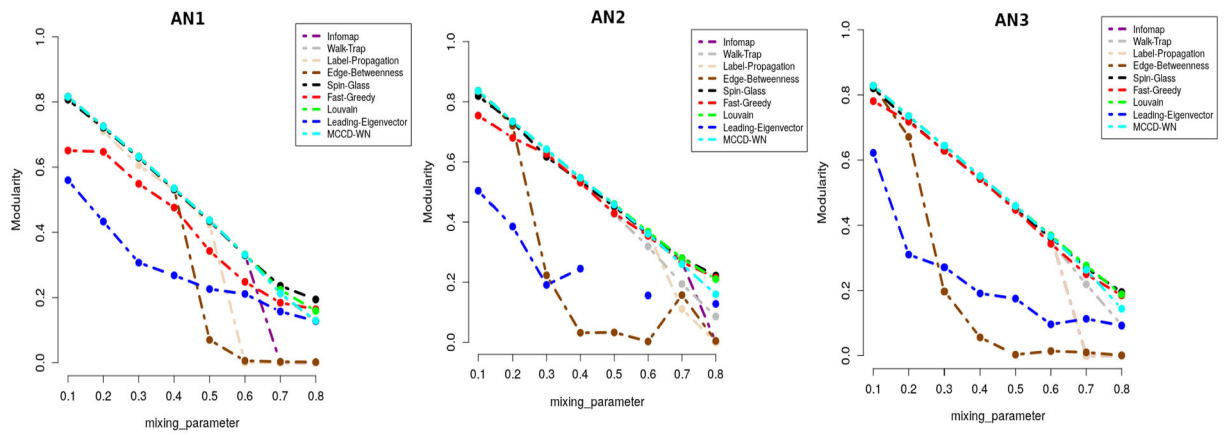

FIGURE 6
Performance comparison in terms of modularity score.

colonization activity and disruption in microbial communities, respectively (Baldassano and Bassett, 2016). In Figure 8B, we demonstrated the number of communities found in ASV co-abundance networks for the healthy and infected cohorts at each time-point.

3.2.3. Microbial families for the ASVs identified as key-driver and members of infected-core-hub communities

Figure 9 shows the interactions between the microbial families of ASVs found to undergo rewiring in the infected co-abundance network and their interaction partners at each of the time-points. In Figure 10A, we can see that at each time-point the majority of the driver ASVs belong to the family *Ruminococcaceae*. Figure 10B shows that the majority of the ASVs in the identified infected-core-hub communities belong to the family *Ruminococcaceae* at Day 0, 21, and 25, whereas at Day 7 most of such ASVs come from the family *Lachnospiraceae*. The microbial families of the ASVs belonging to the infected-core-hub communities may play a crucial role in subduing pathogen colonization.

3.2.4. Changes in the relative proportion of hub ASVs belonging to microbial families

In Figures 10C,D, we showed the relative proportion of hub ASVs assigned to their corresponding family levels in both the healthy and infected cohorts. From these two figures we not only observe a change in the relative proportion of hubs between the healthy and infected cohort at each time-point, but also a change across all the time-points in the infected cohort.

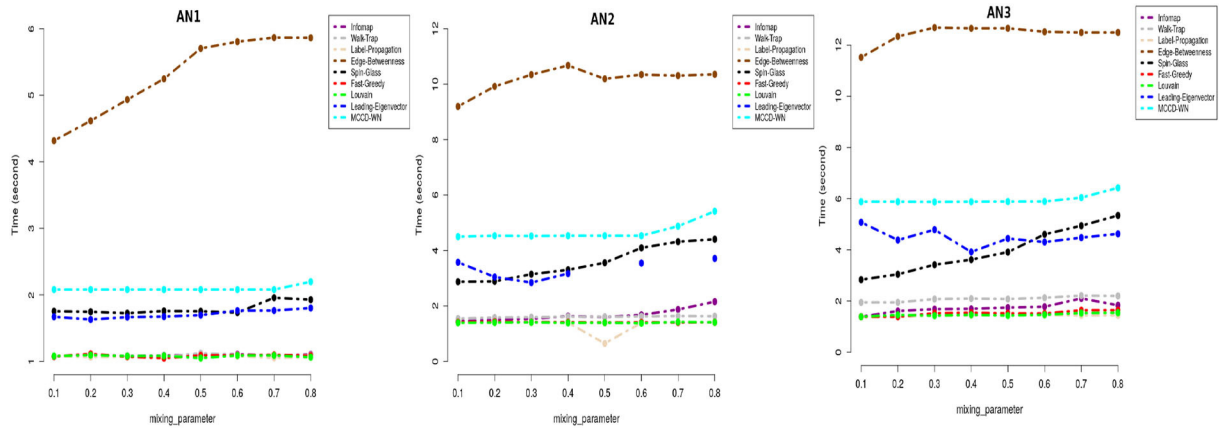

FIGURE 7
Performance comparison in terms of running time (in seconds) on a log-linear scale.

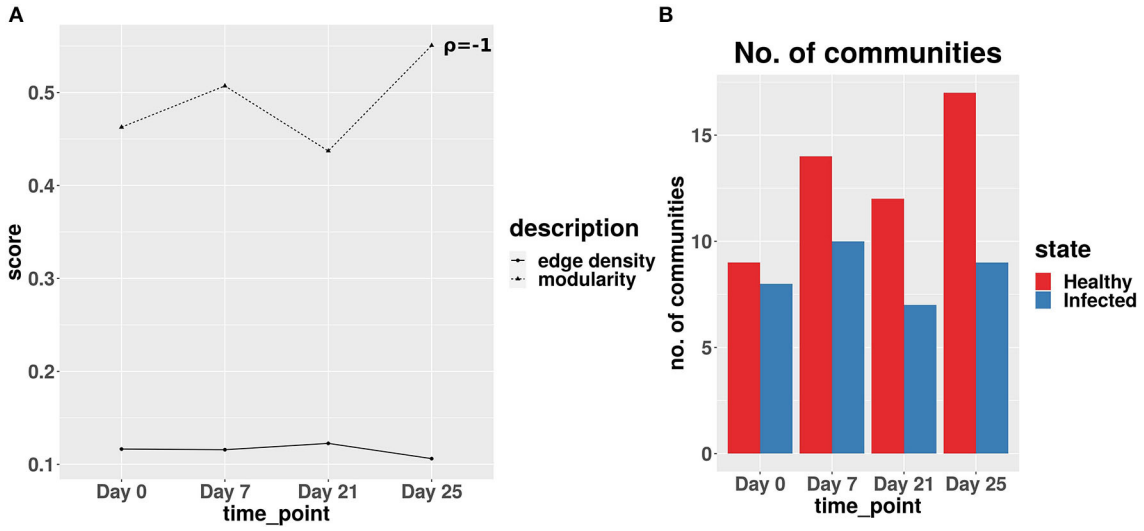

FIGURE 8
(A) Average edge density and modularity scores of the co-abundance networks for infected cohort. ρ denotes the Spearman correlation coefficient between average edge density and modularity scores across all time-points. **(B)** Number of communities obtained for real-life dataset.

3.2.5. Identification of dominant microbial families

In Figure 11, we show the heatmap of the relative abundance of ASVs aggregated at their family levels and selected the microbiome families which contain ASVs having a mean-relative abundance greater than 5%. The selected microbial families are referred to as dominant families in this work and further used for unveiling their functional roles using the metaproteome data.

3.2.6. ASVs belonging to a non-dominant microbial family were found as kinless hubs

In Figure 10D, it is of interest to see that ASVs assigned to the family *Desulfovibrionaceae* are found to act as kinless hubs. Although *Desulfovibrionaceae* was not found as a dominant family in our analysis at day 21, it is known to produce hydrogen sulfide (H_2S), higher concentration of which is reported to increase the seriousness of IAV infection (Santana et al., 2021), while a reduced amount of H_2S is found to protect against the viral infection (Dilek et al., 2020).

3.2.7. Functional roles of the proteins of dominant microbial families

In order to identify the microbial functions associated with both the healthy and infected states and the corresponding protein biomarkers, we performed a Linear discriminant analysis Effect Size (LEfSe) analysis using

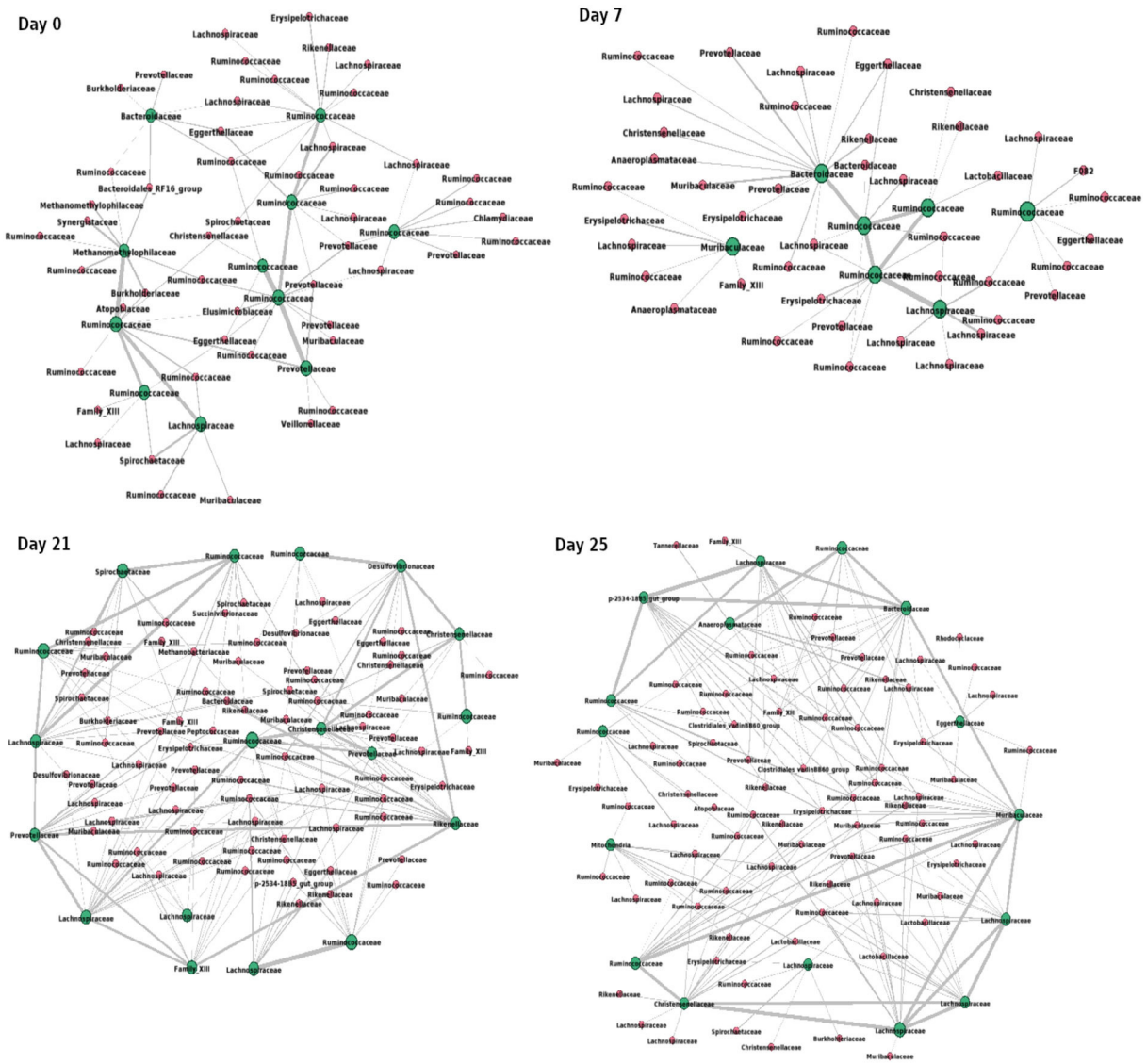

FIGURE 9
Network consisting of driver ASVs and their interaction partners at each of the time-points. Green nodes represent the names of families of corresponding driver ASVs. Thickness of edges is proportional to the corresponding correlation coefficient value. The network is drawn using Gephi tool (Bastian et al., 2009).

the abundance of microbial functions obtained from our metaproteome experiment.

At day 0 (Figure 12) we found family *Ruminococcaceae* as a dominant family. *Ruminococcaceae* have previously been reported to be associated with H1N1-infected animals (Sencio et al., 2020). In addition, LEfSe analysis revealed the association of "Ribosomal proteins synthesis and modification" with the infected cohorts. A previous study reported that the ribosomal proteins not only play an instrumental role to trigger the viral infection by interacting with viral proteins, but also are involved in activating immune pathways against viral infection (Li, 2019).

Besides *Ruminococcaceae*, families *Prevotellaceae* and *Lachnospiraceae* were found to be dominant at day 7 (Figure 13). Both of these families are found to be linked to immune response against Influenza A virus infection (Borey et al., 2021). LEfSe analysis of the functional composition of those microbial families revealed a lower relative abundance for "ATP-proton motive force interconversion" which may indicate the inhibition of viral replication during influenza infection by disrupting the proton motive force (PMF) (Domenech et al., 2020). It is of interest to see that our LEfSe analysis based on the relative abundance of proteins participating in the enriched microbial functions shows a higher relative-abundance of Glutamate

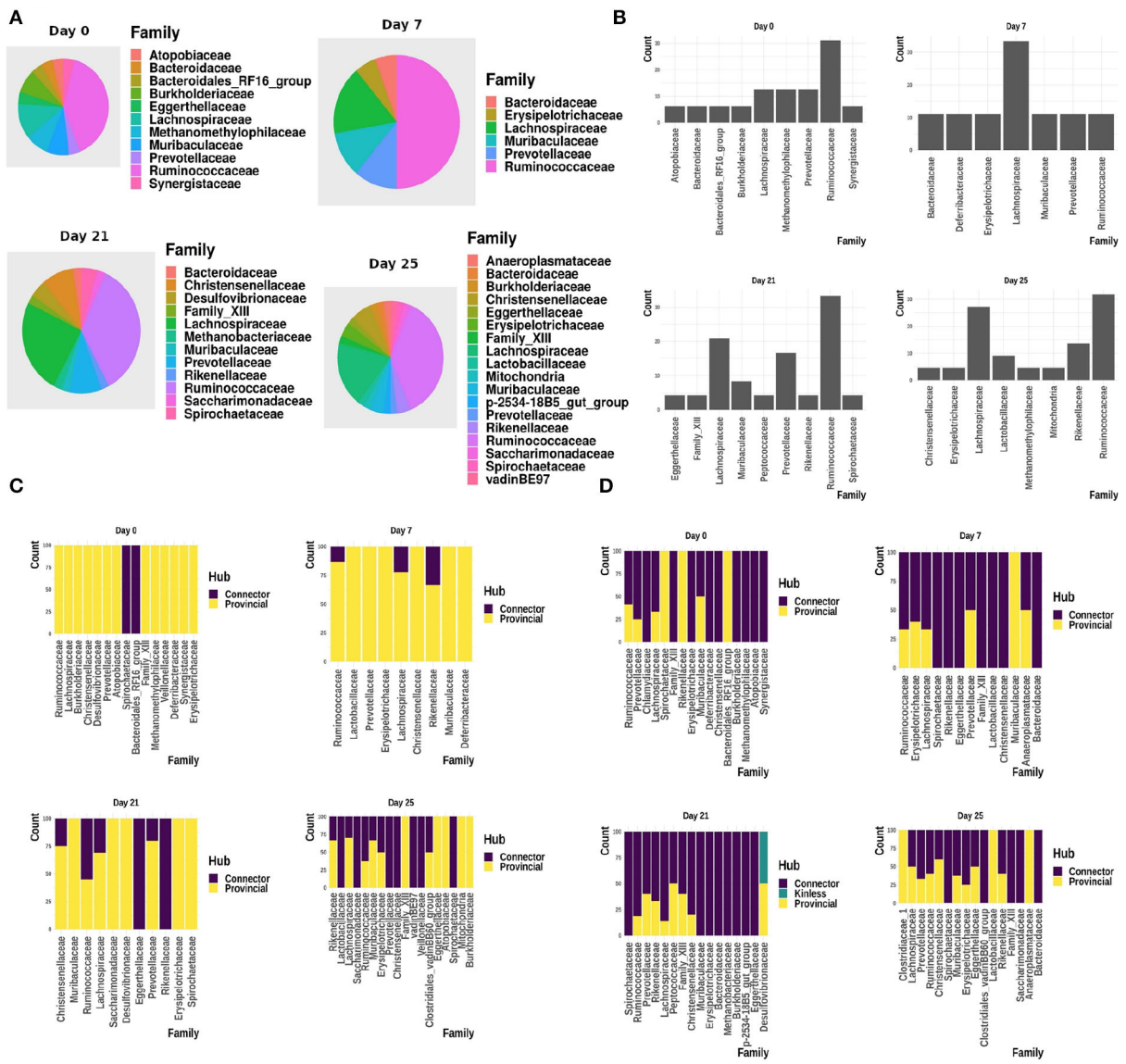

FIGURE 10
(A) Relative proportion of driver ASVs assigned to the corresponding microbiome families at each of the time-points. **(B)** Relative proportion of ASVs belonging to the infected-core-hub communities, assigned to the corresponding microbiome families at each of the time-points. **(C)** Relative proportion of ASVs identified as hubs in healthy co-abundance networks. **(D)** Relative proportion of ASVs identified as hubs in infected co-abundance networks.

dehydrogenase in the infected cohort because a higher level of glutamine is reported to be essential in the immune system cells during infection (de Oliveira et al., 2016). Moreover, a decrease in the relative-abundance of a well known virulence factor flagellin may be caused due to the activation of innate immune response (Hayashi et al., 2001). In addition, chaperonins are found to be prominent from our LEfSe analysis and Young (1990) showed that targeting chaperonins by the immune response plays both the protective and pathogenic roles.

In addition to the *Ruminococcaceae* and *Lachnospiraceae* families, *Spirochaetaceae* is found to be a dominant family at day 21, as shown in Figure 14. Moreover, our network analysis predicted the involvement of *Ruminococcaceae* in subduing pathogen colonization. Interestingly, our prediction is supported by the findings from LEfSe analysis of the functional composition of the family *Ruminococcaceae*. We observed a lower mean relative abundance of "glycolysis" in the infected cohort suggesting a significant reduction in the viral replication (Kohio and Adamson, 2013). Furthermore, LEfSe analysis of the relative abundance of proteins belonging to the function "unclassified" revealed a lower abundance of iron storage protein ferritin in the infected cohort, indicating a recovery from

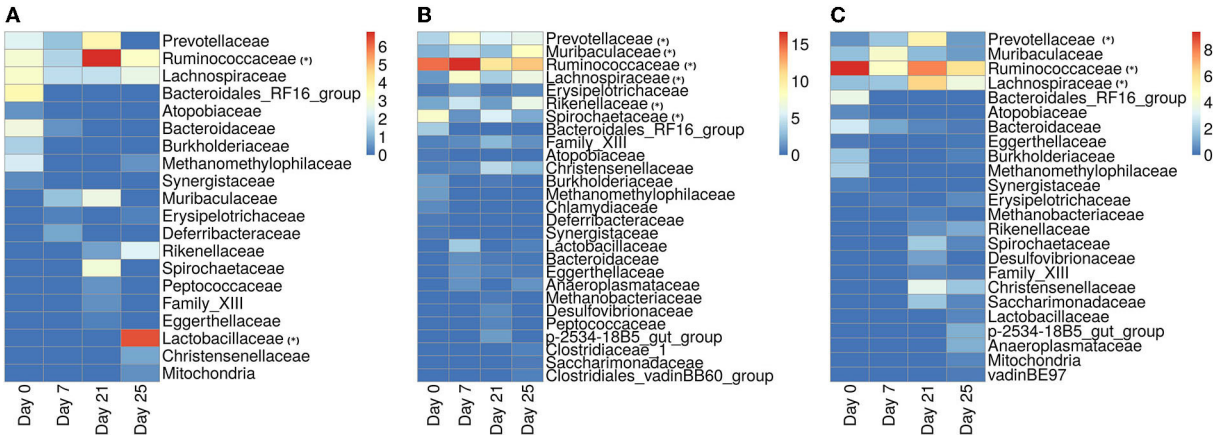

FIGURE 11
Heatmap showing the relative abundance of ASVs aggregated at the family level. **(A)** ASVs belonging to the infected-core-hub communities. **(B)** ASVs identified as hubs in the infected cohort. **(C)** ASVs identified as key-drivers in the infected cohort. Microbial families marked with "*" are identified as dominant families.

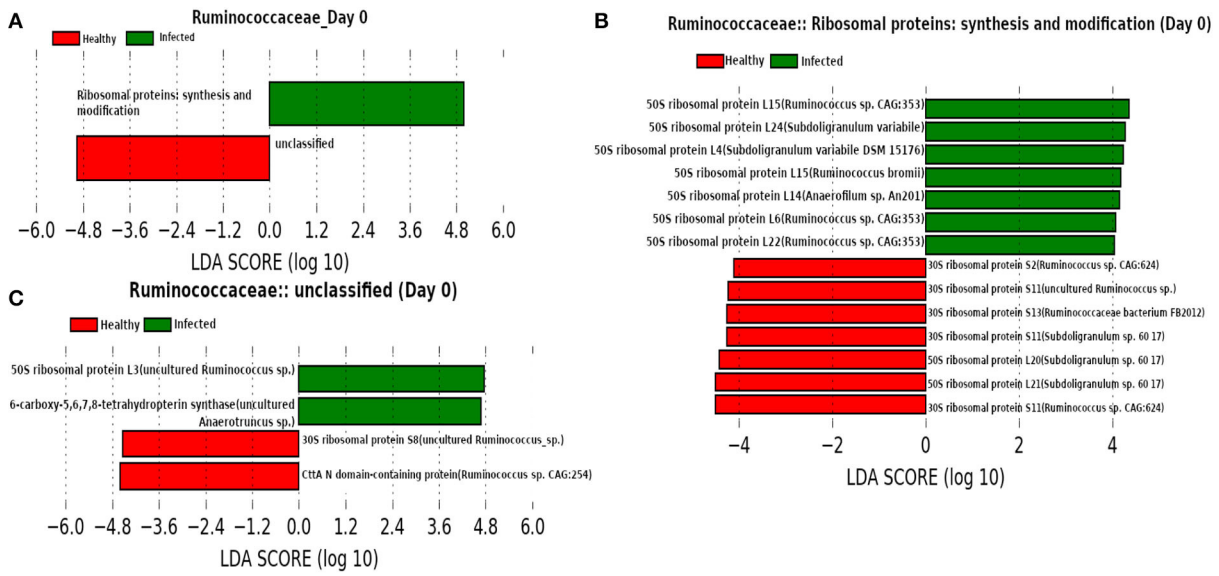

FIGURE 12
(A) Linear discriminant analysis effect size (LEfSe) scores of the enriched functions of the dominant microbiome family, ASVs of which are found to be the member of infected-core-hub community or act as hubs or key-drivers in the infected cohort at day 0. **(B,C)** Linear discriminant analysis effect size (LEfSe) scores of the predicted biomarkers involved in the enriched microbial functions associated with healthy and infected cohorts at day 0.

infection (Lalueza et al., 2020; Perricone et al., 2020). In addition, "glycolysis" and "pentose phosphate pathway" functions of the families *Spirochaetaceae* and *Lachnospiraceae*, respectively, were found to be enriched in the infected cohorts. These findings suggest that glycolysis may play an instrumental role in the activation of the innate immune response, by rising the metabolic flux through the pentose phosphate pathway (Ganeshan and Chawla, 2014). Moreover, *Elongation factor Tu (EF-Tu)*, assigned to the function "translation factors (protein synthesis)" of the family *Lachnospiraceae* is prominent in the infected cohort. The prokaryotic EF-Tu is reported to play critical role in both the enhancement of virulence factor and activation of the host immune system (Harvey et al., 2019). Furthermore, glyceraldehyde 3-phosphate dehydrogenase (GAPDH) is found to be a marker associated with the function "glycolysis." Sheng and Wang (2009) and Awan (2021) reported the role of GAPDH in the immune system and hence, it might be used as a therapeutic target.

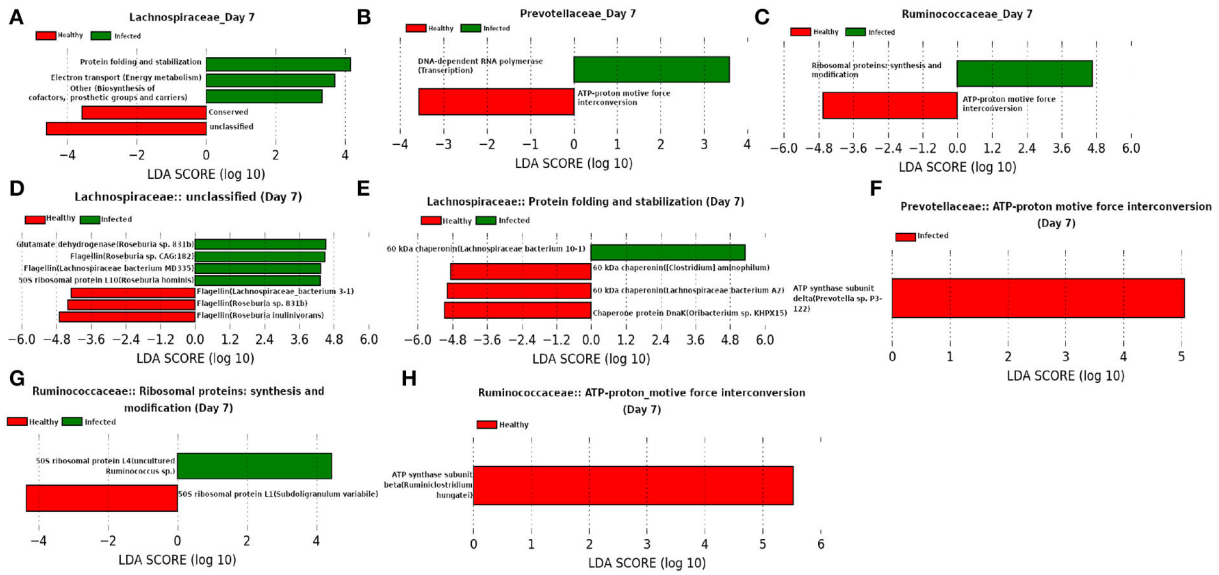

FIGURE 13
(A–C) Linear discriminant analysis effect size (LEfSe) scores of the enriched functions of the dominant microbiome family, ASVs of which are found to be the member of infected-core-hub community or act as hubs or key-drivers in the infected cohort at day 7. **(D–H)** Linear discriminant analysis effect size (LEfSe) scores of the predicted biomarkers involved in the enriched microbial functions associated with healthy and infected cohorts at day 7.

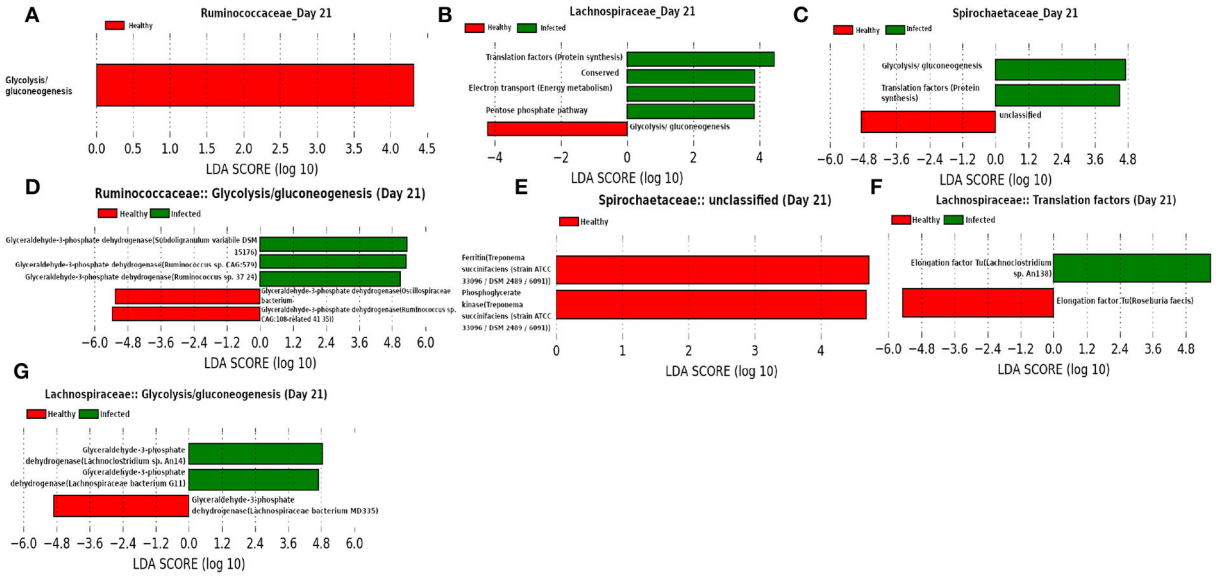

FIGURE 14
(A–C) Linear discriminant analysis effect size (LEfSe) scores of the enriched functions of the dominant microbiome family, ASVs of which are found to be the member of infected-core-hub community or act as hubs or key-drivers in the infected cohort at day 21. **(D–G)** Linear discriminant analysis effect size (LEfSe) scores of the predicted biomarkers involved in the enriched microbial functions associated with healthy and infected cohorts at day 21.

At day 25 (Figure 15) our network analysis based on 16srRNA data predicts the role of *Lactobacillaceae* family in subduing pathogen colonization. LEfSe analysis of relative abundance of the proteins assigned to "unclassified" function reveals the association of surface-layer proteins with the infected cohort. Acosta et al. (2019) and Wakai et al. (2021) reported that a higher concentration of the surface-layer protein of *Lactobacillaceae* family inhibits the viral replication.

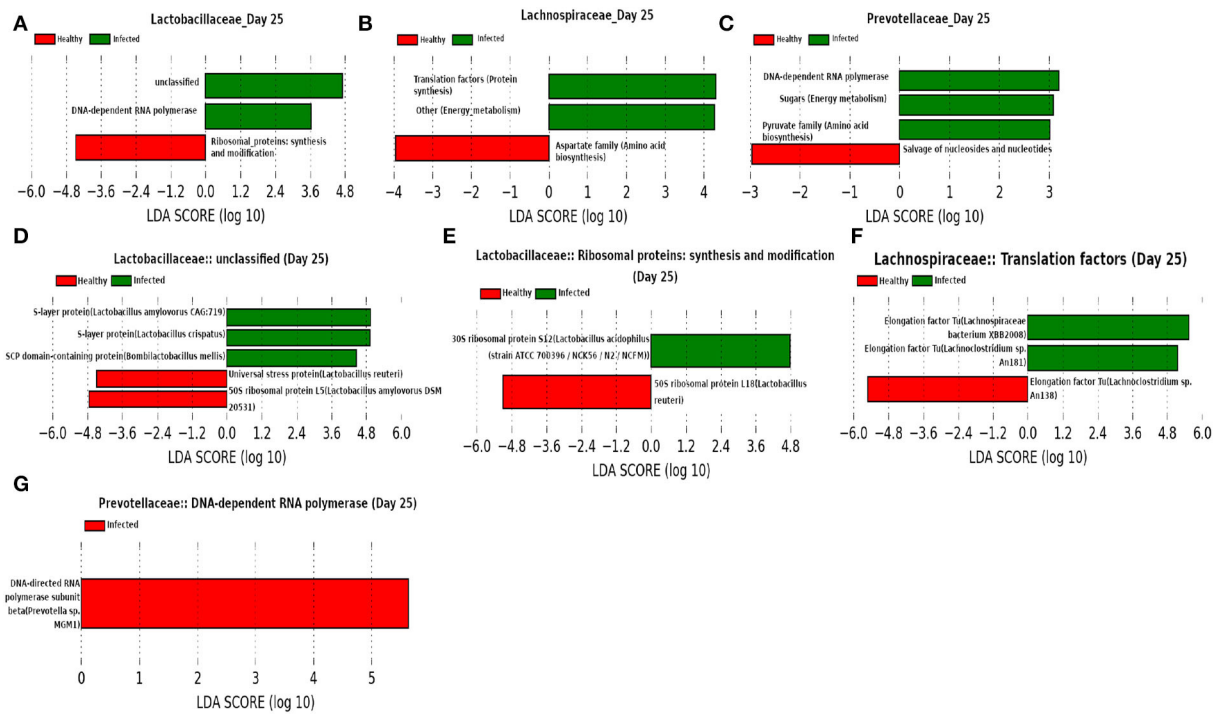

FIGURE 15
(A–C) Linear discriminant analysis effect size (LEfSe) scores of the enriched functions of the dominant microbiome family, ASVs of which are found to be the member of infected-core-hub community or act as hubs or key-drivers in the infected cohort at day 25. **(D–G)** Linear discriminant analysis effect size (LEfSe) scores of the predicted biomarkers involved in the enriched microbial functions associated with healthy and infected cohorts at day 25.

Since the relative abundance of S-layer protein was found to be higher in the infected cohort compared to the healthy cohort, we hypothesize that the S-layer protein might play an instrumental role in combating influenza A infection. Besides, it is of interest to see the function "pyruvate biosynthesis" of family *Prevotellaceae* and "aspartate biosynthesis" of family *Lachnospiraceae* have a relatively higher and lower composition in the infected cohort, respectively. The effect of pyruvate in alleviating influenza A virus infection is reported in Reel and Lupfer (2021), whereas aspartate is known to be crucial for viral genome nucleotide synthesis (Lao-On et al., 2018).

4. Discussion

In this work we proposed a maximal-clique based community detection algorithm to find modules in a weighted undirected network. We compared the performance of the proposed algorithm with some of the existing algorithms using three sets of benchmark networks and found that the proposed algorithm results in either the best or very similar performance. In addition, we applied our algorithm to a microbiome data set containing the abundance of microbial ASVs obtained from 16S rRNA gene sequencing in order to unveil the role of the gut microbiome in the host immune response during IAV infection. Our Network analysis predicts the association of microbial families such as *Ruminococcaceae*, *Lachnospiraceae*, *Spirochaetaceae*, *Prevotellaceae*, *Lactobacillaceae* with the immune response of infected cohort. In particular, we found the role of a low-abundant microbial family *Desulfovibrionaceae* as a kinless hub at day 21 and this finding may indicate its role in initiating the stabilization of gut microbial communities by producing a lower concentration of hydrogen sulfide after the second IAV infection. Moreover, the integration of metaproteome data not only provided the functions of the aforementioned microbial families in the host metabolism closely linked to immune response, but also unveiled the biomarker proteins of those dominant families. At day 25, we observed a prominent role of surface layer proteins of *Lactobacillaceae* family which may inhibit viral infection and thus, lead to fast recovery of infected pigs. Taken together, our results provided insights into the involvement of the gut microbiome and their proteins which might be beneficial for the development of novel antiviral therapy against influenza A viral infection.

Data availability statement

16S rRNA gene sequences are available at European Nucleotide Archive (ENA), with the project number RJEB39963 (accession number ERP123542) for healthy cohort, PRJEB42450 (accession number ERP126308) for infected cohort and the project name "KoInfekt multi-omics-pipeline-swine." Metaproteomics data is available at ProteomeXchange Consortium (submitted via the PRIDE partner repository) with the dataset identifier PXD020775.

Ethics statement

All animal experiments were approved by the State Office for Agriculture, Food Safety and Fishery in Mecklenburg-Western Pomerania (LALFF M-V) with reference number 7221.3-1-035/17.

Author contributions

AB developed the method, implemented the software, conducted the case studies, and drafted the manuscript. AM and HW carried out the 16s rRNA sequencing experiment. LG carried out the metaproteomics experiment. TS, CK, and CS carried out the animal case study. LK did the initial planning together with AB, discussed the results and revised the manuscript. LK, TU, KR, and TCM acquired the funding. All authors read, reviewed, and approved the final manuscript.

Funding

This research was funded by Federal Excellence Initiative of Mecklenburg-Western Pomerania and European Social Fund (ESF) Grant KoInfekt (ESF/14-BM-A55-0014/16). LK and AB acknowledge funding from the European Union (EuCanShare, Grant No. 825903). We acknowledge support for the Article Processing Charge from the DFG and the Open Access Publication Fund of the University of Greifswald.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.979320/full#supplementary-material

References

Acosta, M. P., Geoghegan, E. M., Lepenies, B., Ruzal,1, S., Kielian, M., and Martinez, M. G. (2019). Surface (s) layer proteins of lactobacillus acidophilus block virus infection via dc-sign interaction. *Front. Microbiol.* 10, 810. doi: 10.3389/fmicb.2019.00810

Ahn, Y.-Y., Bagrow, J. P., and Lehmann, S. (2010). Link communities reveal multi-scale complexity in networks. *Nature* 466, 761–764. doi: 10.1038/nature09182

Alvarez, A. J., Sanz-Rodríguez, C. E., and Cabrera, J. L. (2015). Weighting dissimilarities to detect communities in networks. *Philos. Trans. A Math. Phys. Eng. Sci.* 373, 20150108. doi: 10.1098/rsta.2015.0108

Arthur, D., and Vassilvitskii, S. (2007). "K-means++: the advantages of careful seeding," in *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA '07* (Philadelphia, PA: Society for Industrial and Applied Mathematics), 1027–1035.

Awan, A. (2021). Gapdh, interferon γ, and nitric oxide: inhibitors of coronaviruses. *Front. Virol.* 1, 682136. doi: 10.3389/fviro.2021.682136

Baldassano, S. N., and Bassett, D. S. (2016). Topological distortion and reorganized modular structure of gut microbial co-occurrence networks in inflammatory bowel disease. *Sci. Rep.* 6, 26087. doi: 10.1038/srep 26087

Bastian, M., Heymann, S., and Jacomy, M. (2009). "Gephi: an open source software for exploring and manipulating networks," in *International AAAI Conference on Weblogs and Social Media* (San Jose, CA).

Benson, A. R., Gleich, D. F., and Leskovec, J. (2016). Higher-order organization of complex networks. *Network Sci.* 353, 163–166. doi: 10.1126/science.aad9029

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. *J. Stat. Mech.* 2008, P10008. doi: 10.1088/1742-5468/2008/10/P10008

Borey, M., Blanc, F., Lemonnier, G., Leplat, J. J., Jardet, D., Rossignol, M. N., et al. (2021). Links between fecal microbiota and the response to vaccination against influenza a virus in pigs. *NPJ Vaccines* 6, 92. doi: 10.1038/s41541-021-00351-2

Chen, X., Liu, S., Goraya, M. U., Maarouf, M., Huang, S., and Chen, J. L. (2018). Host immune response to influenza a virus infection. *Front. Immunol.* 9, 320. doi: 10.3389/fimmu.2018.00320

Cortes, L., Wopereis, H., Tartiere, A., Piquenot, J., Gouw, J. W., Tims, S., et al. (2019). Metaproteomic and 16s rRNA gene sequencing analysis of the infant fecal microbiome. *Int. J. Mol. Sci.* 20, 1430. doi: 10.3390/ijms20061430

Csardi, G., and Nepusz, T. (2006). The igraph software package for complex network research. *Int. J. complex Syst.* 1695, 1–9.

de Oliveira, D. C., da Silva Lima, F., Sartori, T., Santos, A. C., Rogero, M. M., and Fock, R. A. (2016). Glutamine metabolism and its effects on immune response: molecular mechanism and gene expression. *Nutrire* 41, 14. doi: 10.1186/s41110-016-0016-8

Derényi, I., Palla, G., and Vicsek, T. (2005). Clique percolation in random networks. *Phys. Rev. Lett.* 94, 160202. doi: 10.1103/PhysRevLett.94.160202

Dilek, N., Papapetropoulos, A., Kinsky, T. T., and Szabo, C. (2020). Hydrogen sulfide: an endogenous regulator of the immune system. *Pharmacol. Res.* 161, 105119. doi: 10.1016/j.phrs.2020.105119

Domenech, A., Brochado, A. R., Sender, V., Hentrich, K., Normark, B. H., Typas, A., et al. (2020). Proton motive force disruptors block bacterial competence and horizontal gene transfer. *Cell Host Microbe* 27, 544.e3–555.e3. doi: 10.1016/j.chom.2020.02.002

Elgamal, Z., Singh, P., and Geraghty, P. (2021). The upper airway microbiota, environmental exposures, inflammation, and disease. *Medicina* 57, 823. doi: 10.3390/medicina57080823

Eppstein, D., Löffler, M., and Strash, D. (2010). "Listing all maximal cliques in sparse graphs in near-optimal time," in *Algorithms and Computation*, eds O. Cheong, K. Y. Chwa, and K. Park (Berlin; Heidelberg: Springer), 403–414.

Friedman, J., and Alm, E. J. (2012). Inferring correlation networks from genomic survey data. *PLoS Comput. Biol.* 8, e1002687. doi: 10.1371/journal.pcbi.1002687

Ganeshan, K., and Chawla, A. (2014). Metabolic regulation of immune responses. *Annu. Rev. Immunol.* 32, 609–634. doi: 10.1146/annurev-immunol-032713-120236

Gierse, L. C., Meene, A., Schultz, D., Schwaiger, T., Karte, C., Schroeder, C., et al. (2020). A multi-omics protocol for swine feces to elucidate longitudinal dynamics in microbiome structure and function. *Microorganisms* 8, 1887. doi: 10.3390/microorganisms8121887

Gierse, L. C., Meene, A., Schultz, D., Schwaiger, T., Schroeder, C., Muecke, P., et al. (2021). Influenza a h1n1 induced disturbance of the respiratory and fecal microbiome of german landrace pigs - a multi-omics characterization. *Microbiol. Spectrum* 9, e0018221. doi: 10.1128/Spectrum.00182-21

Girvan, M., and Newman, M. E. J. (2002). Community structure in social and biological networks. *Proc. Natl. Acad. Sci. U.S.A.* 99, 7821–7826. doi: 10.1073/pnas.122653799

Hall, C. V., Lord, A., Betzel, R., Zakrzewski, M., Simms, L. A., Zalesky, A., et al. (2019). Co-existence of network architectures supporting the human gut microbiome. *iScience* 22, 380–391. doi: 10.1016/j.isci.2019.11.032

Harvey, K. L., Jarocki, V. M., Charles, I. G., and Djordjevic, S. P. (2019). The diverse functional roles of elongation factor tu (ef-tu) in microbial pathogenesis. *Front. Microbiol.* 10, 2351. doi: 10.3389/fmicb.2019.02351

Hayashi, F., Smith, K. D., Ozinsky, A., Hawn, T. R., Yi, E. C., Goodlett, D. R., et al. (2001). The innate immune response to bacterial flagellin is mediated by toll-like receptor 5. *Nature* 410, 1099–1103. doi: 10.1038/35074106

Hoefnagels, I., van de Maat, J., van Kampen, J. J. A., van Rossum, A., Obihara, C., Tramper-Stranders, G. A., et al. (2021). The role of the respiratory microbiome and viral presence in lower respiratory tract infection severity in the first five years of life. *Microorganisms* 9, 1446. doi: 10.3390/microorganisms 9071446

Horvath, S. (2011). *Weighted Network Analysis*. New York, NY: Springer-Verlag.

Jing, G., Zhang, Y., Liu, L., Wang, Z., Sun, Z., Knight, R., et al. (2021). A scale-free, fully connected global transition network underlies known microbiome diversity. *mSystems* 6, e0039421. doi: 10.1128/mSystems.00394-21

Kaul, D., Rathnasinghe, R., Ferres, M., Tan, G. S., Barrera, A., Pickett, B. E., et al. (2020). Microbiome disturbance and resilience dynamics of the upper respiratory tract during influenza a virus infection. *Nat. Commun.* 11, 2537. doi: 10.1038/s41467-020-17020-y

Khan, I., Bai, Y., Zha, L., Ullah, N., Ullah, H., Shah, S. R., et al. (2021). Mechanism of the gut microbiota colonization resistance and enteric pathogen infection. *Front. Cell Infect. Microbiol.* 11, 716299. doi: 10.3389/fcimb.2021.716299

Kohio, H. P., and Adamson, A. L. (2013). Glycolytic control of vacuolar-type atpase activity: a mechanism to regulate influenza viral infection. *Virology* 444, 301–309. doi: 10.1016/j.virol.2013.06.026

Kuntal, B. K., Chandrakar, P., Sadhu, S., and Mande, S. S. (2019). 'netshift': a methodology for understanding 'driver microbes' from healthy and disease microbiome datasets. *ISME J.* 13, 442–454. doi: 10.1038/s41396-018-0291-x

Laarhoven, T., and Marchiori, E. (2016). Local network community detection with continuous optimization of conductance and weighted kernel k-means. *J. Mach. Learn. Res.* 17, 5148–5175. doi: 10.48550/arXiv.1601.05775

Lalueza, A., Ayuso, B., Arrieta, E., Trujillo, H., Folgueira, D., Cueto, C., et al. (2020). Elevation of serum ferritin levels for predicting a poor outcome in hospitalized patients with influenza infection. *Clin. Microbiol. Infect.* 26, 1557.e9. doi: 10.1016/j.cmi.2020.02.018

Lancichinetti, A., and Fortunato, S. (2009). Benchmarks for testing community detection algorithms on directed and weighted graphs with overlapping communities. *Phys. Rev. E* 80(1 Pt 2), 016118. doi: 10.1103/PhysRevE.80.016118

Lancichinetti, A., Fortunato, S., and Radicchi, F. (2008). Benchmark graphs for testing community detection algorithms. *Phys. Rev. E* 78(4 Pt 2), 046110. doi: 10.1103/PhysRevE.78.046110

Lao-On, U., Attwood, P. V., and Jitrapakdee, S. (2018). Roles of pyruvate carboxylase in human diseases: from diabetes to cancers and infection. *J. Mol. Med.* 96, 237–247. doi: 10.1007/s00109-018-1622-0

Lee, K. H., Guo, J., Song, Y., Ariff, A., O'Sullivan, M., Hales, B., et al. (2021). Dysfunctional gut microbiome networks in childhood ige-mediated food allergy. *Int. J. Mol. Sci.* 22, 2079. doi: 10.3390/ijms22042079

Li, N., Ma, W. T., Pang, M., Fan, Q. L., and Hua, J. L. (2019). The commensal microbiota and viral infection: a comprehensive review. *Front. Immunol.* 10, 1551. doi: 10.3389/fimmu.2019.01551

Li, S. (2019). Regulation of ribosomal proteins on viral infectiondoi. *Cells* 8, 508. doi: 10.3390/cells8050508

Li, Y., Zhang, G., Feng, Y., and Wu, C. (2015). An entropy-based social network community detecting method and its application to scientometrics. *Scientometrics* 102, 1003–1017. doi: 10.1007/s11192-014-1377-5

Lu, Z., Wahlström, J., and Nehorai, A. (2018). Community detection in complex networks via clique conductance. *Sci. Rep.* 8, 5982. doi: 10.1038/s41598-018-23932-z

Mangangcha, I. R., Malik, M. Z., Kucuk, O., Ali, S., and Singh, R. K. (2020). Kinless hubs are potential target genes in prostate cancer network. *Genomics* 112, 5227–5239. doi: 10.1016/j.ygeno.2020.09.033

Newman, M. E. J. (2004). Detecting community structure in networks. *Eur. Phys. J. B* 38, 321–330. doi: 10.1140/epjb/e2004-00124-y

Newman, M. E. J. (2006a). Finding community structure in networks using the eigenvectors of matrices. *Phys. Rev. E* 74, 036104. doi: 10.1103/PhysRevE.74.036104

Newman, M. E. J. (2006b). Modularity and community structure in networks. *Proc. Natl. Acad. Sci. U.S. A.* 103, 8577–8582. doi: 10.1073/pnas.0601602103

Omar, Y. M., and Plapper, P. (2020). A survey of information entropy metrics for complex networks. *Entropy* 22, 1417. doi: 10.3390/e22121417

Pacheco, G. A., Gálvez, N. M. S., Soto, J. A., Andrade, C. A., and Kalergis, A. M. (2021). Bacterial and viral coinfections with the human respiratory syncytial virus. *Microorganisms* 9, 1293. doi: 10.3390/microorganisms9061293

Page, L., Brin, S., Motwani, R., and Winograd, T. (1999). *The pagerank citation ranking: Bringing order to the web*. Technical Report 1999-66, Stanford InfoLab. Previous number = SIDL-WP-1999-0120.

Perricone, C., Bartoloni, E., Bursi, R., Cafaro, G., Guidelli, G. M., Shoenfeld, Y., et al. (2020). Covid-19 as part of the hyperferritinemic syndromes: the role of iron depletion therapy. *Immunol. Res.* 68, 213–224. doi: 10.1007/s12026-020-09145-5

Pettigrew, M. M., Tanner, W., and Harris, A. D. (2021). The lung microbiome and pneumonia. *J. Infect. Dis.* 223, S241-S245. doi: 10.1093/infdis/jiaa702

Reel, J. M., and Lupfer, C. R. (2021). Sodium pyruvate affects influenza a virus infection in *vivo*. *J. Immunol.* 206(1 Suppl.), 396978. doi: 10.1101/2020.11.25.396978

Reichardt, J., and Bornholdt, S. (2006). Statistical mechanics of community detection. *Phys. Rev. E* 74, 016110. doi: 10.1103/PhysRevE.74.016110

Santana, C. M., Gauger, P., Vetger, A., Magstadt, D., Kim, D. S., Shrestha, D., et al. (2021). Ambient hydrogen sulfide exposure increases the severity of influenza a virus infection in swine. *Arch. Environ. Occupat. Health* 76, 526–538. doi: 10.1080/19338244.2021.1896986

Schwaiger, T., Sehl, J., Karte, C., Schaefer, A., Huehr, J., Mettenleiter, T. C., et al. (2019). Experimental h1n1pdm09 infection in pigs mimics human seasonal influenza infections. *PLoS ONE* 14, e0222943. doi: 10.1371/journal.pone. 0222943

Segata, N., Izard, J., Waldron, L., Gevers, D., Miropolsky, L., Garrett, W. S., et al. (2011). Metagenomic biomarker discovery and explanation. *Genome Biol.* 12, R60. doi: 10.1186/gb-2011-12-6-r60

Sencio, V., Barthelemy, A., Tavares, L. P., Machado, M. G., Soulard, D., Cuinat, C., et al. (2020). Gut dysbiosis during influenza contributes to pulmonary pneumococcal superinfection through altered short-chain fatty acid production. *Cell Rep.* 30, 2934.e6–2947.e6. doi: 10.1016/j.celrep.2020.02.013

Sheng, W. Y., and Wang, T. C. V. (2009). Proteomic analysis of the differential protein expression reveals nuclear gapdh in activated t lymphocytes. *PLoS ONE* 4, e6322. doi: 10.1371/journal.pone.0006322

Shi, Y., Baquerizo, M. D., Li, Y., Yang, Y., Zhu, Y. G., Peñuelas, J., et al. (2020). Abundance of kinless hubs within soil microbial networks are associated with high functional potential in agricultural ecosystems. *Environ. Int.* 142, 105869. doi: 10.1016/j.envint.2020.105869

Stefanska, I., Romanowska, M., Donevski, S., Gawryluk, D., and Brydak, L. B. (2013). Co-infections with influenza and other respiratory viruses. Respiratory regulation-the molecular approach. *Adv. Exp. Med. Biol.* 756, 291–301. doi: 10.1007/978-94-007-4549-0_36

Vardi, Y., and Zhang, C. H. (2000). The multivariate l1-median and associated data depth. *Proc. Natl. Acad. Sci. U.S.A.* 97, 1423–1426. doi: 10.1073/pnas.97.4.1423

Vernocchi, P., Gili, T., Conte, F., Chierico, F. D., Conta, G., Miccheli, A., et al. (2020). Network analysis of gut microbiome and metabolome to discover microbiota-linked biomarkers in patients affected by non-small cell lung cancer. *Int. J. Mol. Sci.* 21, 8730. doi: 10.3390/ijms21228730

Volant, S., Lechat, P., Woringer, P., Motreff, L., Campagne, P., Malabat, C., et al. (2020). Shaman: a user-friendly website for metataxonomic analysis from raw reads to statistical analysis. *BMC Bioinform.* 21, 345. doi: 10.1186/s12859-020-03666-4

Wakai, T., Kano, C., Karsens, H., Kok, J., and Yamamoto, N. (2021). Functional role of surface layer proteins of *Lactobacillus acidophilus* l-92 in stress tolerance and binding to host cell proteins. *Biosci. Microbiota Food Health* 40, 33–42. doi: 10.12938/bmfh.2020-005

Weiszfeld, E., and Plastria, F. (2009). On the point for which the sum of the distances to n given points is minimum. *Ann. Operat. Res.* 167, 7–41. doi: 10.1007/s10479-008-0352-z

Yang, Z., Algesheimer, R., and Tessone, C. J. (2016). A comparative analysis of community detection algorithms on artificial networks. *Sci. Rep.* 6, 30750. doi: 10.1038/srep30750

Yildiz, S., Mazel-Sanchez, B., Kandasamy, M., Manicassamy, B., and Schmolke, M. (2018). Influenza a virus infection impacts systemic microbiota dynamics and causes quantitative enteric dysbiosis. *Microbiome* 6, 9. doi: 10.1186/s40168-017-0386-z

Young, D. B. (1990). Chaperonins and the immune response. *Semin Cell Biol.* 1, 27–35.