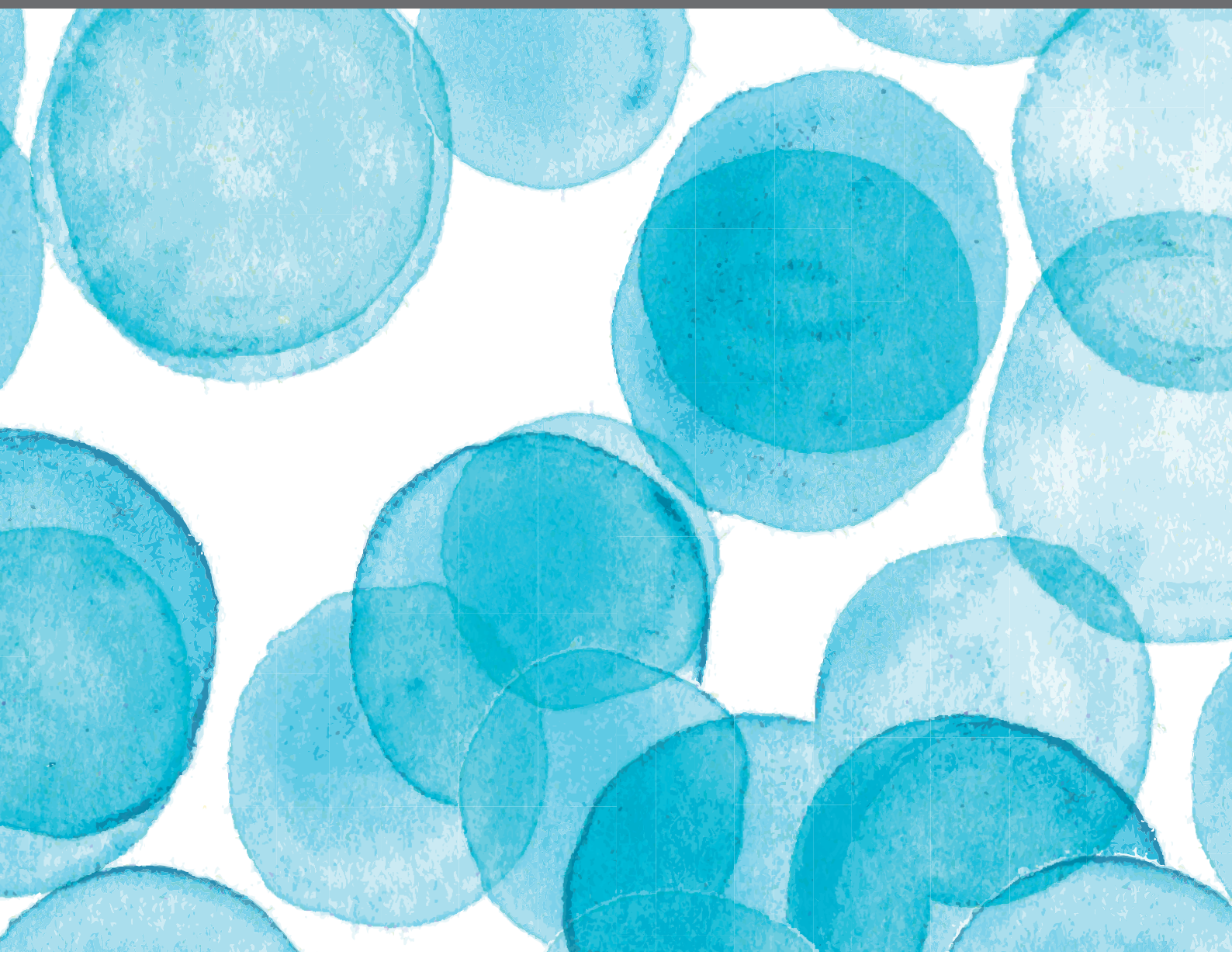
INSIGHTS IN SYSTEMS MICROBIOLOGY: 2021

EDITED BY: George Tsiamis and Qi Zhao
PUBLISHED IN: Frontiers in Microbiology



frontiers Research Topics

frontiers

ISSN 1664-8714
ISBN 978-2-83250-166-5
DOI 10.3389/978-2-83250-166-5

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

INSIGHTS IN SYSTEMS MICROBIOLOGY: 2021

Topic Editors:
George Tsiamis, University of Patras, Greece
Qi Zhao, University of Science and Technology Liaoning, China

Citation: Tsiamis, G., Zhao, Q., (2022). Insights in Systems Microbiology: 2021. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-83250-166-5

Table of Contents

frontiers | Frontiers in Microbiology

TYPE Editorial
PUBLISHED 25 August 2022
DOI 10.3389/fmicb.2022.988296

OPEN ACCESS

EDITED AND REVIEWED BY
Fernando Perez Rodriguez,
University of Cordoba, Spain

*CORRESPONDENCE
George Tsiamis
gtsiamis@upatras.gr

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

RECEIVED 07 July 2022
ACCEPTED 31 July 2022
PUBLISHED 25 August 2022

CITATION
Zhao Q and Tsiamis G (2022) Editorial: Insights in systems microbiology: 2021. *Front. Microbiol.* 13:988296. doi: 10.3389/fmicb.2022.988296

Editorial: Insights in systems microbiology: 2021

Qi Zhao[1] and George Tsiamis[2]*

[1]School of Computer Science and Software Engineering, University of Science and Technology Liaoning, Anshan, China, [2]Department of Environmental Engineering, University of Patras, Agrinio, Greece

KEYWORDS

systems microbiology, multidisciplinary approaches, gut microbiota, drug-microbe association, computational methods

Editorial on the Research Topic
Insights in systems microbiology: 2021

The aim of systems microbiology is to understand life's circuit diagrams and gain knowledge about the relationships between the individual components that build a cellular organism, a community, and an entire ecological niche (Vieites et al., 2009). That means systems microbiology tends to treat the community as a whole, integrating multidisciplinary approaches to finally make a big picture of how a microbial cell or community operates. We are now entering the third decade of the twenty-first century, and, especially in the last years, the achievements made by scientists in the field of microbiology have been exceptional, leading to significant advancements. This Research Topic collects new insights, novel developments, current challenges, latest discoveries, recent advances, and future perspectives in the field of systems microbiology.

We are pleased to note that our Research Topic has attracted contributions from many highly regarded researchers in this field around the world, including from China, the USA, Italy, and Singapore. We received 15 submissions, 10 of which were accepted for publication after rigorous reviews.

In this special issue, three articles were focused on the new findings in gut microbiota. For example, Zhu Y. et al. proposed a new perspective of intestinal microbiota–neural mitochondria interaction as a communicating channel from gut to brain. Such research could help to extend the vision of gut-brain axis regulation and provide additional research directions on the treatment and prevention of responsive neurological disorders. Accumulating evidence is focused on the roles of the gut microbial community in cardiovascular disease, but few studies have unveiled the alterations and further directions of gut microbiota in severe chronic heart failure (CHF) patients. To investigate this deficiency, Sun et al. collected fecal samples from 29 CHF patients diagnosed with NYHA Class III-IV and 30 healthy controls and then analyzed them using bacterial 16S rRNA gene sequencing. As a result, there were many significant

differences between the two groups. Moreover, gut microbiome-based therapeutics have shown promise in ameliorating chronic inflammation. However, they are largely experimental, context- or strain-dependent, and lack a clear mechanistic basis. In this perspective, Koduru et al. reasoned that the future transition toward precision probiotics thus lies in deciphering ligand-receptor interactions, with aryl hydrocarbon receptor (AhR) being a key mediator in managing chronic inflammation.

Three research papers were included in the Research Topic to directly investigate the drug-microbe associations, surfaced enhanced Raman spectra (SERS), and the integration of "multiomic" data with other omics through artificial intelligence algorithms. Zhu B. et al. designed a deep learning-based model named Nearest Neighbor Attention Network (NNAN). The proposed model consists of four components: a similarity network constructor, a nearest-neighbor aggregator, a feature attention block, and a predictor. Under both a cross-validation setting and a realistic potential linkage discovery setting, the empirical comparison of the proposed framework with three state-of-the-art baselines demonstrates that NNAN has significant competitive performance in predicting drug–microbe associations. In another study, Tang et al. used SERS combined with unsupervised and supervised machine learning algorithms to detect 15 bacterial pathogens from clinical samples. According to the results, SERS could accurately identify bacterial pathogens at a general level with comparatively high specificity and sensitivity through the assistance of machine learning methods. Moreover, artificial intelligence (AI) and machine learning (ML) algorithms coupled with other multiomics data (i.e., big data) could help researchers to classify better the patient's molecular characteristics and drive clinicians to identify personalized therapeutic strategies. Here, Paolini et al. highlighted how the integration of "multiomic" data (i.e., miRNAs profiling and microbiota signature) with other omics (i.e., metabolomics, exposomics) analyzed by AI algorithms could improve the diagnostic and prognostic potential of specific biomarkers of disease.

Apart from the novel computational methods, Liu et al. proposed a new family of Trechisporales, Sistotremastraceae fam. Nov. based on the combination of molecular and morphological data, and it is typified by Sistotremastrum. The phylogenetic analyses show that Sistotremastraceae forms a monophyletic lineage with robust support within Trechisporales. Marine Streptomycetes are attracting particular attention as the new producer of novel antibiotics and anti-cancer agents with unusual properties. Shi et al. established the taxonomic status of a new Streptomycetes species isolated from a marine environment, and this new Streptomyces species shows a valuable source of new bioactive secondary metabolites. The great potential to produce novel natural products was evaluated by genomes analysis, compound detection, and antimicrobial activities. Rhoades et al. summarized several host factors and pathways involved in coronavirus infection and are also implicated in neuropsychiatric symptoms. Though several of these host factors are expressed in the central nervous system, they have also provided evidence that their influence on widespread systemic inflammation may play a significant role in the development of long-term psychological symptoms stemming from COVID-19 infection. Wang et al. went through the frontlines of the metaomics techniques and explored their potential applications in clinical diagnoses of human diseases, e.g., infectious diseases, through which we concluded that novel diagnostic methods based on human microbiomes should be achieved in the near future, while the limitations of these techniques such as standard procedures and computational challenges for rapid and accurate analysis of metaomics data in clinical settings were also examined.

Finally, we want to thank all the authors who contributed their original work to our special issue and the reviewers for their valuable comments. We would like to express our sincere gratitude to the editorial office of Frontiers in Microbiology for their excellent support and for providing us with this opportunity to successfully host this hot topic issue.

Author contributions

QZ drafted the manuscript. GT revised the draft. Both authors made a direct and intellectual contribution to the work and approved the final version for publication.

Funding

This study was supported by the Foundation of Education Department of Liaoning Province (Grant No. LJKZ0280).

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.

References

Vieites, J. M., Guazzaroni, M. E., Beloqui, A., Golyshin, P. N., and Ferrer, M. (2009). Metagenomics approaches in systems microbiology. *FEMS Microbiol. Rev.* 33, 236–255. doi: 10.1111/j.1574-6976.2008.00152.x

frontiers in Microbiology

ORIGINAL RESEARCH
published: 31 January 2022
doi: 10.3389/fmicb.2021.813289

Alterations of the Gut Microbiota in Patients With Severe Chronic Heart Failure

Weiju Sun[1†], Debing Du[2†], Tongze Fu[3], Ying Han[4], Peng Li[5*] and Hong Ju[6*]*

[1] Department of Cardiology, The First Affiliated Hospital of Harbin Medical University, Harbin, China, [2] Beidahuang Industry Group General Hospital, Harbin, China, [3] Harbin Medical University, Harbin, China, [4] Department of Cardiology, The Fourth Affiliated Hospital of Harbin Medical University, Harbin, China, [5] National Center for Biomedical Analysis, Beijing, China, [6] Heilongjiang Vocational College of Biology Science and Technology, Harbin, China

Chronic heart failure (CHF) is the final outcome of almost all forms of cardiovascular diseases, remaining the main cause of mortality worldwide. Accumulating evidence is focused on the roles of gut microbial community in cardiovascular disease, but few studies have unveiled the alterations and further directions of gut microbiota in severe CHF patients. Aimed to investigate this deficiency, fecal samples from 29 CHF patients diagnosed with NYHA Class III-IV and 30 healthy controls were collected and then analyzed using bacterial 16S rRNA gene sequencing. As a result, there were many significant differences between the two groups. Firstly, the phylum *Firmicutes* was found to be remarkably decreased in severe CHF patients, and the phylum *Proteobacteria* was the second most abundant phyla in severe CHF patients instead of phylum *Bacteroides* strangely. Secondly, the α diversity indices such as chao1, PD-whole-tree and Shannon indices were significantly decreased in the severe CHF versus the control group, as well as the notable difference in β-diversity between the two groups. Thirdly, our result revealed a remarkable decrease in the abundance of the short-chain fatty acids (SCFA)-producing bacteria including genera *Ruminococcaceae UCG-004*, *Ruminococcaceae UCG-002*, *Lachnospiraceae FCS020 group*, *Dialister* and the increased abundance of the genera in *Enterococcus* and *Enterococcaceae* with an increased production of lactic acid. Finally, the alternation of the gut microbiota was presumably associated with the function including Cell cycle control, cell division, chromosome partitioning, Amino acid transport and metabolism and Carbohydrate transport and metabolism through SCFA pathway. Our findings provide the direction and theoretical knowledge for the regulation of gut flora in the treatment of severe CHF.

Keywords: severe chronic heart failure, gut microbiota, 16S rRNA gene, SCFA, patients

OPEN ACCESS

Edited by:
Qi Zhao,
University of Science and Technology Liaoning, China

Reviewed by:
Juan Wang,
Inner Mongolia University, China
Yanshuo Chu,
University of Texas MD Anderson Cancer Center, United States

****Correspondence:***
Ying Han
flhy81@163.com
Peng Li
58301561@qq.com
Hong Ju
hongju.hit@hotmail.com

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Systems Microbiology, a section of the journal Frontiers in Microbiology

Received: *11 November 2021*
Accepted: *03 December 2021*
Published: *31 January 2022*

Citation:
Sun W, Du D, Fu T, Han Y, Li P and Ju H (2022) Alterations of the Gut Microbiota in Patients With Severe Chronic Heart Failure. Front. Microbiol. 12:813289. doi: 10.3389/fmicb.2021.813289

INTRODUCTION

Chronic heart failure (CHF) is a major health problem worldwide. It is the final outcome of almost all forms of cardiovascular diseases. CHF is recognized not only as a deregulation of hemodynamic disorder and neurocrine activation, but also an uncontrolled elevation of inflammatory responses and oxidative stress (Anker and Von Haehling, 2004; Gajarsa and Kloner, 2011;

Hage et al., 2017; Huang et al., 2020). CHF is still associated with a high rate of hospitalization and a devastating prognosis, despite the recent development of modern combinational therapeutic strategies. Therefore, it is possible that important pathogenic mechanisms have not been targeted by current treatments, such as gut microbiota dysbiosis which have also been implicated to play a role in the development of cardiovascular diseases, including CHF (Brial et al., 2018; Jia et al., 2019; Cheng et al., 2020; Huang et al., 2020; Sanchez-Rodriguez et al., 2020).

The gut microbiota, comprising the trillions of bacteria in the gastrointestinal tract, is essential for maintaining human health in many aspects, digesting the indigestible nutrients of the host, producing vitamins and hormones, shaping the development of the mucosal immune system, and preventing the colonization of pathogenic bacteria (Amoroso et al., 2020; Sanchez-Rodriguez et al., 2020; Deledda et al., 2021). Host-microbiota interactions involving inflammatory and metabolic pathways have been proposed to contribute to the pathogenesis of CHF (Dantzer et al., 2018; Moshkelgosha et al., 2018; Cheng et al., 2019, 2021; Zhang et al., 2019; Kwon et al., 2020). In recent years, several sequencing-based studies have reported that the composition and function of intestinal flora between HF patients and healthy subjects are different. There are some common findings, but also considerable differences between studies (Kamo et al., 2017; Luedde et al., 2017; Cui et al., 2018; Kummen et al., 2018; Mayerhofer et al., 2018, 2020; Iqubal et al., 2020; Khan et al., 2020; Qi et al., 2021). Some computational methods have been applied in the field and other biological data (Long et al., 2021; Lv et al., 2021; Yang et al., 2021). Thus, more studies are still needed to provide detailed information on variations of gut microbial composition and its impacts on CHF, especially the severe CHF.

In order to define a more robust HF-related gut microbiota signature, we conducted this cross-sectional cohorts investigation. In this study, we collected stool samples from severe CHF patients and healthy controls, amplified the variable region of intestinal bacteria 16S rRNA, constructed a DNA library, and then assessed the taxonomic composition of the gut microbiota in patients with severe CHF.

MATERIALS AND METHODS

Study Population and Sample Collection

Chronic heart failure patients (n = 29) were recruited from the First Affiliated Hospital of Harbin Medical University and the Fourth Affiliated Hospital of Harbin Medical University between April 2020 and August 2020, as well as 30 asymptomatic persons undergoing physical examinations as healthy controls. Patients who were recruited represent multiple stages of HF progression, as defined by NYHA Class III-IV. The inclusion criteria were as follows: (1) the subjects had not received antacids, probiotics, antibiotics, or antimicrobial agents within 30 days before sample collection; (2) there was no organic disease of the digestive system; and (3) they had no gastrointestinal surgery. NYHA classification was performed by patients' treating cardiologist and adjudicated by 2 HF specialists who were blinded to the results. All patients with HF were treated according to current HF management guidelines. Associated clinical information was collected from electronic medical records. All participants (or their direct relatives) gave written informed consent, and the First Affiliated Hospital of Harbin Medical University and the Fourth Affiliated Hospital of Harbin Medical University approved all study protocols.

We collected fresh fecal samples (each 2–5 g) from all the participants 1–2 days after admission, then transferred into sterile collecting pipes and frozen at −80°C immediately.

DNA Extraction and 16S rRNA Gene V3-V4 Region Sequencing

The bacterial DNA was extracted from the fecal samples using the Tiangen stool mini kit (Tiangen, Beijing, China) according to the manufacturer's instructions. The extracted DNA from each sample was used as the template to amplify the V3–V4 region of 16S rRNA genes using PCR. PCR amplification, sequencing of the PCR amplicons and quality control of raw data were performed. A sequencing library of the V3–V4 regions of the 16S rRNA gene was prepared as described previously (Han et al., 2021). The purified products were mixed at an equal ratio for sequencing using an Illumina MiSeq system (Illumina Inc., United States).

Statistical Analysis

We evaluated the quality of sequencing data using the Fast-QC software[1] firstly. Next we obtained the clean data for subsequent analysis after removing the Chimera Sequence using QIIME2.[2] Third, operational taxonomic units (OTUs) were delineated at the cutoff of 97% also using QIIME2, and the sequencing results were compared and analyzed to obtain the family and genus annotations of OTUs based on the Silva database.[3] Then α- and β-diversity analyses were performed using QIIME2 fourthly. Shannon-wiener diversity index, Simpson diversity index, the observed OTUs, PD (phylogenetic diversity)-whole-tree and Chao1 index were evaluated. A normalized OTU abundance table was used for the β-diversity analysis, including principal coordinate analysis (PCoA) based on weighted UniFrac, and unweighted UniFrac distances. Next, we performed Lefse analysis to clarify the dominant bacteria. LEfSe is a software for discovering high-dimensional biomarkers and revealing genome characteristics. LEfSe uses linear discriminant analysis (LDA) to estimate the impact of the abundance of each component (species) on the difference effect. At last, the gene function of the sample was inferred based on the species composition obtained by sequencing, and the functional difference between different groups was analyzed using PICRUSt.[4] Subsequently, the Welch's t-test method of two groups was performed using the STAMP software to filter the parts with P-value > 0.05, and Heatmap Plot, PCA plot, and Extented error bar graphs were drawn to reveal significant differences in species abundance between different samples.

[1] http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

[2] http://qiime.org/

[3] https://www.arb-silva.de/

[4] https://picrust.github.io/picrust/index.html

RESULTS

Baseline Characteristics

The baseline characteristics of all the participants are shown in **Table 1**. Patients with CHF were characterized by a greater number of males, increased prevalence of comorbidity with Atrial fibrillation, worsened cardiac functions including larger left ventricular end diastolic diameter (LVEDD), decreased left ventricular ejection fraction (LVEF) and stroke volume (SV), increased E/e', coupled with increased serum Troponin I (TnI) and NT-pro B-type natriuretic peptide (NT-proBNP) levels. Patients with CHF also have worse renal function. Most (21) of the patients were classified to be heart failure patients with reduced ejection fraction, only 1 were classified to be heart failure patients with preserved ejection fraction, and the rest 8 of the patients were classified to be heart failure patients with midrange ejection fraction. All patients with HF were treated according to current HF management guidelines, including diuretics, βblocker, Aldosterone receptor antagonist, Angiotensin converting enzyme inhibitor/Angiotensin receptor antagonist/Angiotensin receptor neprilysin inhibitor, and Sodium-glucose cotransporter 2 inhibitor.

TABLE 1 | Baseline characteristics of the study participants.

Variables	HF patients (*n* = 29)	Healthy controls (*n* = 30)	*P* value
Age, years	60.69 (11.67)	60.0 (9.64)	0.8062
Sex, male	24 (83%)	10 (33%)	<0.0001
BMI (kg/m^2)	24.0 (3.47)	24.9 (3.08)	0.2849
NYHA class (III/IV)	10/19	—	—
HFrEF/HFpEF/HFmrEF	21/1/8	—	—
Hypertension	14 (48%)	11 (37%)	0.3757
Diabetes	10 (34%)	5 (16.7%)	0.1202
Atrial fibrillation	10 (6.7%)	0 (0)	0.0003
Smoking	12 (41.4%)	6 (20%)	0.0769
Echocardiographic parameters			
LVEDD, mm	62.0 (8.67)	43.7 (3.99)	<0.0001
LVEF, %	33.8 (9.1)	63.2 (4.65)	<0.0001
SV, ml	46.7 (11.1)	68.9 (11.94)	<0.0001
E/e'	19.3 (6.5)	13.4 (3.1)	<0.0001
Laboratory parameters			
TnI, ng/dL	55.87 (0.10–88.7)	0.012 (0–0.048)	0.0007
NT-proBNP, pg/mL	4745.7 (1130–16755)	124.0 (25–258)	<0.0001
Leukocyte,10^9/L	7.2 (3.00)	6.7 (1.74)	0.4366
Neutrophils,10^9/L	4.9 (2.51)	4.0 (1.30)	0.0878
Lymphocytes, 10^9/L	1.63 (0.61)	2.1 (0.65)	0.0110
Monocyte, 10^9/L	0.5 (0.18)	0.4 (0.14)	0.0632
Hemoglobin, g/L	140.7 (26.53)	140.8 (17.04)	0.9834
BUN, mg/dl	8.0 (3.04)	5.5 (1.81)	0.0003
Serum creatinine, mg/dl	87.4 (35.16)	67.4 (18.35)	0.0078
Fast glucose	6.5 (3.67)	5.2 (1.30)	0.0698
Cholesterol	4.2 (0.96)	5.0 (0.90)	0.0013
Triglycerides	1.5 (0.92)	1.9 (1.66)	0.2305
HDL-C	0.9 (0.22)	0.91 (0.22)	0.8917
LDL-C	2.5 (0.87)	2.9 (0.80)	0.0592
Treatment diuretics	29 (100%)	—	—
β blocker	27 (93%)	—	—
MRA	29 (100%)	—	—
ACEI/ARB/ARNI	26 (90%)	—	—
SGLT2i	21 (72%)	—	—

Results are presented as median (with standard error or upper and lower quartiles) or% where appropriate. BMI, body mass index; NYHA, New York Heart Association; HFrEF, heart failure with reduced EF; HFpEF, heart failure with preserved EF; HfmrEF, heart failure with midrange EF; LVEDD: left ventricular end diastolic diameter; LVEF: Left ventricular ejection fraction; TnI, Troponin I; NT-proBNP: NT-pro B-type natriuretic peptide; HDL-C, high density lipoprotein-cholesterol; LDL-C, low density lipoprotein-cholesterol; MRA, Aldosterone receptor antagonist; ACEI, Angiotensin converting enzyme inhibitor; ARB, Angiotensin receptor antagonist; ARNI, Angiotensin receptor neprilysin inhibitor; SGLT2i, Sodium-glucose cotransporter 2 inhibitor.

Species Classification

The different distribution of relative abundance of top 19 at the phylum level in the two groups is presented in **Figure 1A**. Sequencing analysis showed that gut microbiota of the two groups were mainly classified into four phyla, including the phyla *Firmicutes*, *Proteobacteria*, *Bacteroidetes*, and *Actinobacteria*. The phylum *Firmicutes* was found with the highest abundance of reads in CHF patients, accounting for 59.5% in total, which was significantly decreased versus that of an abundance of 72.4% in the controls. The second was the phylum *Proteobacteria*, accounting for 21.3% in total, which was much more abundant versus that of an abundance of 6.9% in the controls. Likewise, bacteria belonging to the phyla *Actinobacteria* were more abundant in CHF patients than that in the healthy controls (2.7 vs. 0.9%), While bacteria belonging to the *Bacteroidetes* phyla were slightly less abundant in HF patients than that in the healthy controls (14.9 vs. 17.7%).

At the genus level, the microflora of CHF patients was characterized by less abundant of *Faecalibacterium* (10.5 vs. 22.8%), as well as more abundant of *Escherichia-Shigella* (10.3 vs. 4.6%), *Enterococcus* (7.7 vs. 0.0%), and *Klebsiella* (6.9 vs. 1.1%) than that in the healthy controls (**Figure 1B**).

Analysis of α and β Diversity Index

The α diversity analysis was performed and then chao1 curve, observed-otus curve, PD-whole-tree curve, Shannon-Wiener curve and Simpson curve based on the species annotation information were subsequently obtained by sequencing analysis. As a result, the chao1 and PD-whole-tree indices were significantly decreased in the CHF versus control group, as well as the Shannon indices (**Figures 2A–C**). The taxonomic composition of the metagenomic populations of the gut microflora samples from patients with CHF compared to those from the healthy control group were also analyzed using Principal Coordinate Analysis (PCoA). The differences in β-diversity based on the weighted UniFrac between the HF and healthy control groups were also shown in **Figure 2D**, which indicates that the fecal microbial structure in the CHF group was obviously different than that of the healthy control group in condition of the presence of OTU.

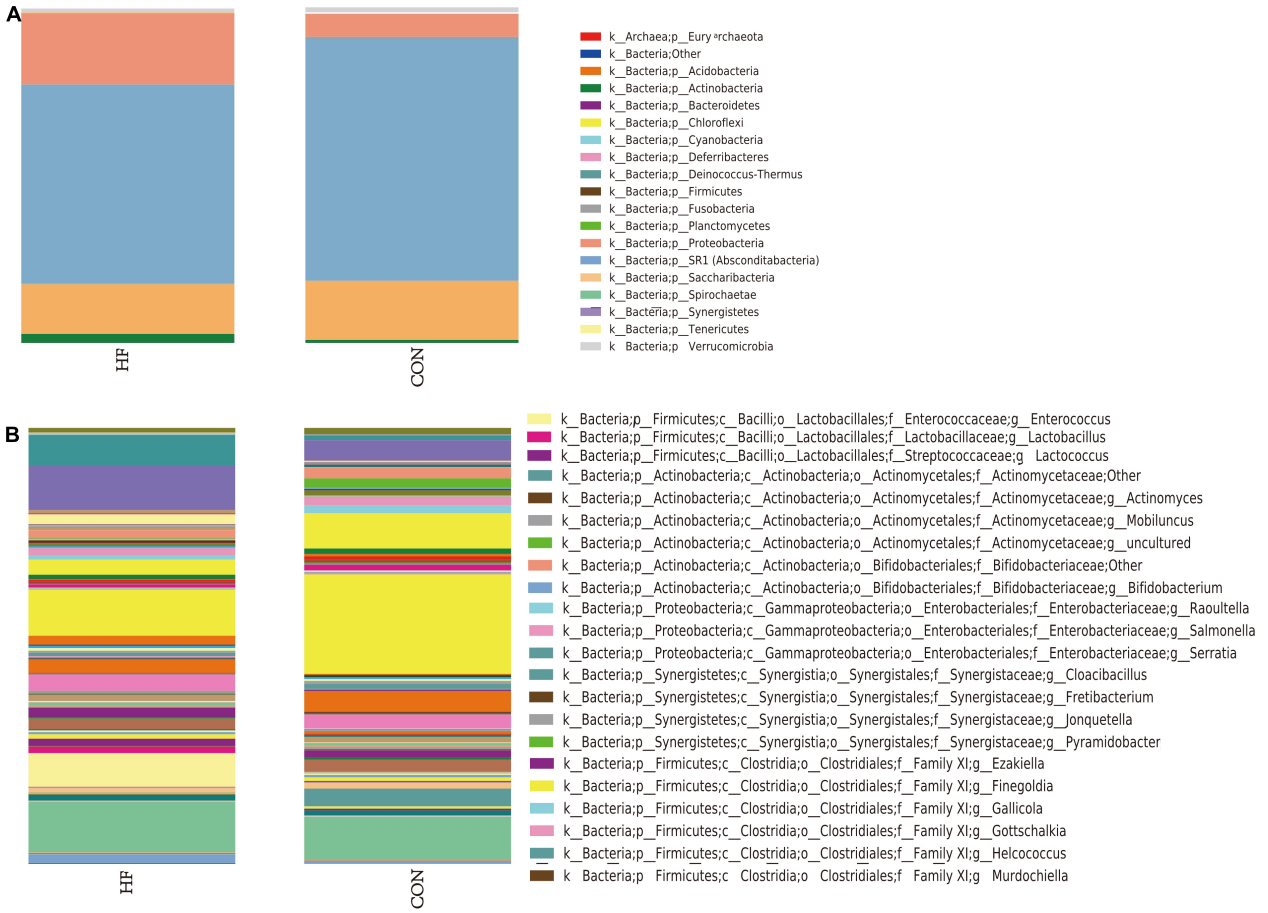


FIGURE 1 | The distribution of relative abundance in the CHF and the healthy control group. Panel **(A)** shows the distribution of relative abundance of top 19 at phylum level in the two groups. Panel **(B)** shows the distribution of relative abundance at the genus level in the two groups. HF, CHF group; CON, the healthy control group.

Analysis of Differential Taxonomy Expression

A differential taxonomy expression analysis was performed using limma algorithms, focusing on differences at the genus level (**Figure 3**). There was a remarkable difference with 152 generus in fecal microflora between the CHF and healthy control group in our result. Among these changes, the decreased abundance of the genera *Ruminococcaceae UCG-004*, *Ruminococcaceae UCG-002*, *Lachnospiraceae FCS020 group*, *Dialister* and the increased abundance of the genera in *Enterococcus* and *Enterococcaceae* were the most notable features (**Figure 3B**).

Analysis of Predictive Function

Based on the closed-reference OUTs, PICRUSt was utilized to predict abundances of the functional category COG orthologs (COs) and KEGG orthologs (KOs). Some of these COs and KOs were indicated to be significantly different in fecal microbiomes between the CHF and healthy control group ($P < 0.05$; **Figure 4**). Furthermore, there were also some meaningful results related with the function including cell cycle control, cell division, chromosome partitioning, Inorganic ion transport and metabolism, translation, ribosomal structure and biogenesis, amino acid transport and metabolism and carbohydrate transport and metabolism.

DISCUSSION

In the current study, bacterial 16S rRNA gene sequencing was applied to confirm the composition and differential expression in gut microbiota between 29 severe CHF patients and 30 healthy controls, resulting in a number of notable differences between these two groups. Firstly, the phylum *Proteobacteria* was significantly more abundant in CHF patients than controls, whereas the phylum *Firmicutes* was found remarkably decreased in CHF patients. Secondly, the α diversity indices significantly decreased in the CHF versus control group, as well as the notable differences in β-diversity between the CHF and healthy control groups, which indicates that the fecal microbial structure in the CHF group was obviously different than that of the healthy control group in

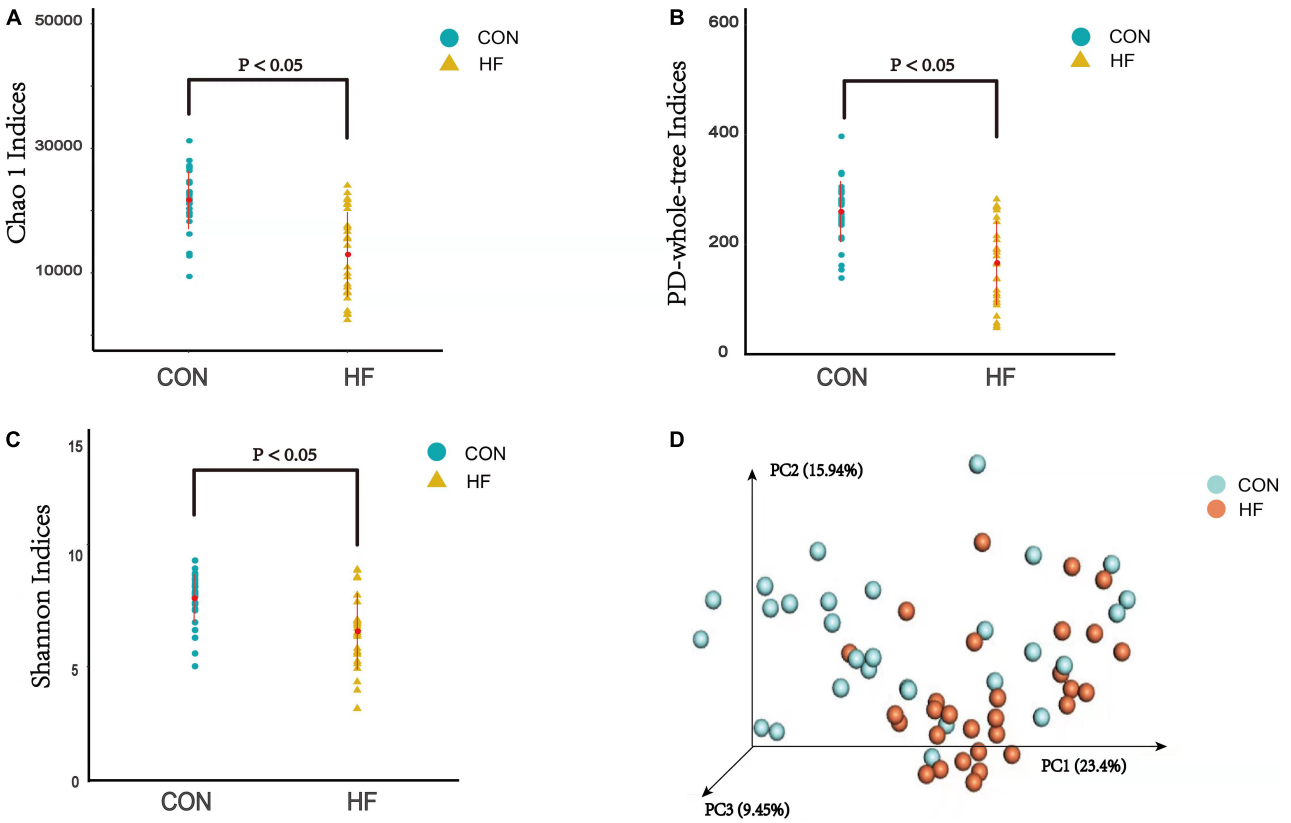


FIGURE 2 | The α-diversity indices and the β-diversity of the gut microflora between the CHF group and the healthy control group. **(A)** PD-whole tree indice; **(B)** Chao1 indices; **(C)** Shannon indices; **(D)** β-diversity based on the weighted UniFrac. HF, CHF group; CON, the healthy control group.

condition of the presence of OTU. Thirdly, the microflora of CHF patients was characterized by the decreased abundance of the genera *Ruminococcaceae UCG-004*, *Ruminococcaceae UCG-002*, *Lachnospiraceae FCS020 group*, *Dialister* and the increased abundance of the genera in *Enterococcus* and *Enterococcaceae*. Finally, the alternation of the gut microbiota was presumably associated with the function including cell cycle control, cell division, chromosome partitioning, inorganic ion transport and metabolism, translation, ribosomal structure and biogenesis, amino acid transport and metabolism and carbohydrate transport and metabolism. To our knowledge, this is the first study to explore the changes in the gut flora of patients with severe CHF.

Tang et al. (2017) put forward the "gut hypothesis of heart failure" for the first time. The hypothesis implies that reduced cardiac output caused by heart failure can lead to decreased intestinal perfusion, mucosal ischemia, and then intestinal mucosal destruction. These changes in the intestinal barrier function, in turn, can lead to increased intestinal permeability, intestinal malnutrition, bacterial translocation and increased circulating endotoxins, resulting in the potential inflammation associated with HF (Nagatomo and Tang, 2015; Tang et al., 2017, 2019; Harikrishnan, 2019). There are also several studies that have reported the composition and function of intestinal flora between HF patients and healthy subjects are different (Kamo et al., 2017; Luedde et al., 2017; Cui et al., 2018; Mayerhofer et al., 2018, 2020). Likewise, we found significant differences in the composition of fecal microbes between CHF patients and healthy controls, suggesting that there is a link between intestinal microflora disorders and CHF. At the phylum level, *Firmicutes* and *Bacteroides* are the two most abundant phyla in the healthy intestine. They are closely related to environmental conditions and may be beneficial or problematic for human and animal health. However, the phylum *Firmicutes* was remarkably decreased in severe CHF patients simultaneously in our study. What is more, the phylum *Proteobacteria* was the second most abundant phyla in severe CHF patients instead of phylum *Bacteroides*. All the members of the phylum *Proteobacteria* are Gram-negative bacteria, with the outer membrane mainly composed by lipopolysaccharide (LPS). Most of the phylum *Proteobacteria* are pathogenic bacteria, as a result, it is considered to be a microbial signature of dysbiosis in gut microbiota (Shin et al., 2015). Furthermore, LPS leaking into the blood through the intestinal wall can stimulate the production of a variety of pro-inflammatory cytokines, which are involved in the apoptosis, hypertrophy and fibrosis of heart cells, playing an important role in the occurrence and development of heart failure (Sandek et al., 2012).

Microbiota diversity has been considered to be a new health biomarker (Shanahan, 2010; Tang et al., 2018; Aponte et al., 2020;

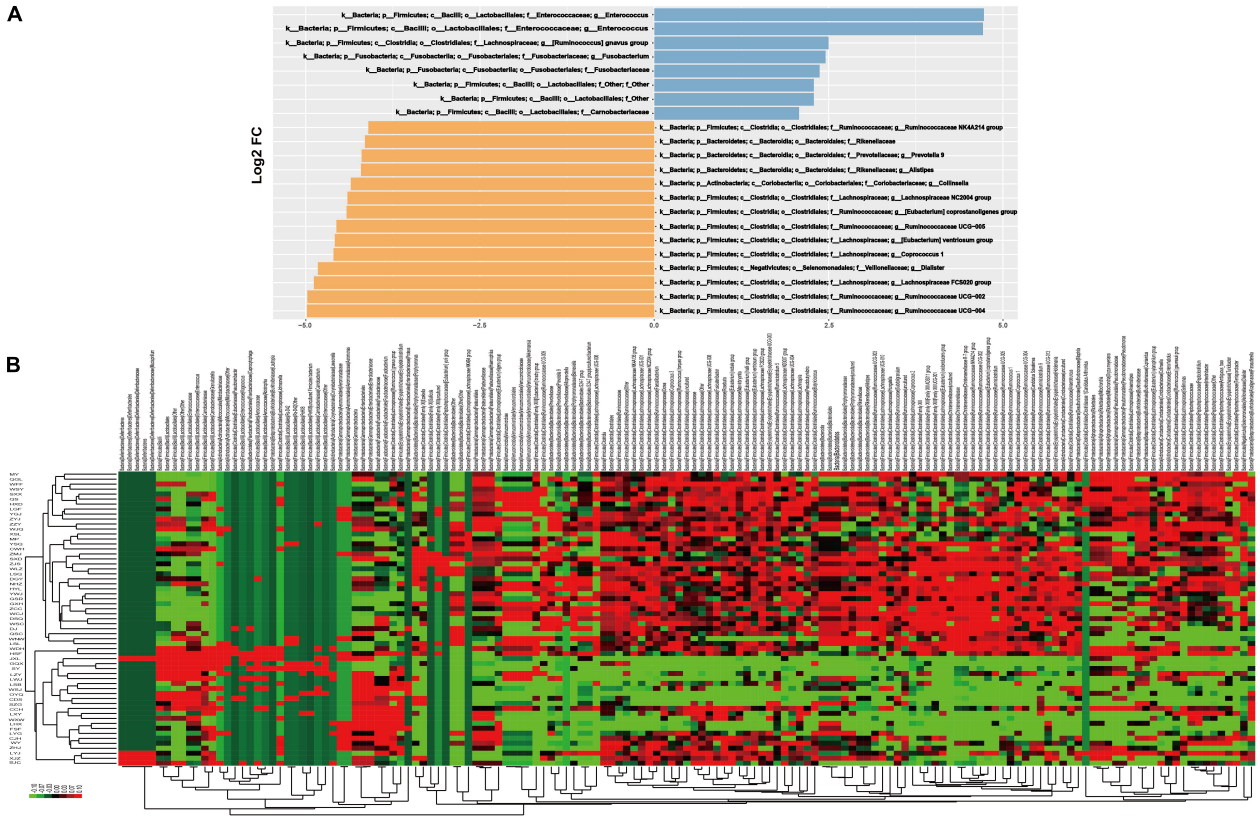


FIGURE 3 | The results of differential taxonomy expression analysis using limma algorithm. Panel **(A)** shows the differential taxonomy expression in the CHF patients vs. healthy control group; **(B)** shows the differential taxonomy expression in every samples.

Nadia and Ramana, 2020; Mulpuru et al., 2021). The α diversity index mainly focuses on the number of species in a local uniform habitat, reflecting the abundance and diversity of the microbial community. As the indicators of the community richness, the chao1 and PD-whole-tree indices were significantly decreased in the CHF patients versus control group, as well as the Shannon indices, which is an important index used to estimate the microbial diversity. β-diversity is a comparative analysis of the microbial community composition of different groups of samples, constituting the overall diversity together with α diversity. These significantly different indexes in the current study revealed that loss of gut flora biodiversity is associated with CHF.

In the further differential classification expression analysis using the limma algorithm, we are more focused on the differences in the genus level. Consistent with other experimental results (Kummen et al., 2018; Mayerhofer et al., 2020), the decreased abundance of the genera *Ruminococcaceae* and *Lachnospiraceae* were also discovered in our study, which are known for their ability to synthesize short-chain fatty acids (SCFA) through the fermentation of dietary polysaccharides. In addition, we firstly reported that the decrease of genra *Dialister* and the increase of the genera in *Enterococcus* and *Enterococcaceae* were also the most notable features in CHF patients. *Dialister* is one of the most representative types of intestinal flora associated with irritable bowel syndrome (Lopetuso et al., 2018). The main products are lactic acid, acetic acid and formic acid, and those are also SCFAs. *Enterococcus* and *Enterococcaceae* (Byappanahalli et al., 2012; Gouba et al., 2020; Xie et al., 2021) are conditional pathogens, causing infections such as urinary tract infections, purulent abdominal infections, sepsis, endocarditis and diarrhea. On the other hand, their microbial preparations can enhance the activity of macrophage cells and promote the immune response. At the same time, due to their metabolism to produce lactic acid, they can form a lactic acid barrier to resist foreign pathogenic microorganisms. The above results suggested that the intestinal flora may participate in the occurrence and development of heart failure through the action of SCFAs. As we all know, SCFAs play an important role in the regulation of inflammation, which is definitely involved in the pathophysiological process of CHF. SCFAs can reduce the production of inflammatory factors by activating GPR41/43 (Li et al., 2018; Weber et al., 2018; Onyszkiewicz et al., 2019). Acetic acid may reduce the production of interleukin-6 and interleukin-8 (Sah et al., 2013; Yu et al., 2015), and butyric acid and propionic acid can reduce the production of interleukin-6 (Esquivel-Rendon et al., 2019; Yue et al., 2020). Butyrate plays an anti-inflammatory role through inducing Foxp3 + Treg cell proliferation and suppressing the generation of Th17 cells by activating G protein-coupled receptor

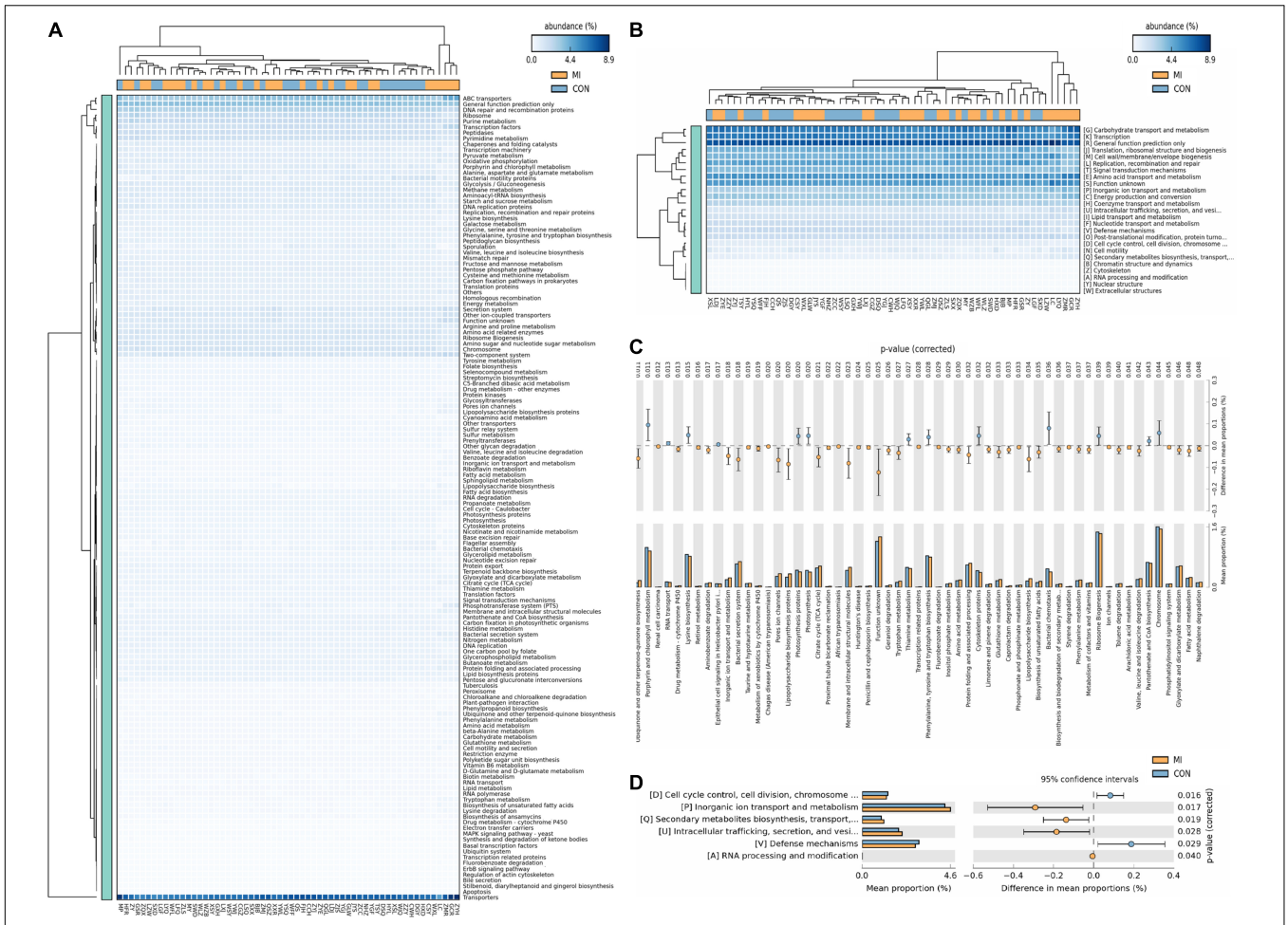


FIGURE 4 | The results of PICRUst based on closed-reference OTU to predict the abundances of functional categories COG orthologs (COs) and KEGG orthologs (KOs). Panels **(A,C)** shows the KOs with significantly different abundances in the fecal microbiome between the CHF group and healthy control group; **(B,D)** shows the COs with significantly different abundances in the fecal microbiome between the CHF group and healthy control group. HF, CHF group; CON, the healthy control group.

43 (Sivaprakasam et al., 2016; Bhaskaran et al., 2018). Other studies have also shown that SCFAs can also improve insulin sensitivity, regulate fat and muscle energy metabolism, and play an important role in the development of diabetes and obesity (Canfora et al., 2015; Muller et al., 2019).

The consequent results from analysis of predictive function using PICRUSt revealed several functional pathways involved in the relationship between gut microbiomes and CHF, including cell cycle control, cell division, chromosome partitioning, inorganic ion transport and metabolism, ribosomal structure and biogenesis, amino acid transport and metabolism and carbohydrate transport and metabolism. The occurrence and development of heart failure are inseparable from the disorders of carbohydrate metabolism, amino acids metabolism and lipid metabolism. The aforementioned differential classification expression analysis has indicated a notable reduction in SCFA-producing bacteria in patients with severe CHF. As a matter of fact, SCFAs can not only modulate the carbohydrate metabolism through activating G protein-coupled receptor and AMP-activated protein kinase, improving insulin sensitivity (Gao et al., 2009; den Besten et al., 2013), but also increasing the production of ketogenic amino acids and ketone bodies (Thevenet et al., 2016; Pujol et al., 2018), which is considered to be one of the energy sources of failing myocardium and closely related to the process of heart failure. Furthermore, as a histone deacetylase inhibitor, SCFAs can partly regulate cell proliferation, apoptosis and differentiation by inhibiting histone deacetylase, as well as exert anti-inflammatory effects (Koh et al., 2016; Makki et al., 2018; Alrafas et al., 2020).

Some limitations should be acknowledged. First, although all participants were from the same region, experienced a normal/routine lifestyle and had similar nutritional patterns, including typical Chinese diets based on carbohydrates versus high-fat diets, and participated in routine levels of general physical activity (e.g., housework and walking), we were still unable to completely account for the influence of diet on gut microbiota. Second, the enrolled cohort was a small sample size and predominantly male in patients with severe CHF.

Further studies informing the generalizability of gut microbiota in patients with CHF are warranted. Third, it is also necessary to address temporality of associations between gut dysbiosis and CHF. Finally, multiple omics data, such as metabolomics and proteomics, will be warranted to confirm the suppose that SCFAs produced by the gut microbiota participating the pathophysiological processes of CHF and explore the exact mechanisms.

In conclusion, the current results firstly exhibited remarkable differencesion the composition and diversity of the gut flora of severe CHF patients and healthy controls using bacterial 16S rRNA gene sequencing. The microflora of severe CHF patients was characterized by the decreased abundance of the SCFA-producing bacteria including genera *Ruminococcaceae UCG-004*, *Ruminococcaceae UCG-002*, *Lachnospiraceae FCS020 group*, *Dialister* and the increased abundance of the genera in *Enterococcus* and *Enterococcaceae* with an increased production of lactic acid. Moreover, the alternation of the gut microbiota was presumably associated with the function including cell cycle control, cell division, chromosome partitioning, amino acid transport and metabolism and carbohydrate transport and metabolism through SCFA pathway. This information may not only improve our understanding of the pathogenesis of severe CHF, but also suggest that the regulation of the composition of gut microbiota may represent a promising therapeutic target.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/supplementary material.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by The First Affiliated Hospital of Harbin Medical University and The Fourth Affiliated Hospital of Harbin Medical University approved all study protocols. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

YH, PL, and HJ conceived and designed the experiments. YH and TF analyzed data and drew the pictures. WS and DD wrote this manuscript. All authors read and approved the final manuscript.

FUNDING

The National Key Research and Development Program of China (2019YFC0118100), National Natural Science Foundation of China (82100353, 81671760, and 81873910), and Excellent youth project of the Fourth Affiliated Hospital of Harbin Medical University (No. HYDSYYXQN202007).

REFERENCES

Alrafas, H. R., Busbee, P. B., Chitrala, K. N., Nagarkatti, M., and Nagarkatti, P. (2020). Alterations in the gut microbiome and suppression of histone deacetylases by resveratrol are associated with attenuation of colonic inflammation and protection against colorectal Cancer. *J. Clin. Med.* 9:1796. doi: 10.3390/jcm9061796

Amoroso, C., Perillo, F., Strati, F., Fantini, M. C., Caprioli, F., and Facciotti, F. (2020). the role of gut microbiota biomodulators on mucosal immunity and intestinal inflammation. *Cells* 9:1234. doi: 10.3390/cells9051234

Anker, S. D., and Von Haehling, S. (2004). Inflammatory mediators in chronic heart failure: an overview. *Heart* 90, 464–470. doi: 10.1136/hrt.2002.007005

Aponte, M., Murru, N., and Shoukat, M. (2020). Therapeutic, prophylactic, and functional use of probiotics: a current perspective. *Front. Microbiol.* 11:562048. doi: 10.3389/fmicb.2020.562048

Bhaskaran, N., Quigley, C., Paw, C., Butala, S., Schneider, E., and Pandiyan, P. (2018). Role of short chain fatty acids in controlling tregs and immunopathology during mucosal infection. *Front. Microbiol.* 9:1995. doi: 10.3389/fmicb.2018.01995

Brial, F., Le Lay, A., Dumas, M. E., and Gauguier, D. (2018). Implication of gut microbiota metabolites in cardiovascular and metabolic diseases. *Cell Mol. Life Sci.* 75, 3977–3990. doi: 10.1007/s00018-018-2901-1

Byappanahalli, M. N., Nevers, M. B., Korajkic, A., Staley, Z. R., and Harwood, V. J. (2012). Enterococci in the environment. *Microbiol. Mol. Biol. Rev.* 76, 685–706.

Canfora, E. E., Jocken, J. W., and Blaak, E. E. (2015). Short-chain fatty acids in control of body weight and insulin sensitivity. *Nat. Rev. Endocrinol.* 11, 577–591. doi: 10.1038/nrendo.2015.128

Cheng, L., Qi, C., Yang, H., Lu, M., Cai, Y., Fu, T., et al. (2021). gutMGene: a comprehensive database for target genes of gut microbes and microbial metabolites. *Nucleic Acids Res.* gkab786. doi: 10.1093/nar/gkab786

Cheng, L., Qi, C., Zhuang, H., Fu, T., and Zhang, X. (2020). gutMDisorder: a comprehensive database for dysbiosis of the gut microbiota in disorders and interventions. *Nucleic Acids Res.* 48, D554–D560.

Cheng, L., Yang, H., Zhao, H., Pei, X., Shi, H., Sun, J., et al. (2019). MetSigDis: a manually curated resource for the metabolic signatures of diseases. *Brief. Bioinform.* 20, 203–209. doi: 10.1093/bib/bbx103

Cui, X., Ye, L., Li, J., Jin, L., Wang, W., Li, S., et al. (2018). Metagenomic and metabolomic analyses unveil dysbiosis of gut microbiota in chronic heart failure patients. *Sci. Rep.* 8:635. doi: 10.1038/s41598-017-18756-2

Dantzer, R., Cohen, S., Russo, S. J., and Dinan, T. G. (2018). Resilience and immunity. *Brain Behav. Immun.* 74, 28–42. doi: 10.1016/j.bbi.2018.08.010

Deledda, A., Annunziata, G., Tenore, G. C., Palmas, V., Manzin, A., and Velluzzi, F. (2021). Diet-derived antioxidants and their role in inflammation, obesity and gut microbiota modulation. *Antioxidants* 10:708. doi: 10.3390/antiox10050708

den Besten, G., Lange, K., Havinga, R., Van Dijk, T. H., Gerding, A., Van Eunen, K., et al. (2013). Gut-derived short-chain fatty acids are vividly assimilated into host carbohydrates and lipids. *Am. J. Physiol. Gastrointest. Liver Physiol.* 305, G900–G910. doi: 10.1152/ajpgi.00265.2013

Esquivel-Rendon, E., Vargas-Mireles, J., Cuevas-Olguin, R., Miranda-Morales, M., Acosta-Mares, P., Garcia-Oscos, F., et al. (2019). Interleukin 6 dependent synaptic plasticity in a social defeat-susceptible prefrontal cortex circuit. *Neuroscience* 414, 280–296. doi: 10.1016/j.neuroscience.2019.07.002

Gajarsa, J. J., and Kloner, R. A. (2011). Left ventricular remodeling in the post-infarction heart: a review of cellular, molecular mechanisms, and therapeutic modalities. *Heart. Fail Rev.* 16, 13–21. doi: 10.1007/s10741-010-9181-7

Gao, Z., Yin, J., Zhang, J., Ward, R. E., Martin, R. J., Lefevre, M., et al. (2009). Butyrate improves insulin sensitivity and increases energy expenditure in mice. *Diabetes* 58, 1509–1517. doi: 10.2337/db08-1637

Gouba, N., Yimagou, E. K., Hassani, Y., Drancourt, M., Fellag, M., and Mbogning Fonkou, M. D. (2020). *Enterococcus burkinafasonensis* sp. nov. isolated from

human gut microbiota. *New Microbes New Infect.* 36:100702. doi: 10.1016/j.nmni.2020.100702

Hage, C., Michaelsson, E., Linde, C., Donal, E., Daubert, J. C., Gan, L. M., et al. (2017). Inflammatory biomarkers predict heart failure severity and prognosis in patients with heart failure with preserved ejection fraction: a holistic proteomic approach. *Circ. Cardiovasc. Genet.* 10:e001633. doi: 10.1161/CIRCGENETICS.116.001633

Han, Y., Gong, Z., Sun, G., Xu, J., Qi, C., Sun, W., et al. (2021). Dysbiosis of gut microbiota in patients with acute myocardial infarction. *Front. Microbiol.* 12:680101. doi: 10.3389/fmicb.2021.680101

Harikrishnan, S. (2019). Diet, the gut microbiome and heart failure. *Card Fail Rev.* 5, 119–122. doi: 10.15420/cfr.2018.39.2

Huang, S. Y., Xiang, X., Qiu, L., Wang, L. Y., Zhu, B. H., Guo, R. Q., et al. (2020). Transfection of TGF-beta shRNA by using ultrasound-targeted microbubble destruction to inhibit the early adhesion repair of rats wounded achilles tendon in vitro and in vivo. *Curr. Gene Ther.* 20, 71–81. doi: 10.2174/1566523220666200516165828

Iqubal, A., Iqubal, M. K., Khan, A., Ali, J., Baboota, S., and Haque, S. E. (2020). Gene therapy, a novel therapeutic tool for neurological disorders: current progress, challenges and future prospective. *Curr. Gene Ther.* 20, 184–194. doi: 10.2174/1566523220999200716111502

Jia, Q., Xie, Y., Lu, C., Zhang, A., Lu, Y., Lv, S., et al. (2019). Endocrine organs of cardiovascular diseases: gut microbiota. *J. Cell Mol. Med.* 23, 2314–2323. doi: 10.1111/jcmm.14164

Kamo, T., Akazawa, H., Suda, W., Saga-Kamo, A., Shimizu, Y., Yagi, H., et al. (2017). Dysbiosis and compositional alterations with aging in the gut microbiota of patients with heart failure. *PLoS One* 12:e0174099. doi: 10.1371/journal.pone.0174099

Khan, A., Zahra, A., Mumtaz, S., Fatmi, M. Q., and Khan, M. J. (2020). Integrated in-silico analysis to study the role of microRNAs in the detection of chronic kidney diseases. *Curr. Bioinform.* 15, 144–154. doi: 10.2174/1574893614666190923115032

Koh, A., De Vadder, F., Kovatcheva-Datchary, P., and Backhed, F. (2016). From dietary fiber to host physiology: short-chain fatty acids as key bacterial metabolites. *Cell* 165, 1332–1345. doi: 10.1016/j.cell.2016.05.041

Kummen, M., Mayerhofer, C. C. K., Vestad, B., Broch, K., Awoyemi, A., Storm-Larsen, C., et al. (2018). Gut microbiota signature in heart failure defined from profiling of 2 independent cohorts. *J. Am. Coll. Cardiol.* 71, 1184–1186. doi: 10.1016/j.jacc.2017.12.057

Kwon, E., Cho, M., Kim, H., and Son, H. S. (2020). A study on host tropism determinants of influenza virus using machine learning. *Curr. Bioinform.* 15, 121–134. doi: 10.2174/1574893614666191104160927

Li, M., Van Esch, B., Henricks, P. A. J., Folkerts, G., and Garssen, J. (2018). The anti-inflammatory effects of short chain fatty acids on lipopolysaccharide- or tumor necrosis factor alpha-stimulated endothelial cells via activation of GPR41/43 and inhibition of HDACs. *Front. Pharmacol.* 9:533. doi: 10.3389/fphar.2018.00533

Long, J., Yang, H., Yang, Z., Jia, Q., Liu, L., Kong, L., et al. (2021). Integrated biomarker profiling of the metabolome associated with impaired fasting glucose and type 2 diabetes mellitus in large-scale Chinese patients. *Clin. Transl. Med.* 11:e432. doi: 10.1002/ctm2.432

Lopetuso, L. R., Petito, V., Graziani, C., Schiavoni, E., Paroni Sterbini, F., Poscia, A., et al. (2018). Gut microbiota in health, diverticular disease, irritable bowel syndrome, and inflammatory bowel diseases: time for microbial marker of gastrointestinal disorders. *Dig. Dis.* 36, 56–65. doi: 10.1159/000477205

Luedde, M., Winkler, T., Heinsen, F. A., Ruhlemann, M. C., Spehlmann, M. E., Bajrovic, A., et al. (2017). Heart failure is associated with depletion of core intestinal microbiota. *ESC Heart Fail* 4, 282–290. doi: 10.1002/ehf2.12155

Lv, H., Dao, F. Y., Zulfiqar, H., and Lin, H. (2021). DeepIPs: comprehensive assessment and computational identification of phosphorylation sites of SARS-CoV-2 infection using a deep learning-based approach. *Brief. Bioinform.* 22:bbab244. doi: 10.1093/bib/bbab244

Makki, K., Deehan, E. C., Walter, J., and Backhed, F. (2018). The impact of dietary fiber on gut microbiota in host health and disease. *Cell Host Microbe* 23, 705–715. doi: 10.1016/j.chom.2018.05.012

Mayerhofer, C. C. K., Awoyemi, A. O., Moscavitch, S. D., Lappegard, K. T., Hov, J. R., Aukrust, P., et al. (2018). Design of the GutHeart-targeting gut microbiota to treat heart failure-trial: a Phase II, randomized clinical trial. *ESC Heart Fail* 5, 977–984. doi: 10.1002/ehf2.12332

Mayerhofer, C. C. K., Kummen, M., Holm, K., Broch, K., Awoyemi, A., Vestad, B., et al. (2020). Low fibre intake is associated with gut microbiota alterations in chronic heart failure. *ESC Heart Fail* 7, 456–466. doi: 10.1002/ehf2.12596

Moshkelgosha, S., Masetti, G., Berchner-Pfannschmidt, U., Verhasselt, H. L., Horstmann, M., Diaz-Cano, S., et al. (2018). Gut microbiome in BALB/c and C57BL/6J mice undergoing experimental thyroid autoimmunity associate with differences in immunological responses and thyroid function. *Horm. Metab. Res.* 50, 932–941. doi: 10.1055/a-0653-3766

Muller, M., Hernandez, M. A. G., Goossens, G. H., Reijnders, D., Holst, J. J., Jocken, J. W. E., et al. (2019). Circulating but not faecal short-chain fatty acids are related to insulin sensitivity, lipolysis and GLP-1 concentrations in humans. *Sci. Rep.* 9:12515. doi: 10.1038/s41598-019-48775-0

Mulpuru, V., Semwal, R., Varadwaj, P. K., and Mishra, N. (2021). HAMP: a knowledgebase of antimicrobial peptides from human microbiome. *Curr. Bioinform.* 16, 534–540. doi: 10.2174/1574893615999200802041228

Nadia, and Ramana, J. (2020). The human OncoBiome database: a database of cancer microbiome datasets. *Curr. Bioinform.* 15, 472–477. doi: 10.2174/1574893614666190902152727

Nagatomo, Y., and Tang, W. H. (2015). Intersections between microbiome and heart failure: revisiting the gut hypothesis. *J. Card Fail* 21, 973–980. doi: 10.1016/j.cardfail.2015.09.017

Onyszkiewicz, M., Gawrys-Kopczynska, M., Konopelski, P., Aleksandrowicz, M., Sawicka, A., Kozniewska, E., et al. (2019). Butyric acid, a gut bacteria metabolite, lowers arterial blood pressure via colon-vagus nerve signaling and GPR41/43 receptors. *Pflugers Arch.* 471, 1441–1453. doi: 10.1007/s00424-019-02322-y

Pujol, J. B., Christinat, N., Ratinaud, Y., Savoia, C., Mitchell, S. E., and Dioum, E. H. M. (2018). Coordination of GPR40 and ketogenesis signaling by medium chain fatty acids regulates beta cell function. *Nutrients* 10:473. doi: 10.3390/nu10040473

Qi, C., Wang, P., Fu, T., Lu, M., Cai, Y., Chen, X., et al. (2021). A comprehensive review for gut microbes: technologies, interventions, metabolites and diseases. *Brief. Funct. Genomics* 20, 42–60. doi: 10.1093/bfgp/elaa029

Sah, S. K., Kim, B. H., Park, G. T., Kim, S., Jang, K. H., Jeon, J. E., et al. (2013). Novel isonahocol E(3) exhibits anti-inflammatory and anti-angiogenic effects in endothelin-1-stimulated human keratinocytes. *Eur. J. Pharmacol.* 720, 205–211. doi: 10.1016/j.ejphar.2013.09.077

Sanchez-Rodriguez, E., Egea-Zorrilla, A., Plaza-Diaz, J., Aragon-Vela, J., Munoz-Quezada, S., Tercedor-Sanchez, L., et al. (2020). The gut microbiota and its implication in the development of atherosclerosis and related cardiovascular diseases. *Nutrients* 12:605. doi: 10.3390/nu12030605

Sandek, A., Bjarnason, I., Volk, H. D., Crane, R., Meddings, J. B., Niebauer, J., et al. (2012). Studies on bacterial endotoxin and intestinal absorption function in patients with chronic heart failure. *Int. J. Cardiol.* 157, 80–85. doi: 10.1016/j.ijcard.2010.12.016

Shanahan, F. (2010). Probiotics in perspective. *Gastroenterology* 139, 1808–1812. doi: 10.1053/j.gastro.2010.10.025

Shin, N. R., Whon, T. W., and Bae, J. W. (2015). *Proteobacteria*: microbial signature of dysbiosis in gut microbiota. *Trends Biotechnol.* 33, 496–503. doi: 10.1016/j.tibtech.2015.06.011

Sivaprakasam, S., Prasad, P. D., and Singh, N. (2016). Benefits of short-chain fatty acids and their receptors in inflammation and carcinogenesis. *Pharmacol. Ther.* 164, 144–151. doi: 10.1016/j.pharmthera.2016.04.007

Tang, W., Wan, S., Yang, Z., Teschendorff, A. E., and Zou, Q. (2018). Tumor origin detection with tissue-specific miRNA and DNA methylation markers. *Bioinformatics* 34, 398–406. doi: 10.1093/bioinformatics/btx622

Tang, W. H., Kitai, T., and Hazen, S. L. (2017). Gut microbiota in cardiovascular health and disease. *Circ. Res.* 120, 1183–1196.

Tang, W. H. W., Li, D. Y., and Hazen, S. L. (2019). Dietary metabolism, the gut microbiome, and heart failure. *Nat. Rev. Cardiol.* 16, 137–154. doi: 10.1038/s41569-018-0108-7

Thevenet, J., De Marchi, U., Domingo, J. S., Christinat, N., Bultot, L., Lefebvre, G., et al. (2016). Medium-chain fatty acids inhibit mitochondrial metabolism in astrocytes promoting astrocyte-neuron lactate and ketone body shuttle systems. *FASEB J.* 30, 1913–1926. doi: 10.1096/fj.201500182

Weber, G. J., Foster, J., Pushpakumar, S. B., and Sen, U. (2018). Altered microRNA regulation of short chain fatty acid receptors in the hypertensive kidney is

normalized with hydrogen sulfide supplementation. *Pharmacol. Res.* 134, 157–165. doi: 10.1016/j.phrs.2018.06.012

Xie, A., Song, J., Lu, S., Liu, Y., Tang, L., and Wen, S. (2021). Influence of diet on the effect of the probiotic lactobacillus paracasei in rats suffering from allergic asthma. *Front. Microbiol.* 12:737622. doi: 10.3389/fmicb.2021.737622

Yang, H., Luo, Y., Ren, X., Wu, M., He, X., Peng, B., et al. (2021). Risk prediction of diabetes: big data mining with fusion of multifarious physical examination indicators. *Inform. Fusion* 75, 140–149. doi: 10.1016/j.inffus.2021.02.015

Yu, Y., Jia, T. Z., Cai, Q., Jiang, N., Ma, M. Y., Min, D. Y., et al. (2015). Comparison of the anti-ulcer activity between the crude and bran-processed *Atractylodes lancea* in the rat model of gastric ulcer induced by acetic acid. *J. Ethnopharmacol.* 160, 211–218. doi: 10.1016/j.jep.2014.10.066

Yue, Y., He, Z., Zhou, Y., Ross, R. P., Stanton, C., Zhao, J., et al. (2020). Lactobacillus plantarum relieves diarrhea caused by enterotoxin-producing *Escherichia coli* through inflammation modulation and gut microbiota regulation. *Food Funct.* 11, 10362–10374. doi: 10.1039/d0fo02670k

Zhang, X. Y., Shi, S. P., Shen, J., Zhao, M. Y., and He, Q. N. (2019). Functional immunoregulation by heme oxygenase 1 in juvenile autoimmune diseases. *Curr. Gene Ther.* 19, 110–116. doi: 10.2174/1566523219666190710092935

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.

frontiers in Microbiology

PERSPECTIVE
published: 24 February 2022
doi: 10.3389/fmicb.2022.798917

Interactions Between Intestinal Microbiota and Neural Mitochondria: A New Perspective on Communicating Pathway From Gut to Brain

Yao Zhu[1], Ying Li[2], Qiang Zhang[1], Yuanjian Song[1], Liang Wang[3,4] and Zuobin Zhu[1]**

[1] Xuzhou Engineering Research Center of Medical Genetics and Transformation, Key Laboratory of Genetic Foundation and Clinical Application, Department of Genetics, Xuzhou Medical University, Xuzhou, China, [2] Medical Technology College, Xuzhou Medical University, Xuzhou, China, [3] Department of Bioinformatics, School of Medical Informatics and Engineering, Xuzhou Medical University, Xuzhou, China, [4] Jiangsu Key Laboratory of New Drug Research and Clinical Pharmacy, School of Pharmacy, Xuzhou Medical University, Xuzhou, China*

Many studies shown that neurological diseases are associated with neural mitochondrial dysfunctions and microbiome composition alterations. Since mitochondria emerged from bacterial ancestors during endosymbiosis, mitochondria, and bacteria had analogous genomic characteristics, similar bioactive compounds and comparable energy metabolism pathways. Therefore, it is necessary to rationalize the interactions of intestinal microbiota with neural mitochondria. Recent studies have identified neural mitochondrial dysfunction as a critical pathogenic factor for the onset and progress of multiple neurological disorders, in which the non-negligible role of altered gut flora composition was increasingly noticed. Here, we proposed a new perspective of intestinal microbiota – neural mitochondria interaction as a communicating channel from gut to brain, which could help to extend the vision of gut-brain axis regulation and provide additional research directions on treatment and prevention of responsive neurological disorders.

Keywords: intestinal microbiome, mitochondria, microbiota-gut-brain axis, brain, gut

OPEN ACCESS

Edited by:
George Tsiamis,
University of Patras, Greece

Reviewed by:
Xin Zhang,
Ningbo University, China
Bing Gu,
Guangdong Provincial People's Hospital, China

****Correspondence:***
Zuobin Zhu
zhuzuobin@xzhmu.edu.cn
Liang Wang
healthscience@foxmail.com

Specialty section:
This article was submitted to Systems Microbiology, a section of the journal Frontiers in Microbiology

Received: *20 October 2021*
Accepted: *03 February 2022*
Published: *24 February 2022*

Citation:
Zhu Y, Li Y, Zhang Q, Song Y, Wang L and Zhu Z (2022) Interactions Between Intestinal Microbiota and Neural Mitochondria: A New Perspective on Communicating Pathway From Gut to Brain. Front. Microbiol. 13:798917. doi: 10.3389/fmicb.2022.798917

INTRODUCTION

Human body is a super organism composed of own cells and resident microorganisms. In the long-term co-evolutionary process, human gut microbes, and the hosts constantly selected and adapted to each other, bringing about a close symbiotic relationship presently (Moeller et al., 2016). While microbiota exist in many body sites such as oral cavity, vagina, airways, and skin, etc., we focused only on the gut microbiota in this study as its interplay with systemic health is the most extensively documented (van de Guchte et al., 2018). In consideration of its distributive peculiarity, the gut microbiota was primarily proposed to have specific interactions with the host digestive system, which served as the main study topic for past decades. In recent years, a mass of research has identified that gut microbiota and corresponding bacterial metabolites can target the brain through various pathways, such as nervous conduction (enteric nerve, vagus nerve, etc.) (Fulling et al., 2019), hypothalamic–pituitary–adrenal (HPA) axis (McNeilly et al., 2010), and enteric endocrine and immune response (Fung, 2020; Morais et al., 2020), etc., (**Figure 1**). However, the specific regulatory mechanisms in these channels remain largely unclear.

Brain is one of the most energy-consuming organ in the body (Karbowski, 2007). Neural mitochondria can not only provide energy for maintaining brain homeostasis, but are also central regulators of cognitive function as well as fate determinant for neural stem cells (Khacho et al., 2016; Iwata et al., 2020). It has been widely reported that mitochondrial dysfunction can accelerate senescence of neural cells and facilitate the onset of multiple neurological diseases (Nguyen et al., 2014). In parallelly, a large amount of evidence confirmed that gut microbiota composition played critical roles in regulating the physiological and pathological functions of the brain. Therefore, considering the common ancestries, similar mechanisms, similar goals, and similar structures between gut microbiota and mitochondria (Franco-Obregon and Gilbert, 2017), is it possible that neural mitochondria are direct targets of intestinal microflora and function as key mediators regulating gut-brain interaction?

MITOCHONDRIA HAVE A CLOSE RELATIONSHIP WITH BACTERIA

In Sagan (1967), first proposed the hypothesis that mitochondria evolved from bacteria. Currently, determining the nature of the bacterial lineages that gave rise to mitochondrial ancestors is still a hotly debated topic (**Figure 1**). Although many studies have already showed that mitochondrion originated from within the bacterial phylum *Alphaproteobacteria*, the phylogenetic relationship of the mitochondrial endosymbiont to extant *Alphaproteobacteria* is yet unclear (Fan et al., 2020), while other studies support the idea that mitochondria evolved from an ancestor related to *Rickettsiales* (Andersson et al., 1998; Wang and Wu, 2015). It is true that mitochondria have a bacterial origin and do share many proteins that mediate similar or even the same metabolic processes (Karlberg et al., 2000). Besides that, the use of antibiotics such as quinolones, aminoglycosides and poplar polysaccharide antibiotics can induce mitochondria dysfunction due to similarities in their structures with bacteria (Kalghatgi et al., 2013; Lleonart et al., 2017). For instance, that quinolones target mtDNA topoisomerases (Gootz et al., 1990) and bacterial gyrases (Wolfson and Hooper, 1985), aminoglycosides target both mitochondrial (Hutchin and Cortopassi, 1994) and bacterial ribosomes (Davis, 1987). Reversibly, mitochondrial-targeted antioxidants could also function as effective antibiotics (Nazarov et al., 2017). These studies suggested that mitochondria have a close relationship with bacteria, which indicates a possibility of information exchange between gut microbiota and mitochondria.

INTESTINAL MICROBIOTA DIRECTLY REGULATE BRAIN FUNCTION THROUGH INTESTINAL EPITHELIUM AND GASTROINTESTINAL NERVE

Intestinal epithelium and gastrointestinal nerves are the first sites of interactions between microbes and hosts (**Figure 2**). Moreover, many toxins produced by gut microbiota can lead to mitochondrial dysfunctions. For example, when the host was infected by pathogenic bacteria, mitochondria will be activated by lipopolysaccharides and other toxins released by the gut microbiota, inducing the accumulation of mitochondrial reactive oxygen species (ROS), which sequentially mediate intestinal inflammation (Mills et al., 2016). In addition, toxins secreted by certain species of *Clostridium* could inhibit the mitochondrial ATP-sensitive potassium channels, leading to mitochondrial membrane hyperpolarization, cell apoptosis, and intestinal epithelial barrier disruption (Matarrese et al., 2007; Berger et al., 2016; Crakes et al., 2019). The increased intestinal permeability enabled the translocation of damaging substances or pathogens into the intestinal epithelium and gastrointestinal nerves. Vagus nerve, an important link in the gut-brain axis, is able to sense microbial metabolites through its afferents, which transfers gut information to the central nervous system (Bonaz et al., 2018; Yu et al., 2020). Moreover, mitochondria are an important source of damaged cells release endogenous messengers (DAMPs), the release of these mitochondrial ROS have role of signaling in neuroinflammation and neurodegenerative diseases (Hsieh and Yang, 2013). DAMPs can also activate the innate immune system (Taanman, 1999), while innate immunity further reacts to different insults that may challenge the integrity of the central nervous system (CNS; Liu et al., 2014).

INTESTINAL MICROBIOTA REMOTE CONTROL THE PHYSIOLOGICAL AND PATHOLOGICAL FUNCTIONS OF THE BRAIN THROUGH THE MITOCHONDRIAL PATHWAY

Experimental evidence has been showing that metabolites produced by gut microbes interact directly with brain mitochondria, providing new insights into the communication pathway from gut to brain. In specificity, intestinal infection caused by lipopolysaccharides from gram-negative bacteria could trigger mitochondrial antigen presentation (MitAP) in autoimmune CD8 + T cells, which could enter the brain to attack dopamine neurons and cause a sharp decrement in the density of dopaminergic axon expansion in the striatum, leading to movement disorders like Parkinson's disease (PD) in mice (Matheoud et al., 2019). Effects of pathogenic bacteria on mitochondria can also include morphological and functional changes. Listeriolysin O (LLO) secreted by *Listeria monocytogenes*, which can cause CNS infection, can inserted into the plasma membrane, causing calcium influx and indirectly leading to mitochondrial fission (Stavru et al., 2011). *Helicobacter pylori*, an important pathogen that causes chronic gastritis in various areas of stomach and duodenum, could also cause nervous system inflammation and even promote the occurrence of Alzheimer's disease (AD). However, the relationship between *Helicobacter pylori* infection and neurological inflammation or AD remains unclear. It has been reported that toxin VacA secreted by *Helicobacter pylori* can migrate across the blood-brain barrier (Suzuki et al., 2019), which could also be

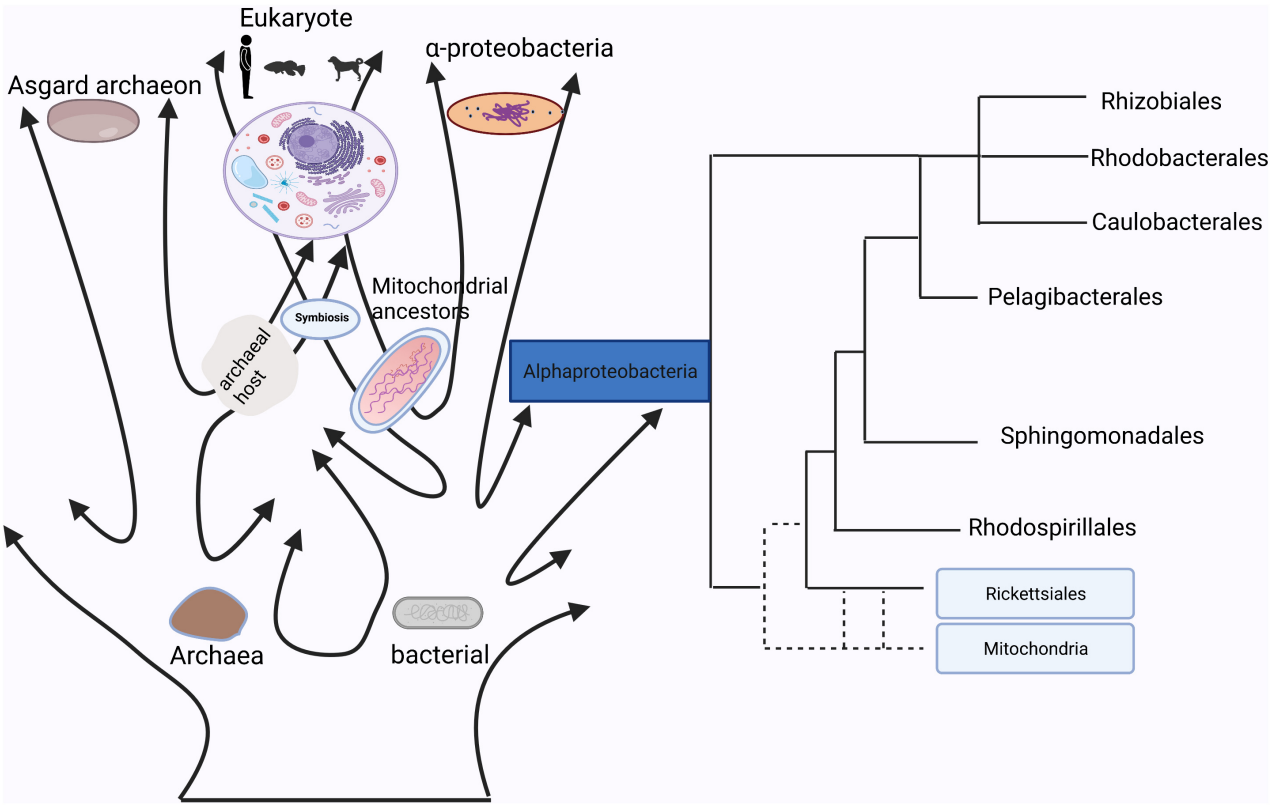


FIGURE 1 | The homology between mitochondria and bacteria.

inserted in the mitochondrial inner membrane, causing calcium influx and thus indirectly leading to mitochondrial fission (Foo et al., 2010). Since it was proposed that altered balance in mitochondrial fission might be an important mechanism of neuronal dysfunction in the brain tissue of AD patients (Wang et al., 2009; Manczak and Reddy, 2012), it could provide novel insights into how *Helicobacter pylori* targets neural mitochondria and how mitochondrial and neuronal dysfunctions evolve. A recent study also showed that aberrant mitochondria functionality could be a key mediator for the effects of the intestinal microbiota on the progression of depression (Chen and Vitetta, 2020). These studies supported the role that mitochondria plays as an emerging target for bacteria-induced neurological diseases.

A large amount of evidence suggest that gut microbiota can also remotely regulate the mitochondrial function of brain tissue through the various metabolites they produced (**Figure 3**). Short-chain fatty acids (SCFA) generated by gut microbiota can cross the highly selective semipermeable blood-brain barrier (Luu and Visekruna, 2019; Melbye et al., 2019) and directly enter the mitochondria for further oxidative metabolism (Chen and Vitetta, 2020). In addition, supplementation of propionic acid (PA), could defer the progression of Multiple sclerosis (MS) and brain atrophy (Duscha et al., 2020). Since PA can improve mitochondrial function and morphology in competent regulatory T (Treg) cells, and can enter the brain directly, it is rationally speculated that the protective effect of PA on brain tissue may be achieved by improving neural mitochondrial function. Folate produced by gut flora (mainly *Escherichia coli*) could regulate mitochondrial respiration and play an important role in the early development of the nervous system by activating the mammalian target of rapamycin (mTOR) signaling pathway (Silva et al., 2017). Isoallolithocholic acid (IsoalloLCA), distinct derivatives of lithocholic acid, which is also closely related to nervous system diseases, can also induce the production of mitochondrial ROS (Hang et al., 2019). The gut microbiota metabolites 4-(trimethylammonio) pentanoate valerate and 3-methyl-4-(trimethylammonio) butanoate could enter the brain tissue and inhibit the oxidation of mitochondrial fatty acids so as to mediate gut-brain communication (Hulme et al., 2020). Another study found that trimethylamine-N-oxide can also increase mitochondrial damage and superoxide production in mice, thereby accelerating the aging of neurons in the hippocampus of mice, causing and exacerbating aging-related cognitive impairment (Li et al., 2018). Therefore, it is of great theoretical significance to clarify the underlying pathways that gut microbiota metabolites affect brain function by regulating mitochondrial bioactivities, and to reveal how gut microbiota regulate the neuronal functions through dietary metabolism. These studies may provide new drug targets for the ontological enteric treatment of encephalopathy.

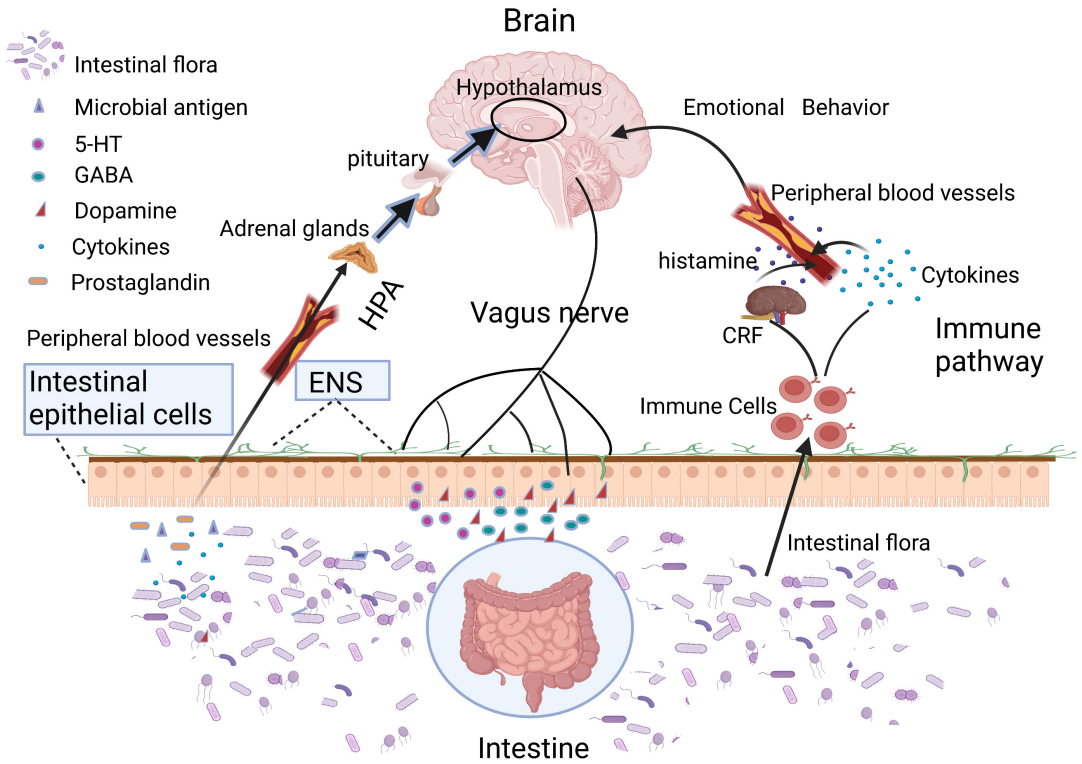


FIGURE 2 | Intestinal microbiota or its metabolites can directly regulate brain function through HPA axis, vagus nerve, and immune pathway.

MITOCHONDRIA PLAY AN IMPORTANT ROLE IN GUT MICROBIOTA–NEUROTRANSMITTERS-BRAIN COMMUNICATION

Another strategy of gut microbiota affecting the host's nervous system is to regulate the host's neurotransmitter levels, such as gamma-aminobutyric acid (GABA), serotonin (5-HT), and dopamine (DA). These neurotransmitters have been found to be closely in mitochondrial function. For example, GABA can pass through mitochondrial membrane, and regulate citric acid cycle reaction. The distribution mode of GABA is believed to play a critical role in regulating its cytoplasm levels; Conversely, increased mitochondrial activity can reduce GABAergic signaling, resulting in defective social behavior (Kanellopoulos et al., 2020). Recent studies have also shown that 5-HT could promote mitochondrial biogenesis, which is involved in reducing toxic ROS in neurons, protecting buffered neurons from damages caused by cellular stress (Fanibunda et al., 2019). Dopamine has been reported to be associated with mitochondrial dysfunction. Experimental evidence suggests that high dopamine concentrations induce striatal mitochondrial dysfunction through a decrease in mitochondrial respiratory control and loss of membrane potential (Czerniczyniec et al., 2010) and promotes mitochondrial complex I inhibition and leads to neuropsychiatric disorders (Ben-Shachar et al., 2004; Brenner-Lavie et al., 2008). These studies suggested that mitochondria also play an important role in gut microbiota-neurotransmitters-brain communication.

MITOCHONDRIA PLAY AN IMPORTANT ROLE IN SIGNAL TRANSMISSION FROM BRAIN TO GUT

Reciprocally, mitochondria can also regulate the intestinal microbiota. Studies have shown that mitochondria play an important role in the innate immune response to pathogen infection (Lobet et al., 2015). In addition, mitochondria dysfunction also involves in the regulation of the gut epithelial barrier, allowing transepithelial flux of *Escherichia coli* (Wang et al., 2014). In addition, mitochondrial variants can affect the diversity and composition of intestinal microbiota (Evaldson et al., 1980; Ma et al., 2014; Yardeni et al., 2019). Moreover, mitochondrial chaperone HSP-60 in the neurons regulates anti-bacterial immunity *via* p38 MAP kinase signaling (Jeong et al., 2017). Clinical studies have shown that a large proportion of patients with neurodegenerative diseases, such as Alzheimer's disease (AD; Haran et al., 2019; Sochocka et al., 2019), Parkinson's disease (PD; Brudek, 2019; Dumitrescu et al., 2021), and Huntington's disease (HD; Du et al., 2020; Wasser et al., 2020), often suffer from intestinal inflammation simultaneously. A recent report found that

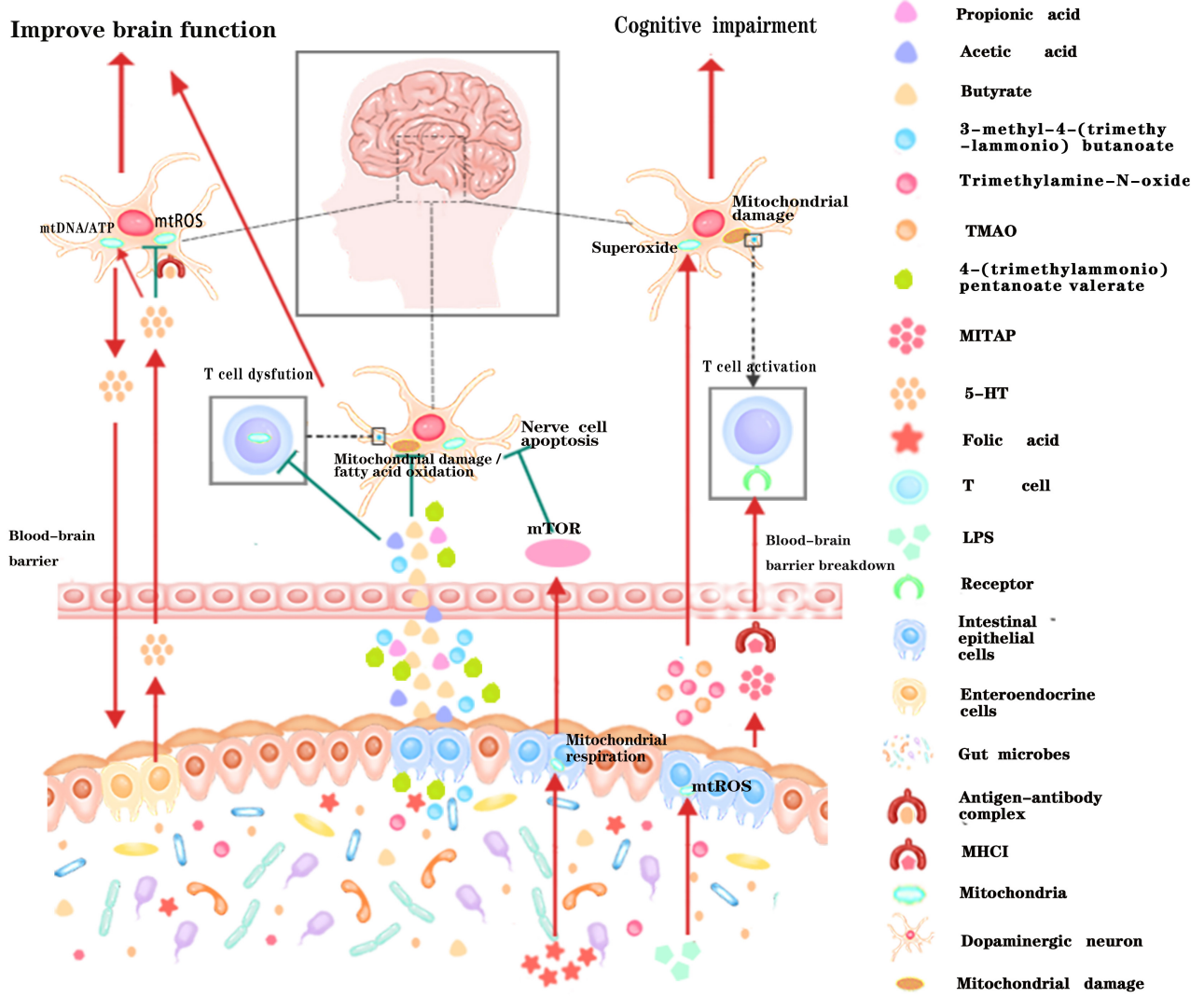


FIGURE 3 | Schematic illustration of the dialog between intestinal microbiota and neural mitochondria across the blood-brain barrier. Metabolites or active small molecules secreted by intestinal microbiota cross the blood-brain barrier and modulate mitochondrial functions through mTOR signaling pathway, ROS signaling pathway, immune pathway, or directly acting on neural mitochondria to influence brain functions. Red arrows indicate enhancement while the green T arrows indicate inhibition.

artificial expression of HD-causing protein PolyQ40 in nerve cells of *Caenorhabditis elegans* can induce mitochondrial unfolded protein response in the intestine (Zhang et al., 2018). These results further indicate that mitochondria play an important role in signal transmission between the brain and the gut. Another study found that 5-HT can regulate the colonization of *Turicibacter sanguinis* in the intestine, and *Turicibacter sanguinis* can affect the expression of multiple pathways including lipid and cholesterol metabolism in intestinal (Fung et al., 2019). Combined with the evidence that 5-HT can promote mitochondrial biogenesis and the level of ROS produced by host mitochondria can regulate the diversity of gut flora (Fanibunda et al., 2019), we can deduce that neuron secreted 5-HT regulate the gut flora by regulating the function of intestinal mitochondria. These results suggest that mitochondria mediate two-way communication between gut and brain.

CONCLUSION

The understanding of intestinal system has been revolutionized over the past decades, especially in regarding to its physiological and pathological interconnection with brain function. The crosstalk between gut microbiota and central nervous system, which is also known as the microbiota-gut-brain axis have been well elucidated from numerous studies. Numbers of evidence has confirmed that mitochondria can be regulated by the composition of gut microbiota, and mitochondria are also closely related to the physiological and pathological state of the nervous

system. In addition, although there are few studies on how gut microbiota directly regulate the physiological and pathological functions of the brain through the mitochondrial pathway, or how the nervous system regulates the composition of gut microbiota through the mitochondrial pathway, the evidence had been gradually reported in recent years. This perspective proposed a hypothetical model about cross-talk between the intestinal microbiome and the neural mitochondria based on the previously known fact that the mitochondria and the bacteria have the evolutionary homology. Symbiont and pathobiont bacteria have the influence to control the neuronal mitochondrial activity. We highlighted the new role of mitochondria in dialog with gut microbiota across the blood-brain barrier, which is one of the important ways of communicating between the brain and gut (**Figure 3**). The new perspective not only expands our understanding of the brain-gut interaction mechanism, but also provides a new treatment strategy targeting the gut microbiota-mitochondria-brain communication which has the potential to treat a variety of nervous system diseases or digestive system diseases, and may also have a profound impact on future medical treatment.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding authors.

AUTHOR CONTRIBUTIONS

YZ performed the statistical analysis. ZZ, LW, YL, QZ, and YS wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the National Natural Science Foundation of China (81902040 and 81701390), the Jiangsu Province Postgraduate Research and Practice Innovation Project (KYCX20-2464 and KYCX21-2642), the Natural Science Foundation of Jiangsu Province (BK20170250), the Xuzhou Science and Technology Innovation Project (KC19057). Fusion Innovation Project of Xuzhou Medical University (XYRHCX). LW would like to thank the financial support of National Natural Science Foundation of China (Nos. 31900022 and 32171281), the Natural Science Foundation of Jiangsu Province (No. BK20180997), the Young Science and Technology Innovation Team of Xuzhou Medical University (No. TD202001) and the Jiang-Su Qing-Lan Project (2020).

ACKNOWLEDGMENTS

Thanks are due to Prof. Shi Huang for fruitful discussions.

REFERENCES

Andersson, S. G., Zomorodipour, A., Andersson, J. O., Sicheritz-Ponten, T., Alsmark, U. C., Podowski, R. M., et al. (1998). The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. *Nature* 396, 133–140. doi: 10.1038/24094

Ben-Shachar, D., Zuk, R., Gazawi, H., and Ljubuncic, P. (2004). Dopamine toxicity involves mitochondrial complex i inhibition: implications to dopamine-related neuropsychiatric disorders. *Biochem. Pharmacol.* 67, 1965–1974. doi: 10.1016/j.bcp.2004.02.015

Berger, E., Rath, E., Yuan, D., Waldschmitt, N., Khaloian, S., Allgauer, M., et al. (2016). Mitochondrial function controls intestinal epithelial stemness and proliferation. *Nat. Commun.* 7:13171. doi: 10.1038/ncomms13171

Bonaz, B., Bazin, T., and Pellissier, S. (2018). The vagus nerve at the interface of the microbiota-gut-brain axis. *Front. Neurosci.* 12:49. doi: 10.3389/fnins.2018.00049

Brenner-Lavie, H., Klein, E., Zuk, R., Gazawi, H., Ljubuncic, P., and Ben-Shachar, D. (2008). Dopamine modulates mitochondrial function in viable SH-SY5Y cells possibly via its interaction with complex I: relevance to dopamine pathology in schizophrenia. *Biochim. Biophys. Acta* 1777, 173–185. doi: 10.1016/j.bbabio.2007.10.006

Brudek, T. (2019). Inflammatory bowel diseases and Parkinson's disease. *J. Parkinsons Dis.* 9, S331–S344.

Chen, J., and Vitetta, L. (2020). Mitochondria could be a potential key mediator linking the intestinal microbiota to depression. *J. Cell. Biochem.* 121, 17–24. doi: 10.1002/jcb.29311

Crakes, K. R., Santos Rocha, C., Grishina, I., Hirao, L. A., Napoli, E., Gaulke, C. A., et al. (2019). PPARalpha-targeted mitochondrial bioenergetics mediate repair of intestinal barriers at the host-microbe intersection during SIV infection. *Proc. Natl. Acad. Sci. U.S.A.* 3, 24819–24829. doi: 10.1073/pnas.1908977116

Czerniczyniec, A., Bustamante, J., and Lores-Arnaiz, S. (2010). Dopamine modifies oxygen consumption and mitochondrial membrane potential in striatal mitochondria. *Mol. Cell. Biochem.* 341, 251–257. doi: 10.1007/s11010-010-0456-z

Davis, B. D. (1987). Mechanism of bactericidal action of aminoglycosides. *Microbiol. Rev.* 51, 341–350.

Du, G., Dong, W., Yang, Q., Yu, X., Ma, J., Gu, W., et al. (2020). Altered gut microbiota related to inflammatory responses in patients with huntington's disease. *Front. Immunol.* 11:603594. doi: 10.3389/fimmu.2020.603594

Dumitrescu, L., Marta, D., Danau, A., Lefter, A., Tulba, D., Cozma, L., et al. (2021). Serum and fecal markers of intestinal inflammation and intestinal barrier permeability are elevated in Parkinson's disease. *Front. Neurosci.* 15:689723. doi: 10.3389/fnins.2021.689723

Duscha, A., Gisevius, B., Hirschberg, S., Yissachar, N., Stangl, G. I., Eilers, E., et al. (2020). Propionic acid shapes the multiple sclerosis disease course by an immunomodulatory mechanism. *Cell* 180, 1067–1080.e16. doi: 10.1016/j.cell.2020.02.035

Evaldson, G., Carlstrom, G., Lagrelius, A., Malmborg, A. S., and Nord, C. E. (1980). Microbiological findings in pregnant women with premature rupture of the membranes. *Med. Microbiol. Immunol.* 168, 283–297. doi: 10.1007/BF02121812

Fan, L., Wu, D., Goremykin, V., Xiao, J., Xu, Y., Garg, S., et al. (2020). Phylogenetic analyses with systematic taxon sampling show that mitochondria branch within *Alphaproteobacteria*. *Nat. Ecol. Evol.* 4, 1213–1219. doi: 10.1038/s41559-020-1239-x

Fanibunda, S. E., Deb, S., Maniyadath, B., Tiwari, P., Ghai, U., Gupta, S., et al. (2019). Serotonin regulates mitochondrial biogenesis and function in rodent cortical neurons via the 5-HT2A receptor and SIRT1-PGC-1alpha axis. *Proc. Natl. Acad. Sci. U.S.A.* 116:11028. doi: 10.1073/pnas.1821332116

Foo, J. H., Culvenor, J. G., Ferrero, R. L., Kwok, T., Lithgow, T., and Gabriel, K. (2010). Both the p33 and p55 subunits of the *Helicobacter pylori* VacA toxin are targeted to mammalian mitochondria. *J. Mol. Biol.* 401, 792–798. doi: 10.1016/j.jmb.2010.06.065

Franco-Obregon, A., and Gilbert, J. A. (2017). The Microbiome-mitochondrion connection: common ancestries, common mechanisms, common goals. *mSystems* 2:e00018-17. doi: 10.1128/mSystems.00018-17

Fulling, C., Dinan, T. G., and Cryan, J. F. (2019). Gut microbe to brain signaling: what happens in vagus. *Neuron* 101, 998–1002. doi: 10.1016/j.neuron.2019.02.008

Fung, T. C. (2020). The microbiota-immune axis as a central mediator of gut-brain communication. *Neurobiol. Dis.* 136:104714. doi: 10.1016/j.nbd.2019.104714

Fung, T. C., Vuong, H. E., Luna, C. D. G., Pronovost, G. N., Aleksandrova, A. A., Riley, N. G., et al. (2019). Intestinal serotonin and fluoxetine exposure modulate bacterial colonization in the gut. *Nat. Microbiol.* 4, 2064–2073. doi: 10.1038/s41564-019-0540-4

Gootz, T. D., Barrett, J. F., and Sutcliffe, J. A. (1990). Inhibitory effects of quinolone antibacterial agents on eucaryotic topoisomerases and related test systems. *Antimicrob. Agents Chemother.* 34, 8–12. doi: 10.1128/AAC.34.1.8

Hang, S., Paik, D., Yao, L., Kim, E., Trinath, J., Lu, J., et al. (2019). Bile acid metabolites control T_H17 and T_{reg} cell differentiation. *Nature* 576, 143–148.

Haran, J. P., Bhattarai, S. K., Foley, S. E., Dutta, P., Ward, D. V., Bucci, V., et al. (2019). Alzheimer's disease microbiome is associated with dysregulation of the anti-inflammatory P-glycoprotein pathway. *mBio* 10:e00632-19. doi: 10.1128/mBio.00632-19

Hsieh, H. L., and Yang, C. M. (2013). Role of redox signaling in neuroinflammation and neurodegenerative diseases. *Biomed. Res. Int.* 2013:484613. doi: 10.1155/2013/484613

Hulme, H., Meikle, L. M., Strittmatter, N., van der Hooft, J. J. J., Swales, J., Bragg, R. A., et al. (2020). Microbiome-derived carnitine mimics as previously unknown mediators of gut-brain axis communication. *Sci. Adv.* 6:eaax6328. doi: 10.1126/sciadv.aax6328

Hutchin, T., and Cortopassi, G. (1994). Proposed molecular and cellular mechanism for aminoglycoside ototoxicity. *Antimicrob. Agents Chemother.* 38, 2517–2520. doi: 10.1128/AAC.38.11.2517

Iwata, R., Casimir, P., and Vanderhaeghen, P. (2020). Mitochondrial dynamics in postmitotic cells regulate neurogenesis. *Science* 369, 858–862. doi: 10.1126/science.aba9760

Jeong, D. E., Lee, D., Hwang, S. Y., Lee, Y., Lee, J. E., Seo, M., et al. (2017). Mitochondrial chaperone HSP-60 regulates anti-bacterial immunity via p38 MAP kinase signaling. *EMBO J.* 36, 1046–1065. doi: 10.15252/embj.201694781

Kalghatgi, S., Spina, C. S., Costello, J. C., Liesa, M., Morones-Ramirez, J. R., Slomovic, S., et al. (2013). Bactericidal antibiotics induce mitochondrial dysfunction and oxidative damage in Mammalian cells. *Sci. Transl. Med.* 5:192ra185. doi: 10.1126/scitranslmed.3006055

Kanellopoulos, A. K., Mariano, V., Spinazzi, M., Woo, Y. J., McLean, C., Pech, U., et al. (2020). Aralar sequesters GABA into hyperactive mitochondria, causing social behavior deficits. *Cell* 180, 1178–1197.e20 doi: 10.1016/j.cell.2020.02.044

Karbowski, J. (2007). Global and regional brain metabolic scaling and its functional consequences. *BMC Biol.* 5:18. doi: 10.1186/1741-7007-5-18

Karlberg, O., Canback, B., Kurland, C. G., and Andersson, S. G. (2000). The dual origin of the yeast mitochondrial proteome. *Yeast* 17, 170–187. doi: 10.1002/1097-0061(20000930)17:3<170::AID-YEA25>3.0.CO;2-V

Khacho, M., Clark, A., Svoboda, D. S., Azzi, J., MacLaurin, J. G., Meghaizel, C., et al. (2016). Mitochondrial dynamics impacts stem cell identity and fate decisions by regulating a nuclear transcriptional program. *Cell Stem Cell.* 19, 232–247. doi: 10.1016/j.stem.2016.04.015

Li, D., Ke, Y., Zhan, R., Liu, C., Zhao, M., Zeng, A., et al. (2018). Trimethylamine-N-oxide promotes brain aging and cognitive impairment in mice. *Aging Cell* 17:e12768. doi: 10.1111/acel.12768

Liu, H. Y., Chen, C. Y., and Hsueh, Y. P. (2014). Innate immune responses regulate morphogenesis and degeneration: roles of Toll-like receptors and Sarm1 in neurons. *Neurosci. Bull.* 30, 645–654. doi: 10.1007/s12264-014-1445-5

Lleonart, M. E., Grodzicki, R., Graifer, D. M., and Lyakhovich, A. (2017). Mitochondrial dysfunction and potential anticancer therapy. *Med. Res. Rev.* 37, 1275–1298. doi: 10.1002/med.21459

Lobet, E., Letesson, J. J., and Arnould, T. (2015). Mitochondria: a target for bacteria. *Biochem. Pharmacol.* 94, 173–185. doi: 10.1016/j.bcp.2015.02.007

Luu, M., and Visekruna, A. (2019). Short-chain fatty acids: bacterial messengers modulating the immunometabolism of T cells. *Eur. J. Immunol.* 49, 842–848. doi: 10.1002/eji.201848009

Ma, J., Coarfa, C., Qin, X., Bonnen, P. E., Milosavljevic, A., Versalovic, J., et al. (2014). mtDNA haplogroup and single nucleotide polymorphisms structure human microbiome communities. *BMC Genomics* 15:257. doi: 10.1186/1471-2164-15-257

Manczak, M., and Reddy, P. H. (2012). Abnormal interaction between the mitochondrial fission protein Drp1 and hyperphosphorylated tau in Alzheimer's disease neurons: implications for mitochondrial dysfunction and neuronal damage. *Hum. Mol. Genet.* 21, 2538–2547. doi: 10.1093/hmg/dds072

Matarrese, P., Falzano, L., Fabbri, A., Gambardella, L., Frank, C., Geny, B., et al. (2007). Clostridium difficile toxin B causes apoptosis in epithelial cells by thrilling mitochondria. Involvement of ATP-sensitive mitochondrial potassium channels. *J. Biol. Chem.* 282, 9029–9041. doi: 10.1074/jbc.M607614200

Matheoud, D., Cannon, T., Voisin, A., Penttinen, A. M., Ramet, L., Fahmy, A. M., et al. (2019). Intestinal infection triggers Parkinson's disease-like symptoms in Pink1(-/-) mice. *Nature* 571, 565–569. doi: 10.1038/s41586-019-1405-y

McNeilly, A. D., Macfarlane, D. P., O'Flaherty, E., Livingstone, D. E., Mitic, T., McConnell, K. M., et al. (2010). Bile acids modulate glucocorticoid metabolism and the hypothalamic-pituitary-adrenal axis in obstructive jaundice. *J. Hepatol.* 52, 705–711. doi: 10.1016/j.jhep.2009.10.037

Melbye, P., Olsson, A., Hansen, T. H., Sondergaard, H. B., and Bang Oturai, A. (2019). Short-chain fatty acids and gut microbiota in multiple sclerosis. *Acta Neurol. Scand.* 139, 208–219. doi: 10.1111/ane.13045

Mills, E. L., Kelly, B., Logan, A., Costa, A. S. H., Varma, M., Bryant, C. E., et al. (2016). Succinate dehydrogenase supports metabolic repurposing of mitochondria to drive inflammatory macrophages. *Cell* 167, 457–470.e13. doi: 10.1016/j.cell.2016.08.064

Moeller, A. H., Caro-Quintero, A., Mjungu, D., Georgiev, A. V., Lonsdorf, E. V., Muller, M. N., et al. (2016). Cospeciation of gut microbiota with hominids. *Science* 353, 380–382. doi: 10.1126/science.aaf3951

Morais, L. H., Schreiber, HLt, and Mazmanian, S. K. (2020). The gut microbiota-brain axis in behaviour and brain disorders. *Nat. Rev. Microbiol.* 19, 241–255 doi: 10.1038/s41579-020-00460-0

Nazarov, P. A., Osterman, I. A., Tokarchuk, A. V., Karakozova, M. V., Korshunova, G. A., Lyamzaev, K. G., et al. (2017). Mitochondria-targeted antioxidants as highly effective antibiotics. *Sci. Rep.* 7:1394. doi: 10.1038/s41598-017-00802-8

Nguyen, T. T., Oh, S. S., Weaver, D., Lewandowska, A., Maxfield, D., Schuler, M. H., et al. (2014). Loss of Miro1-directed mitochondrial movement results in a novel murine model for neuron disease. *Proc. Natl. Acad. Sci. U.S.A.* 111, E3631–E3640. doi: 10.1073/pnas.1402449111

Sagan, L. (1967). On the origin of mitosing cells. *J. Theor. Biol.* 14, 255–274.

Silva, E., Rosario, F. J., Powell, T. L., and Jansson, T. (2017). Mechanistic target of rapamycin is a novel molecular mechanism linking folate availability and cell function. *J. Nutr.* 147, 1237–1242. doi: 10.3945/jn.117.248823

Sochocka, M., Donskow-Lysoniewska, K., Diniz, B. S., Kurpas, D., Brzozowska, E., and Leszek, J. (2019). The gut microbiome alterations and inflammation-driven pathogenesis of Alzheimer's disease-a critical review. *Mol. Neurobiol.* 56, 1841–1851. doi: 10.1007/s12035-018-1188-4

Stavru, F., Bouillaud, F., Sartori, A., Ricquier, D., and Cossart, P. (2011). *Listeria monocytogenes* transiently alters mitochondrial dynamics during infection. *Proc. Natl. Acad. Sci. U.S.A.* 108, 3612–3617. doi: 10.1073/pnas.1100126108

Suzuki, H., Ataka, K., Asakawa, A., Cheng, K. C., Ushikai, M., Iwai, H., et al. (2019). *Helicobacter pylori* vacuolating cytotoxin A causes anorexia and anxiety via hypothalamic urocortin 1 in mice. *Sci. Rep.* 9:6011. doi: 10.1038/s41598-019-42163-4

Taanman, J. W. (1999). The mitochondrial genome: structure, transcription, translation and replication. *Biochim. Biophys. Acta.* 1410, 103–123. doi: 10.1016/s0005-2728(98)00161-3

van de Guchte, M., Blottiere, H. M., and Dore, J. (2018). Humans as holobionts: implications for prevention and therapy. *Microbiome* 6:81. doi: 10.1186/s40168-018-0466-8

Wang, A., Keita, A. V., Phan, V., McKay, C. M., Schoultz, I., Lee, J., et al. (2014). Targeting mitochondria-derived reactive oxygen species to reduce epithelial barrier dysfunction and colitis. *Am. J. Pathol.* 184, 2516–2527. doi: 10.1016/j.ajpath.2014.05.019

Wang, X., Su, B., Lee, H. G., Li, X., Perry, G., Smith, M. A., et al. (2009). Impaired balance of mitochondrial fission and fusion in Alzheimer's disease. *J. Neurosci.* 29, 9090–9103. doi: 10.1523/JNEUROSCI.1357-09.2009

Wang, Z., and Wu, M. (2015). An integrated phylogenomic approach toward pinpointing the origin of mitochondria. *Sci. Rep.* 5:7949. doi: 10.1038/srep07949

Wasser, C. I., Mercieca, E. C., Kong, G., Hannan, A. J., McKeown, S. J., Glikmann-Johnston, Y., et al. (2020). Gut dysbiosis in Huntington's disease: associations among gut microbiota, cognitive performance and clinical outcomes. *Brain Commun.* 2:fcaa110. doi: 10.1093/braincomms/fcaa110

Wolfson, J. S., and Hooper, D. C. (1985). The fluoroquinolones: structures, mechanisms of action and resistance, and spectra of activity in vitro. *Antimicrob. Agents Chemother.* 28, 581–586. doi: 10.1128/AAC.28.4.581

Yardeni, T., Tanes, C. E., Bittinger, K., Mattei, L. M., Schaefer, P. M., Singh, L. N., et al. (2019). Host mitochondria influence gut microbiome diversity: a role for ROS. *Sci. Signal.* 12:eaaw3159 doi: 10.1126/scisignal.aaw3159

Yu, C. D., Xu, Q. J., and Chang, R. B. (2020). Vagal sensory neurons and gut-brain signaling. *Curr. Opin. Neurobiol.* 62, 133–140. doi: 10.1016/j.conb.2020.03.006

Zhang, Q., Wu, X., Chen, P., Liu, L., Xin, N., Tian, Y., et al. (2018). The mitochondrial unfolded protein response is mediated cell-non-autonomously by retromer-dependent Wnt signaling. *Cell* 174, 870–883. doi: 10.1016/j.cell.2018.06.029

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.

frontiers in Microbiology

PERSPECTIVE
published: 03 March 2022
doi: 10.3389/fmicb.2022.846555

Systems Biology of Gut Microbiota-Human Receptor Interactions: Toward Anti-inflammatory Probiotics

Lokanand Koduru[1], Meiyappan Lakshmanan[2,3], Shawn Hoon[1], Dong-Yup Lee[4], Yuan Kun Lee[5,6] and Dave Siak-Wei Ow[2]**

[1] *Institute of Molecular and Cell Biology, Agency for Science, Technology and Research (A*STAR), Singapore, Singapore,* [2] *Bioprocessing Technology Institute, Agency for Science, Technology and Research (A*STAR), Singapore, Singapore,* [3] *Bioinformatics Institute, Agency for Science, Technology and Research (A*STAR), Singapore, Singapore,* [4] *School of Chemical Engineering, Sungkyunkwan University, Suwon, South Korea,* [5] *Department of Microbiology and Immunology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore,* [6] *Department of Surgery, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore*

The incidence and prevalence of inflammatory disorders have increased globally, and is projected to double in the next decade. Gut microbiome-based therapeutics have shown promise in ameliorating chronic inflammation. However, they are largely experimental, context- or strain-dependent and lack a clear mechanistic basis. This hinders precision probiotics and poses significant risk, especially to individuals with pre-existing conditions. Molecules secreted by gut microbiota act as ligands to several health-relevant receptors expressed in human gut, such as the G-protein coupled receptors (GPCRs), Toll-like receptor 4 (TLR4), pregnane X receptor (PXR), and aryl hydrocarbon receptor (AhR). Among these, the human AhR expressed in different tissues exhibits anti-inflammatory effects and shows activity against a wide range of ligands produced by gut bacteria. However, different AhR ligands induce varying host responses and signaling in a tissue/organ-specific manner, which remain mostly unknown. The emerging systems biology paradigm, with its powerful *in silico* tool repertoire, provides opportunities for comprehensive and high-throughput strain characterization. In particular, combining metabolic models with machine learning tools can be useful to delineate tissue and ligand-specific signaling and thus their causal mechanisms in disease and health. The knowledge of such a mechanistic basis is indispensable to account for strain heterogeneity and actualize precision probiotics.

Keywords: systems biology, gut microbiota, inflammatory disorders, probiotics, postbiotics

OPEN ACCESS

Edited by:
Qi Zhao,
University of Science and Technology Liaoning, China

Reviewed by:
Liang Wang,
Xuzhou Medical University, China

****Correspondence:***
Lokanand Koduru
kodurul@imcb.a-star.edu.sg
Dave Siak-Wei Ow
dave_ow@bti.a-star.edu.sg

Specialty section:
This article was submitted to Systems Microbiology, a section of the journal Frontiers in Microbiology

Received: *31 December 2021*
Accepted: *11 February 2022*
Published: *03 March 2022*

Citation:
Koduru L, Lakshmanan M, Hoon S, Lee D-Y, Lee YK and Ow DS-W (2022) Systems Biology of Gut Microbiota-Human Receptor Interactions: Toward Anti-inflammatory Probiotics. Front. Microbiol. 13:846555. doi: 10.3389/fmicb.2022.846555

INTRODUCTION

The human gut microbiome has attracted attention in the past decade for its role in inflammatory disorders (Kamada et al., 2013). Altering the gut microbiome profile through probiotic supplementation has been suggested as a potential strategy for ameliorating the symptoms associated with such diseases including inflammatory bowel disease (IBD) and colorectal cancer (Hendler and Zhang, 2018). The commercial probiotic formulations largely exist as generic supplements labeled with health claims that often remain unsubstantiated. Moreover, the existence of non-specific labels affects public perception on their proposed efficacy in treating health

conditions, thereby posing risk to the probiotic market over a long run. Hence, precision probiotics with well-defined activities that target specific health conditions are imperative. The development of precision probiotics is not a trivial task, as it requires a deeper understanding of the complex molecular cross talk between gut microbiota and host. One way to address this problem is to first map possible interactions between the two in the gut environment and subsequently trace the host effects downstream to the interaction. Several human receptors have been shown to interact with the gut microbiota, including the G-protein coupled receptors (GPCRs), Toll-like receptors (TLRs), pregnane X receptor (PXR), and aryl hydrocarbon receptor (AhR) (Mohandas and Vairappan, 2017; Chang and Kao, 2019; Chen et al., 2019). The activity of each of these receptors brings about varied molecular responses that affect health. In simpler terms, the "precision" aspect of probiotics can be gauged by their net conceivable interactions with the host receptors.

Tracing the physiological effects of gut microbiota, particularly during probiotic supplementation, and the small molecules they produce is a daunting task. In particular, the pleiotropy resulting from the interaction of multiple gut microbial ligand-host receptor combinations is hard to discern using classical molecular biology techniques. Integration of multiple omics datasets has been instrumental in explaining complex biological phenomena. Further, genome-scale metabolic models (GEMs) serve as convenient tools to represent complex molecular networks in a meaningful manner, especially when constrained using omics datasets. Therefore, the future of gut microbiome therapeutics will greatly benefit from a systematic amalgamation of multiomics, GEMs and machine learning tools.

GUT MICROBIOTA-HUMAN RECEPTOR INTERACTIONS

Gut microbiota interact with the host either through direct association with the intestinal epithelial cells or *via* secretion of small molecules. Although direct interactions are possible, they are limited to human cells in the gut lining. On the other hand, small molecules produced by the microbiota in gut environment not only interact with the host cell lining in intestine but also enter the blood circulation and access receptors expressed in nearly all human organs. Such interactions play a wide array of roles in health and disease, ranging from intestinal homeostasis, dysbiosis, inflammation and cancer. The human receptors such as GPCRs, PXR, and TLRs have been shown to play an important role in gut-microbiome mediated immunomodulation including their anti-inflammatory roles (Valentini et al., 2014; Venkatesh et al., 2014; Chen et al., 2019; Melhem et al., 2019). However, AhR is more often implicated in governing the overall gut homeostasis and inflammation (Zelante et al., 2013; Marinelli et al., 2019). Several studies have clearly demonstrated causal links between anti-inflammatory outcomes in chronic inflammatory conditions and treatment with AhR ligands or the gut microbial strains synthesizing them (Takamura et al., 2011; Fukumoto et al., 2014; Lamas et al., 2016; Abdulla et al., 2021; Zhao et al., 2021). AhR is a transcription factor, originally thought to function as a xenobiotic sensor, is now known to be an important regulator of immunity, stem cell maintenance and cellular differentiation (Gutiérrez-Vázquez and Quintana, 2018). Several signaling mechanisms downstream to AhR activation have been elucidated (Agus et al., 2018; Roager and Licht, 2018). Different gut microbial species and strains exhibit varied AhR modulating capabilities *via* their secretory metabolites. For example, short chain fatty acids (SCFAs) produced by gut microbiota bring about AhR activity through multiple mechanisms, including enhancement of AhR responsiveness to ligands (Jin et al., 2017), acting as a direct ligand to AhR (Marinelli et al., 2019), and increasing the expression of AhR (Yang et al., 2020). Each of these studies have indicated that the SCFA-mediated mechanisms eventually converge at promoting AhR activity and bring about anti-inflammatory effects. Similarly, tryptophan and indole derivatives activate the AhR resulting in anti-inflammatory outcomes (Lamas et al., 2016; Özçam et al., 2019). Despite their promise, the ligand- and tissue-specific effects of AhR signaling, which confound clinical studies, is not completely understood; not all AhR ligands confer anti-inflammatory phenotype (Denison et al., 2011). Therefore, characterizing strains and their tissue-specific AhR-mediated inflammation is an indispensable requirement for their use as therapeutic modalities.

ROLE OF SYSTEMS BIOLOGY

Gut is a complex environment characterized by innumerable biotic and abiotic interactions. The chaotic regimen consisting of diet, microbiota, host secretions and gut barrier render the ligand-receptor interactions largely untraceable using conventional "top-down" experimental techniques (Veiga et al., 2020). However, the emergent network properties of such systems of interacting components have always been tractable using mathematical modeling techniques. Systems biology, a discipline that treats organisms as an integration of one or more such networks, plays an important role in today's biological research, especially studying complex systems such as the human gut. An aspect that makes a strong case for the use of systems biology has been its tremendous success in studying metabolic networks using constraint-based modeling approach. Large-scale metabolic model reconstruction efforts have been undertaken in the past to facilitate studies deciphering interactions within and between gut microbiota and human intestinal cells (Diener et al., 2020; Heinken et al., 2021). These models largely rely on biochemical capabilities reflected by genomic data and taxonomic abundances estimated from metagenomic information, and therefore, the depth of interactions prevailing in the gut extend beyond what is usually captured by them. Fecal and serum metabolomic and transcriptomic datasets when integrated with such metabolic models can systematically reveal metabolic correlations of the microbiome. Furthermore, the use of machine learning tools in combination with omics integrated metabolic models can help derive mechanistic insights on potential health effects of specific strains (Yang et al., 2019; Lewis and Kemp, 2021). A list of publicly available systems

TABLE 1 | Survey of systems biology tools and resources to study gut microbiome-host receptor interactions.

Name of the tool/resource	URL	References
Ligand-receptor pairs in humans and mice		
Compiled by Lewis Lab at UCSD	https://github.com/LewisLabUCSD/Ligand-Receptor-Pairs	Armingol et al., 2021
CellTalkDB	http://tcm.zju.edu.cn/celltalkdb/	Shao et al., 2021
Database of available microbiome and host genome-scale models		
BiGGModels	http://bigg.ucsd.edu/	Norsigian et al., 2020
BioModels	https://www.ebi.ac.uk/biomodels/	Malik-Sheriff et al., 2020
AGORA	https://www.vmh.life/	Magnúsdóttir et al., 2017
Human model (Recon)	https://www.vmh.life/	Noronha et al., 2019
metaGEM	https://github.com/franciscozorrilla/metaGEM	Zorrilla et al., 2021
Metabolic Atlas	https://metabolicatlas.org/	Wang et al., 2021
Tools to reconstruct microbiome and host genome-scale models		
RAVEN Toolbox	https://github.com/SysBioChalmers/RAVEN	Wang et al., 2018
COBRA toolbox/COBRAPy	http://opencobra.github.io/	Heirendt et al., 2019
gapSeq	https://github.com/jotech/gapseq	Zimmermann et al., 2021
CarveMe	https://carveme.readthedocs.io/en/latest/index.html	Machado et al., 2018
ModelSEED	https://modelseed.org/	Seaver et al., 2021
Memote	https://memote.io/	Lieven et al., 2020
Tools to simulate genome-scale metabolism of microbiome and hosts and their interactions		
RAVEN Toolbox	https://github.com/SysBioChalmers/RAVEN	Wang et al., 2018
COBRA toolbox/COBRAPy	http://opencobra.github.io/	Heirendt et al., 2019
MICOM	https://github.com/micom-dev/micom	Diener et al., 2020
CASINO		Shoaie et al., 2015
SteadyCom	https://github.com/maranasgroup/SteadyCom	Chan et al., 2017
COMETS	https://github.com/segrelab/comets	Harcombe et al., 2014
CODY	https://github.com/JunGeng-Sysbio-Chalmers/CODY1.0_SourceCode	Geng et al., 2021

biology tools that can aid in such an integrated analysis is shown in **Table 1**.

PRECISION ANTI-INFLAMMATORY PROBIOTICS: NEED FOR A TWO-PRONGED APPROACH

The current trend in systems biology has been to predict and quantify immunomodulatory effects of specific probiotic strains and metabolites that they produce in the gut environment (dos Santos et al., 2010; Fiocchi and Iliopoulos, 2021). Availability of quantitative information on microbiota-induced immunomodulation is an essential requirement to realize precision probiotics. Ligand-receptor interactions form the direct link between microbiota and immunomodulation in humans. Therefore, a systematic approach toward designing precision probiotics must first take into account the sum total interactions brought about by each bacterial strain or microbial consortia *via* the small molecule ligands they secrete and their possible receptors in humans. In this regard, GEMs can greatly simplify the evaluation of various strains for the biosynthesis of potential ligands. Well-curated models can further provide quantitative estimates of the ligand production, which may serve as proxies for ligand-receptor activities. For instance, we have reconstructed and used GEMs for 18 lactic acid bacterial strains to systematically evaluate their probiotic capabilities (Koduru et al., 2021). This study constitutes one of the first attempts to theoretically quantify the probiotic potential of strains. Specifically, we relied on three important metrics based on constraint-based flux analysis to evaluate the strains. First, we estimated their capacities to biosynthesize metabolites, commonly referred to as "postbiotics," with anti-inflammatory properties such as SCFAs, tryptophan and indole, and pro-inflammatory metabolites such as lipoteichoic acids and branched chain amino acids. Subsequently, we evaluated their ability to persist in the gut environment under various dietary regimens. Finally, we assessed the nature of their potential interactions with beneficial gut commensals. These estimates can be prudently used to recommend strains to manage inflammation.

The second important requirement of precision probiotics involves understanding the nature of pleiotropic effects resulting from the cumulative downstream signaling of multiple ligand-receptor interactions. To this end, (Stein et al., 2018) have developed a computational approach to design optimal microbial consortia for induction of regulatory T-cells (Tregs), which are immune cells with an anti-inflammatory effect. The authors combined data on the contribution of 11 Clostridia strains in Treg induction with an ecological model to identify and rank different stable combinations for their ability to promote Treg activation. Importantly, they introduce and validate a computational metric called Treg Induction Score (TrIS) to facilitate the systematic selection

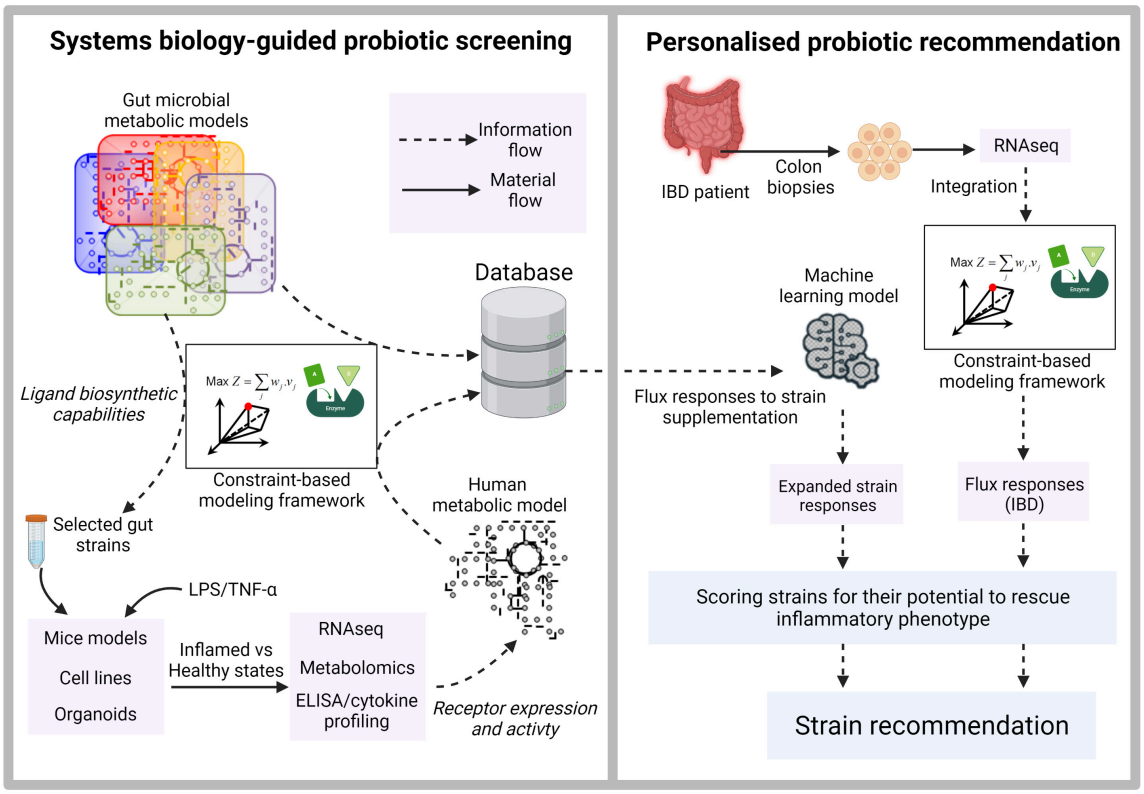


FIGURE 1 | Illustration of systems biology-guided personalized probiotic recommendation. The first stage consists of constraint-based model-driven screening of gut microbial strains for the production of diverse ligands capable of binding to human receptors. The secretory metabolites from selected strains are then used to treat mice models, cell lines or organoids. Their inflamed and healthy states are quantified using inflammatory cytokine profiling. Paired RNAseq and metabolomic datasets derived from each probiotic supplementation scenario provide information on the expression and activity of the target receptors and serve as constraints to derive context-specific models from a generic human genome-scale metabolic network. The context-specific models are used to derive flux responses corresponding to the supplementation of each probiotic strain. The second stage involves training a machine learning model using the extensive flux response information derived in the first stage. RNAseq data derived from real IBD patients for whom the probiotic recommendations are to be made is used to generate a context-specific model and the corresponding metabolic flux distribution. This flux distribution is then used as an input to the trained machine learning model to rank strains based on their ability to rescue the inflammatory phenotype.

of immunomodulatory strains. Interestingly, at least part of AhR's anti-inflammatory effects can be attributed to its *in vivo* role as a potent Treg inducer (Goettel et al., 2016; Ye et al., 2017), an angle that was not pursued by Stein et al. (2018). In this regard, characterization of each of the 11 Clostridia strains for their potential to synthesize AhR ligands could further provide a thorough mechanistic basis of their Treg inducing capabilities, as well as, potentially help fine-tuning the microbial consortia for optimal activity. Therefore, while such studies constitute big steps in the direction of precision probiotics, integrating them with metabolic models and machine learning tools can yield insights into metabolites beyond the extensively investigated SCFAs, potentially paving ways to decipher the functional aspects of "large metabolic unknowns" of the gut microbial world. Furthermore, the precision aspect can be strengthened by gaining individual-specific understanding of the host physiology. An illustration of a systematic approach to integrate multiple omics datasets, metabolic models and machine learning to screen and make personalized probiotic recommendations targeting IBD is shown in **Figure 1**. Although the focus here is on the microbiome-host interactions in gut environment, a similar approach can be extended to such interactions in skin, oral and vaginal microbiome.

CONCLUSION

The conventional top-down approach of probiotic discovery largely relies on correlations between microbiota composition in gut or fermentative food and health or disease (Veiga et al., 2020). Although this approach has been successful in recommending candidate strains to promote gut health, the absence of a mechanistic basis limits their application to "generic" probiotic supplements. Inflammatory disorders such as IBD and colorectal cancer, being complex manifestations of unhealthy gut and microbiota dysbiosis extensively depend on host-genetic, age, diet and other environmental factors, require precision probiotic targeting. In addition, although gut microbial metabolites such as the AhR ligands have shown tremendous

potential in ameliorating inflammation, their exact role might vary depending on the cells, tissue and organs expressing AhR (Cannon et al., 2022). The bottom-up approach guided by systems biology proposed here involves a two-pronged strategy – comprehensive characterization of receptor activating potential of strains and cell/tissue-specific pleiotropic signaling effects downstream to receptor activation in humans. We reason that the future transition toward precision probiotics thus lies in deciphering ligand-receptor interactions, with AhR being a key mediator in managing chronic inflammation.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding authors.

AUTHOR CONTRIBUTIONS

LK drafted the manuscript. ML, SH, D-YL, YL, and DO edited and proof-read the manuscript. All authors made a substantial contribution to the work and approved it for publication.

FUNDING

This research was supported by the Agency for Science, Technology and Research (A*STAR), Singapore, and was also funded by the Industry Alignment Fund (Pre-Positioning) Food Structure Engineering for Nutrition and Health (FSENH) programme (H18/01/a0/C11) and by the Singapore Food Story R&D Programme theme 2: Future Foods- Alternative proteins (W20W2D0005). D-YL is supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2020R1A2C2007192).

REFERENCES

Abdulla, O. A., Neamah, W., Sultan, M., Alghetaa, H. K., Singh, N., Busbee, P. B., et al. (2021). The Ability of AhR Ligands to Attenuate Delayed Type Hypersensitivity Reaction Is Associated With Alterations in the Gut Microbiota. *Front. Immunol.* 12:2505. doi: 10.3389/fimmu.2021.684727

Agus, A., Planchais, J., and Sokol, H. (2018). Gut Microbiota Regulation of Tryptophan Metabolism in Health and Disease. *Cell Host Microbe* 23, 716–724. doi: 10.1016/j.chom.2018.05.003

Armingol, E., Officer, A., Harismendy, O., and Lewis, N. E. (2021). Deciphering cell–cell interactions and communication from gene expression. *Nat. Rev. Genet.* 22, 71–88. doi: 10.1038/s41576-020-00292-x

Cannon, A. S., Nagarkatti, P. S., and Nagarkatti, M. (2022). Targeting AhR as a Novel Therapeutic Modality against Inflammatory Diseases. *Int. J. Mol. Sci.* 23:288. doi: 10.3390/ijms23010288

Chan, S. H. J., Simons, M. N., and Maranas, C. D. (2017). SteadyCom: predicting microbial abundances while ensuring community stability. *PLoS Comput. Biol.* 13:e1005539. doi: 10.1371/journal.pcbi.1005539

Chang, C.-S., and Kao, C.-Y. (2019). Current understanding of the gut microbiota shaping mechanisms. *J. Biomed. Sci.* 26:59. doi: 10.1186/s12929-019-0554-5

Chen, H., Nwe, P.-K., Yang, Y., Rosen, C. E., Bielecka, A. A., Kuchroo, M., et al. (2019). A forward chemical genetic screen reveals gut microbiota metabolites that modulate host physiology. *Cell* 177, 1217–1231. doi: 10.1016/j.cell.2019.03.036

Denison, M. S., Soshilov, A. A., He, G., Degroot, D. E., and Zhao, B. (2011). Exactly the same but different: promiscuity and diversity in the molecular mechanisms of action of the aryl hydrocarbon (dioxin) receptor. *Toxicol. Sci.* 124, 1–22. doi: 10.1093/toxsci/kfr218

Diener, C., Gibbons, S. M., and Resendis-Antonio, O. (2020). MICOM: metagenome-scale modeling to infer metabolic interactions in the gut microbiota. *MSystems* 5, e606–e619. doi: 10.1128/mSystems.00606-19

dos Santos, V. M., Müller, M., and de Vos, W. M. (2010). Systems biology of the gut: the interplay of food, microbiota and host at the mucosal interface. *Curr. Opin. Biotechnol.* 21, 539–550. doi: 10.1016/j.copbio.2010.08.003

Fiocchi, C., and Iliopoulos, D. (2021). IBD Systems Biology Is Here to Stay. *Inflamm. Bowel Dis.* 27, 760–770. doi: 10.1093/ibd/izaa343

Fukumoto, S., Toshimitsu, T., Matsuoka, S., Maruyama, A., Oh-Oka, K., Takamura, T., et al. (2014). Identification of a probiotic bacteria-derived activator of the aryl hydrocarbon receptor that inhibits colitis. *Immunol. Cell Biol.* 92, 460–465. doi: 10.1038/icb.2014.2

Geng, J., Ji, B., Li, G., López-Isunza, F., and Nielsen, J. (2021). CODY enables quantitatively spatiotemporal predictions on in vivo gut microbial variability induced by diet intervention. *Proc. Natl. Acad. Sci.* 118:e2019336118. doi: 10.1073/pnas.2019336118

Goettel, J., Gandhi, R., Kenison, J., Yeste, A., Murugaiyan, G., Sambanthamoorthy, S., et al. (2016). AHR activation is protective against colitis driven by T cells in humanized mice. *Cell Rep.* 17, 1318–1329. doi: 10.1016/j.celrep.2016.09.082

Gutiérrez-Vázquez, C., and Quintana, F. J. (2018). Regulation of the Immune Response by the Aryl Hydrocarbon Receptor. *Immunity* 48, 19–33. doi: 10.1016/j.immuni.2017.12.012

Harcombe, W. R., Riehl, W. J., Dukovski, I., Granger, B. R., Betts, A., Lang, A. H., et al. (2014). Metabolic resource allocation in individual microbes determines ecosystem interactions and spatial dynamics. *Cell Rep.* 7, 1104–1115. doi: 10.1016/j.celrep.2014.03.070

Heinken, A., Hertel, J., and Thiele, I. (2021). Metabolic modelling reveals broad changes in gut microbial metabolism in inflammatory bowel disease patients with dysbiosis. *NPJ Syst. Biol. Appl.* 7:19. doi: 10.1038/s41540-021-00178-6

Heirendt, L., Arreckx, S., Pfau, T., Mendoza, S. N., Richelle, A., Heinken, A., et al. (2019). Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v. 3.0. *Nat. Protoc.* 14, 639–702. doi: 10.1038/s41596-018-0098-2

Hendler, R., and Zhang, Y. (2018). Probiotics in the Treatment of Colorectal Cancer. *Medicines* 5:101. doi: 10.3390/medicines5030101

Jin, U.-H., Cheng, Y., Park, H., Davidson, L. A., Callaway, E. S., Chapkin, R. S., et al. (2017). Short chain fatty acids enhance aryl hydrocarbon (Ah) responsiveness in mouse colonocytes and Caco-2 human colon cancer cells. *Sci. Rep.* 7:10163. doi: 10.1038/s41598-017-10824-x

Kamada, N., Seo, S. U., Chen, G. Y., and Núñez, G. (2013). Role of the gut microbiota in immunity and inflammatory disease. *Nat. Rev. Immunol.* 13, 321–335. doi: 10.1038/nri3430

Koduru, L., Lakshmanan, M., Lim, P.-Y., Ho, P.-L., Banu, M., Park, D.-S., et al. (2021). Systematic evaluation of genome-wide metabolic landscapes in lactic acid bacteria reveals diet-induced and strain-specific probiotic idiosyncrasies. *BioRxiv* [preprint]. doi: 10.1101/2021.06.20.449192

Lamas, B., Richard, M. L., Leducq, V., Pham, H. P., Michel, M. L., da Costa, G., et al. (2016). CARD9 impacts colitis by altering gut microbiota metabolism of tryptophan into aryl hydrocarbon receptor ligands. *Nat. Med.* 22, 598–605. doi: 10.1038/nm.4102

Lewis, J. E., and Kemp, M. L. (2021). Integration of machine learning and genome-scale metabolic modeling identifies multi-omics biomarkers for radiation resistance. *Nat. Commun.* 12:2700. doi: 10.1038/s41467-021-22989-1

Lieven, C., Beber, M. E., Olivier, B. G., Bergmann, F. T., Ataman, M., Babaei, P., et al. (2020). MEMOTE for standardized genome-scale metabolic model testing. *Nat. Biotechnol.* 38, 272–276.

Machado, D., Andrejev, S., Tramontano, M., and Patil, K. R. (2018). Fast automated reconstruction of genome-scale metabolic models for microbial species and communities. *Nucleic Acids Res.* 46, 7542–7553. doi: 10.1093/nar/gky537

Magnúsdóttir, S., Heinken, A., Kutt, L., Ravcheev, D. A., Bauer, E., Noronha, A., et al. (2017). Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota. *Nat. Biotechnol.* 35, 81–89. doi: 10.1038/nbt.3703

Malik-Sheriff, R. S., Glont, M., Nguyen, T. V. N., Tiwari, K., Roberts, M. G., Xavier, A., et al. (2020). BioModels—15 years of sharing computational models in life science. *Nucleic Acids Res.* 48, D407–D415. doi: 10.1093/nar/gkz1055

Marinelli, L., Martin-Gallausiaux, C., Bourhis, J. M., Béguet-Crespel, F., Blottière, H. M., and Lapaque, N. (2019). Identification of the novel role of butyrate as AhR ligand in human intestinal epithelial cells. *Sci. Rep.* 9:643. doi: 10.1038/s41598-018-37019-2

Melhem, H., Kaya, B., Ayata, C. K., Hruz, P., and Niess, J. H. (2019). Metabolite-sensing G protein-coupled receptors connect the diet-microbiota-metabolites axis to inflammatory bowel disease. *Cells* 8:450. doi: 10.3390/cells8050450

Mohandas, S., and Vairappan, B. (2017). Role of pregnane X-receptor in regulating bacterial translocation in chronic liver diseases. *World J. Hepatol.* 9:1210. doi: 10.4254/wjh.v9.i32.1210

Noronha, A., Modamio, J., Jarosz, Y., Guerard, E., Sompairac, N., Preciat, G., et al. (2019). The Virtual Metabolic Human database: integrating human and gut microbiome metabolism with nutrition and disease. *Nucleic Acids Res.* 47, D614–D624. doi: 10.1093/nar/gky992

Norsigian, C. J., Pusarla, N., McConn, J. L., Yurkovich, J. T., Dräger, A., Palsson, B. O., et al. (2020). BiGG Models 2020: multi-strain genome-scale models and expansion across the phylogenetic tree. *Nucleic Acids Res.* 48, D402–D406. doi: 10.1093/nar/gkz1054

Özçam, M., Tocmo, R., Oh, J. H., Afrazi, A., Mezrich, J. D., Roos, S., et al. (2019). Gut symbionts Lactobacillus reuteri R2lc and 2010 encode a polyketide synthase cluster that activates the mammalian aryl hydrocarbon receptor. *Appl. Environ. Microbiol.* 85, e1661–e1668. doi: 10.1128/AEM.01661-18

Roager, H. M., and Licht, T. R. (2018). Microbial tryptophan catabolites in health and disease. *Nat. Commun.* 9:3294. doi: 10.1038/s41467-018-05470-4

Seaver, S. M. D., Liu, F., Zhang, Q., Jeffryes, J., Faria, J. P., Edirisinghe, J. N., et al. (2021). The ModelSEED Biochemistry Database for the integration of metabolic annotations and the reconstruction, comparison and analysis of metabolic models for plants, fungi and microbes. *Nucleic Acids Res.* 49, D575–D588.

Shao, X., Liao, J., Li, C., Lu, X., Cheng, J., and Fan, X. (2021). CellTalkDB: a manually curated database of ligand–receptor interactions in humans and mice. *Brief. Bioinform.* 22:bbaa269. doi: 10.1093/bib/bbaa269

Shoaie, S., Ghaffari, P., Kovatcheva-Datchary, P., Mardinoglu, A., Sen, P., Pujos-Guillot, E., et al. (2015). Quantifying diet-induced metabolic changes of the human gut microbiome. *Cell Metab.* 22, 320–331. doi: 10.1016/j.cmet.2015.07.001

Stein, R. R., Tanoue, T., Szabady, R. L., Bhattarai, S. K., Olle, B., Norman, J. M., et al. (2018). Computer-guided design of optimal microbial consortia for immune system modulation. *Elife* 7, e30916. doi: 10.7554/eLife.30916

Takamura, T., Harama, D., Fukumoto, S., Nakamura, Y., Shimokawa, N., Ishimaru, K., et al. (2011). Lactobacillus bulgaricus OLL1181 activates the aryl hydrocarbon receptor pathway and inhibits colitis. *Immunol. Cell Biol.* 89, 817–822. doi: 10.1038/icb.2010.165

Valentini, M., Piermattei, A., di Sante, G., Migliara, G., Delogu, G., and Ria, F. (2014). Immunomodulation by gut microbiota: role of Toll-like receptor expressed by T cells. *J. Immunol. Res.* 2014:586939. doi: 10.1155/2014/586939

Veiga, P., Suez, J., Derrien, M., and Elinav, E. (2020). Moving from probiotics to precision probiotics. *Nat. Microbiol.* 5, 878–880. doi: 10.1038/s41564-020-0721-1

Venkatesh, M., Mukherjee, S., Wang, H., Li, H., Sun, K., Benechet, A. P., et al. (2014). Symbiotic bacterial metabolites regulate gastrointestinal barrier function via the xenobiotic sensor PXR and Toll-like receptor 4. *Immunity* 41, 296–310. doi: 10.1016/j.immuni.2014.06.014

Wang, H., Marcišauskas, S., Sánchez, B. J., Domenzain, I., Hermansson, D., Agren, R., et al. (2018). RAVEN 2.0: a versatile toolbox for metabolic network reconstruction and a case study on Streptomyces coelicolor. *PLoS Comput. Biol.* 14:e1006541. doi: 10.1371/journal.pcbi.1006541

Wang, H., Robinson, J. L., Kocabas, P., Gustafsson, J., Anton, M., Cholley, P.-E., et al. (2021). Genome-scale metabolic network reconstruction of model animals as a platform for translational research. *Proc. Natl. Acad. Sci.* 118:e2102344118. doi: 10.1073/pnas.2102344118

Yang, J. H., Wright, S. N., Hamblin, M., McCloskey, D., Alcantar, M. A., Schrübbers, L., et al. (2019). A white-box machine learning approach for revealing antibiotic mechanisms of action. *Cell* 177, 1649–1661. doi: 10.1016/j.cell.2019.04.016

Yang, W., Yu, T., Huang, X., Bilotta, A. J., Xu, L., Lu, Y., et al. (2020). Intestinal microbiota-derived short-chain fatty acids regulation of immune cell IL-22 production and gut immunity. *Nat. Commun.* 11:4457. doi: 10.1038/s41467-020-18262-6

Ye, J., Qiu, J., Bostick, J. W., Ueda, A., Schjerven, H., Li, S., et al. (2017). The aryl hydrocarbon receptor preferentially marks and promotes gut regulatory T cells. *Cell Rep.* 21, 2277–2290. doi: 10.1016/j.celrep.2017.10.114

Zelante, T., Iannitti, R. G., Cunha, C., DeLuca, A., Giovannini, G., Pieraccini, G., et al. (2013). Tryptophan catabolites from microbiota engage aryl hydrocarbon receptor and balance mucosal reactivity via interleukin-22. *Immunity* 39, 372–385. doi: 10.1016/j.immuni.2013.08.003

Zhao, C., Hu, X., Bao, L., Wu, K., Feng, L., Qiu, M., et al. (2021). Aryl hydrocarbon receptor activation by Lactobacillus reuteri tryptophan metabolism alleviates *Escherichia coli*-induced mastitis in mice. *PLoS Pathog.* 17:e1009774. doi: 10.1371/journal.ppat.1009774

Zimmermann, J., Kaleta, C., and Waschina, S. (2021). gapseq: informed prediction of bacterial metabolic pathways and reconstruction of accurate metabolic models. *Genom. Biol.* 22:81. doi: 10.1186/s13059-021-02295-1

Zorrilla, F., Buric, F., Patil, K. R., and Zelezniak, A. (2021). metaGEM: reconstruction of genome scale metabolic models directly from metagenomes. *Nucleic Acids Res.* 49, e126–e126. doi: 10.1093/nar/gkab815

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.

frontiers | Frontiers in Microbiology

REVIEW
published: 04 April 2022
doi: 10.3389/fmicb.2022.845559

Impact of SARS-CoV-2 on Host Factors Involved in Mental Disorders

*Raina Rhoades, Sarah Solomon, Christina Johnson and Shaolei Teng**

Department of Biology, Howard University, Washington, DC, United States

COVID-19, caused by SARS-CoV-2, is a systemic illness due to its multiorgan effects in patients. The disease has a detrimental impact on respiratory and cardiovascular systems. One early symptom of infection is anosmia or lack of smell; this implicates the involvement of the olfactory bulb in COVID-19 disease and provides a route into the central nervous system. However, little is known about how SARS-CoV-2 affects neurological or psychological symptoms. SARS-CoV-2 exploits host receptors that converge on pathways that impact psychological symptoms. This systemic review discusses the ways involved by coronavirus infection and their impact on mental health disorders. We begin by briefly introducing the history of coronaviruses, followed by an overview of the essential proteins to viral entry. Then, we discuss the downstream effects of viral entry on host proteins. Finally, we review the literature on host factors that are known to play critical roles in neuropsychiatric symptoms and mental diseases and discuss how COVID-19 could impact mental health globally. Our review details the host factors and pathways involved in the cellular mechanisms, such as systemic inflammation, that play a significant role in the development of neuropsychological symptoms stemming from COVID-19 infection.

Keywords: SARS-CoV-2, mental disorders, depression, schizophrenia, psychosis

OPEN ACCESS

Edited by:
Qi Zhao,
University of Science and Technology Liaoning, China

Reviewed by:
Nhat Tu Le,
Houston Methodist Research Institute, United States
Hongbao Cao,
George Mason University, United States

****Correspondence:***
Shaolei Teng
shaolei.teng@howard.edu

Specialty section:
This article was submitted to Systems Microbiology, a section of the journal Frontiers in Microbiology

Received: *29 December 2021*
Accepted: *14 February 2022*
Published: *04 April 2022*

Citation:
Rhoades R, Solomon S, Johnson C and Teng S (2022) Impact of SARS-CoV-2 on Host Factors Involved in Mental Disorders. Front. Microbiol. 13:845559. doi: 10.3389/fmicb.2022.845559

INTRODUCTION

Post-acute COVID-19 Syndrome, also known as long-COVID, is a significant concern for global public health. The symptoms of long COVID range from length recovery from organ damage, persistent symptoms lasting up to 6 weeks, to a patient presenting as asymptomatic or experiencing a period of healing only to see a return of symptoms that persist from 3 to 6 months, and even sudden death up to 12 months post-infection (Raveendran and Misra, 2021). Neurological and neuropsychiatric symptoms have also been observed in one-third of patients after COVID-19 infection (Schou et al., 2021). These symptoms include depression, anxiety, cognitive deficits, "brain fog," and fatigue, which have been reported in conjunction with infection by MERS-CoV and SARS-CoV and previous pandemics such as the Spanish Flu (Schou et al., 2021; Stefano, 2021).

Studies have shown that some coronaviruses can invade the brainstem via a synapse-connected route from the lungs and airways (Zhang et al., 2020). A few central mechanisms have been proposed to explain neurological symptoms related to SARS-CoV-2 infection. The first theory, the "indirect attack theory," proposes that neurological effects are due to the immune impacts triggered by infections, i.e., the cytokine storm (Wu et al., 2020b). The second suggests that the virus gains entry to the central nervous system (CNS) via the olfactory pathway, or peripheral route, and demonstrated in animal models of encephalitis induced by corona viral infection. The

reports of deficits in taste, smell, and psychiatric symptoms following coronavirus infection are consistent with the peripheral route or the olfactory pathway as a route of entry for the SARS-CoV-2 virus (Acharya et al., 2020; Butowt and von Bartheld, 2020). Anosmia and other deficits in sensation are features of several mental disorders, including post-traumatic stress disorder (PTSD), major depression disorder (MDD), SCZ, bipolar disorder (BPD), and neurodegenerative disorders. The third route of entry for SARS-CoV-2 into the CNS involves exosomes. Several studies have noted that the cytokine storm cannot explain CNS damage since the inflammatory markers seen in SARS-CoV-2 cases are less abundant than seen in other cases of a viral infection such as H1N1 influenza. Additionally, the lack of detected viral gene expression in the CNS casts doubt on the hypothesis that the cytokine storm is causing or leading contributor to the neurological damage and neuropsychiatric symptoms seen in some SARS-CoV-2 cases. Exosomes have been previously reported to aid in viral pathogenesis (Estrada, 2021).

Viral entry of the coronavirus is mediated by the spike (S) protein, which has two subunits, S1 and S2. The S1 component binds to the host cell receptor, and the S2 subunit mediates the fusion of the virus with the host's cell membrane. The key to the entry of the SARS-CoV-2 virus into host cells is the angiotensin-converting enzyme 2 (ACE2) receptor, which is expressed in many tissues, including the respiratory system and neurons, and brain endothelium (Hamming et al., 2004; Sheraton et al., 2020). However, viral entry is also dependent on the priming of the S protein by host proteases such as transmembrane serine protease 2 (TMPRSS2) and FURIN. Several virion components linked to the pathology of coronaviruses have also been previously linked to mental health disorders. Coronavirus proteins such as the envelope (E) and nucleocapsid (N) proteins have also been demonstrated to bind to post-synaptic density-95 (PSD-95) and retinoic acid-inducible gene-1 (RIG-1) proteins. The envelope protein of SARS CoV-2 has also been reported to have a PSD-95 binding motif. PSD-95 is a scaffolding protein that plays an essential role in excitatory neurons and viral pathogenesis (Javier and Rice, 2011). Previous investigations have shown that the N protein of SARS-CoV-1, 90% similar to that of SARS-CoV-2, halts cell cycle progression *in vitro* (Li et al., 2005a,b; Dutta et al., 2020). Additionally, the SARS-CoV-2 N protein has been shown to possess a RIG-1 binding domain and inhibit RIG-1-like pathways (Oh and Shin, 2021). Rig-1 is a gene that recognizes viral infection, such as in *Toxoplasma gondii*. The Rig-1 gene has also been found to be associated with schizophrenia (SCZ) diagnosis (Carter, 2009). Additionally, the N-protein activates the cyclooxygenase-2 (COX-2) promoter. Thus, it plays a role in increased inflammation associated with coronavirus infection (Yan et al., 2006). Host receptor ACE2 serves as the point of entry for SARS-CoV-2 via the attachment of the S glycoprotein (Krassowski et al., 2018). A genome-wide association study of 1980 patients infected with SARS-CoV-2 found two loci 3p21.31 and 9q34.2 with genome-wide significance to be associated with severe symptoms. The significant association at the 3p21.31 locus was driven by solute carrier family 6 member 20 (SLC6A20), leucine zipper transcription factor-like 1 (LZTFL1), C-C chemokine receptor 1 (CCR1), FYVE coiled-coil domain-containing protein 1 (FYCO1), CXC motif chemokine receptor 6 (CXCR6), and X-C motif chemokine receptor 1 (XCR1), and the gene contributing to the significant association in the 9q34.2 locus was the histo-blood group ABO system transferase (ABO) (Severe Covid-19 GWAS Group et al., 2020). Additionally, five genes that seem to facilitate infection of the SARS-CoV2 virus are glycogen synthase kinase 3 beta (GSK-3β), furin protease, TMPRSS2, a disintegrin and metalloprotease 17 (ADAM17), and neuropilin-1 (Heurich et al., 2014; Cantuti-Castelvetri et al., 2020; Coutard et al., 2020; Nowak and Walkowiak, 2020).

With this in mind, we must now consider how these viral pathways can activate mental health disorders, as links between infectious disease and mental health disorders have been previously reported. Increased risk of developing SCZ, for example, has been linked to several contagious agents such as *Chlamydia* spp., *T. gondii*, Human Herpesvirus, and Cytomegalovirus (Arias et al., 2012). Coronavirus infection could lead to injury and inflammation, the exacerbation of neuropsychiatric symptoms. Studies of the olfactory epithelia have demonstrated its utility in studying psychiatric disorders as well as neurodevelopmental processes. Deficits in olfactory functioning have been reported in depression and other affective disorders (Taalman et al., 2017; Kamath et al., 2018). Therefore, several proteins affected by coronavirus infection, such as ACE2 and dipeptidyl peptidase 4 (DPP4), are enriched in the epithelia of the respiratory tract (Hamming et al., 2004; Jia et al., 2005; Solerte et al., 2020). Previous work has also reported several genes that may be related to increased susceptibility or resistance to SARS-CoV-2 infection (Wei et al., 2021). And there are several suggested mechanisms by which SARS-CoV-2 may affect the CNS, such as "viral encephalitis, systemic inflammation, organ dysfunction, and cerebrovascular change" (Heneka et al., 2020). This suggests that investigating genes enriched in the respiratory tract or found to be important in SARS-CoV-2 infection may help to understand how coronavirus infections may impact mental health (Wei et al., 2021).

Depression is among the top five leading causes of disability worldwide. Mental health disorders have a significant impact on the global economy, costing as much as 2.5 trillion dollars per year and rising (GBD 2016 Disease and Injury Incidence and Prevalence Collaborators, 2017; The Lancet Global Health, 2020). Therefore, particularly imperative to understand how infectious diseases might be converging with social, economic, and life stressors that are perturbed during global pandemics. Fear, social isolation, anxiety, sleep disturbances, unemployment, and housing insecurity can compound ongoing or predisposed mental health issues. For example, it has been reported in Wuhan, China, that more than half of the residents experienced symptoms of depression and or anxiety (Clemente-Suárez et al., 2021). Many recovered COVID-19 patients have been reported to experience neurological symptoms such as parkinsonism, intracranial hemorrhaging, and strokes. Long-term psychological symptoms such as dementia, anxiety, and psychosis have also been reported (Taquet et al., 2021a). A retrospective cohort study of 62,354 patients showed that hazard ratios for psychiatric diagnoses were higher than influenza, skin infections, and

respiratory tract infections for the first 14–90 days following COVID-19 diagnosis (Taquet et al., 2021b). In a retrospective study of 236,379 patients, the authors found that the incidence of neurological and psychological symptoms in the 6 months following COVID-19 diagnosis was between 33 and 62%. Many of these patients were diagnosed with these symptoms for the first time, with an estimated incidence of 1–84% (Taquet et al., 2021a). Additional studies have found that patients with long-COVID have exhibited imbalance, vertigo, hallucinations, headaches, memory deficits, and depression (Mehandru and Merad, 2022).

Despite the production of several SAR-CoV-2 vaccines, the SARS-CoV-2 virus will likely become endemic (Shaman and Galanti, 2020; Veldhoen and Simas, 2021). We, therefore, must study and develop an understanding of how infectious diseases like SARS-CoV-2 may contribute to long-term conditions such as mental health. The following review aims to highlight genes perturbed by a corona viral infection that are also implicated in mental disorders, emphasizing the effects of the SARS-CoV-2 virus. We begin by discussing host proteins vital to viral entry, a discussion of host proteins and factors that are affected downstream. Finally, we conclude by discussing how these host proteins relate to the etiology of mental health disorders (**Figure 1**).

SARS-CoV-2 STRUCTURAL PROTEINS AND THEIR ROLES IN VIRAL ENTRY

SARS-CoV-2 exploits several proteins, including host proteases and host receptors, to gain entry to cells. The S protein, by which the virion enters host cells, must be cleaved by host proteases. Once the S protein has been primed, the protein can then bind host receptors, and the virion can then fuse with the host membrane. These host proteins vary in their spatial-temporal expression, but they each play a role in inflammatory responses, among other physiological effects. Understanding the functional functions of these entry proteins is crucial in understanding their role in the SARS-CoV-2 infection (**Table 1**).

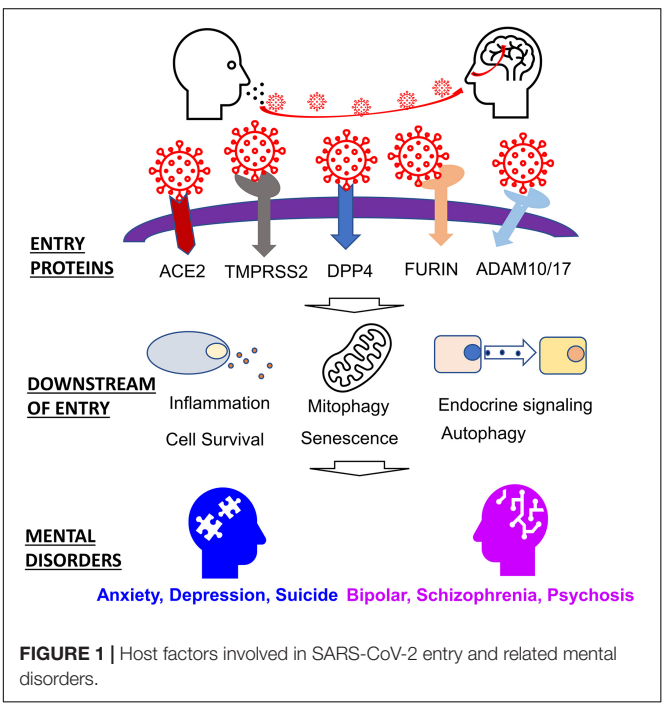


FIGURE 1 | Host factors involved in SARS-CoV-2 entry and related mental disorders.

Host Proteases

The SARS-CoV-2 S protein must be primed by host proteases before it can bind to host receptors and infect cells. The host proteases that have been identified in helping aid in the binding of the S protein to host receptors include TMPRSS2, ADAM10/17, and Furin (Hussain et al., 2020). TMPRSS2 was found to increase viral entry into host cells significantly and is expressed in astrocytes and oligodendrocytes (Heurich et al., 2014; Dong et al., 2020). Previous work has demonstrated that camostat mesylate, a TMPRSS2 inhibitor, resulted in the blockage of SARS-CoV-2 into TMPRSS2^{+} cells (Hoffmann et al., 2020). ADAM-10/17 (A disintegrin and metalloprotease 10 and 17) are proteases that cleave the extracellular domain of ACE2. However, they are reportedly less efficient than TMPRSS2 (Heurich et al., 2014; Aljohmani and Yildiz, 2020). FURIN also aids SARS-CoV-2 entry. The FURIN protein is an endoprotease and is expressed in hippocampal and cortical neurons (Yang et al., 2018). FURIN cleaves proteins within a specific motif (R/K)-(2X) n-(R/K) and plays a role in priming the SARS-CoV at the S1/S2 site (Coutard et al., 2020). This cleavage allows the virus to shed the spike protein and enter the host cell. The use of the protease is thought to be a key component of the pathogenicity of many viruses, including SARS-CoV-2 infection (Coutard et al., 2020; Fitzgerald, 2020). Of the known pathogenic beta coronaviruses, only the SARS-CoV2, MERS-CoV, and HCoV-OC43 viruses possess the FURIN cleavage complex motif. This protease also plays a role in apoptosis, inflammation of the vasculature, and lipid metabolism (Liu et al., 2020).

Essential Host Proteins That Interact With SARS-CoV-2

ACE2 is a part of the renin-angiotensin-aldosterone system (RAAS), and it is the principal host receptor used by SARS-CoV-2 (Motaghinejad and Gholami, 2020). The RAAS functions to maintain blood pressure by regulating fluid and electrolyte balance and vascular diameter (Wiese et al., 2020). The SARS-CoV-2 infection leads to the downregulation of ACE2, leading to what is referred to as Angiotensin II intoxication (Sfera et al., 2020; Wiese et al., 2020). ACE2 is expressed throughout the epithelia of the respiratory tract. However, the expression of ACE2 can be described as a gradient, where it is highest in the proximal nasal epithelia and attenuates as one proceeds to the epithelia of the lower respiratory tract (Hou et al., 2020). Within the central nervous system, the ACE2 receptor is expressed in both neurons and glial cells (Venkatakrishnan et al., 2020). It is also important to note that the expression of ACE2 and TMPRSS2 also increases with age, according to an investigation of temporal expression profiles in mice at ages 2 months and 2 years (Bilinska et al., 2020).

TABLE 1 | Host genes involved in SARS-CoV-2 infection and mental disorders.

Gene	Symptom/disorder	Summary	References
Entry proteins			
TMPRSS2	Depression	TMPRSS2 is implicated in depression associated with prostate cancer.	Rice et al., 2018; Wang et al., 2020
ADAM-10/17	Schizophrenia, depression, bipolar disorder, and conduct disorder	Increased levels of ADAM17 are associated with the diagnosis of schizophrenia in post-mortem brain tissue and CSF. A SNP located in ADAM10 was significantly associated with conduct disorder.	Jian et al., 2011; Qian et al., 2016; Hoseth et al., 2017
FURIN	Alzheimer's disease, Schizophrenia	rs4702 was significantly associated with schizophrenia was detected both by GWAS and eQTL analyses. This SNP is also associated with reduced FURIN and BDNF expression.	Fromer et al., 2016; Hou et al., 2018
ACE2	Anxiety, depression, cognitive impairment	ACE2 is implicated in the dysregulation of the HPA axis following SARS-CoV-2 infection.	Steenblock et al., 2020
DPP4	PTSD, depression, other neuropsychiatric illnesses	NPY is a ligand for the DPP4 or CD26 receptor and has been a proposed biomarker for these illnesses.	Canneva et al., 2015; Gołyszny and Obuchowicz, 2019
XCR1	Traumatic brain injury	XCR1 expression increased significantly in the thalamus and hippocampus beginning 24 h post-injury	Ciechanowska et al., 2020
HMGB1	Schizophrenia	SCZ patients exhibited increased expression variability in HMGB1 and several other genes.	Huang et al., 2020
	Bipolar disorder	Serum levels of HMGB1 were significantly high in the bipolar patients compared to the controls.	Marie-Claire et al., 2019
Neuropilin	Major depressive disorder	Authors found increased NRP-1 expression in the post-mortem PFC samples from patients diagnosed with MDD than in controls.	Goswami et al., 2013
Downstream of entry: inflammation			
GSK-3β	Schizophrenia and bipolar disorder	Increased levels of GSK-3β were found in nasal biopsies of bipolar patients and the blood, serum, and CSF of patients with SCZ.	Narayan et al., 2014; Mohammadi et al., 2018
HLA	Schizophrenia	Several HLA genes, including HLA-A10, HLA-B, and HLA-DRB1, have been linked to SCZ	Carter, 2009
TLR (7/8)	Depression	Increased mRNA expression of TLR3 and TLR4 in the brains of depressed non-suicidal and suicidal subjects	Pandey et al., 2014
Interleukins	Autism Schizophrenia	IL-23 and IL-17 are implicated in immune dysregulation seen in patients with schizophrenia and experimental models of autism.	Debnath and Berk, 2017; Alves de Lima et al., 2020
CXCR6	Anxiety	Meningeal γδ T cells expressing CXCR6 were shown to influence anxiety in mice	Alves de Lima et al., 2020
CCR1	Bipolar and schizophrenia	Greater expression of CCR1 and 28 other genes were found in patients diagnosed with schizophrenia when compared to patients diagnosed with bipolar disorder	de Baumont et al., 2015
Downstream of entry: cell survival			
Histone complex H3.3	Depression	H3.3 was found to be elevated in the Nucleus Accumbens of depressed humans.	Lepack et al., 2016
SWI/SNF complex genes	Major depressive disorder and schizophrenia	The SWI/SN subunit, BRM (SMARCA2), has been associated with self-reported MDD and schizophrenia.	Amare et al., 2020; Wu et al., 2020a
ARID1A/B	Craniofacial abnormalities	Mutations in ARID1A are associated with craniofacial abnormalities, while mutations in ARID1B are associated with autism spectrum disorder and SCZ.	Son and Crabtree, 2014; Pagliaroli and Trizzino, 2021
BDNF	Schizophrenia	Decreased BDNF expression has been associated with schizophrenia.	Bar-Yosef et al., 2019; Suchanek-Raif et al., 2020
	Anxiety, Major Depressive Disorder	A common SNP of BDNF, rs62265, is a missense mutation that has been associated with anxiety, major depression and suicide, and neurodegenerative disease, as has dysregulation of mTOR signaling	Dincheva et al., 2016; Youssef et al., 2018
SLC6A20	Schizophrenia and schizoaffective disorder	Hyperprolinemia has been reported in conjunction with SCZ and schizoaffective disorder.	Jacquet et al., 2005; Clelland et al., 2011
Downstream of entry: autophagy			
FYCO1	Senescence	A significant decrease in FYCO1 expression was associated with senescence	Cheng et al., 2007
	Neurodegenerative disorders	FYCO1 is involved in the clearance of α-synuclein aggregates.	Saridaki et al., 2018
CTSB/L	Alzheimer's disease Traumatic brain injury	Increased levels of cathepsin B in the cytosol, plasma, and CSF have been associated with cognitive dysfunction in Alzheimer's and traumatic brain injury.	Hook et al., 2020
CALM/CaMKII	Schizophrenia	Calmodulin levels were reportedly altered in postmortem lysates taken from ACC, CC, and the temporal lobe in patients with schizophrenia.	Vidal-Domènech et al., 2020

(Continued)

TABLE 1 | (Continued)

Gene	Symptom/disorder	Summary	References
Downstream of entry: endocrine signaling			
Estrogen receptor	Schizophrenia	Increased polymorphisms in ERα have been associated with SCZ. And circulating levels of estrogen have been associated with psychosis.	Min et al., 2012; Brzezinski-Sinai and Brzezinski, 2020
Androgen receptor	Depression Bipolar	Increased levels of AR expression were reported in patients with bipolar disorder.	
SHBG pathway	Depression	Positively and statistically significantly associated with depression risk ($p = 0.003$) in all women.	Colangelo et al., 2012
	Schizophrenia	In a study of schizophrenic male patients and a group of undiagnosed adults, both treated and untreated patients had lower serum levels of SHBG than undiagnosed controls.	Costa et al., 2006
Mental disorders			
TGF-beta	Schizophrenia and psychosis	TGF-Beta plays a role in the immune-inflammatory response and the compensatory immune-regulatory reflex system, which contribute to the etiology of schizophrenia.	Roomruangwong et al., 2020

DPP4 is a ubiquitously expressed serine protease that plays a role in inflammation energy metabolism and has also been reported as a marker of senescence (Klemann et al., 2016; Kim et al., 2017; Shao et al., 2020; Rohmann et al., 2021). The DPP4 protein is widely expressed in many cell types throughout the CNS, including dopaminergic neurons, macrophages, and glia (Venkatakrishnan et al., 2020). Although it is primarily known as the host receptor utilized by the MERS-CoV virus, previous work has shown that SARS-CoV-2 may also use as a point of entry. A protein docking simulation and subsequent analysis of free energy binding found that SARS-CoV2 bound firmly to DPP4 (Li et al., 2020). It is worth noting that the RAAS system and the DPP4 receptor are dysregulated in diabetes, a risk factor in severe COVID illness (Valencia et al., 2020).

Chemokine Receptors

Lymphopenia is one of the symptoms seen in patients with COVID-19. This observation has led to the notion that SARS-CoV-2 might also utilize other receptors, like XCR1, to facilitate T-Cell entry (Mobini et al., 2021). A structural study of binding affinity found that XCR1, in addition to several chemokine and immune receptors, had a higher binding affinity for the SARS-CoV-2 S protein than ACE2. XCR1 and other chemokine receptors are present in many types of immune cells. The XCR1 gene is upregulated in response to traumatic brain injury (Mobini et al., 2021). Several other chemokines, as well as their receptors, have been linked to prognostic outcomes in SARS-CoV and MERS-CoV infection (Khalil et al., 2021).

Neuropilin

Neuropilin is a host receptor that concretizes the overlapping impacts of SARS-CoV-2 infection as it plays a role in the inflammatory response, angiogenesis, and nerve growth, as well as synaptogenesis (Cai and Reed, 1999; Mayi et al., 2021). Investigators who used x-ray crystallography were able to demonstrate that SARS-CoV-2 spike protein cleaved at the furin site was able to bind with neuropilin (NRP1) (Daly et al., 2020). Neuropilin is known to bind proteins cleaved by FURIN protease. In an investigation of host cell entry, the authors used HEK-293 T cells transfected with plasmids to permit the expression of ACE2 and NRP1 (Cantuti-Castelvetri et al., 2020). Furthermore, comparative analysis of postmortem olfactory epithelium from COVID-19 patients and uninfected controls showed that SARS-CoV-2 could infect NRP1 positive cells of the olfactory epithelium (Cantuti-Castelvetri et al., 2020). Although the levels of ACE2 in the cells of the olfactory epithelium were relatively low, the authors found that expression levels of high levels of NRP1 and oligodendrocyte transcription factor (OLIG2), a marker for neuronal progenitors of the olfactory tract, were higher by comparison (Cantuti-Castelvetri et al., 2020).

HOST MECHANISMS ACTIVATED BY SARS-CoV-2 INFECTION

Once SARS-CoV-2 begins to proliferate and spread, innate immunity is deployed as T lymphocytes, and dendritic cells are activated by pattern recognition receptors like toll-like receptors (TLRs) (Bai et al., 2021). However, this innate immunity is overcome by viral suppressors of RNAi (VSRs) (Bai et al., 2021). This leads to the release of inflammatory factors, which in severe cases may lead to a cytokine storm, resulting in tissue damage to organs such as the lungs and heart (Mortaz et al., 2020). These inflammatory factors and cytokines cause adaptive immune cell activation as CD4+ T-cells to act as antigen-presenting cells, and CD8+ T cells are deployed to kill infected cells (Mortaz et al., 2020). Viruses like SARS-CoV-2 have evolved methods of evading host immunity and usurping cellular machinery involved in cell survival, senescence, autophagy, mitophagy, etc., to enable their proliferation (Alcock and Masters, 2021). These mechanisms are further impacted by age as well as hormone signaling. In this section, we explore the effects of SARS-CoV-2 entry and genes involved in the downstream process (**Table 1**).

Inflammation

The TLRs are molecular pattern recognition receptors that help to monitor the external cellular environment for pathogenic-associated molecular patterns (PAMPs) and damage-associated

molecular patterns (DAMPS) (Lim and Staudt, 2013; Kumar, 2019; Liu et al., 2019a). The activation of TLRs following SARS-CoV-2 infection can incite a cytokine storm within the respiratory endothelia. However, it is also capable of activating glial cells of the CNS, releasing several inflammatory factors such as interleukin-1 (IL-1), IL-6, IL-12, C-X-C motif chemokine ligand 10 (CXCL10), C-C motif ligand 3 (CCL3), CCL5, CCL2, TNF-alpha, CXCR6, XCR1, and CCR1, causing chronic inflammation and brain damage (Bouças et al., 2020; Coperchini et al., 2020; Jakhmola et al., 2020; Wu et al., 2020b; Khanmohammadi and Rezaei, 2021). Several of these chemokines and inflammatory factors are expressed in astrocytes, glia, neurons, neural stem cells, and oligodendrocytes (Sowa and Tokarski, 2021). The cytokine storm, particularly the release of TNF-alpha, then leads to the suppression of B-cells and thus antibody production (Kumar et al., 2021). One host protein that is critically involved in the cytokine storm is GSK-3β.GSK-3β is a serine-threonine kinase involved in the inflammatory response to infectious disease and plays a role in the phosphorylation of the SARS-CoV-2 N-protein. Inhibition of GSK-3β by drugs such as lithium has been demonstrated to reduce viral replication and enhance immune response (Taylor et al., 2016; de Souza et al., 2020; Rana et al., 2021). Human Leukocyte Antigen (HLA) also plays a key role in genes regulating the immune response to pathogens through antigen presentation. However, the effect of HLA variants on susceptibility and resistance in coronavirus infection is less evident in the case of SAR-CoV-2 infection (Saulle et al., 2021). For example, the HLA-A*24:02 allele was reported to be both a contributing factor to susceptibility and resistance to SARS-CoV-2 infection in separate investigations (Saulle et al., 2021).

Chromatin Remodeling

The pro-inflammatory High Mobility Group Box 1 (HMGB1) is a non-histone protein that also provides an entry point for SARS-CoV-2 (Andersson et al., 2020). HMGB1 is involved in organizing chromatin but acts as a damage signal when released by cells, such as neurons and glia, under conditions of stress or inflammation (Paudel et al., 2018). When necrotic cells release DAMP and PAMP molecules in the extracellular milieu, they can bind with HMGB1. These complexes of HMGB1 and DAMP and PAMP signals are then taken up by the cell through endocytosis and translocated to lysosomes. This activity leads to increased proinflammatory effects by breaking down the lysosomal membrane and releasing cytokines and other factors into the cytosol (Andersson et al., 2020). The extent to which chloroquine compounds may provide some benefit in COVID infections is that they might prevent the transfer of PAMPs and DAMPs containing SARS-CoV-2 RNA to the cytosol (Andersson et al., 2020).

Previous reports have demonstrated correlations between severe SARS-CoV-2 infection and cell cycle arrest in the S/G2 phase (Suryawanshi et al., 2021). For example, the C-terminus of the E-protein of the SARS-CoV and SARS-CoV-2 shares a very similar motif to the N-terminus of histone 3 (Gordon et al., 2020). Recently several proteins involved chromatin remodeling were identified in a genome-wide CRISPR screen in Vero-E6 cells infected with SARS-CoV-2, MERS-CoV, bat HKU5 expressing the SARS-CoV-1 S protein, and the vesicular stomatitis virus expressing the SARS-CoV-1 S protein. The authors found that AT-rich interactive domain-containing protein 1A (ARID1A) was a pro-viral gene in the case of infection by SARS-CoV-2 and MERS-CoV viruses (Wei et al., 2021). ARID1A/B is a component of the mammalian BRG1/BRM (BAF) complex, involved in chromatin remodeling and cell cycle arrest (Shigetomi et al., 2011; Pagliaroli and Trizzino, 2021). ARID1A is ubiquitously expressed in neural stem progenitor cells and throughout the brain (Liu et al., 2021). Another cellular component found to be perturbed by SARS-CoV-2 infection was the SWI/SNF (SWItch/Sucrose Non-Fermentable) complex, which is responsible for ATP-dependent chromatin remodeling. Interference with cell cycle progression allows the SARS-CoV-2 to hijack cellular machinery to increase viral replication (Kumar et al., 2021).

Cell Survival

Bone-derived neurotrophic factor (BDNF) is a growth factor that plays a role in neurotransmission and neuroplasticity. It is expressed throughout the brain, including in astrocytes, Schwann cells, and neurons (Sakharnova and Vedunova, 2012). BDNF binds to tyrosine kinase B (Trk B), initiating a signal cascade that leads to the activation of the mechanistic target of rapamycin (mTOR), which promotes survival, growth, and differentiation of neurons (Bar-Yosef et al., 2019). SARS-CoV-2 has been demonstrated to enhance mTOR complex 1 (mTORC1) activity (Bar-Yosef et al., 2019). Calmodulin is not only an essential regulator of cellular activity, including apoptosis, neurotransmitter release, etc. (Yu et al., 2002; Ando et al., 2013; Schweitzer et al., 2021). Solute carrier family six-member 20 (SLC6A20) plays a role in the regulation of glycine as well as *N*-methyl-D-aspartate (NMDA) signaling (Bae et al., 2021).

Senescence and Mitophagy

SARS-CoV-2, like many other viruses, is thought to induce senescence in host cells through the increased binding of Angiotensin II (ANGII) to the Angiotensin II Type 1 receptor. ANGII acts as a toxin with respect to the host's cells' mitochondria through activation of nicotinamide adenine dinucleotide phosphate (NADPH) oxidase and the creation of reactive oxygen species (ROS), H_2O_s (Chang et al., 2020). This increase leads to the formation of hydroxyl radicals that cause DNA damage and the activation of poly ADP-ribose polymerases (PARPs), which are DNA damage sensors and deplete stores of NAD+ and exacerbate both the dysfunction of mitochondria. The depletion of NAD+ also results in the reduced mitophagy the increased formation of ROS, which in turn activates ADAM17 and inhibits nitric oxide (NO) synthesis (Dikalov and Nazarewicz, 2013; Chang et al., 2020; Sfera et al., 2020). ADAM17 is also a metalloprotease that has been reported to prime the SARS-CoV-2 spike protein (Heurich et al., 2014).

Autophagy

Autophagy plays an essential role in the homeostatic balance between cell survival and cell death. Previous work has shown that coronaviruses MERS-CoV and SARS-CoV can

prevent autophagosomes from binding to lysosomes (Randhawa et al., 2020). The SARS-CoV-2 infection has been shown to reduce zinc finger FYVE and coiled-coil domain-containing autophagy adaptor 1 (FYCO1) expression, which participates in autophagosome maturation through the Rab7 effector protein, a late endosomal GTPase (Cheng et al., 2007; Pankiv et al., 2010; Randhawa et al., 2020). FYCO1 is expressed in several different cell types within the cortex (Mestres et al., 2020). The cysteine proteases cathepsin B (CTSB) and cathepsin L (CTSL) have also been implicated, alongside TMPRSS2, in the activation of the S proteins of the SARS-CoV-1, SARS-Cov-2, and MERS-CoV coronaviruses. These proteases are found in endosomes/lysosomes and participate in autophagy and apoptosis (Pišlar et al., 2020). Cathepsins consists of serine, aspartic, and cysteine proteases and are ubiquitously expressed (Vidak et al., 2019). Although the cysteine cathepsins are primarily located within the lysosome, where the acidic environment maintains their stability, the excess secretion of cathepsins is associated with inflammatory responses and disease (Huang et al., 2006; Gomes et al., 2020; Pišlar et al., 2020). Previous research has demonstrated host cell entry of corona-pseudoviruses via CTSL dependent endocytosis, and cysteine protease inhibitors effectively blocked viral entry (Simmons et al., 2005, 2011; Zhou et al., 2011; Rabaan et al., 2017). Much like ACE2 and TMPRSS2, CTSB/L is enriched in the lungs (Darbani, 2020). However, the gene expression of the CTSB/L in the cortex and cerebellum was greater relative to the gene expression of ACE2 and TMPRSS2, which were nearly undetectable in the same tissue (Darbani, 2020).

Endocrine Signaling

Testosterone levels have emerged as a risk factor for severe SARS-CoV-2 infection, and sex hormone signaling genes have been identified in previous investigations as potential targets in the treatment of SARS-CoV-2. Androgen receptors (ARs) are expressed through the CNS; however, the cortical expression of the AR is higher relative to other structures (Schumacher et al., 2021). The receptor influences the expression of ACE2 and TMPRSS2. Previous investigations of the effects of anti-androgenic drugs on the expression of genes related to the pathogenesis of SARS found that AR is a transcriptional regulator of ACE2, Furin, and TMPRSS2 (Samuel et al., 2020; Wambier et al., 2020). The TMPRSS2 gene is a target of the androgen receptor, which enhances transcription of TMPRSS2 (Clinckemalie et al., 2013; Samuel et al., 2020). It is, therefore, worth noting that hyperandrogenism in women has been associated with a greater risk of severe complications related to COVID-19 infection (Moradi et al., 2020). Previous investigations have demonstrated that the estrogen receptor (ER) is expressed by all neural cells and plays a role in resistance to infection and influences cytokine and macrophage activity (Seli and Arici, 2002; Villa et al., 2016). Interventions targeting estrogen and estradiol have been proposed as potential treatments for SARS-CoV-2 (Hussman, 2020). Sex-binding globulin (SHBG) is produced and secreted by the liver, and it binds sex hormones such as testosterone, and estrogen, thus regulating their levels in the bloodstream (Colangelo et al., 2012). An observational study of COVID-19 patients found lower SHBG levels in patients who died.

POTENTIAL MECHANISMS OF SARS-CoV-2 MEDIATED MENTAL DISORDERS

Several of the host proteins genes that contribute to the pathobiology of SARS-CoV-2 infection, such as those involved in chromatin remodeling, are critical in the development of the central nervous system (Moffat et al., 2019; Torres-Berrío et al., 2019; Pagliaroli and Trizzino, 2021). Other host proteases and cellular receptors are involved in neurodevelopment, cellular proliferation, neurotransmitter release, sympathetic nervous system activation, neuroinflammation, etc. (Seidah, 2011). For example, factors involved in chromatin remodeling such as SWI/SNF and HMGB1 the SWI/SNF complex are important to embryonic and neurodevelopment. Dysfunction in genes associated with this complex are associated with neuropsychiatric disorders, neurodegenerative disorders, and intellectual disability (Marballi et al., 2014; Son and Crabtree, 2014; Vogel-Ciernia and Wood, 2014; Gozes, 2017; Paudel et al., 2018). Meanwhile, markers of neuroinflammation like XCR1 and CCXR1 are also implicated in stress, infection, and traumatic brain injury (Ciechanowska et al., 2020). These conditions lead to the presence of damage signals or antigens that can thereby be recognized by receptors such as toll-like receptors. The binding of these signal molecules then initiates signaling pathways, which lead to increased expression of inflammatory cytokines. This, in turn, leads to the activation of the hypothalamic-pituitary-adrenal (HPA) axis and sympathetic nervous system and the release of adrenaline, epinephrine, etc. (Canneva et al., 2015). The SARS-Cov-2 infection has also been known to trigger Guillain–Barre Syndrome, an autoimmune disorder characterized by demyelination of peripheral nerve axons (Toscano et al., 2020). Neuroinflammation and autoimmune disorders such as rheumatoid arthritis and celiac disease have been linked to mental health disorders such as BPD, SCZ, and psychosis (Eaton et al., 2010; Bergink et al., 2014; Dasdemir et al., 2016; Goldsmith et al., 2016; Hong et al., 2017; Milenkovic et al., 2019). A study of a large cohort of 3.57 million births linked to the Psychiatric Care Register in Denmark found that the relative risk for individuals diagnosed with an autoimmune disorder to be diagnosed with SCZ was 1.2 (Eaton et al., 2010). In this section, we examine the host factors that play critical roles in the etiology of mental disorders (**Table 1**).

Anxiety, Depression, and Suicide

SARS-CoV-2 entry protein, ACE2, exert neuroinhibitory influence within brain regions such as the middle temporal gyrus and posterior cingulate cortex (Chen et al., 2020). Angiotensin (Ang) 1–7, a product of ACE2, decreases the synthesis and reuptake of noradrenaline and increases its uptake (Gironacci et al., 2014). ACE-2 and Mas protein regulate brain function and release neurotrophic factors, like BDNF (Steenblock et al., 2020). This factor has several critical roles, including the formation,

development, and inhibition of degeneration of the neurons. It also plays a role in stabilizing mood and in cognitive function. Decreases in ACE-2 activity or expression have been known to disturb normal neurologic functions. This inhibition of ACE2 and subsequent decrease in BDNF leads to neurodegeneration and may cause mental disorders such as anxiety, depression, and cognitive impairment (Steenblock et al., 2020). It is important to note that the AR regulates the expression of ACE2 and TMPRSS2. Both AR and TMPRSS2 are implicated in prostate cancer, and some data suggests that there may be an association between prostate cancer and depression and anxiety (Newby et al., 2015; Rice et al., 2018; Wang et al., 2020).

Other host proteins such as neuropilin and DPP4 also relate to depressive symptoms. The expression of neuropilin in olfactory epithelia seems to be related to major mental disorders such as MDD. In one investigation, the authors found a significantly higher expression of NRP1 in post-mortem samples from the PFC of patients diagnosed with MDD when compared to controls ($p < 0.001$) (Goswami et al., 2013). Similarly, the expression of neuropeptide Y (NPY), a ligand for the DPP4 receptor, has been proposed as a biomarker for diagnosing PTSD, depression, and other neuropsychiatric illnesses (Canneva et al., 2015; Schmeltzer et al., 2016; Gołyszny and Obuchowicz, 2019). NPY has anxiolytic effects, and in an investigation, NPY immunoreactivity was significantly decreased in the cerebral spinal fluid (CSF) of unmedicated patients with persistent unipolar depression (Heilig et al., 2004). There is currently a clinical trial underway to investigate the value of Vildagliptin, an anti-diabetic drug, as adjunctive therapy to the SSRI, Escitalopram, and PDE3 inhibitor, Cilostazol, for the treatment of MDD (Clinical Trial ID: NCT04410341). It is also worth noting that cathepsins play a role in processing proneuropeptides like neuropeptide Y and have been found to be moderately associated with higher cognitive function following exercise training (Funkelstein et al., 2008, 2012; Moon et al., 2016).

Neuroinflammatory and immune responses are known to contribute to the development of mental disorders. One investigation of postmortem tissue taken from the dorsolateral prefrontal cortex (DLPFC) found increased mRNA expression of TLR3 and TLR4 and the increased presence of pro-inflammatory factors in the brains of depressed non-suicidal, and suicidal subjects (Pandey et al., 2014). Increased expression of TLR3 also results in reduced expression of disrupted in schizophrenia 1 (DISC1), which leads to aberrant neuronal morphology (Chen et al., 2017). In fact, previous research has shown that treatment with endotoxin to stimulate inflammatory cytokines or even treatment with inflammatory kinases themselves can lead to symptoms of depression in people who were previously undiagnosed (Bonaccorso et al., 2002; Anttila et al., 2018). Another investigation of the unfolded protein response in rats found increased expression of TLRs 2, 4, 7, and 9 as well as inflammatory cytokines within the hippocampus (Timberlake et al., 2019). In patients with hepatitis C, interferon-alpha (IFN-α) treatment can lead to clinical symptoms of depression, which can be alleviated by antidepressant therapy. This finding suggests that depression is caused by inflammation, and typical presentations of depression may have some similar underlying mechanisms.

Expression of inflammatory markers, such as chemokine CXCR6, by meningeal γδ T cells, has been shown to influence anxiety in mice (Alves de Lima et al., 2020). Mice deficient in CXCR6 have been demonstrated to have fewer γδ T cells than controls. The γδ T cells, in turn, release IL-17, a gene implicated in autism spectrum disorder (ASD) and SCZ (Debnath and Berk, 2017; Alves de Lima et al., 2020). γδ T cell-deficient mice demonstrated reduced anxiety behavior in the open field test. The authors showed that these cells could control anxiety behavior through IL-17 signaling (Alves de Lima et al., 2020).

Sex hormones and neuroimmune responses play converging roles in the etiology of mental disorders (Kokkosis and Tsirka, 2020). Lower testosterone is a predictor of depression symptoms in men, while higher levers of free testosterone in serum have been linked to manic episodes in men (Ozcan and Banoglu, 2003; Sankar and Hampson, 2012). Anti-androgenic therapies have been considered as a potential treatment for vulnerable populations (Bravaccini et al., 2021). A greater risk of depressive symptoms was positively associated with SHBG in a study of depressive disorders in post-menopausal women. SHBG was positively and statistically associated with depression risk ($p = 0.003$) in all women (Colangelo et al., 2012). In a study of schizophrenic male patients and a group of undiagnosed adults, the authors found that both treated and untreated patients had lower serum levels of SHBG (33.3 and 26.6 nmol/L) than undiagnosed controls (48.4 nmol/L, $p < 0.05$) (Costa et al., 2006).

Bipolar Disorder, Schizophrenia, and Psychosis

Several infections, such as cytomegalovirus, herpes simplex virus, and parasitic infection by *T. gondii*, have been noted for interacting with the HLA system and for their association with affective disorders like BPD and SCZ (Parks et al., 2018). SCZ has been linked to several polymorphisms in the major histocompatibility complex (MHC) or the HLA system through several GWAS (Parks et al., 2018). HLA genes are expressed in astrocytes and microglia within the brain, although primarily in microglia (Tian et al., 2012). Previous investigations have identified HLA-B*4601, HLA-B*0703, HLA-B*4601, HLA-C*0801, and HLA-DRB1*1202 as alleles associated with severe illness following SARS-CoV-2 infection (Lin et al., 2003; Ng et al., 2004; Morsy and Morsy, 2021). The HLA-DRB*0301 and HLA-Cw*1502 alleles were associated with a reduced frequency of severe infection. These clusters, namely HLA-B, HLA-DRB1, HLA-C, HLA-DRA, HLA-DQA, HLA-DQB, HLA-DPB, have been associated with mental health disorders, i.e., SCZ, BPD, and PTSD (Carter, 2009). A gene-wide association study of 13,4982 cases and 663 controls found significant associations between the Notch 4 intronic variant rs3131296 and HLA alleles: HLA-DRB1*0301 and HLA-B*0801 (R^2 = 0.86 and 0.81, respectively) (Stefansson et al., 2009). In a separate investigation of molecular pathways underlying SCZ and BPD, the authors found that patients demonstrate more variation in the HLA-C and HLA-DRA genes than would be expected by chance (Marco et al., 2015). It is also worth noting that PTSD has also been found to be associated with

HLA alleles (HLA-B*5801, HLA-C*0701, HLA-DQA1*0101, HLA-DQB1*0501, and HLA-DPB1*1701) in a case-control study of 403 diagnosed patients with 369 individuals who had been exposed to trauma (Katrinli et al., 2019).

Differential levels of cytokines and cytokine receptors have been found between the first episode, acute relapse of psychosis, and post-treatment patients diagnosed with SCZ (Capuzzi et al., 2017). A previous meta-analysis of first-episode psychosis, acute relapse, and post-treatment SCZ patients compared effect sizes of blood levels of inflammatory markers (cytokines, cytokine receptors, and antagonists) (Miller et al., 2011). Significant differences between the effect sizes of several inflammatory markers were found between post-treatment patients diagnosed with SCZ and first-episode psychosis and acutely relapsed patients (Miller et al., 2011). These cytokines and receptors include including IL-6, IL-12, TNF-α, IL-1β, IL-8, transforming growth factor-β (TGF-β), IL-1RA, IFN-γ, sIL-2R, and IL-10 (Miller et al., 2011). It is worth noting that levels of IL-6 and TNF-α were significantly correlated with survival in SARS-CoV-2 in a previous investigation (Del Valle et al., 2020). Increased levels of circulating IL-1β, IL-12, IL-6, CXCL10, CCL2 have been reported in severe cases of SARS-CoV-2 infection (Coperchini et al., 2020). TGF-beta plays a role in regulating immune response and plays a role in the development of mental disorders like SCZ and symptoms such as psychosis (Sanjabi et al., 2017; Roomruangwong et al., 2020). Increased expression of chemokines such as CCR1 has been shown in postmortem brain tissue in patients diagnosed with SCZ compared to patients diagnosed with BPD (de Baumont et al., 2015).

Metalloproteases ADAM10/17 have been implicated in neurodegenerative disorders. They play a role in the proteolysis of the amyloid precursor protein (APP) and several other proteins (Vincent and Govitrapong, 2011; Qian et al., 2016). Similarly, another protein affected by SARS-CoV-2 infection, FYCO1, has been linked to neurodegenerative disorders, neuropsychiatric disorders, and senescence in age-accelerated mice (Cheng et al., 2007; Saridaki et al., 2018). However, ADAM10/17 have also been linked to SCZ, depression, BPD, and conduct disorder, a condition that has been found to be comorbid with mood disorders (Jian et al., 2011; Marballi et al., 2012; Qian et al., 2016; Hoseth et al., 2017; Yuan et al., 2017; Pantazopoulos et al., 2020). A family-based association study found 20 variants associated significantly associated with conduct disorder; among these single nucleotide polymorphisms (SNPs), rs383902 was located within ADAM10 ($p = 0.00036$) (Jian et al., 2011). In one investigation of postmortem brain tissue from BA9, using ANCOVA analysis, investigators found a significant difference in ADAM17 expression between the control and bipolar groups and levels observed in the schizophrenic group ($p < 0.007$). The authors also found a significant negative correlation between levels of neuregulin-1 (NRG-1) and ADAM17 in Broca's area 9 samples taken from the post-mortem tissues of patients diagnosed with SCZ (Marballi et al., 2012). Similarly, Hoseth et al. (2017) found greater mRNA expression of ADAM17 in the plasma of SCZ patients vs. that seen in controls (Hoseth et al., 2017). ADAM10/17 influences glutamatergic signaling, which is also impacted by the SLC6A20 transporter protein. In a GWAS of NMDA receptors and the detection of their coagonists in cerebrospinal fluid, the authors found that a missense variant in SLC6S20 as associated with increased L-proline levels in CSF, thus demonstrating that SLC6A20 plays a role in the trafficking of proline to the CSF (Luykx et al., 2015). Hyperprolinemia has been previously reported in conjunction with SCZ and schizoaffective disorder (Jacquet et al., 2005; Clelland et al., 2011).

Neurotransmission may also be related to the expression of SARS-CoV-2 entry protease FURIN. FURIN was found among several genes linked to comorbidity SCZ and cardiometabolic illness, which gives insight into the etiology of these conditions (Liu et al., 2020). Several studies underscore the importance of furin in the CNS, as it has been linked to Alzheimer's disease (AD) and SCZ (Scamuffa et al., 2006; Schrode et al., 2019; Yang et al., 2020). In a GWAS of 49 ancestry matched non-overlapping case-controls and 1,235 parent affected offspring trios, the authors found 108 loci that were significantly associated with SCZ (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014). Of those, Fromer et al. (2016) found nineteen of the SCZ risk loci were enriched for eQTLs. However, only eight involved a single gene; among them was the gene encoding furin protease. The authors found that furin expression was downregulated by the risk variant rs4702 (GG to AA allelic conversion), a 3′ UTR variant, which was both the most significant SNP detected by GWAS and eQTL analyses (Fromer et al., 2016). The rs4702 SNP results in the alteration in the binding site for miR-338-3p. miR-338-3p is an mRNA that is expressed predominantly in mature neurons within the dentate gyrus (Howe et al., 2017). The authors noted that cells in which miR-338-3p was effectively knocked down showed aberrations in the number of primary dendrites as well as the angles of their extension from the soma (Howe et al., 2017). Interestingly enough, the rs4702 variant, which is associated with SCZ, may be protective against SARS-CoV-2 infection, as cells expressing rs4702 had reduced levels of vRNA relative to cells expressing the normative allele (AA) (Dobrindt et al., 2021). In a separate investigation, the rs4702 specific reduction in the expression of FURIN and BDNF was "mediated" by miR-338-3p (Hou et al., 2018). BDNF is a member of the BDNF-mTORC1, which helps to regulate synaptic plasticity, glutamatergic signaling, monoaminergic signaling, and autophagy. The SARS-CoV-2 infection has been demonstrated to increase the activity of mTORC in Vero kidney epithelial cells 24 h post-infection (Mullen et al., 2021). A common SNP of BDNF, rs62265, is a missense mutation that has been associated with anxiety, major depression, suicide, and neurodegenerative disease, as has dysregulation of mTOR signaling (Dincheva et al., 2016; Youssef et al., 2018; Bar-Yosef et al., 2019). Additionally, decreased BDNF expression has also been associated with SCZ (Suchanek-Raif et al., 2020). Epigenetic regulation of BDNF has also been demonstrated to play a role in mental illness, as methylation of genes associated with SCZ, like BDNF, has been linked to psychosis (Gavin et al., 2010).

Increased levels of GSK-3β in blood, serum, and CSF have been associated with SCZ (Mohammadi et al., 2018). The GSK-3β inhibitor lithium, which is utilized as a treatment for

psychiatric disorders such as SCZ and BPD, has been shown to inhibit infection by several viruses, including coronaviruses (Murru et al., 2020). One investigation examined gene expression in 12 BPD patients and ten controls following two laser microdissections of the olfactory epithelia: one pretreatment with lithium in the second after 6 weeks of daily lithium treatment. The BPD patients demonstrated greater levels of GSK-3β than controls (Narayan et al., 2014). Lithium has been shown to inhibit GSK-3β, and similar to those studies the authors found that GSK-3β was reduced in the second microdissection samples taken from BPD patients following 6 weeks of daily lithium treatment (Harrison et al., 2007; Narayan et al., 2014; Zhao et al., 2017).

Genes responsible for chromatin remodeling are implicated in SCZ as well. The SWI/SNF complex protein ARID1A is typically associated with craniofacial abnormalities. However, mutations in the ARID1B gene have been associated with intellectual disability, ASD, and SCZ as well (Son and Crabtree, 2014; Pagliaroli and Trizzino, 2021). SWI/SNF-related matrix-associated actin-dependent regulator of chromatin, family a, member 2 gene (SMARCA2), encoding the SWI/SNF subunit, Brahma (BRM), has been associated with self-reported MDD and SCZ (Amare et al., 2020; Wu et al., 2020a). In an investigation of bivariate analyses of genome-wide association study results relating to depression combined with MDD, BPD, and SCZ, the authors found that the SMARCA2 gene and the SWI/SNF gene set were enriched. This indicated the role of epigenetic mechanisms in the etiology of complex mental health disorders (Amare et al., 2020). In a separate investigation, drugs capable of inducing psychosis were found to reduce BRM expression, while anti-psychotics led to increased expression of BRM (Koga et al., 2009).

Further demonstrating the importance of chromatin remodeling proteins in the etiology underlying SCZ is the HMGB1 protein. An investigation of gene expression in post-mortem tissue from 212 patients with SCZ and 214 undiagnosed controls found 87 genes demonstrated expression variability, including HMGB1 (Huang et al., 2020). In a separate study, serum levels of HMGB1 were significantly higher in bipolar patients than in undiagnosed controls. The authors found that serum levels of HMGB1 were significantly higher in the bipolar patients compared to the controls137. A systematic review of the literature also found increased levels of HMGB1 in conjunction with several studies of mouse models of depression (Zhang et al., 2019). One drug, minocycline, was found to reduce depressive-like symptoms in a mouse model of depression. This reduction was associated with a significant decrease in the translocation of HGMB1 from neurons and microglia (Wang et al., 2020).

Sex hormones have also been shown to play a role in the risk of developing SCZ (Kokkosis and Tsirka, 2020). Women with polycystic ovarian syndrome (PCOS) have also been demonstrated to be at greater risk of developing psychiatric disorders such as bipolar disorder and SCZ (Owens et al., 2019). An investigation of androgen receptor expression among individuals diagnosed with SCZ, BPD, and undiagnosed controls (n = 35, 31, and 34, respectively) found increased expression of AR among individuals diagnosed with bipolar disorder relative to individuals diagnosed with SCZ and control volunteers. No significant differences were observed in 5α-reductase between the experimental groups. However, a small but significant correlation was found between bipolar disorder and 5α-reductase expression (r = 0.422, p < 0.01) (Owens et al., 2019). Hormones have been demonstrated to affect neuropeptides involved in stress and anxiety, like oxytocin and corticotropin-releasing hormone (CRH) (Wang and Wang, 2021). Previous research has linked decreased oxytocin and oxytocin receptor levels to first-episode SCZ and bipolar II disorder (Liu et al., 2019b; Wei et al., 2020). The therapeutic use of OXT has been proposed as a treatment to protect against cardiovascular damage caused by SARS-CoV-2 infection (Wang and Wang, 2021).

DISCUSSION

Several host genes affected by SARS-CoV-2 infection are implicated in mental disorders and neuropsychiatric symptoms. Of the host genes perturbed by the coronavirus spike protein, many are involved in innate and adaptive immunity, stress response, cell cycle regulation, and other biological functions. These genes have also been implicated in mental disorders such as depression, SCZ, and bipolar disorder. Other components of the SARS-CoV-2 virion, such as E and N proteins on host proteins PSD95 and RIG-1, also relate to neuropsychiatric symptoms (Javorsky et al., 2021; Oh and Shin, 2021).

Several other genes that are dysregulated in mental disorders, such as DISC1, phosphodiesterase 4B (PDE4B), and neurexin-1 (NRXN1), could also be impacted by SARS-CoV-2 infection and contribute to neurotropism and inflammation in the CNS. We previously noted that increased TLR3 signaling leads to reduced DISC1 expression and aberrant neurogenesis. A recent transcriptomics study of publicly available datasets demonstrated that DISC1 is downregulated by COVID-19 (Alqutami et al., 2021). Though the exact role that DISC1 plays in complex mental disorders is unclear, DISC1 is an important component in the formation of the immune synapse. DISC1 forms a complex with Girdin and dynein that allows for the translocation of the microtubule-organizing center to the synapse; however, in DISC1 knock-out cell lines, the MTOC fails to translocate to the immune synapse (Maskalenko et al., 2020). The DISC1 pathway is a massive multi-step pathway of 203 genes that can be subdivided in the interactome and regulome (Teng et al., 2017). DISC1 and the DISC1 pathway genes like PDE4B and NRXN1 are implicated in several mental health disorders (Millar et al., 2007; Korth, 2009; Hu et al., 2019). PDE4B is found in the DISC1 interactome, and differential expression of PDE4B has also been noted in relation to COVID-19 infection (Alqutami et al., 2021). PDE4B has been shown to regulate cytokine signaling pathways (Lugnier et al., 2021; Moolamalla et al., 2021). Several adjunct therapies for the treatment of SARS-CoV-2 symptoms have been identified that target PDE4B (Lugnier et al., 2021; Moolamalla et al., 2021). Studies of microRNAs as potential targets of treatments for viral infection have shown that miR-1290 is upregulated in SARS-CoV-2 infection, and this is predicted to result in downregulation of NRXN1 expression (Chen and Wang, 2020; Guterres et al., 2020).

Long-lasting pulmonary symptoms, pain, fatigue, and other symptoms stemming from coronavirus infection have been documented throughout the literature. However, currently, no studies have investigated the mechanisms concerning the long-lasting mental health symptoms or disorders that might result from COVID-19 infection. However, several publications have enumerated observations of long-COVID neuropsychiatric symptoms and life stressors that affect mental health (Crook et al., 2021). One published review listed several probable risk factors related to PTSD and psychological dysfunction, including isolation, loss of a loved one, disability, and occupation (Boyraz, 2020). An investigation of1,427 United States adults reported the percentage of respondents reporting depressive symptoms increased from 27.8% in early 2020 to 32.8% just 1 year later (Ettman et al., 2021).

It is unclear what the precise causes of long COVID or neuropsychiatric symptoms could be the result of neuroinvasion by coronavirus in the brain and CNS or could result from systemic inflammation or a combination of both. There are conflicting studies regarding the specific ability of coronaviruses to cross the blood-brain barrier and infect the CNS or to be transmitted from neuron to neuron via the olfactory bulb (Thye et al., 2022). However, clinical observations of anosmia and encephalitis would suggest that SARS-CoV-2 and other coronaviruses are capable of both (Mondelli and Pariante, 2021). The precise mechanisms leading to long-term psychological sequelae are yet elusive. Some investigators have concluded that there may be myriad factors contributing to long COVID cases, including prolonged inflammation, ischemia, neuroinvasion, prolonged sedation, etc. (Alonso-Lana et al., 2020; Song et al., 2020). Given the comorbidity between autoimmune disorders and mental disorders and observations of increased levels of pro-inflammatory factors in the absence of encephalitis, it seems that inflammation is likely the underlying cause (Soria et al., 2018; Alonso-Lana et al., 2020; Proal and VanElzakker, 2021).

Several animal models suitable to the study of COVID-19 are currently available. Among these models are rhesus macaques, ferrets, mice expressing the human ACE2 receptor, and Golden Syrian hamsters (Jia et al., 2021). All these animal models feature pathological symptoms related to human pathological symptoms encountered with COVID-19. These symptoms include mild to moderate pneumonia, increased inflammatory markers, and weight loss. However, only two of these models are commonly utilized to investigate behavioral phenotypes: rhesus macaques and mice. Rhesus macaques are animal models that are used to investigate mental disorders such as anxiety. Many studies utilize transgenic mice to investigate obsessive-compulsive disorder (OCD), depression, SCZ, and ASD.

In this review, we summarized several host factors and pathways that are involved in coronavirus infection and are also implicated in neuropsychiatric symptoms. Though several of these host factors are expressed in the CNS, we have also provided evidence that their influence on widespread systemic inflammation may play a significant role in the development of long-term psychological symptoms stemming from COVID-19 infection. We've highlighted several cellular mechanisms that are impacted by SARS-CoV-2 infection and connected them to complex mental disorders such as MDD, SCZ, and BPD. We have elucidated the connection between DISC1 and DISC1 pathway proteins such as NRXN1 and PDE4B to viral infection as well as to mental disorders.

Future work should focus on the mechanisms by which infectious diseases like COVID-19 may impact mental illnesses of neuropsychiatric symptoms. This knowledge could contribute to interventions to lessen the effects of infection on the central nervous system or inform the development of treatments for existing mental disorders. Some of the host factors described here are already being investigated for their potential use as therapies or co-therapies for mental illness symptoms. However, further investigation is necessary to determine what impact coronavirus and other flu-like infections may have on mental symptoms and disorders. These investigations could elucidate the biological changes underlying the etiology of complex mental illnesses like SCZ, BPD, and depression.

AUTHOR CONTRIBUTIONS

RR and ST: conceptualization and writing—review and editing. RR, SS, CJ, and ST: writing—original draft preparation. ST: supervision and funding acquisition. All authors contributed to the article and approved the submitted version.

FUNDING

This research was supported by the Howard University startup funds [U100193] and National Science Foundation [DBI 2000296 and IIS 1924092]. This project was supported (in part) by the National Institute on Minority Health and Health Disparities of the National Institutes of Health [2U54MD007597]. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

REFERENCES

Acharya, A., Kevadiya, B. D., Gendelman, H. E., and Byrareddy, S. N. (2020). SARS-CoV-2 infection leads to neurological dysfunction. *J. Neuroimmune Pharmacol.* 15, 167–173. doi: 10.1007/s11481-020-09924-9

Alcock, J., and Masters, A. (2021). Cytokine storms, evolution and COVID-19. *Evol. Med. Public Health* 9, 83–92. doi: 10.1093/EMPH/EOAB005

Aljohmani, A., and Yildiz, D. (2020). A disintegrin and metalloproteinase—control elements in infectious diseases. *Front. Cardiovasc. Med.* 7:608281. doi: 10.3389/fcvm.2020.608281

Alonso-Lana, S., Marquié, M., Ruiz, A., and Boada, M. (2020). Cognitive and neuropsychiatric manifestations of COVID-19 and effects on elderly individuals with dementia. *Front. Aging Neurosci.* 12:588872. doi: 10.3389/FNAGI.2020.588872

Alqutami, F., Senok, A., and Hachim, M. (2021). COVID-19 transcriptomic atlas: a comprehensive analysis of COVID-19 related transcriptomics datasets. *Front. Genet.* 12:755222. doi: 10.3389/FGENE.2021.755222

Alves de Lima, K., Rustenhoven, J., Da Mesquita, S., Wall, M., Salvador, A. F., Smirnov, I., et al. (2020). Meningeal γδ T cells regulate anxiety-like behavior

via IL-17a signaling in neurons. *Nat. Immunol.* 21, 1421–1429. doi: 10.1038/s41590-020-0776-4

Amare, A. T., Vaez, A., Hsu, Y.-H., Direk, N., Kamali, Z., Howard, D. M., et al. (2020). Bivariate genome-wide association analyses of the broad depression phenotype combined with major depressive disorder, bipolar disorder or schizophrenia reveal eight novel genetic loci for depression. *Mol. Psychiatry* 25, 1420–1429. doi: 10.1038/s41380-018-0336-6

Andersson, U., Ottestad, W., and Tracey, K. J. (2020). Extracellular HMGB1: a therapeutic target in severe pulmonary inflammation including COVID-19? *Mol. Med.* 26:42. doi: 10.1186/S10020-020-00172-4

Ando, K., Kudo, Y., Aoyagi, K., Ishikawa, R., Igarashi, M., and Takahashi, M. (2013). Calmodulin-dependent regulation of neurotransmitter release differs in subsets of neuronal cells. *Brain Res.* 1535, 1–13. doi: 10.1016/J.BRAINRES.2013.08.018

Anttila, V., Bulik-Sullivan, B., Finucane, H. K., Walters, R. K., Bras, J., Duncan, L., et al. (2018). Analysis of shared heritability in common disorders of the brain. *Science* 360:eaa8757. doi: 10.1126/science.aap8757

Arias, I., Sorlozano, A., Villegas, E., Luna, J., de, D., McKenney, K., et al. (2012). Infectious agents associated with schizophrenia: a meta-analysis. *Schizophr. Res.* 136, 128–136. doi: 10.1016/j.schres.2011.10.026

Bae, M., Roh, J. D., Kim, Y., Kim, S. S., Han, H. M., Yang, E., et al. (2021). SLC6A20 transporter: a novel regulator of brain glycine homeostasis and NMDAR function. *EMBO Mol. Med.* 13:e12632. doi: 10.15252/emmm.202012632

Bai, Z., Cao, Y., Liu, W., and Li, J. (2021). The SARS-CoV-2 nucleocapsid protein and its role in viral structure, biological functions, and a potential target for drug or vaccine mitigation. *Viruses* 13:1115. doi: 10.3390/V13061115

Bar-Yosef, T., Damri, O., and Agam, G. (2019). Dual role of autophagy in diseases of the central nervous system. *Front. Cell. Neurosci.* 13:196. doi: 10.3389/fncel.2019.00196

Bergink, V., Gibney, S. M., and Drexhage, H. A. (2014). Autoimmunity, inflammation, and psychosis: a search for peripheral markers. *Biol. Psychiatry* 75, 324–331. doi: 10.1016/j.biopsych.2013.09.037

Bilinska, K., Jakubowska, P., Von Bartheld, C. S., and Butowt, R. (2020). Expression of the SARS-CoV-2 entry proteins, ACE2 and TMPRSS2, in cells of the olfactory epithelium: identification of cell types and trends with age. *ACS Chem. Neurosci.* 11, 1555–1562. doi: 10.1021/acschemneuro.0c00210

Bonaccorso, S., Marino, V., Puzella, A., Pasquini, M., Biondi, M., Artini, M., et al. (2002). Increased depressive ratings in patients with hepatitis C receiving interferon-α-based immunotherapy are related to interferon-α-induced changes in the serotonergic system. *J. Clin. Psychopharmacol.* 22, 86–90. doi: 10.1097/00004714-200202000-00014

Bouças, A. P., Rheinheimer, J., and Lagopoulos, J. (2020). Why severe COVID-19 patients are at greater risk of developing depression: a molecular perspective. *Neuroscientist* [Online ahead of print]. doi: 10.1177/1073858420967892

Boyraz, G. (2020). Coronavirus disease (COVID-19) and traumatic stress: probable risk factors and correlates of posttraumatic stress disorder critical thinking view project. *J. Loss Trauma* 25, 503–522. doi: 10.1080/15325024.2020.1763556

Bravaccini, S., Fonzi, E., Tebaldi, M., Angeli, D., Martinelli, G., Nicolini, F., et al. (2021). Estrogen and androgen receptor inhibitors: unexpected allies in the fight against COVID-19. *Cell Transplant.* 30:963689721991477. doi: 10.1177/0963689721991477

Brzezinski-Sinai, N. A., and Brzezinski, A. (2020). Schizophrenia and sex hormones: what is the link? *Front. Psychiatry* 11:693. doi: 10.3389/fpsyt.2020.00693

Butowt, R., and von Bartheld, C. S. (2020). Anosmia in COVID-19: underlying mechanisms and assessment of an olfactory route to brain infection. *Neuroscientist* 27, 582–603. doi: 10.1177/1073858420956905

Cai, H., and Reed, R. R. (1999). Cloning and characterization of neuropilin-1-interacting protein: a PSD- 95/Dlg/ZO-1 domain-containing protein that interacts with the cytoplasmic domain of neuropilin-1. *J. Neurosci.* 19, 6519–6527. doi: 10.1523/jneurosci.19-15-06519.1999

Canneva, F., Golub, Y., Distler, J., Dobner, J., Meyer, S., and von Hörsten, S. (2015). DPP4-deficient congenic rats display blunted stress, improved fear extinction and increased central NPY. *Psychoneuroendocrinology* 53, 195–206. doi: 10.1016/j.psyneuen.2015.01.007

Cantuti-Castelvetri, L., Ojha, R., Pedro, L. D., Djannatian, M., Franz, J., Kuivanen, S., et al. (2020). Neuropilin-1 facilitates SARS-CoV-2 cell entry and infectivity. *Science* 370, 856–860. doi: 10.1126/science.abd2985

Capuzzi, E., Bartoli, F., Crocamo, C., Clerici, M., and Carrà, G. (2017). Acute variations of cytokine levels after antipsychotic treatment in drug-naïve subjects with a first-episode psychosis: a meta-analysis. *Neurosci. Biobehav. Rev.* 77, 122–128. doi: 10.1016/j.neubiorev.2017.03.003

Carter, C. J. (2009). Schizophrenia susceptibility genes directly implicated in the life cycles of pathogens: cytomegalovirus, influenza, herpes simplex, rubella, and *Toxoplasma gondii*. *Schizophr. Bull.* 35, 1163–1182. doi: 10.1093/schbul/sbn054

Chang, R., Mamun, A., Dominic, A., and Le, N.-T. (2020). SARS-CoV-2 mediated endothelial dysfunction: the potential role of chronic oxidative stress. *Front. Physiol.* 11:605908. doi: 10.3389/FPHYS.2020.605908

Chen, C., Liu, H., and Hsueh, Y. (2017). TLR 3 downregulates expression of schizophrenia gene Disc1 via MYD 88 to control neuronal morphology. *EMBO Rep.* 18, 169–183. doi: 10.15252/embr.201642586

Chen, R., Wang, K., Yu, J., Howard, D., French, L., Chen, Z., et al. (2020). The spatial and cell-type distribution of SARS-CoV-2 receptor ACE2 in the human and mouse brains. *Front. Neurol.* 11:573095. doi: 10.3389/fneur.2020.573095

Chen, Y., and Wang, X. (2020). miRDB: an online database for prediction of functional microRNA targets. *Nucleic Acids Res.* 48, D127–D131. doi: 10.1093/NAR/GKZ757

Cheng, X. R., Zhou, W. X., Zhang, Y. X., Zhou, D. S., Yang, R. F., and Chen, L. F. (2007). Differential gene expression profiles in the hippocampus of senescence-accelerated mouse. *Neurobiol. Aging* 28, 497–506. doi: 10.1016/j.neurobiolaging.2006.02.004

Ciechanowska, A., Popiolek-Barczyk, K., Ciapała, K., Pawlik, K., Oggioni, M., Mercurio, D., et al. (2020). Traumatic brain injury in mice induces changes in the expression of the XCL1/XCR1 and XCL1/ITGA9 axes. *Pharmacol. Rep.* 72, 1579–1592. doi: 10.1007/s43440-020-00187-y

Clelland, C. L., Read, L. L., Baraldi, A. N., Bart, C. P., Pappas, C. A., Panek, L. J., et al. (2011). Evidence for association of hyperprolinemia with schizophrenia and a measure of clinical outcome. *Schizophr. Res.* 131, 139–145. doi: 10.1016/j.schres.2011.05.006

Clemente-Suárez, V. J., Martínez-González, M. B., Benitez-Agudelo, J. C., Navarro-Jiménez, E., Beltran-Velasco, A. I., Ruisoto, P., et al. (2021). The impact of the COVID-19 pandemic on mental disorders. A critical review. *Int. J. Environ. Res. Public Health* 18:10041. doi: 10.3390/IJERPH181910041

Clinckemalie, L., Spans, L., Dubois, V., Laurent, M., Helsen, C., Joniau, S., et al. (2013). Androgen regulation of the TMPRSS2 gene and the effect of a SNP in an androgen response element. *Mol. Endocrinol.* 27, 2028–2040. doi: 10.1210/me.2013-1098

Colangelo, L. A., Craft, L. L., Ouyang, P., Liu, K., Schreiner, P. J., Michos, E. D., et al. (2012). Association of sex hormones and sex hormone-binding globulin with depressive symptoms in postmenopausal women: the multi-ethnic study of atherosclerosis. *Menopause* 19, 877–885. doi: 10.1097/gme.0b013e3182432de6

Coperchini, F., Chiovato, L., Croce, L., Magri, F., and Rotondi, M. (2020). The cytokine storm in COVID-19: an overview of the involvement of the chemokine/chemokine-receptor system. *Cytokine Growth Factor Rev.* 53, 25–32. doi: 10.1016/j.cytogfr.2020.05.003

Costa, A. M. N., de Lima, M. S., Tosta, J., Filho, S. R., de Oliveira, I. R., Sena, E. P., et al. (2006). Hormone profile in acute psychotic disorders: a cross-sectional comparison of serum hormone concentrations in treated and untreated male patients with schizophrenia. *Curr. Ther. Res.* 67, 350–363. doi: 10.1016/j.curtheres.2006.10.003

Coutard, B., Valle, C., de Lamballerie, X., Canard, B., Seidah, N. G., and Decroly, E. (2020). The spike glycoprotein of the new coronavirus 2019-nCoV contains a furin-like cleavage site absent in CoV of the same clade. *Antiviral Res.* 176:104742. doi: 10.1016/j.antiviral.2020.104742

Crook, H., Raza, S., Nowell, J., Young, M., and Edison, P. (2021). Long covid-mechanisms, risk factors, and management. *BMJ* 374:n1648. doi: 10.1136/BMJ.N1648

Daly, J. L., Simonetti, B., Klein, K., Chen, K. E., Williamson, M. K., Antón-Plágaro, C., et al. (2020). Neuropilin-1 is a host factor for SARS-CoV-2 infection. *Science* 370, 861–865. doi: 10.1126/science.abd3072

Darbani, B. (2020). The expression and polymorphism of entry machinery for covid-19 in human: juxtaposing population groups, gender, and different tissues. *Int. J. Environ. Res. Public Health* 17:3433. doi: 10.3390/ijerph17103433

Dasdemir, S., Kucukali, C. I., Bireller, E. S., Tuzun, E., and Cakmakoglu, B. (2016). Chemokine gene variants in schizophrenia. *Nord. J. Psychiatry* 70, 407–412. doi: 10.3109/08039488.2016.1141981

de Baumont, A., Maschietto, M., Lima, L., Carraro, D. M., Olivieri, E. H., Fiorini, A., et al. (2015). Innate immune response is differentially dysregulated between

bipolar disease and schizophrenia. *Schizophr. Res.* 161, 215–221. doi: 10.1016/j.schres.2014.10.055

de Souza, A., Tavora, F. A., Mahalingam, D., Munster, P. N., Safran, H. P., El-Deiry, W. S., et al. (2020). Commentary: GSK-3 inhibition as a therapeutic approach against SARs CoV2: dual benefit of inhibiting viral replication while potentiating the immune response. *Front. Immunol.* 11:595289. doi: 10.3389/FIMMU.2020.595289

Debnath, M., and Berk, M. (2017). Functional implications of the IL-23/IL-17 immune axis in schizophrenia. *Mol. Neurobiol.* 54, 8170–8178. doi: 10.1007/s12035-016-0309-1

Del Valle, D. M., Kim-Schulze, S., Huang, H. H., Beckmann, N. D., Nirenberg, S., Wang, B., et al. (2020). An inflammatory cytokine signature predicts COVID-19 severity and survival. *Nat. Med.* 26, 1636–1643. doi: 10.1038/s41591-020-1051-9

Dikalov, S. I., and Nazarewicz, R. R. (2013). Angiotensin II-induced production of mitochondrial reactive oxygen species: potential mechanisms and relevance for cardiovascular disease. *Antioxid. Redox Signal.* 19, 1085–1094. doi: 10.1089/ars.2012.4604

Dincheva, I., Lynch, N. B., and Lee, F. S. (2016). The role of BDNF in the development of fear learning. *Depress. Anxiety* 33, 907–916. doi: 10.1002/da.22497

Dobrindt, K., Hoagland, D. A., Seah, C., Kassim, B., O'Shea, C. P., Iskhakova, M., et al. (2021). Common genetic variation in humans impacts in vitro susceptibility to SARS-CoV-2 infection. *Stem Cell Rep.* 16, 505–518. doi: 10.1016/j.stemcr.2021.02.010

Dong, M., Zhang, J., Ma, X., Tan, J., Chen, L., Liu, S., et al. (2020). ACE2, TMPRSS2 distribution and extrapulmonary organ injury in patients with COVID-19. *Biomed. Pharmacother.* 131:110678. doi: 10.1016/J.BIOPHA.2020.110678

Dutta, N. K., Mazumdar, K., and Gordy, J. T. (2020). The nucleocapsid protein of SARS–CoV-2: a target for vaccine development. *J. Virol.* 94:e00647-20. doi: 10.1128/jvi.00647-20

Eaton, W. W., Pedersen, M. G., Nielsen, P. R., and Mortensen, P. B. (2010). Autoimmune diseases, bipolar disorder, and non-affective psychosis. *Bipolar Disord.* 12, 638–646. doi: 10.1111/j.1399-5618.2010.00853.x

Estrada, E. (2021). Cascading from SARS-CoV-2 to Parkinson's disease through protein-protein interactions. *Viruses* 13:897. doi: 10.3390/v13050897

Ettman, C. K., Cohen, G. H., Abdalla, S. M., Sampson, L., Trinquart, L., Castrucci, B. C., et al. (2021). Persistent depressive symptoms during COVID-19: a national, population-representative, longitudinal study of U.S. adults. *Lancet Reg. Health Am.* 5:100091. doi: 10.1016/J.LANA.2021.100091

Fitzgerald, K. (2020). Furin protease: from SARS CoV-2 to anthrax, diabetes, and hypertension. *Perm. J.* 24:20.187. doi: 10.7812/TPP/20.187

Fromer, M., Roussos, P., Sieberts, S. K., Johnson, J. S., Kavanagh, D. H., Perumal, T. M., et al. (2016). Gene expression elucidates functional impact of polygenic risk for schizophrenia. *Nat. Neurosci.* 19, 1442–1453. doi: 10.1038/nn.4399

Funkelstein, L., Lu, W. D., Koch, B., Mosier, C., Toneff, T., Taupenot, L., et al. (2012). Human cathepsin V protease participates in production of enkephalin and NPY neuropeptide neurotransmitters. *J. Biol. Chem.* 287, 15232–15241. doi: 10.1074/JBC.M111.310607

Funkelstein, L., Toneff, T., Hwang, S. R., Reinheckel, T., Peters, C., and Hook, V. (2008). Cathepsin L participates in the production of neuropeptide Y in secretory vesicles, demonstrated by protease gene knockout and expression. *J. Neurochem.* 106, 384–391. doi: 10.1111/j.1471-4159.2008.05408.x

Gavin, D. P., Chase, K., Matrisciano, F., Grayson, D. R., Guidotti, A., and Sharma, R. P. (2010). Growth arrest and DNA-damage-inducible, beta (GADD45b)-mediated DNA demethylation in major psychosis. *Neuropsychopharmacology* 37, 531–542. doi: 10.1038/npp.2011.221

GBD 2016 Disease and Injury Incidence and Prevalence Collaborators (2017). Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet* 390, 1211–1259. doi: 10.1016/S0140-6736(17)32154-2

Gironacci, M. M., Cerniello, F. M., Longo Carbajosa, N. A., Goldstein, J., and Cerrato, B. D. (2014). Protective axis of the renin-angiotensin system in the brain. *Clin. Sci.* 127, 295–306. doi: 10.1042/CS20130450

Goldsmith, D. R., Rapaport, M. H., and Miller, B. J. (2016). A meta-analysis of blood cytokine network alterations in psychiatric patients: comparisons between schizophrenia, bipolar disorder and depression. *Mol. Psychiatry* 21, 1696–1709. doi: 10.1038/mp.2016.3

Gołyszny, M., and Obuchowicz, E. (2019). Are neuropeptides relevant for the mechanism of action of SSRIs? *Neuropeptides* 75, 1–17. doi: 10.1016/j.npep.2019.02.002

Gomes, C. P., Fernandes, D. E., Casimiro, F., da Mata, G. F., Passos, M. T., Varela, P., et al. (2020). Cathepsin L in COVID-19: from pharmacological evidences to genetics. *Front. Cell. Infect. Microbiol.* 10:589505. doi: 10.3389/fcimb.2020.589505

Gordon, D. E., Jang, G. M., Bouhaddou, M., Xu, J., Obernier, K., White, K. M., et al. (2020). A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* 583, 459–468. doi: 10.1038/s41586-020-2286-9

Goswami, D. B., Jernigan, C. S., Chandran, A., Iyo, A. H., May, W. L., Austin, M. C., et al. (2013). Gene expression analysis of novel genes in the prefrontal cortex of major depressive disorder subjects. *Prog. Neuropsychopharmacol. Biol. Psychiatry* 43, 126–133. doi: 10.1016/j.pnpbp.2012.12.010

Gozes, I. (2017). Sexual divergence in activity-dependent neuroprotective protein impacting autism, schizophrenia, and Alzheimer's disease. *J. Neurosci. Res.* 95, 652–660. doi: 10.1002/jnr.23808

Guterres, A., de Azeredo Lima, C. H., Miranda, R. L., and Gadelha, M. R. (2020). What is the potential function of microRNAs as biomarkers and therapeutic targets in COVID-19? *Infect. Genet. Evol.* 85:104417. doi: 10.1016/J.MEEGID.2020.104417

Hamming, I., Timens, W., Bulthuis, M., Lely, A., Navis, G., and van Goor, H. (2004). Tissue distribution of ACE2 protein, the functional receptor for SARS coronavirus. A first step in understanding SARS pathogenesis. *J. Pathol.* 203, 631–637. doi: 10.1002/PATH.1570

Harrison, S. M., Tarpey, I., Rothwell, L., Kaiser, P., and Hiscox, J. A. (2007). Lithium chloride inhibits the coronavirus infectious bronchitis virus in cell culture. *Avian Pathol.* 36, 109–114. doi: 10.1080/03079450601156083

Heilig, M., Zachrisson, O., Thorsell, A., Ehnvall, A., Mottagui-Tabar, S., Sjögren, M., et al. (2004). Decreased cerebrospinal fluid neuropeptide Y (NPY) in patients with treatment refractory unipolar major depression: preliminary evidence for association with preproNPY gene polymorphism. *J. Psychiatr. Res.* 38, 113–121. doi: 10.1016/S0022-3956(03)00101-8

Heneka, M. T., Golenbock, D., Latz, E., Morgan, D., and Brown, R. (2020). Immediate and long-term consequences of COVID-19 infections for the development of neurological disease. *Alzheimers Res. Ther.* 12:69. doi: 10.1186/s13195-020-00640-3

Heurich, A., Hofmann-Winkler, H., Gierer, S., Liepold, T., Jahn, O., and Pohlmann, S. (2014). TMPRSS2 and ADAM17 cleave ACE2 differentially and only proteolysis by TMPRSS2 augments entry driven by the severe acute respiratory syndrome coronavirus spike protein. *J. Virol.* 88, 1293–1307. doi: 10.1128/jvi.02202-13

Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., et al. (2020). SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell* 181, 271–280.e8. doi: 10.1016/j.cell.2020.02.052

Hong, S., Lee, E. E., Martin, A. S., Soontornniyomkij, B., Soontornniyomkij, V., Achim, C. L., et al. (2017). Abnormalities in chemokine levels in schizophrenia and their clinical correlates. *Schizophr. Res.* 181, 63–69. doi: 10.1016/j.schres.2016.09.019

Hook, V., Yoon, M., Mosier, C., Ito, G., Podvin, S., Head, B. P., et al. (2020). Cathepsin B in neurodegeneration of Alzheimer's disease, traumatic brain injury, and related brain disorders. *Biochim. Biophys. Acta Proteins Proteom.* 1868:140428. doi: 10.1016/j.bbapap.2020.140428

Hoseth, E. Z., Ueland, T., Dieset, I., Birnbaum, R., Shin, J. H., Kleinman, J. E., et al. (2017). A study of TNF pathway activation in schizophrenia and bipolar disorder in plasma and brain tissue. *Schizophr. Bull.* 43, 881–890. doi: 10.1093/schbul/sbw183

Hou, Y., Liang, W., Zhang, J., Li, Q., Ou, H., Wang, Z., et al. (2018). Schizophrenia-associated rs4702 G allele-specific downregulation of FURIN expression by miR-338-3p reduces BDNF production. *Schizophr. Res.* 199, 176–180. doi: 10.1016/j.schres.2018.02.040

Hou, Y. J., Okuda, K., Edwards, C. E., Martinez, D. R., Asakura, T., Dinnon, K. H., et al. (2020). SARS-CoV-2 reverse genetics reveals a variable infection gradient in the respiratory tract. *Cell* 182, 429–446.e14. doi: 10.1016/j.cell.2020.05.042

Howe, J. R., Li, E. S., Streeter, S. E., Rahme, G. J., Chipumuro, E., Russo, G. B., et al. (2017). MIR-338-3p regulates neuronal maturation and suppresses glioblastoma proliferation. *PLoS One* 12:e0177661. doi: 10.1371/journal.pone.0177661

Hu, Z., Xiao, X., Zhang, Z., and Li, M. (2019). Genetic insights and neurobiological implications from NRXN1 in neuropsychiatric disorders. *Mol. Psychiatry* 24, 1400–1414. doi: 10.1038/s41380-019-0438-9

Huang, G., Osorio, D., Guan, J., Ji, G., and Cai, J. J. (2020). Overdispersed gene expression in schizophrenia. *NPJ Schizophr.* 6:9. doi: 10.1038/s41537-020-0097-5

Huang, I. C., Bosch, B. J., Li, F., Li, W., Kyoung, H. L., Ghiran, S., et al. (2006). SARS coronavirus, but not human coronavirus NL63, utilizes cathepsin L to infect ACE2-expressing cells. *J. Biol. Chem.* 281, 3198–3203. doi: 10.1074/jbc.M508381200

Hussain, M., Jabeen, N., Amanullah, A., Baig, A. A., Aziz, B., Shabbir, S., et al. (2020). Molecular docking between human tmprss2 and sars-cov-2 spike protein: conformation and intermolecular interactions. *AIMS Microbiol.* 6, 350–360. doi: 10.3934/microbiol.2020021

Hussman, J. P. (2020). Cellular and molecular pathways of COVID-19 and potential points of therapeutic intervention. *Front. Pharmacol.* 11:1169. doi: 10.3389/FPHAR.2020.01169

Jacquet, H., Demily, C., Houy, E., Hecketsweiler, B., Bou, J., Raux, G., et al. (2005). Hyperprolinemia is a risk factor for schizoaffective disorder. *Mol. Psychiatry* 10, 479–485. doi: 10.1038/sj.mp.4001597

Jakhmola, S., Indari, O., Chatterjee, S., and Jha, H. C. (2020). SARS-CoV-2, an underestimated pathogen of the nervous system. *SN Compr. Clin. Med.* 2, 2137–2146. doi: 10.1007/s42399-020-00522-7

Javier, R. T., and Rice, A. P. (2011). Emerging theme: cellular PDZ proteins as common targets of pathogenic viruses. *J. Virol.* 85, 11544–11556. doi: 10.1128/JVI.05410-11

Javorsky, A., Humbert, P. O., and Kvansakul, M. (2021). Structural basis of coronavirus E protein interactions with human PALS1 PDZ domain. *Commun. Biol.* 4:724. doi: 10.1038/s42003-021-02250-7

Jia, H. P., Look, D. C., Shi, L., Hickey, M., Pewe, L., Netland, J., et al. (2005). ACE2 receptor expression and severe acute respiratory syndrome coronavirus infection depend on differentiation of human airway epithelia. *J. Virol.* 79, 14614–14621. doi: 10.1128/jvi.79.23.14614-14621.2005

Jia, W., Wang, J., Sun, B., Zhou, J., Shi, Y., and Zhou, Z. (2021). The mechanisms and animal models of SARS-CoV-2 infection. *Front. Cell Dev. Biol.* 9:578825. doi: 10.3389/FCELL.2021.578825

Jian, X. Q., Wang, K. S., Wu, T. J., Hillhouse, J. J., and Mullersman, J. E. (2011). Association of ADAM10 and CAMK2A polymorphisms with conduct disorder: evidence from family-based studies. *J. Abnorm. Child Psychol.* 39, 773–782. doi: 10.1007/s10802-011-9524-4

Kamath, V., Paksarian, D., Cui, L., Moberg, P. J., Turetsky, B. I., and Merikangas, K. R. (2018). Olfactory processing in bipolar disorder, major depression, and anxiety. *Bipolar Disord.* 20, 547–555. doi: 10.1111/bdi.12625

Katrinli, S., Lori, A., Kilaru, V., Carter, S., Powers, A., Gillespie, C. F., et al. (2019). Association of HLA locus alleles with posttraumatic stress disorder. *Brain Behav. Immun.* 81, 655–658. doi: 10.1016/j.bbi.2019.07.016

Khalil, B. A., Elemam, N. M., and Maghazachi, A. A. (2021). Chemokines and chemokine receptors during COVID-19 infection. *Comput. Struct. Biotechnol. J.* 19, 976–988. doi: 10.1016/J.CSBJ.2021.01.034

Khanmohammadi, S., and Rezaei, N. (2021). Role of Toll-like receptors in the pathogenesis of COVID-19. *J. Med. Virol.* 93, 2735–2739. doi: 10.1002/JMV.26826

Kim, K. M., Noh, J. H., Bodogai, M., Martindale, J. L., Yang, X., Indig, F. E., et al. (2017). Identification of senescent cell surface targetable protein DPP4. *Genes Dev.* 31, 1529–1534. doi: 10.1101/GAD.302570.117

Klemann, C., Wagner, L., Stephan, M., and von Hörsten, S. (2016). Cut to the chase: a review of CD26/dipeptidyl peptidase-4's (DPP4) entanglement in the immune system. *Clin. Exp. Immunol.* 185, 1–21. doi: 10.1111/CEI.12781

Koga, M., Ishiguro, H., Yazaki, S., Horiuchi, Y., Arai, M., Niizato, K., et al. (2009). Involvement of SMARCA2/BRM in the SWI/SNF chromatin-remodeling complex in schizophrenia. *Hum. Mol. Genet.* 18, 2483–2494. doi: 10.1093/hmg/ddp166

Kokkosis, A. G., and Tsirka, S. E. (2020). Neuroimmune mechanisms and sex/gender-dependent effects in the pathophysiology of mental disorders. *J. Pharmacol. Exp. Ther.* 375, 175–192. doi: 10.1124/JPET.120.266163

Korth, C. (2009). DISCopathies: brain disorders related to DISC1 dysfunction. *Rev. Neurosci.* 20, 321–330. doi: 10.1515/REVNEURO.2009.20.5-6.321

Krassowski, M., Paczkowska, M., Cullion, K., Huang, T., Dzneladze, I., Ouellette, B. F. F., et al. (2018). ActiveDriverDB: human disease mutations and genome variation in post-translational modification sites of proteins. *Nucleic Acids Res.* 46, D901–D910. doi: 10.1093/nar/gkx973

Kumar, A., Prasoon, P., Kumari, C., Pareek, V., Faiq, M. A., Narayan, R. K., et al. (2021). SARS-CoV-2-specific virulence factors in COVID-19. *J. Med. Virol.* 93, 1343–1350. doi: 10.1002/JMV.26615

Kumar, V. (2019). Toll-like receptors in the pathogenesis of neuroinflammation. *J. Neuroimmunol.* 332, 16–30. doi: 10.1016/j.jneuroim.2019.03.012

Lepack, A. E., Bagot, R. C., Peña, C. J., Loh, Y.-H. E., Farrelly, L. A., Lu, Y., et al. (2016). Aberrant H3.3 dynamics in NAc promote vulnerability to depressive-like behavior. *Proc. Natl. Acad. Sci. U.S.A.* 113, 12562–12567. doi: 10.1073/pnas.1608270113

Li, F. Q., Xiao, H., Tam, J. P., and Liu, D. X. (2005a). Sumoylation of the nucleocapsid protein of severe acute respiratory syndrome coronavirus. *FEBS Lett.* 579, 2387–2396. doi: 10.1016/j.febslet.2005.03.039

Li, Y. H., Li, J., Liu, X. E., Wang, L., Li, T., Zhou, Y. H., et al. (2005b). Detection of the nucleocapsid protein of severe acute respiratory syndrome coronavirus in serum: comparison with results of other viral markers. *J. Virol. Methods* 130, 45–50. doi: 10.1016/j.jviromet.2005.06.001

Li, Y., Zhang, Z., Yang, L., Xin, S., Cao, P., and Lu, J. (2020). The MERS-CoV receptor DPP4 as a candidate binding target of the SARS-CoV-2 spike. *iScience* 23:101160. doi: 10.1016/j.isci.2020.101160

Lim, K. H., and Staudt, L. M. (2013). Toll-like receptor signaling. *Cold Spring Harb. Perspect. Biol.* 5:a011247. doi: 10.1101/cshperspect.a011247

Lin, M., Tseng, H. K., Trejaut, J. A., Lee, H. L., Loo, J. H., Chu, C. C., et al. (2003). Association of HLA class I with severe acute respiratory syndrome coronavirus infection. *BMC Med. Genet.* 4:9. doi: 10.1186/1471-2350-4-9

Liu, H., Sun, Y., Zhang, X., Li, S., Hu, D., Xiao, L., et al. (2020). Integrated analysis of summary statistics to identify pleiotropic genes and pathways for the comorbidity of schizophrenia and cardiometabolic disease. *Front. Psychiatry* 11:256. doi: 10.3389/fpsyt.2020.00256

Liu, J. F., Wu, R., and Li, J. X. (2019a). Toll of mental disorders: TLR-mediated function of the innate immune system. *Neurosci. Bull.* 35, 771–774. doi: 10.1007/s12264-018-00335-8

Liu, Y., Tao, H., Yang, X., Huang, K., Zhang, X., and Li, C. (2019b). Decreased serum oxytocin and increased homocysteine in first-episode schizophrenia patients. *Front. Psychiatry* 10:217. doi: 10.3389/fpsyt.2019.00217

Liu, X., Dai, S. K., Liu, P. P., and Liu, C. M. (2021). Arid1a regulates neural stem/progenitor cell proliferation and differentiation during cortical development. *Cell Prolif.* 54:e13124. doi: 10.1111/CPR.13124

Lugnier, C., Al-Kuraishy, H. M., and Rousseau, E. (2021). PDE4 inhibition as a therapeutic strategy for improvement of pulmonary dysfunctions in Covid-19 and cigarette smoking. *Biochem. Pharmacol.* 185:114431. doi: 10.1016/J.BCP.2021.114431

Luykx, J. J., Bakker, S. C., Visser, W. F., Verhoeven-Duif, N., Buizer-Voskamp, J. E., Den Heijer, J. M., et al. (2015). Genome-wide association study of NMDA receptor coagonists in human cerebrospinal fluid and plasma. *Mol. Psychiatry* 20, 1557–1564. doi: 10.1038/mp.2014.190

Marballi, K., Cruz, D., Thompson, P., and Walss-Bass, C. (2012). Differential neuregulin 1 cleavage in the prefrontal cortex and hippocampus in schizophrenia and bipolar disorder: preliminary findings. *PLoS One* 7:e36431. doi: 10.1371/journal.pone.0036431

Marballi, K. K., McCullumsmith, R. E., Yates, S., Escamilla, M. A., Leach, R. J., Raventos, H., et al. (2014). Global signaling effects of a schizophrenia-associated missense mutation in neuregulin 1: an exploratory study using whole genome and novel kinome approaches. *J. Neural Transm.* 121, 479–490. doi: 10.1007/S00702-013-1142-6

Marco, C., Antonio, D., Antonina, S., Alessandro, S., and Concetta, C. (2015). Genes involved in pruning and inflammation are enriched in a large mega-sample of patients affected by schizophrenia and bipolar disorder and controls. *Psychiatry Res.* 228, 945–949. doi: 10.1016/j.psychres.2015.06.013

Marie-Claire, C., Courtin, C., Curis, E., Bouaziz-Amar, E., Laplanche, J. L., Jacob, A., et al. (2019). Increased plasma levels of high mobility group box 1 protein in patients with bipolar disorder: a pilot study. *J. Neuroimmunol.* 334:576993. doi: 10.1016/J.JNEUROIM.2019.576993

Maskalenko, N., Nath, S., Ramakrishnan, A., Anikeeva, N., Sykulev, Y., and Poenie, M. (2020). The DISC1–Girdin complex – a missing link in signaling to the T cell cytoskeleton. *J. Cell Sci.* 133:jcs242875. doi: 10.1242/jcs.242875

Mayi, B. S., Leibowitz, J. A., Woods, A. T., Ammon, K. A., Liu, A. E., and Raja, A. (2021). The role of neuropilin-1 in COVID-19. *PLoS Pathog.* 17:e1009153. doi: 10.1371/journal.ppat.1009153

Mehandru, S., and Merad, M. (2022). Pathological sequelae of long-haul COVID. *Nat. Immunol.* 23, 194–202. doi: 10.1038/S41590-021-01104-Y

Mestres, I., Einsiedel, M., Möller, J., and Cardoso De Toledo, B. (2020). Smad anchor for receptor activation nuclear localization during development identifies Layers V and VI of the neocortex. *J. Comp. Neurol.* 528, 2161–2173. doi: 10.1002/cne.24881

Milenkovic, V. M., Stanton, E. H., Nothdurfter, C., Rupprecht, R., and Wetzel, C. H. (2019). The role of chemokines in the pathophysiology of major depressive disorder. *Int. J. Mol. Sci.* 20:2283. doi: 10.3390/ijms20092283

Millar, K. J., Mackie, S., Clapcote, S. J., Murdoch, H., Pickard, B. S., Christie, S., et al. (2007). Disrupted in schizophrenia 1 and phosphodiesterase 4B: towards an understanding of psychiatric illness. *J. Physiol.* 584, 401–405. doi: 10.1113/JPHYSIOL.2007.140210

Miller, B. J., Buckley, P., Seabolt, W., Mellor, A., and Kirkpatrick, B. (2011). Meta-analysis of cytokine alterations in schizophrenia: clinical status and antipsychotic effects. *Biol. Psychiatry* 70, 663–671. doi: 10.1016/j.biopsych.2011.04.013

Min, J. A., Kim, J. J., Pae, C. U., Kim, K. H., Lee, C. U., Lee, C., et al. (2012). Association of estrogen receptor genes and schizophrenia: a preliminary study. *Prog. Neuro Psychopharmacol. Biol. Psychiatry* 36, 1–4. doi: 10.1016/j.pnpbp.2011.09.012

Mobini, S., Chizari, M., Mafakher, L., Rismani, E., and Rismani, E. (2021). Structure-based study of immune receptors as eligible binding targets of coronavirus SARS-CoV-2 spike protein. *J. Mol. Graph. Model.* 108:107997. doi: 10.1016/J.JMGM.2021.107997

Moffat, J. J., Jung, E. M., Ka, M., Smith, A. L., Jeon, B. T., Santen, G. W. E., et al. (2019). The role of ARID1B, a BAF chromatin remodeling complex subunit, in neural development and behavior. *Prog. Neuropsychopharmacol. Biol. Psychiatry* 89, 30–38. doi: 10.1016/j.pnpbp.2018.08.021

Mohammadi, A., Rashidi, E., and Amooeian, V. G. (2018). Brain, blood, cerebrospinal fluid, and serum biomarkers in schizophrenia. *Psychiatry Res.* 265, 25–38. doi: 10.1016/j.psychres.2018.04.036

Mondelli, V., and Pariante, C. M. (2021). What can neuroimmunology teach us about the symptoms of long-COVID? *Oxf. Open Immunol.* 2, 1–5. doi: 10.1093/OXFIMM/IQAB004

Moolamalla, S. T. R., Balasubramanian, R., Chauhan, R., Priyakumar, U. D., and Vinod, P. K. (2021). Host metabolic reprogramming in response to SARS-CoV-2 infection: a systems biology approach. *Microb. Pathog.* 158:105114. doi: 10.1016/J.MICPATH.2021.105114

Moon, H. Y., Becke, A., Berron, D., Becker, B., Sah, N., Benoni, G., et al. (2016). Running-induced systemic cathepsin B secretion is associated with memory function. *Cell Metab.* 24, 332–340. doi: 10.1016/j.cmet.2016.05.025

Moradi, F., Enjezab, B., and Ghadiri-Anari, A. (2020). The role of androgens in COVID-19. *Diabetes Metab. Syndr.* 14, 2003–2006. doi: 10.1016/j.dsx.2020.10.014

Morsy, S., and Morsy, A. (2021). Epitope mimicry analysis of SARS-COV-2 surface proteins and human lung proteins. *J. Mol. Graph. Model.* 105:107836. doi: 10.1016/j.jmgm.2021.107836

Mortaz, E., Tabarsi, P., Varahram, M., Folkerts, G., and Adcock, I. M. (2020). The immune response and immunopathology of COVID-19. *Front. Immunol.* 11:2037. doi: 10.3389/FIMMU.2020.02037

Motaghinejad, M., and Gholami, M. (2020). Possible neurological and mental outcomes of COVID-19 infection: a hypothetical role of ACE-2\Mas\BDNF signaling pathway. *Int. J. Prev. Med.* 11:84. doi: 10.4103/ijpvm.IJPVM_114_20

Mullen, P. J., Garcia, G., Purkayastha, A., Matulionis, N., Schmid, E. W., Momcilovic, M., et al. (2021). SARS-CoV-2 infection rewires host cell metabolism and is potentially susceptible to mTORC1 inhibition. *Nat. Commun.* 12:1876. doi: 10.1038/s41467-021-22166-4

Murru, A., Manchia, M., Hajek, T., Nielsen, R. E., Rybakowski, J. K., Sani, G., et al. (2020). Lithium's antiviral effects: a potential drug for CoViD-19 disease? *Int. J. Bipolar Disord.* 8:21. doi: 10.1186/s40345-020-00191-4

Narayan, S., McLean, C., Sawa, A., Lin, S. Y., Rai, N., Hipolito, M. M. S., et al. (2014). Olfactory neurons obtained through nasal biopsy combined with laser-capture microdissection: a potential approach to study treatment response in mental disorders. *J. Vis. Exp.* 94:51853. doi: 10.3791/51853

Newby, T. A., Graff, J. N., Ganzini, L. K., and McDonagh, M. S. (2015). Interventions that may reduce depressive symptoms among prostate cancer patients: a systematic review and meta-analysis. *Psychooncology* 24, 1686–1693. doi: 10.1002/pon.3781

Ng, M. H. L., Lau, K. M., Li, L., Cheng, S. H., Chan, W. Y., Hui, P. K., et al. (2004). Association of human-leukocyte-antigen class I (B*0703) and class II (DRB1*0301) genotypes with susceptibility and resistance to the development of severe acute respiratory syndrome. *J. Infect. Dis.* 190, 515–518. doi: 10.1086/421523

Nowak, J. K., and Walkowiak, J. (2020). Lithium and coronaviral infections. A scoping review. *F1000Res.* 9:93. doi: 10.12688/f1000research.22299.2

Oh, S. J., and Shin, O. S. (2021). SARS-CoV-2 nucleocapsid protein targets RIG-I-like receptor pathways to inhibit the induction of interferon response. *Cells* 10:530. doi: 10.3390/cells10030530

Owens, S. J., Purves-Tyson, T. D., Webster, M. J., and Shannon Weickert, C. (2019). Evidence for enhanced androgen action in the prefrontal cortex of people with bipolar disorder but not schizophrenia or major depressive disorder. *Psychiatry Res.* 280:112503. doi: 10.1016/j.psychres.2019.112503

Ozcan, M. E., and Banoglu, R. (2003). Gonadal hormones in schizophrenia and mood disorders. *Eur. Arch. Psychiatry Clin. Neurosci.* 253, 193–196. doi: 10.1007/s00406-003-0424-7

Pagliaroli, L., and Trizzino, M. (2021). The evolutionary conserved SWI/SNF subunits ARID1A and ARID1B are key modulators of pluripotency and cell-fate determination. *Front. Cell Dev. Biol.* 9:643361. doi: 10.3389/fcell.2021.643361

Pandey, G. N., Rizavi, H. S., Ren, X., Bhaumik, R., and Dwivedi, Y. (2014). Toll-like receptors in the depressed and suicide brain. *J. Psychiatr. Res.* 53, 62–68. doi: 10.1016/j.jpsychires.2014.01.021

Pankiv, S., Alemu, E. A., Brech, A., Bruun, J. A., Lamark, T., Øvervatn, A., et al. (2010). FYCO1 is a Rab7 effector that binds to LC3 and PI3P to mediate microtubule plus end - directed vesicle transport. *J. Cell Biol.* 188, 253–269. doi: 10.1083/jcb.200907015

Pantazopoulos, H., Katsel, P., Haroutunian, V., Chelini, G., Klengel, T., and Berretta, S. (2020). Molecular signature of extracellular matrix pathology in schizophrenia. *Eur. J. Neurosci.* 53, 3960–3987. doi: 10.1111/ejn.15009

Parks, S., Avramopoulos, D., Mulle, J., McGrath, J., Wang, R., Goes, F. S., et al. (2018). HLA typing using genome wide data reveals susceptibility types for infections in a psychiatric disease enriched sample. *Brain Behav. Immun.* 70, 203–213. doi: 10.1016/j.bbi.2018.03.001

Paudel, Y. N., Shaikh, M. F., Chakraborti, A., Kumari, Y., Aledo-Serrano, Á., Aleksovska, K., et al. (2018). HMGB1: a common biomarker and potential target for TBI, neuroinflammation, epilepsy, and cognitive dysfunction. *Front. Neurosci.* 12:628. doi: 10.3389/FNINS.2018.00628

Pišlar, A., Mitrovic, A., Sabotiè, J., Fonovic, U. P., Nanut, M. P., Jakoš, T., et al. (2020). The role of cysteine peptidases in coronavirus cell entry and replication: the therapeutic potential of cathepsin inhibitors. *PLoS Pathog.* 16:e1009013. doi: 10.1371/journal.ppat.1009013

Proal, A. D., and VanElzakker, M. B. (2021). Long COVID or post-acute sequelae of COVID-19 (PASC): an overview of biological factors that may contribute to persistent symptoms. *Front. Microbiol.* 12:698169. doi: 10.3389/FMICB.2021.698169

Qian, M., Shen, X., and Wang, H. (2016). The distinct role of ADAM17 in APP proteolysis and microglial activation related to Alzheimer's disease. *Cell. Mol. Neurobiol.* 36, 471–482. doi: 10.1007/s10571-015-0232-4

Rabaan, A. A., Alahmed, S. H., Bazzi, A. M., and Alhani, H. M. (2017). A review of candidate therapies for middle east respiratory syndrome from a molecular perspective. *J. Med. Microbiol.* 66, 1261–1274. doi: 10.1099/jmm.0.000565

Rana, A. K., Rahmatkar, S. N., Kumar, A., and Singh, D. (2021). Glycogen synthase kinase-3: a putative target to combat severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) pandemic. *Cytokine Growth Factor Rev.* 58, 92–101. doi: 10.1016/J.CYTOGFR.2020.08.002

Randhawa, P. K., Scanlon, K., Rappaport, J., and Gupta, M. K. (2020). Modulation of autophagy by SARS-CoV-2: a potential threat for cardiovascular system. *Front. Physiol.* 11:611275. doi: 10.3389/FPHYS.2020.611275

Raveendran, A. V., and Misra, A. (2021). Post COVID-19 syndrome ("Long COVID") and diabetes: challenges in diagnosis and management. *Diabetes Metab. Syndr.* 15:102235. doi: 10.1016/J.DSX.2021.102235

Rice, S. M., Oliffe, J. L., Kelly, M. T., Cormie, P., Chambers, S., Ogrodniczuk, J. S., et al. (2018). Depression and prostate cancer: examining comorbidity and male-specific symptoms. *Am. J. Mens Health* 12, 1864–1872. doi: 10.1177/1557988318784395

Rohmann, N., Schlicht, K., Geisler, C., Hollstein, T., Knappe, C., Krause, L., et al. (2021). Circulating sDPP-4 is increased in obesity and insulin resistance but is not related to systemic metabolic inflammation. *J. Clin. Endocrinol. Metab.* 106, E592–E601. doi: 10.1210/CLINEM/DGAA758

Roomruangwong, C., Noto, C., Kanchanatawan, B., Anderson, G., Kubera, M., Carvalho, A. F., et al. (2020). The role of aberrations in the immune-inflammatory response system (IRS) and the compensatory immune-regulatory reflex system (CIRS) in different phenotypes of schizophrenia: the IRS-CIRS theory of schizophrenia. *Mol. Neurobiol.* 57, 778–797. doi: 10.1007/S12035-019-01737-Z

Sakharnova, T. A., and Vedunova, M. V. (2012). Brain-derived neurotrophic factor (BDNF) and its role in the functioning of the central nervous system. *Neurochem. J.* 6, 269–277. doi: 10.1134/S1819712412030129

Samuel, R. M., Majd, H., Richter, M. N., Ghazizadeh, Z., Zekavat, S. M., Navickas, A., et al. (2020). Androgen signaling regulates SARS-CoV-2 receptor levels and is associated with severe COVID-19 symptoms in men. *Cell Stem Cell* 27, 876–889.e12. doi: 10.1016/j.stem.2020.11.009

Sanjabi, S., Oh, S. A., and Li, M. O. (2017). Regulation of the immune response by TGF-β: from conception to autoimmunity and infection. *Cold Spring Harb. Perspect. Biol.* 9:a022236. doi: 10.1101/CSHPERSPECT.A022236

Sankar, J. S., and Hampson, E. (2012). Testosterone levels and androgen receptor gene polymorphism predict specific symptoms of depression in young men. *Gend. Med.* 9, 232–243. doi: 10.1016/j.genm.2012.05.001

Saridaki, T., Nippold, M., Dinter, E., Roos, A., Diederichs, L., Fensky, L., et al. (2018). FYCO1 mediates clearance of α-synuclein aggregates through a Rab7-dependent mechanism. *J. Neurochem.* 146, 474–492. doi: 10.1111/JNC.14461

Saulle, I., Vicentini, C., Clerici, M., and Biasin, M. (2021). Antigen presentation in SARS-CoV-2 infection: the role of class I HLA and ERAP polymorphisms. *Hum. Immunol.* 82, 551–560. doi: 10.1016/J.HUMIMM.2021.05.003

Scamuffa, N., Calvo, F., Chrétien, M., Seidah, N. G., and Khatib, A. (2006). Proprotein convertases: lessons from knockouts. *FASEB J.* 20, 1954–1963. doi: 10.1096/fj.05-5491rev

Schizophrenia Working Group of the Psychiatric Genomics Consortium (2014). Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 511, 421–427. doi: 10.1038/nature13595

Schmeltzer, S. N., Herman, J. P., and Sah, R. (2016). Neuropeptide Y (NPY) and posttraumatic stress disorder (PTSD): a translational update. *Exp. Neurol.* 284(Pt B), 196–210. doi: 10.1016/j.expneurol.2016.06.020

Schou, T. M., Joca, S., Wegener, G., and Bay-Richter, C. (2021). Psychiatric and neuropsychiatric sequelae of COVID-19 – A systematic review. *Brain Behav. Immun.* 97, 328–348. doi: 10.1016/J.BBI.2021.07.018

Schrode, N., Ho, S. M., Yamamuro, K., Dobbyn, A., Huckins, L., Matos, M. R., et al. (2019). Synergistic effects of common schizophrenia risk variants. *Nat. Genet.* 51, 1475–1485. doi: 10.1038/s41588-019-0497-5

Schumacher, M., Ghoumari, A., Mattern, C., Bougnères, P., and Traiffort, E. (2021). Testosterone and myelin regeneration in the central nervous system. *Androg. Clin. Res. Ther.* 2, 231–251. doi: 10.1089/ANDRO.2021.0023

Schweitzer, K. S., Crue, T., Nall, J. M., Foster, D., Sajuthi, S., Correll, K. A., et al. (2021). Influenza virus infection increases ACE2 expression and shedding in human small airway epithelial cells. *Eur. Respir. J.* 58:2003988. doi: 10.1183/13993003.03988-2020

Seidah, N. G. (2011). What lies ahead for the proprotein convertases? *Ann. N. Y. Acad. Sci.* 1220, 149–161. doi: 10.1111/j.1749-6632.2010.05883.x

Seli, E., and Arici, A. (2002). Sex steroids and the immune system. *Immunol. Allergy Clin. North Am.* 22, 407–433. doi: 10.1016/S0889-8561(02)00017-6

Severe Covid-19 GWAS Group, Ellinghaus, D., Degenhardt, F., Bujanda, L., Buti, M., Albillos, A., et al. (2020). Genomewide association study of severe Covid-19 with respiratory failure. *N. Engl. J. Med.* 383, 1522–1534. doi: 10.1056/nejmoa2020283

Sfera, A., Osorio, C., Jafri, N., Diaz, E. L., and Campo Maldonado, J. E. (2020). Intoxication with endogenous angiotensin II: a COVID-19 hypothesis. *Front. Immunol.* 11:1472. doi: 10.3389/fimmu.2020.01472

Shaman, J., and Galanti, M. (2020). Will SARS-CoV-2 become endemic? *Science* 370, 527–529. doi: 10.1126/SCIENCE.ABE5960

Shao, S., Xu, Q. Q., Yu, X., Pan, R., and Chen, Y. (2020). Dipeptidyl peptidase 4 inhibitors and their potential immune modulatory functions. *Pharmacol. Ther.* 209:107503. doi: 10.1016/J.PHARMTHERA.2020.107503

Sheraton, M., Deo, N., Kashyap, R., and Surani, S. (2020). A review of neurological complications of COVID-19. *Cureus* 12:e8192. doi: 10.7759/cureus.8192

Shigetomi, H., Oonogi, A., Tsunemi, T., Tanase, Y., Yamada, Y., Kajihara, H., et al. (2011). The role of components of the chromatin modification machinery in carcinogenesis of clear cell carcinoma of the ovary (Review). *Oncol. Lett.* 2, 591–597. doi: 10.3892/OL.2011.316

Simmons, G., Bertram, S., Glowacka, I., Steffen, I., Chaipan, C., Agudelo, J., et al. (2011). Different host cell proteases activate the SARS-coronavirus spike-protein for cell-cell and virus-cell fusion. *Virology* 413, 265–274. doi: 10.1016/j.virol.2011.02.020

Simmons, G., Gosalia, D. N., Rennekamp, A. J., Reeves, J. D., Diamond, S. L., and Bates, P. (2005). Inhibitors of cathepsin L prevent severe acute respiratory syndrome coronavirus entry. *Proc. Natl. Acad. Sci. U.S.A.* 102, 11876–11881. doi: 10.1073/pnas.0505577102

Solerte, S. B., Di Sabatino, A., Galli, M., and Fiorina, P. (2020). Dipeptidyl peptidase-4 (DPP4) inhibition in COVID-19. *Acta Diabetol.* 57, 779–783. doi: 10.1007/s00592-020-01539-z

Son, E. Y., and Crabtree, G. R. (2014). The role of BAF (mSWI/SNF) complexes in mammalian neural development. *Am. J. Med. Genet. C Semin. Med. Genet.* 166, 333–349. doi: 10.1002/ajmg.c.31416

Song, J. W., Zhang, C., Fan, X., Meng, F. P., Xu, Z., Xia, P., et al. (2020). Immunological and inflammatory profiles in mild and severe cases of COVID-19. *Nat. Commun.* 11:3410. doi: 10.1038/S41467-020-17240-2

Soria, V., Uribe, J., Salvat-Pujol, N., Palao, D., Menchón, J. M., and Labad, J. (2018). Psychoneuroimmunology of mental disorders. *Rev. Psiquiatr. Salud Ment.* 11, 115–124. doi: 10.1016/J.RPSM.2017.07.006

Sowa, J. E., and Tokarski, K. (2021). Cellular, synaptic, and network effects of chemokines in the central nervous system and their implications to behavior. *Pharmacol. Rep.* 73, 1595–1625. doi: 10.1007/S43440-021-00323-2

Steenblock, C., Todorov, V., Kanczkowski, W., Eisenhofer, G., Schedl, A., Wong, M. L., et al. (2020). Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and the neuroendocrine stress axis. *Mol. Psychiatry* 25, 1611–1617. doi: 10.1038/s41380-020-0758-9

Stefano, G. B. (2021). Historical insight into infections and disorders associated with neurological and psychiatric sequelae similar to long COVID. *Med. Sci. Monit.* 27:e931447. doi: 10.12659/MSM.931447

Stefansson, H., Ophoff, R. A., Steinberg, S., Andreassen, O. A., Cichon, S., Rujescu, D., et al. (2009). Common variants conferring risk of schizophrenia. *Nature* 460, 744–747. doi: 10.1038/nature08186

Suchanek-Raif, R., Raif, P., Kowalczyk, M., Paul-Samojedny, M., Zieliñska, A., Kucia, K., et al. (2020). An analysis of five TrkB gene polymorphisms in schizophrenia and the interaction of its haplotype with rs6265 BDNF gene polymorphism. *Dis. Markers* 2020:4789806. doi: 10.1155/2020/4789806

Suryawanshi, R. K., Koganti, R., Agelidis, A., Patil, C. D., and Shukla, D. (2021). Dysregulation of cell signaling by SARS-CoV-2. *Trends Microbiol.* 29, 224–237. doi: 10.1016/J.TIM.2020.12.007

Taalman, H., Wallace, C., and Milev, R. (2017). Olfactory functioning and depression: a systematic review. *Front. Psychiatry* 8:190. doi: 10.3389/fpsyt.2017.00190

Taquet, M., Geddes, J. R., Husain, M., Luciano, S., and Harrison, P. J. (2021a). 6-month neurological and psychiatric outcomes in 236 379 survivors of COVID-19: a retrospective cohort study using electronic health records. *Lancet Psychiatry* 8, 416–427. doi: 10.1016/S2215-0366(21)00084-5

Taquet, M., Luciano, S., Geddes, J. R., and Harrison, P. J. (2021b). Bidirectional associations between COVID-19 and psychiatric disorder: retrospective cohort

studies of 62 354 COVID-19 cases in the USA. *Lancet Psychiatry* 8, 130–140. doi: 10.1016/S2215-0366(20)30462-4

Taylor, A., Harker, J. A., Chanthong, K., Stevenson, P. G., Zuniga, E. I., and Rudd, C. E. (2016). Glycogen synthase kinase 3 inactivation drives T-bet-mediated downregulation of co-receptor PD-1 to enhance CD8+ cytolytic T cell responses. *Immunity* 44, 274–286. doi: 10.1016/J.IMMUNI.2016.01.018

Teng, S., Thomson, P. A., Mccarthy, S., Kramer, M., Muller, S., Lihm, J., et al. (2017). Rare disruptive variants in the DISC1 interactome and regulome: association with cognitive ability and schizophrenia. *Mol. Psychiatry* 23, 1270–1277. doi: 10.1038/mp.2017.115

The Lancet Global Health (2020). Mental health matters. *Lancet Glob. Health* 8:e1352. doi: 10.1016/S2214-109X(20)30432-0

Thye, A. Y.-K., Law, J. W.-F., Tan, L. T.-H., Pusparajah, P., Ser, H.-L., Thurairajasingam, S., et al. (2022). Psychological symptoms in COVID-19 patients: insights into pathophysiology and risk factors of long COVID-19. *Biology* 11:61. doi: 10.3390/BIOLOGY11010061

Tian, L., Ma, L., Kaarela, T., and Li, Z. (2012). Neuroimmune crosstalk in the central nervous system and its significance for neurological diseases. *J. Neuroinflammation* 9:155. doi: 10.1186/1742-2094-9-155

Timberlake, M., Roy, B., and Dwivedi, Y. (2019). A novel animal model for studying depression featuring the induction of the unfolded protein response in hippocampus. *Mol. Neurobiol.* 56, 8524–8536. doi: 10.1007/s12035-019-01687-6

Torres-Berrío, A., Issler, O., Parise, E. M., and Nestler, E. J. (2019). Unraveling the epigenetic landscape of depression: focus on early life stress. *Dialogues Clin. Neurosci.* 21, 341–357. doi: 10.31887/DCNS.2019.21.4/enestler

Toscano, G., Palmerini, F., Ravaglia, S., Ruiz, L., Invernizzi, P., Cuzzoni, M. G., et al. (2020). Guillain–Barré syndrome associated with SARS-CoV-2. *N. Engl. J. Med.* 382, 2574–2576. doi: 10.1056/nejmc2009191

Valencia, I., Peiró, C., Lorenzo, Ó., Sánchez-Ferrer, C. F., Eckel, J., and Romacho, T. (2020). DPP4 and ACE2 in diabetes and COVID-19: therapeutic targets for cardiovascular complications? *Front. Pharmacol.* 11:1161. doi: 10.3389/FPHAR.2020.01161

Veldhoen, M., and Simas, J. P. (2021). Endemic SARS-CoV-2 will maintain post-pandemic immunity. *Nat. Rev. Immunol.* 21, 131–132. doi: 10.1038/s41577-020-00493-9

Venkatakrishnan, A. J., Puranik, A., Anand, A., Zemmour, D., Yao, X., Wu, X., et al. (2020). Knowledge synthesis of 100 million biomedical documents augments the deep expression profiling of coronavirus receptors. *eLife* 9:e58040. doi: 10.7554/ELIFE.58040

Vidak, E., Javoršek, U., Vizovišek, M., and Turk, B. (2019). Cysteine cathepsins and their extracellular roles: shaping the microenvironment. *Cells* 8:264. doi: 10.3390/cells8030264

Vidal-Domènech, F., Riquelme, G., Pinacho, R., Rodriguez-Mias, R., Vera, A., Monje, A., et al. (2020). Calcium-binding proteins are altered in the cerebellum in schizophrenia. *PLoS One* 15:e0230400. doi: 10.1371/journal.pone.0230400

Villa, A., Vegeto, E., Poletti, A., and Maggi, A. (2016). Estrogens, neuroinflammation, and neurodegeneration. *Endocr. Rev.* 37, 372–402. doi: 10.1210/ER.2016-1007

Vincent, B., and Govitrapong, P. (2011). Activation of the α-secretase processing of AβPP as a therapeutic approach in Alzheimer's disease. *J. Alzheimers Dis.* 24, 75–94. doi: 10.3233/JAD-2011-110218

Vogel-Ciernia, A., and Wood, M. A. (2014). Neuron-specific chromatin remodeling: a missing link in epigenetic mechanisms underlying synaptic plasticity, memory, and intellectual disability disorders. *Neuropharmacology* 80, 18–27. doi: 10.1016/j.neuropharm.2013.10.002

Wambier, C. G., Goren, A., Vaño-Galván, S., Ramos, P. M., Ossimetha, A., Nau, G., et al. (2020). Androgen sensitivity gateway to COVID-19 disease severity. *Drug Dev. Res.* 81, 771–776. doi: 10.1002/ddr.21688

Wang, B., Huang, X., Pan, X., Zhang, T., Hou, C., Su, W. J., et al. (2020). Minocycline prevents the depressive-like behavior through inhibiting the release of HMGB1 from microglia and neurons. *Brain Behav. Immun.* 88, 132–143. doi: 10.1016/J.BBI.2020.06.019

Wang, S. C., and Wang, Y. F. (2021). Cardiovascular protective properties of oxytocin against COVID-19. *Life Sci.* 270:119130. doi: 10.1016/j.lfs.2021.119130

Wei, J., Alfajaro, M. M., DeWeirdt, P. C., Hanna, R. E., Lu-Culligan, W. J., Cai, W. L., et al. (2021). Genome-wide CRISPR screens reveal host factors critical for SARS-CoV-2 infection. *Cell* 184, 76–91.e13. doi: 10.1016/j.cell.2020.10.028

Wei, S. Y., Tseng, H. H., Chang, H. H., Lu, T. H., Chang, W. H., Chiu, N. T., et al. (2020). Dysregulation of oxytocin and dopamine in the corticostriatal circuitry in bipolar II disorder. *Transl. Psychiatry* 10:281. doi: 10.1038/s41398-020-00972-6

Wiese, O. J., Allwood, B. W., and Zemlin, A. E. (2020). COVID-19 and the renin-angiotensin system (RAS): a spark that sets the forest alight? *Med. Hypotheses* 144:110231. doi: 10.1016/j.mehy.2020.110231

Wu, S., Ge, Y., Li, X., Yang, Y., Zhou, H., Lin, K., et al. (2020a). BRM-SWI/SNF chromatin remodeling complex enables functional telomeres by promoting co-expression of TRF2 and TRF1. *PLoS Genet.* 16:e1008799. doi: 10.1371/journal.pgen.1008799

Wu, Y., Xu, X., Chen, Z., Duan, J., Hashimoto, K., Yang, L., et al. (2020b). Nervous system involvement after infection with COVID-19 and other coronaviruses. *Brain Behav. Immun.* 87, 18–22. doi: 10.1016/j.bbi.2020.03.031

Yan, X., Hao, Q., Mu, Y., Timani, K. A., Ye, L., Zhu, Y., et al. (2006). Nucleocapsid protein of SARS-CoV activates the expression of cyclooxygenase-2 by binding directly to regulatory elements for nuclear factor-kappa B and CCAAT/enhancer binding protein. *Int. J. Biochem. Cell Biol.* 38, 1417–1428. doi: 10.1016/j.biocel.2006.02.003

Yang, Y., He, M., Tian, X., Guo, Y., Liu, F., Li, Y., et al. (2018). Transgenic overexpression of furin increases epileptic susceptibility. *Cell Death Dis.* 9:1058. doi: 10.1038/S41419-018-1076-X

Yang, Z., Zhou, D., Li, H., Cai, X., Liu, W., Wang, L., et al. (2020). The genome-wide risk alleles for psychiatric disorders at 3p21.1 show convergent effects on mRNA expression, cognitive function, and mushroom dendritic spine. *Mol. Psychiatry* 25, 48–66. doi: 10.1038/s41380-019-0592-0

Youssef, M. M., Underwood, M. D., Huang, Y. Y., Hsiung, S. C., Liu, Y., Simpson, N. R., et al. (2018). Association of BDNF Val66MET polymorphism and brain BDNF levels with major depression and suicide. *Int. J. Neuropsychopharmacol.* 21, 528–538. doi: 10.1093/ijnp/pyy008

Yu, W., Niwa, T., Miura, Y., Horio, F., Teradaira, S., Ribar, T. J., et al. (2002). Calmodulin overexpression causes Ca2+ -dependent apoptosis of pancreatic β cells, which can be prevented by inhibition of nitric oxide synthase. *Lab. Invest.* 82, 1229–1239. doi: 10.1097/01.lab.0000027921.01548.c5

Yuan, X. Z., Sun, S., Tan, C. C., Yu, J. T., and Tan, L. (2017). The role of ADAM10 in Alzheimer's disease. *J. Alzheimers Dis.* 58, 303–322. doi: 10.3233/JAD-170061

Zhang, H., Ding, L., Shen, T., and Peng, D. (2019). HMGB1 involved in stress-induced depression and its neuroinflammatory priming role: a systematic review. *Gen. Psychiatry* 32:100084. doi: 10.1136/GPSYCH-2019-100084

Zhang, Y., Geng, X., Tan, Y., Li, Q., Xu, C., Xu, J., et al. (2020). New understanding of the damage of SARS-CoV-2 infection outside the respiratory system. *Biomed. Pharmacother.* 127:110195. doi: 10.1016/j.biopha.2020.110195

Zhao, F. R., Xie, Y. L., Liu, Z. Z., Shao, J. J., Li, S. F., Zhang, Y. G., et al. (2017). Lithium chloride inhibits early stages of foot-and-mouth disease virus (FMDV) replication in vitro. *J. Med. Virol.* 89, 2041–2046. doi: 10.1002/jmv.24821

Zhou, Y., Agudelo, J., Lu, K., Goetz, D. H., Hansell, E., Chen, Y. T., et al. (2011). Inhibitors of SARS-CoV entry – Identification using an internally-controlled dual envelope pseudovirion assay. *Antiviral Res.* 92, 187–194. doi: 10.1016/j.antiviral.2011.07.016

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.

frontiers | Frontiers in Microbiology

ORIGINAL RESEARCH
published: 08 April 2022
doi: 10.3389/fmicb.2022.843417

Rapid Discrimination of Clinically Important Pathogens Through Machine Learning Analysis of Surface Enhanced Raman Spectra

Jia-Wei Tang[1†], Jia-Qi Li[1†], Xiao-Cong Yin[2†], Wen-Wen Xu[1], Ya-Cheng Pan[3], Qing-Hua Liu[4], Bing Gu[2,5], Xiao Zhang[1]* and Liang Wang[1,6]**

[1] Department of Bioinformatics, School of Medical Informatics and Engineering, Xuzhou Medical University, Xuzhou, China, [2] Department of Laboratory Medicine, School of Medical Technology, Xuzhou Medical University, Xuzhou, China, [3] Department of Basic Medicine and Biological Science, Soochow University, Suzhou, China, [4] State Key Laboratory of Quality Research in Chinese Medicines, Macau University of Science and Technology, Macao, Macau SAR, China, [5] Laboratory Medicine, Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences, Guangzhou, China, [6] Jiangsu Key Laboratory of New Drug Research and Clinical Pharmacy, School of Pharmacy, Xuzhou Medical University, Xuzhou, China

OPEN ACCESS

Edited by:
George Tsiamis,
University of Patras, Greece

Reviewed by:
Qian Du,
GNS Healthcare, United States
Spiros Papakostas,
Aristotle University of Thessaloniki, Greece

***Correspondence:**
Bing Gu
gb20031129@163.com.
Xiao Zhang
changshui@homtail.com
Liang Wang
healthscience@foxmail.com

[†] These authors have contributed equally to this work

Specialty section:
This article was submitted to Systems Microbiology, a section of the journal Frontiers in Microbiology

Received: 25 December 2021
Accepted: 14 March 2022
Published: 08 April 2022

Citation:
Tang J-W, Li J-Q, Yin X-C, Xu W-W, Pan Y-C, Liu Q-H, Gu B, Zhang X and Wang L (2022) Rapid Discrimination of Clinically Important Pathogens Through Machine Learning Analysis of Surface Enhanced Raman Spectra. Front. Microbiol. 13:843417. doi: 10.3389/fmicb.2022.843417

With its low-cost, label-free and non-destructive features, Raman spectroscopy is becoming an attractive technique with high potential to discriminate the causative agent of bacterial infections and bacterial infections *per se*. However, it is challenging to achieve consistency and accuracy of Raman spectra from numerous bacterial species and phenotypes, which significantly hinders the practical application of the technique. In this study, we analyzed surfaced enhanced Raman spectra (SERS) through machine learning algorithms in order to discriminate bacterial pathogens quickly and accurately. Two unsupervised machine learning methods, K-means Clustering (K-Means) and Agglomerative Nesting (AGNES) were performed for clustering analysis. In addition, eight supervised machine learning methods were compared in terms of bacterial predictions *via* Raman spectra, which showed that convolutional neural network (CNN) achieved the best prediction accuracy (99.86%) with the highest area (0.9996) under receiver operating characteristic curve (ROC). In sum, machine learning methods can be potentially applied to classify and predict bacterial pathogens *via* Raman spectra at general level.

Keywords: surface enhanced Raman spectra, bacterial pathogen, machine learning, convolutional neural network, long short-term memory

HIGHLIGHTS

- Surface-enhanced Raman spectroscopy (SERS) has the potential to be used for detecting bacterial pathogens in clinical settings.
- Pretreatment of SERS spectra facilitates the increment of signal-to-noise ratio and increases the accuracy of bacterial detection rate.
- All machine learning algorithms showed their capacities in pathogen clustering and species discrimination via SERS spectra.
- Convolutional neural network showed highest accuracy and robustness in discriminating bacterial pathogens in terms of SERS spectral analysis.

INTRODUCTION

Infectious diseases frequently cause major public health threats and risks due to long-standing, emerging, and re-emerging bacterial pathogens (Bloom and Cadarette, 2019), while rapid and accurate identification of the causing bacterial agents could greatly improve therapeutical effectiveness and reduce host mortality (Caliendo et al., 2013). Although conventional methods are reliable and accurate in clinical diagnosis of bacterial infections, they mainly rely on culture-based testing and biochemical analysis that yield results in days or up to weeks after sampling, not even mentioning the fastidious and viable but non-culturable (VBNC) bacterial species under laboratory conditions (Järvinen et al., 2009). Recently, matrix-assisted laser desorption-ionization time of flight mass spectrometry (MALDI-TOF MS) is emerging as an important tool in bacterial identification in clinical laboratories due to its rapidity, reliability, and cost-effectiveness (Singhal et al., 2015). However, MALDI-TOF MS also suffers disadvantages such as lack of a complete spectra database for known bacteria and inaccuracy of bacterial discrimination at genus, species and sub-species levels like *Shigella* and *Escherichia coli*, etc. (Sloan et al., 2017). Thus, advanced and diverse detection methods should be developed in order to facilitate the rapid and accurate diagnosis of bacterial infections in clinical settings.

Raman spectroscopy is a non-destructive chemical analysis based on interactions between the light and chemical bonds within a material, which could generate detailed fingerprinting spectra for a particular biological sample (Orlando et al., 2021). Cumulative studies show that Raman spectroscopy (RS) has the potential to rapidly analyze clinical samples and efficiently identify bacterial species in simple procedures (Ho et al., 2019). However, because of the intrinsically weak signal of Raman effect, surface enhanced Raman spectroscopy (SERS) has been developed for the analysis of biological samples, which not only greatly improves the detection capacity of bacterial pathogens but also opens new directions for the detection of analytes at very low concentrations (Kuhar et al., 2018). Due to the complexity of the raw Raman spectral data, traditional statistical methods are not sufficient for data analysis and pattern recognition (Wang et al., 2021), which hinders the application of Raman spectroscopy in the field of infectious diseases. With the assistance of advanced computational methods like machine learning methods, it would be possible for the promising technique to overcome current challenges and gradually realize its real-world applications in clinical laboratories for the detection of bacterial pathogens.

In this study, we analyzed a group of 15 bacterial pathogens belonging to different genera through SERS spectra. All the Raman spectra were processed to calculate average Raman spectrum for each bacterial genus, together with the corresponding characteristic peaks. In order to discriminate these bacteria efficiently, two representative unsupervised machine learning methods, K-means and agglomerative nesting (AGNES) were then applied to the spectral data for bacterial clustering. Moreover, three classic supervised machine learning methods, Random Forest (RF), Decision Tree (DT), and Support Vector Machine (SVM), together with five deep learning algorithms, Multilayer Perceptron (MLP), Convolutional Neural Network (CNN), Recurrent Neural Networks (RNN), Gate Recurrent Unit (GRU), and Long Short-term Memory (LSTM) were implemented to all the bacterial SERS spectra, the results of which were compared in terms of their prediction accuracies, sensitivities and specificities. In order to evaluate the robustness of supervised machine learning methods, artificial noise signals were added to the SERS spectra and the prediction accuracies were compared among algorithms, which revealed that CNN had the best capacity in terms of noise interference during species predictions. In sum, this study confirmed that Raman spectroscopy has the potential in clustering, discriminating and predicting bacterial pathogens from different bacterial genera in clinical settings with the assistance of machine learning algorithms.

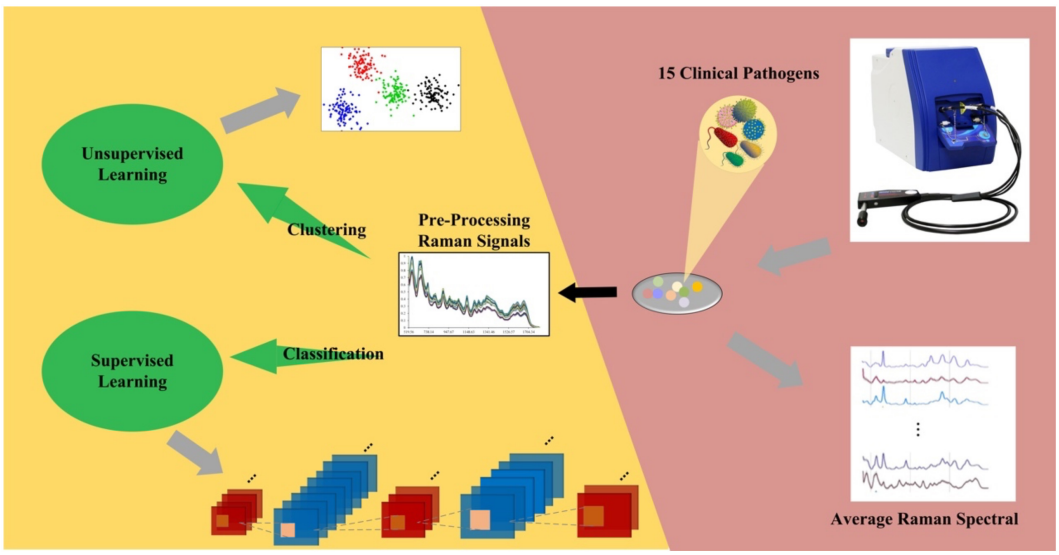
MATERIALS AND METHODS

Bacterial Strains and Chemical Materials

A total of 15 clinical bacterial pathogens studied in this experiment were directly isolated from clinical samples and cultured on Columbia blood agar (35°C, 18–24 h) at the Department of Laboratory Medicine, Affiliated Hospital of Xuzhou Medical University, which included six isolates of *Achromobacter xylosoxidans* (n = 610), nine isolates of *Burkholderia cepacian* (n = 600), 4 isolates of *Chryseobacterium indologenes* (n = 690), one isolate of *Corynebacterium glucuronolyticum* (n = 600), seven isolates of *Elizabethkingia meningoseptica* (n = 601), 20 isolates of *E. coli* (n = 38), five strains of *Micrococcus luteus* (n = 601), five isolates of *Moraxella catarrhalis* (n = 600), seven isolates of *Morganella morganii* (n = 130), five isolates of *Myroides odoratimimus* (n = 610), two isolates of *Neisseria flavescens* (n = 601), three isolates of *Providencia rettgeri* (n = 601), two isolates of *Pseudomonas putica* (n = 100), 18 isolates of *Serratia marcescens* (n = 569), and nine isolates of *Vibrio parahaemolyticus* (n = 600). The letter n represents the total number of SERS spectra for all the isolates of the same bacterial species. All bacterial pathogens were cultured, isolated, and then identified using MALDI-TOF MS and stored in freezer (Thermo Fisher Scientific, Waltham, MA, United States) at −80°C. Morphological, physiological, and clinical features of these bacterial pathogens were summarized in **Supplementary Table 1**. During analysis, all the strains were thawed, inoculated onto and cultivated on standard Columbia Blood Agar (CBA) for 24 h at 37°C. Single colonies were then randomly selected and mixed with negatively-charged silver nitrate nanoparticle ($AgNO_3$) substrate for SERS study. For the preparation of the $AgNO_3$ substrate, please refer to the procedures described by Tang et al. (2021).

Surface-Enhanced Raman Spectroscopy

A single bacterial colony was mixed with 15 μL phosphate buffer saline (PBS) *via* vortexing, which was then mixed with 15 μL negatively-charged silver nanoparticle substrate solution. The mixed solution was placed onto silicon wafer and left on clean bench to air-dry completely. The commercial



GRAPHICAL ABSTRACT |

Raman spectrometer i-Raman® Plus BWS 465-785H (B&W Tek, Plainsboro, United States) was used to measure the dried spot on the silicon wafer. Measurement parameters were set as following: excitation light source wavelength at 785 nm; laser power: 20 mW. Detection parameters were set as following: (1) spectrum acquisition: 5 seconds; (2) detector type: high quantum efficiency CCD array; (3) Raman shift range: 65-2800 cm^{-1}; (4) resolution: less than $<3.5\ cm^{-1}$ at 912 nm. Finally, signals ranged from 519.56–1800.81 cm^{-1} were captured, which consisted of 657 value points in total for each spectrum.

Average Raman Spectra and Characteristic Peaks

Removal of Outlier Raman Spectra

In this study, all the SERS spectral data for each clinical isolate were obtained from Raman spectrometer *via* the software BWSpec 4.02 (B&W Tek, United States). We identified outliers in the raw SERS spectra *via* variance contribution rate. The procedure was implemented through the PCount() function in the mvoutlier package of the R programming language (Filzmoser et al., 2008). **Supplementary Table 2** showed the number of SERS spectra before and after outlier analysis.

Average Raman Spectra

By calculating the repeated Raman intensity of all samples under the same Raman shift for each bacterium, the average value of the intensity at the Raman shift was obtained, and then the average intensities at all the Raman shifts were calculated to generate the average Raman spectrum of the particular bacterial species. By following this procedure, 15 average Raman spectral curves were obtained, and the standard deviation (SD) of each average spectral line was calculated. Both average SERS spectra and 20% SD band were visualized by the origin software (OriginLab, United States). The width of the error band shows the reproducibility of Raman spectra for each bacterial species.

Identification of Characteristic Peaks

In order to discriminate the differences among different bacterial species through the SERS spectra, average Raman spectral curves was preprocessed through software LabSpec 6 (HORIBA Scientific, Japan), which included spectral smoothing, denoising, baseline correction, and normalization. After that, characteristic peaks in each average Raman spectrum were then identified. The specific parameter settings were *Degree* = 4, *Size* = 5 and *Height* = 50. The *smooth* function was used to denoise the spectrum. For the baseline correction, the parameters were set as *type* = Polynom, *Degree* = 6, *Attach* = No, and then the *Auto* function was applied to perform baseline fitting. Finally, the LabSpec software was used to fit the characteristic peaks. *GaussLoren* function was used with parameters set to *Level* = 13% and *Size* = 19 while other parameters were set by default. The normalization operation was performed for better comparing the curves of different bacterial species. The function *search* was used for the identification of characteristic peaks.

Surface-Enhanced Raman Spectroscopy Spectral Preprocessing for Machine Learning Analysis

Before machine learning analyses, raw SERS spectra excluding outliers require a series of pre-processing steps, which includes spectral smoothing and denoising, baseline correction and normalization. Through preprocessing of SERS spectra, data quality was significantly improved and data dimensionalities were reduced, which greatly facilitated further statistical analysis of Raman spectra *via* supervised and unsupervised machine learning algorithms.

Smoothing and Denoising of Raman Spectra

Curve smoothing and denoising were performed to remove noise signals in SERS spectra caused by dark current and fluctuation of the external environment in order to improve the signal-to-noise ratio (SNR). There were a variety of filtering algorithms that could effectively reduce noise interference in the Raman spectra, which included moving window averaging and Savitzky-Golay filters and so on. In this study, Savitzky-Golay filtering method in the Unscrambler® X software was used, which was a weighted average method highlighting the effect of the center point and calculating the filter value with a window. In particular, it is noteworthy that the number of points on both the left and right of the center point was set to three while the derivative order was set to two during data analysis procedures.

Baseline Correction and Spectral Normalization

Due to the noise interference in Raman spectroscopy, it was necessary to perform baseline correction on Raman spectra. Commonly used baseline correction methods include polynomial fitting based on least squares and asymmetric least squares (Baek et al., 2015). In this study, we used the *Baseline function* under *Transform* in the software Unscrambler® X to perform baseline correction of previously smoothed SERS spectra. For parameter setting, *Rows* and *Cols* were set to *All*, and the *Method* was set to *Baseline offset*. In addition, each SERS spectrum in this study contains 657 Raman shifts. In order to remove the influences of signal intensities in the spectral data among different samples of the same species, max-min normalization by column for each spectrum was used.

Machine Learning Analysis

Unsupervised Machine Learning

This study used two clustering algorithms to evaluate whether SERS spectral data belonging to 15 bacterial species were separable. SERS spectra were first pre-processed *via* removal of abnormal spectra, curve smoothing and denoising, baseline correction, and normalization as described above. Principal component analysis (PCA) was then applied to identify the principal components according to their cumulative contribution values, and only principal components with contribution value greater than 95% were retained. That is, the top m principal components whose cumulative variance contribution rate reached 95% were kept. Two clustering algorithms, K-means and AGNES were used to analyze the dimensionality-reduced SERS spectral data *via* Python *sklearn* package. The n_cluster parameter of the two clustering algorithms was preset to 15, and the linkage parameter of AGNES was set to *ward*, which minimized the sum of squared distances between all clusters.

Supervised Machine Learning

This study constructed and compared three classical supervised machine learning algorithms (RF, DT, SVM) and five deep learning algorithms (MLP, CNN, Simple RNN, GRU, LSTM) in terms of their capacities in predicting Raman spectra into different bacterial species, through which the best model(s) were identified. The pre-processed SERS spectra for each bacterial species were divided into training set, validation set and test set by following the ratio of 6:2:2 while the labels in the dataset were converted into the one-hot encoding form. In particular, one-hot encoding mainly uses N-bit state registers to encode N states. Each state has its own independent register bits, and only one bit is valid at any time. In simple terms, it is the representation of a categorical variable as a binary vector. Except for the integer index representing the variable, all other values are 0, while the variable is marked as 1.

Deep learning methods included two models, CNN and RNN, in which MLP had a special CNN network structure while SimpleRNN and GRU were simplified LSTM model (**Figure 1**). In particular, CNN model consisted of one input layer, six convolutional layers with convolution kernel sizes of 5*1 and 3*1, three maximum pooling layers, one fully connected layer and a *softmax* output layer that achieved 15-dimensional outputs. As for MLP, its network structure consisted of one input layer, four fully connected layers and one output layer. The activation function was selected as *relu*. In specificity, the framework of MLP neural network model in this study included an input layer, three hidden layers and an output layer. Each hidden layer is paired with a dropout layer. The rate of dropout is 0.2 while the activation function of each hidden layer is used *relu*. For the output layer, units are set to 15, and the activation function selects *softmax* for multi-classification. Dropout layer was mainly used to prevent curve overfitting and enhance the generalization ability of the model. As for LSTM, SimpleRNN and GRU, the three RNN modes were composed of three RNN layers, two Dropout layers and a fully connected layer. All the machine learning models were fitted and trained on the training dataset, and their optimized parameters were listed in **Supplementary Table 3**.

Accuracy rates (ACC) were compared in terms of prediction abilities of different supervised machine learning algorithms on SERS spectra. In order to verify the reliability of ACC, 5-fold cross-validation was also performed, which could eliminate the adverse effects of unbalanced data divisions. Moreover, this study also used F1-score (F1) as an additional metric, which was equivalent to the harmonic average of Precision (Pre) and Recall. The larger F1 was, the better the model performance was. Similar to accuracy and recall rate, receiver operating characteristic (ROC) curve also functioned as a measurement of the model quality. The optimal machine learning model was then selected based on all the above-mentioned indicators. In order to assess how many bacterial genera were misjudged as other genera, this study also used a confusion matrix of CNN to quickly visualize the proportion of misclassified genera.

Robustness of Machine Learning Models

Addition of noises to existing Raman spectral data could evaluate the performance of different models in terms of their predicting capacities. In this study, we added artificial noise with different signal-to-noise ratios (SNRs) to the pre-processed data set in order to test model robustness. The specific process was as follows: set different values of SNR in advance (SNR = 1, 2, 3, 5, 15, 25, 35); generate the required input noise D randomly (Formula 1); calculate SNR through dividing the signal power (Power of Signal, PS, Formula 2) with noise power (Power of

Nosie, PN, Formula 3); calculate the required noise ND (Formula 4) through PN and D; and the noisy signal (NS, Formula 5) was finally obtained.

$$D = d_{\max} - \frac{\sum_{i=1}^{n} d_i}{n} \left(d_i \in (0,\ 1)\right) \tag{1}$$

$$PS = \sum |x|^2 \tag{2}$$

$$PN = \frac{PS}{10^{\frac{SNR}{10}}} \tag{3}$$

$$ND = \frac{\sqrt{PN}}{\sqrt{\frac{\sum_{i=1}^{n}(D-\mu)^2}{n}}} * D \tag{4}$$

$$NS = ND + x \tag{5}$$

In the above-listed formula, d_i was the randomly generated signal, d_{max} was the maximum value of the randomly generated noise signal, x was the original Raman spectral signal data, n was the number of random signals generated, and μ was the arithmetic mean of n random signals.

RESULTS

Average Surface-Enhanced Raman Spectroscopy Spectra and Characteristic Peaks

In this study, we calculated mean signal intensity at each Raman shift for a specific bacterial species to generate average SERS spectra, with the addition of 20% standard error (SE) band. The thinner the error band, the smaller the standard deviation, and the higher the reproducibility of Raman spectrum. After that, the software LabSpec was used to perform curve smoothing and denoising, baseline correction, and normalization operations on each average Raman spectrum. According to the results, all bacteria showed smooth and distinct spectral distributions with observable signal peaks at different Raman shifts (**Figure 2A**). Through analyzing average SERS spectra *via* the LabSpec software, it was shown that different bacterial pathogens had their own species-specific combinations of characteristic peaks (**Figure 2B**). In addition, according to previous studies, characteristic peaks of Raman spectra could be matched to metabolites (Tang et al., 2021; Wang et al., 2021). The corresponding metabolites of all the characteristic peaks in the SERS spectra of the 15 bacterial pathogens were found in literature and are presented in **Supplementary Table 4**.

Clustering of Pathogenic Bacteria

Two common unsupervised machine learning algorithms, K-means and AGNES, were used to classify the SERS spectral data into different groups. According to previous studies, K-means algorithm has already been successfully applied to analyse Raman spectra of biological samples. As for AGNES, it is a hierarchical clustering method that divides data into different sets through successive fusion of a single object, which has also been widely used in biological sample analysis. However, these two methods were rarely used for Raman spectral analysis. In this study, the clustering results of K-means and AGNES were shown in the form a scattering dot diagram in **Figure 3**. In order to quantify the clustering effects of the two methods, the metric Rand Index was used to evaluate the performance of the two algorithms. K-means algorithm achieved the highest score that was only 27.4%. The possible reason might be due to that data in the same Raman spectrum had large differences between the maximum and minimum intensities, and the number of samples of different bacterial pathogens were unevenly distributed, which made K-means unable to fit correctly.

Prediction of Pathogenic Bacteria

Eight supervised machine learning algorithms were compared in this study to explore their predictive abilities in the identification of bacterial pathogens through the analysis of their SERS spectra, which included CNN, DT, GRU, LSTM, MLP, RF, SimpleRNN, and SVM. A total of four evaluation indicators, which were accuracy (ACC), precision (Pre), Recall and F1, together with 5-fold cross-validation were used to evaluate the performance of the eight algorithms. According to the results summarized in **Table 1**, CNN had the highest prediction accuracy (99.86%), and its five-fold cross-validation was the most robust with the overall accuracy of 99.47%.

In order to evaluate the diagnostic abilities of a supervised machine learning algorithm, a probability curve called receiver operating characteristic (ROC) curve was drawn at various threshold settings, through which sensitivities (true positive rate, TPR) and specificities (false positive rate, FPR) for different values of a continuous test were visualized (Hoo et al., 2017). In the ROC curve, upper left corner indicated higher TPR and lower FPR. Thus, regions in the ROC curves closer to the upper left corner had larger sum of sensitivity and specificity. In order to quantify TPR and FPR, the index area under the curve (AUC) was calculated, according to which, the larger the AUC value, the better the performance of a supervised machine learning model. According to the results in **Figure 4**, it was apparent that CNN had the highest AUC that was closely followed by LSTM.

Confusion matrix is an evaluation table to quantify the classification performance of machine learning algorithms by using true class and predicted class. Each row of the matrix represents the probability that the model predicts a true sample, and each column represents the probability that the model predicts an incorrect sample. Since CNN model achieved the best performance in bacterial identification in this study, we calculated its confusion matrix to provide further classification details (**Figure 5**). According to the matrix, most of the bacterial pathogens could be accurately discriminated by the CNN model with 100% accuracy.

Robustness of Machine Learning Methods

Noise sources could greatly compromise the quality of Raman spectra, causing issues in computational analysis of the spectral

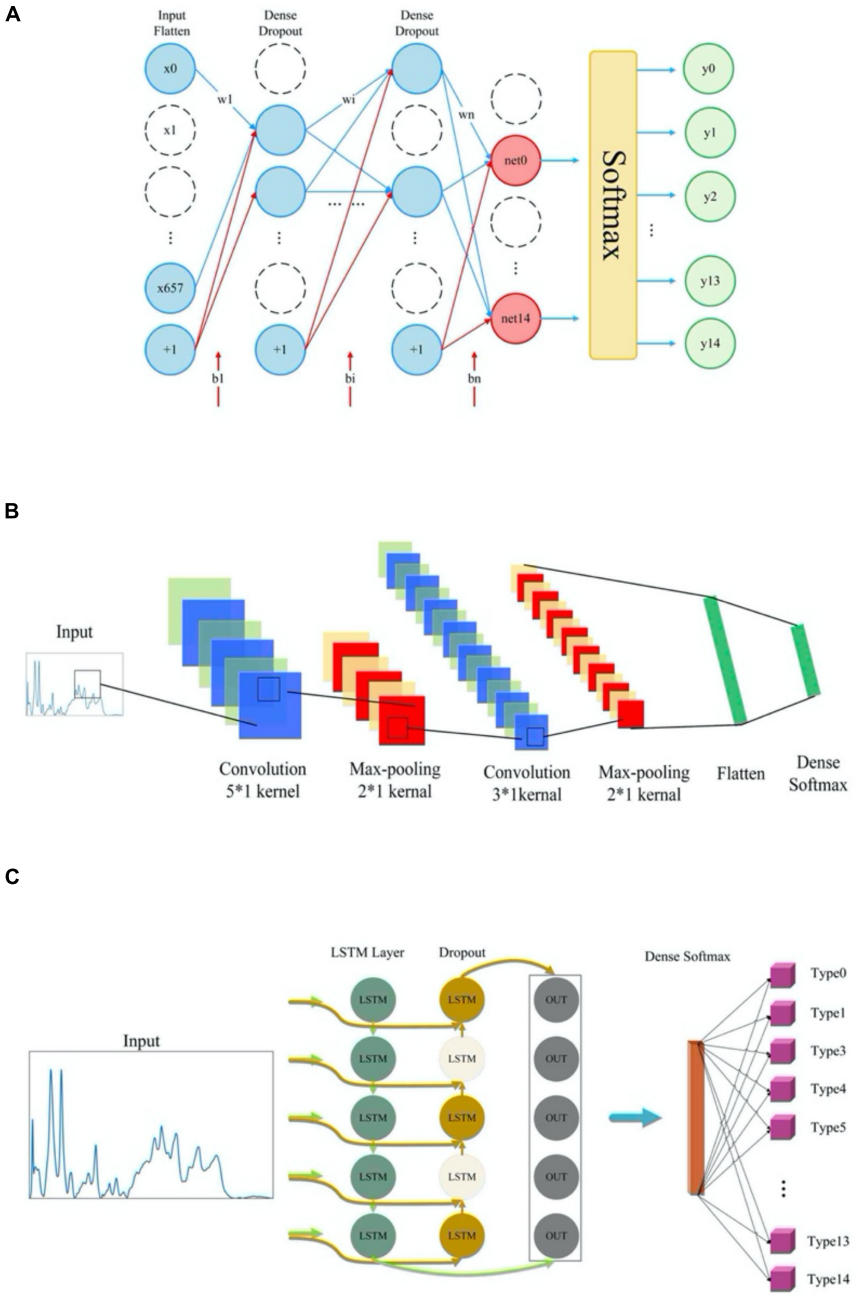


FIGURE 1 | Schematic illustration of the network structures of deep learning algorithms. **(A)** Multi-layer perceptron model architecture. **(B)** Convolutional neural network (CNN) model architecture. **(C)** Network structure of three recurrent neural networks (RNN) models.

data and leading to inaccurate determination of specimen composition (Tuchin et al., 2017). However, it is not known to what degree that noises could influence the performance of machine learning algorithms. In order to check the impacts of noises on Raman spectral analysis, artificial noise intensity was added to the pre-processed SERS spectra at 1, 2, 3, 5, 15, 25, and 35 dB, respectively. Eight different machine learning algorithms were used to analyze the modified spectra, and the effects of noises on these models were assessed *via* prediction accuracy. According to the results shown in **Figure 6**, CNN maintained a consistently high and stable prediction accuracy at different noise intensities, which was followed by LSTM and GRU. The prediction accuracies of three algorithms were kept above 95%, showing good and stable performance during spectral data analysis. It could also be seen that the prediction accuracies of RF and DT models were less than 75% when SNR was equal to 1 or 2, which indicated that performance robustness of RF and DT models were poor with low SNRs.

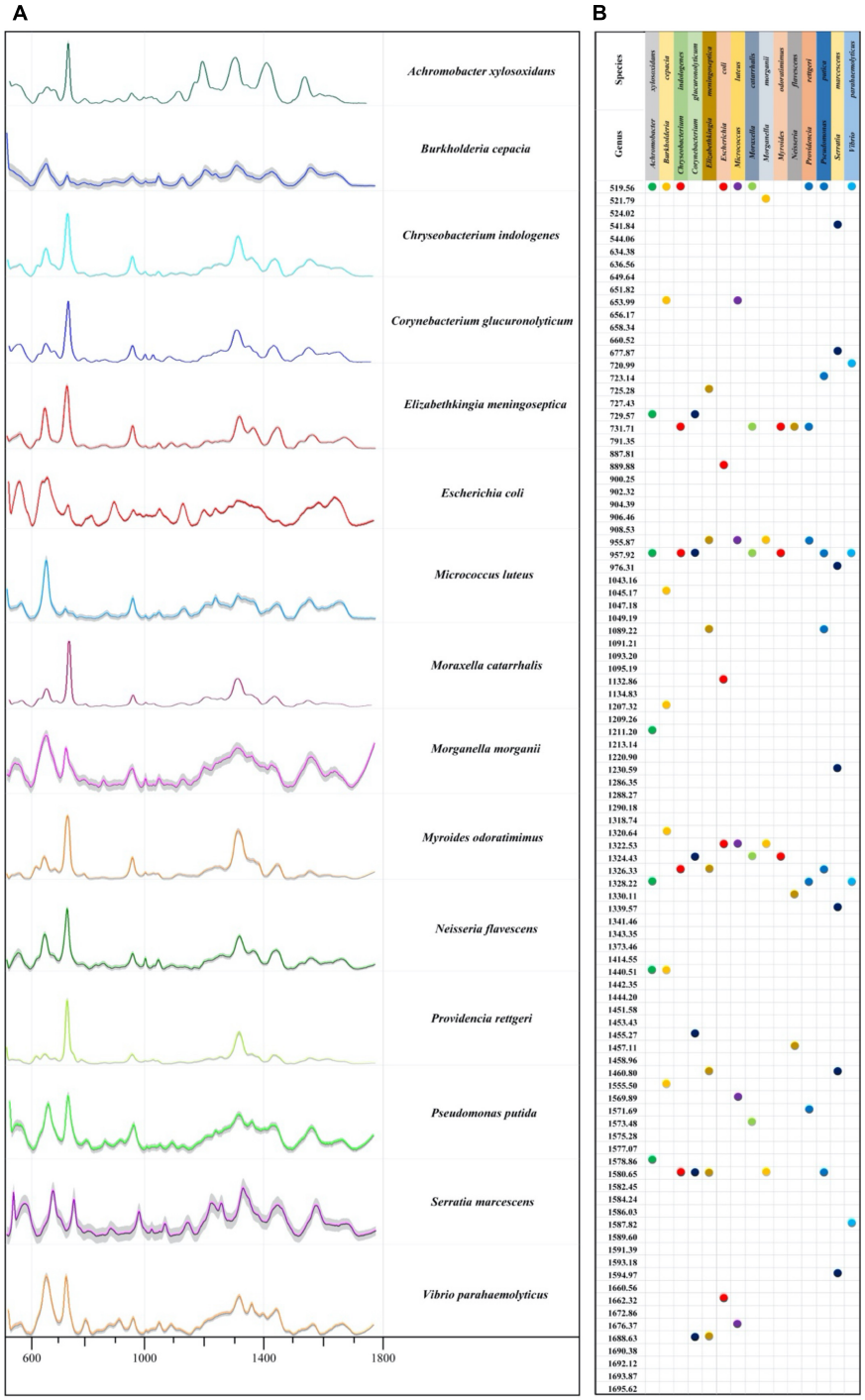


FIGURE 2 | Average Raman spectra and the corresponding characteristic peaks of 15 bacterial pathogens isolated from clinical samples. **(A)** Average surfaced enhanced Raman spectra (SERS) spectra of 15 different clinical bacterial pathogens. Shaded part in each spectrum was 20% error band. **(B)** Dot plot distribution of characteristic peaks in the Raman spectra for 15 bacterial pathogens.

DISCUSSION

Traditional methods for the detection of bacterial pathogens mainly rely on culture and biochemical tests to perform bacterial discrimination and phenotypic profiling, which, in spite of their high accuracies, are generally time-consuming and laborious while advanced molecular methods such as PCR and ELISA require specially designed primers or antibodies

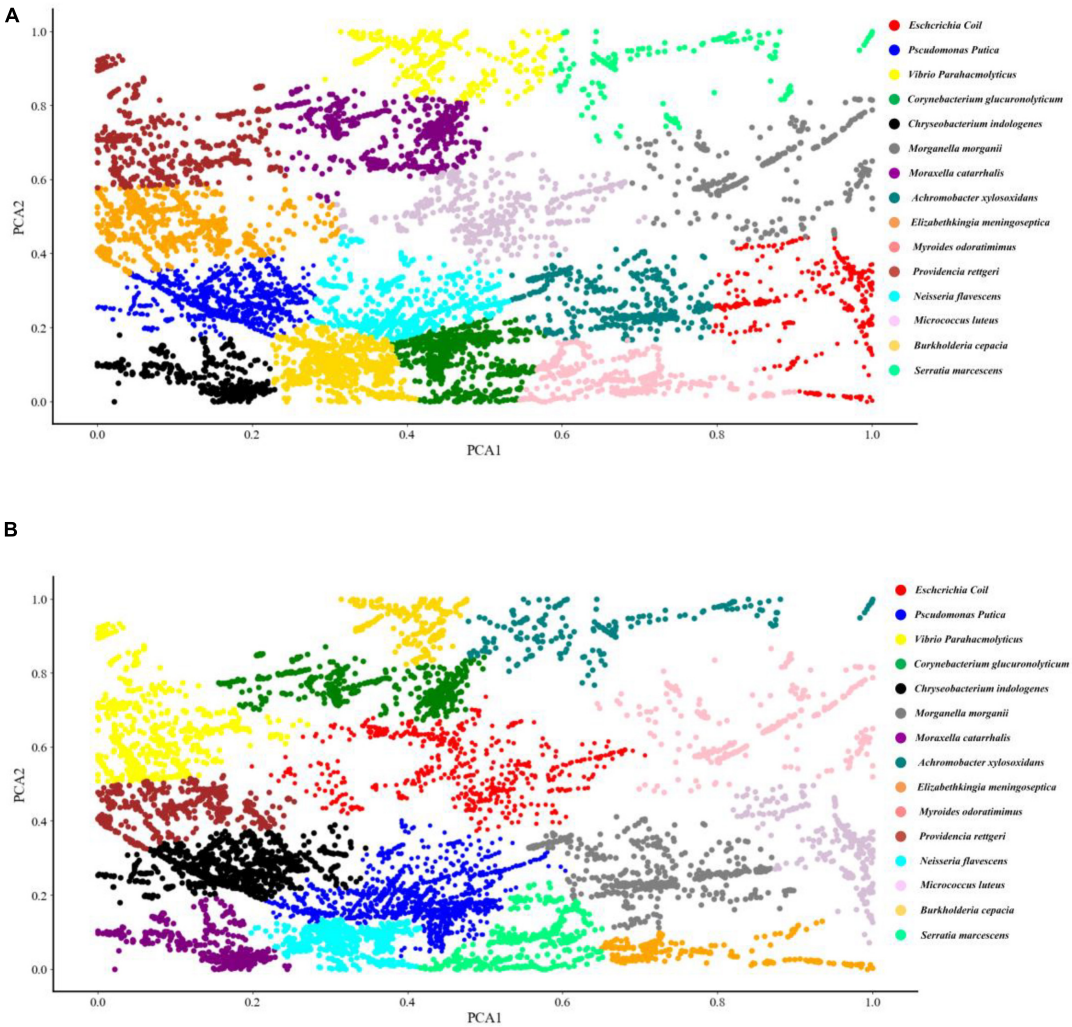


FIGURE 3 | Schematic illustration of the clustering results of 15 bacterial pathogens used in this study *via* K-means and agglomerative nesting (AGNES). **(A)** K-means. **(B)** AGNES. Dots with different colors represented different bacterial pathogens as indicated by the figure legends on the right.

and have relatively high costs and/or false positive rates (Wang et al., 2021). For the newly developed high-throughput sequencing technology, although the cost of sequencing has dropped significantly, complex sample preparation procedures and data analysis processes have limited its wide application in clinical laboratories for routine diagnosis of bacterial pathogens (Deurenberg et al., 2017). As a sensitive, low-cost, label-free, and non-destructive technology of biological sample analysis, Raman spectroscopy has potential in promoting fast diagnosis of bacterial pathogens, though it is still considered as a novel technology and is too arbitrary to ascertain that RS can be applied in clinical settings at any time soon because a gap between basic research and clinical implementation still exists for the methodology (Wang et al., 2021).

Although SERS spectra have higher signal intensity and data quality than Raman spectra, they still need to be preprocessed in order to improve the performance of computational analysis (Xiong and Ye, 2014; Tang et al., 2021). In addition, due to the complexity of Raman spectral data, the classical linear method is no longer suitable for its spectral data analysis (Lussier et al., 2020). In this study, we aimed to understand

TABLE 1 | Comparative analysis of the predicative capabilities of eight machine learning algorithms on surfaced enhanced Raman spectra (SERS) spectral data belonging to 15 bacterial pathogens.

Algorithms	ACC	Pre	Recall	F1	5-Fold CV
CNN	99.86%	99.91%	99.91%	99.93%	99.47%
LSTM	98.87%	98.87%	92.20%	98.74%	96.76%
RF	98.71%	98.77%	98.80%	98.77%	98.35%
GRU	98.61%	97.91%	97.93%	97.92%	89.68%
SVM	97.30%	97.30%	97.08%	97.28%	97.93%
SimpleRNN	96.43%	96.91%	95.89%	95.91	83.63%
DT	96.01%	97.96%	97.53%	97.95%	97.48%
MLP	95.17%	96.07%	95.54%	95.86%	96.84%

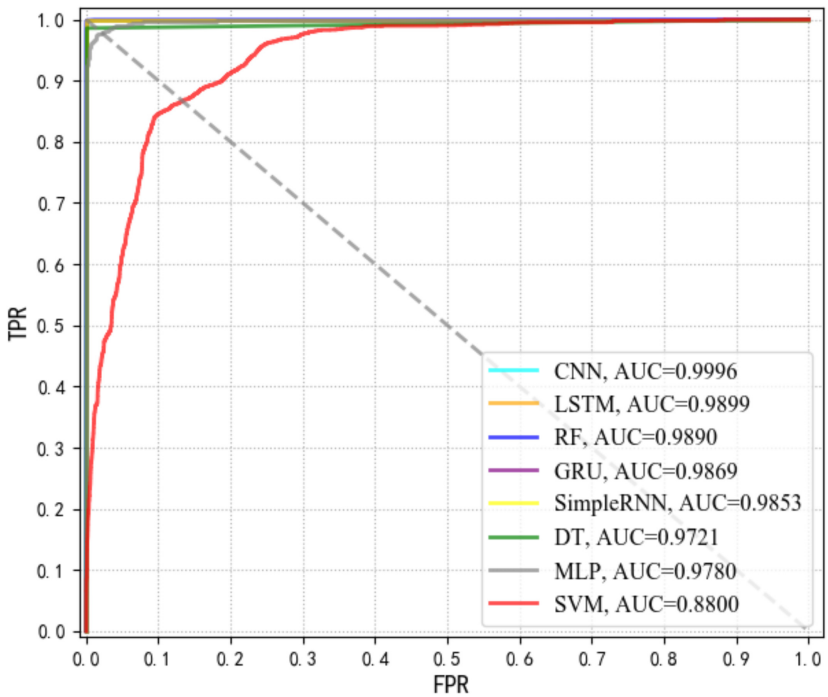


FIGURE 4 | Comparison of receiver operating characteristic curve (ROC) curves *via* area under the curve (AUC) values for eight supervised machine learning algorithms.

the intrinsic differences among Raman spectra belonging to 15 bacterial pathogens through comparing the classification and prediction abilities of both unsupervised and supervised machine learning algorithms. Previously, PCA combined with HCA was successfully applied to cluster *Staphylococcus aureus* and *E. coli* into different groups (Boardman et al., 2016). In addition, Weng et al. (2018) applied K-means to urine samples for automatic filter of dynamic spectra and rapid detection of drugs in urine, while Geng et al. (2021) used hierarchical clustering analysis (HCA) to differentiate neural stem cells accurately through label-free Raman spectroscopy. In this study, we used two clustering algorithms, KMeans and AGNES, to classify the 15 bacterial pathogens. Pathogenic bacteria were clustered in **Figure 3**. However, due to the uneven distribution of the spectral data of different bacterial pathogens and the complex SERS spectral data, the clustering effects were not ideal, indicating that more advanced calculation methods were needed for further investigations. In addition to this, more data corresponding to distinct isolates and a more even representation of isolate variability within each species are issues that should be addressed in future research.

As for the prediction of bacterial species, we used three traditional machine learning algorithms (RF, DT and SVM) and five deep learning algorithms (CNN, GRU, LSTM, MLP, and SimpleRNN) to analyze SERS spectral data. Although many supervised machine learning algorithms were applied to the analysis of Raman spectra (Riva et al., 2021), few studies systematically compared the classification and prediction of multiple algorithms among clinical pathogens belonging to different genera. Previously, Tang et al. compared 10 supervised machine learning algorithms in terms of performance on 2,752 SERS spectra from 117 *Staphylococcus* strains belonging to nine clinically important *Staphylococcus* species, according to which all supervised machine learning models achieved good prediction results while CNN topped all other models and accurately predicted *Staphylococcus* species with the highest accuracy at 98.21% (Tang et al., 2021). In this study, our results suggested that the deep learning algorithm CNN had the best performance on SERS spectra (accuracy = 99.86%) for the prediction of bacterial species at general level (**Table 1**).

When dealing with low-dimensional data in small volumes, it is convenient to pick-up outlier values through inter-quartile range (IQR) analysis. However, the method is time-consuming and labor-intensive when applied to large-scale data and is not suitable for high-dimensional data analysis. Common methods for processing high-dimensional data include Mahalanobis distance, robust Mahalanobis distance and principal component measurement method (PCout). In particular, the Mahalanobis distance method could evaluate whether a spectrum is an outlier or not by comparing the distances of all the corresponding points between the tested spectra and all other spectra one by one (De Maesschalck et al., 2000). However, the method is not robust because individual outliers will cause the mean vector and covariance matrix to shift toward wrong direction, leading to abnormal Mahalanobis distance and misidentified outliers; in contrast, robust Mahalanobis distance method constructs a robust mean and covariance matrix estimator through iteration to identify

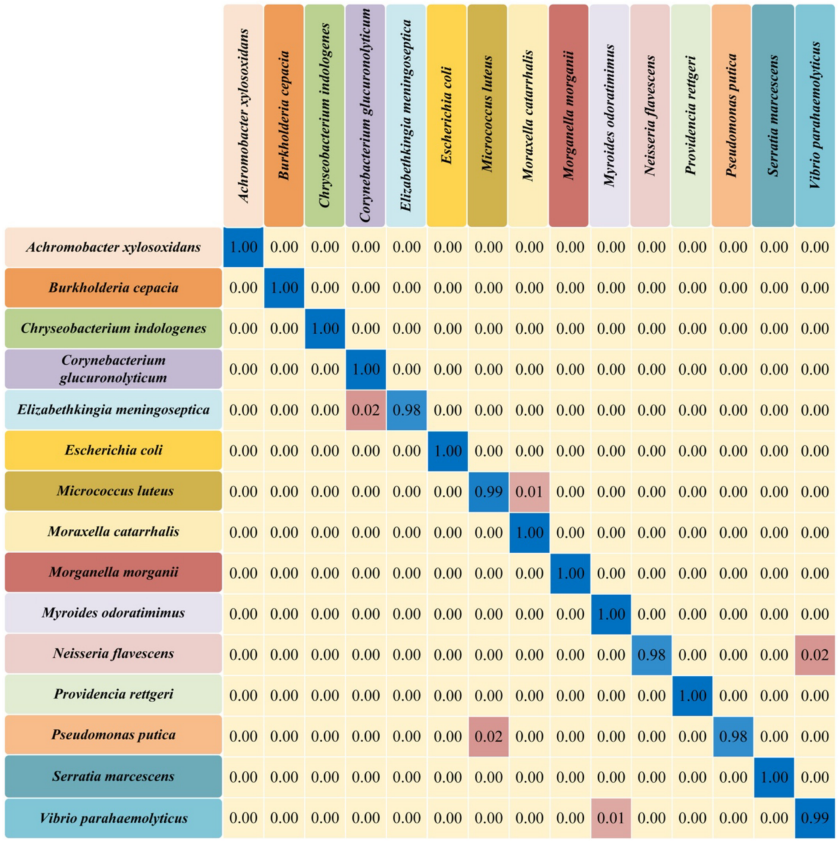


FIGURE 5 | Confusion matrix of the convolutional neural network (CNN) model for 15 different bacterial pathogens. The rows in the confusion matrix represented the true categories of predictions, while the columns represented the categories of the incorrect predictions. The probability of correct prediction (diagonal) and the probability of incorrect prediction (off-diagonal) were all present in the matrix.

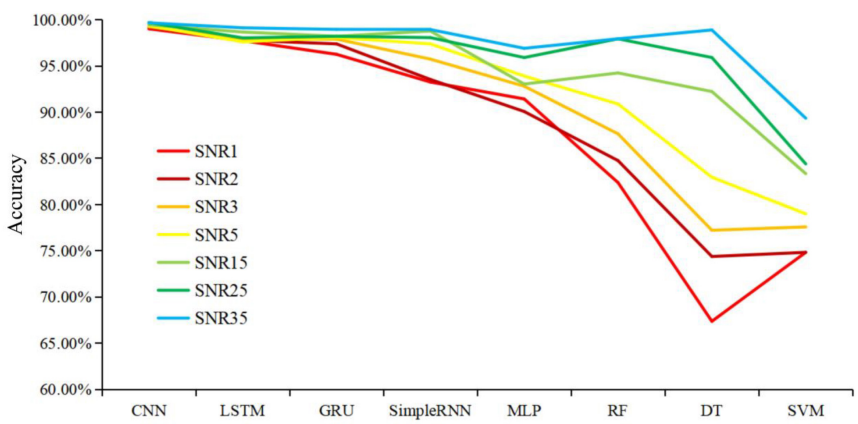


FIGURE 6 | Schematic illustration of the influences of different signal-to-noise ratio on the prediction accuracies of eight machine learning algorithms. X-axis shows different machine learning models. Y-axis represents prediction accuracy of machine learning algorithms under different signal-to-noise ratios. Lines with different colors represent noises intensities. The smaller the signal-to-noise ratio (SNR) value, the more noise added to the spectra and the worse the data quality.

outliers, which is able to solve the problem (Cabana et al., 2019). As the dimensionality increases, the distribution of data in the coordinate system will become increasingly sparse, leading to mis-judgement of real-data and increases and insufficiency of outlier identification through distance methods. Thus, in this study, we recruited PCount() function in the mvoutlier package of the R programming language for outlier identification and removal.

It should also be noted that SERS spectral preprocessing was important in reducing the noise signals and improving the predictive ability of the model. Noise in signals was unavoidable for Raman spectroscopy because of factors such as fluctuations of environmental conditions, sample contaminations, and background fluorescence, etc., leading to the generation of abnormal spectra data that compromised data quality (Xiong and Ye, 2014; Tuchin et al., 2017). Thus, in this study, artificial noises were added to SERS spectra to objectively evaluate the robustness of the model. According to the result in **Figure 6**, it was found that CNN maintained a consistently high and stable prediction accuracy at different noise intensities, indicating that the CNN had strong robustness in classification and prediction of different pathogens. Further research should focus on directly identifying bacterial pathogens from clinical samples such sputum, blood and urine, etc., which is very challenging and will greatly facilitate the application of Raman spectroscopy in the clinical settings.

CONCLUSION

Raman spectroscopy has been widely investigated in terms of its capacities in rapid diagnosis of bacterial pathogens such as species discriminations, antibiotic resistance detections and toxin identifications, etc. However, there is no rationale to claim that Raman spectroscopy is applicable for microbiologists and clinicians in real-world situations because a wide gap still exists between basic research and clinical implementation. In this study, we used surface enhanced Raman spectroscopy combined with unsupervised and supervised machine learning algorithms to detect 15 bacterial pathogens sourced from clinical samples. According to the results, SERS could accurately identify bacterial pathogens at general level with comparatively high specificity and sensitivity through the assistance of machine learning methods. Comparative analyses of all the supervised machine learning algorithms used in this study revealed that the deep learning algorithm CNN had the best prediction performance. In addition, CNN also topped other algorithms in terms of robustness when dealing with SERS spectra with artificially added noises. However, there are still many machine learning algorithms that have not been explored and should be investigated in future studies. For example, when the sample datasets are limited, a generative adversarial network can be used to amplify data amount while for datasets with more Raman shifts and higher dimensions, wavelength selection could be used, which is conducive to identify and select important bands for down-stream analysis. Moreover, standardized Raman spectroscopy database with reproducible spectra for clinically important pathogens should also be constructed, which could greatly improve the implementation of Raman spectroscopy in clinical environments. Taken together, Raman spectroscopy is a promising technique with potential for label-free detection and non-invasive identification of clinical pathogens, which is worthy of extensive explorations in future studies.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding authors.

AUTHOR CONTRIBUTIONS

LW, XZ, and BG conceived and designed the experiments and provided the platform and resources. LW was responsible for project administration. J-WT, J-QL, X-CY, W-WX, Y-CP, and Q-HL carried out the experimental work. LW and J-WT performed the data analysis. LW, J-WT, and Q-HL wrote and revised the manuscript. All authors read and approved the final manuscript.

FUNDING

LW was financially supported by National Natural Science Foundation of China (No. 31900022), Young Science and Technology Innovation Team of Xuzhou Medical University (No. TD202001), and Jiangsu Qinglan Project (2020). BG appreciated the financial supports by the National Natural Science Foundation of China (81871734 and 82072380), Key R&D Program of Jiangsu Province (BE2020646), and Research Foundation for Advanced Talents of Guandong Provincial People's Hospital (KJ012021097).

ACKNOWLEDGMENTS

We greatly appreciate the reviewers whose comments significantly improved the quality of the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.843417/full#supplementary-material

REFERENCES

Baek, S.-J., Park, A., Ahn, Y.-J., and Choo, J. (2015). Baseline correction using asymmetrically reweighted penalized least squares smoothing. *Anal.* 140, 250–257. doi: 10.1039/c4an01061b

Bloom, D. E., and Cadarette, D. (2019). Infectious disease threats in the twenty-first century: strengthening the global response. *Front. Immunol.* 10:549.

Boardman, A. K., Wong, W. S., Premasiri, W. R., Ziegler, L. D., Lee, J. C., Miljkovic, M., et al. (2016). Rapid detection of bacteria from blood with surface-enhanced raman spectroscopy. *Anal. Chem.* 88, 8026–8035. doi: 10.1021/acs.analchem.6b01273

Cabana, E., Lillo, R. E., and Laniado, H. (2019). Multivariate outlier detection based on a robust Mahalanobis distance with shrinkage estimators. *Statist. Pap.* 62, 1583–1609. doi: 10.1007/s00362-019-01148-1

Caliendo, A. M., Gilbert, D. N., Ginocchio, C. C., Hanson, K. E., May, L., Quinn, T. C., et al. (2013). Better tests, better care: improved diagnostics for infectious diseases. *Clin. Infect. Dis.* 57, S139–S170. doi: 10.1093/cid/cit578

De Maesschalck, R., Jouan-Rimbaud, D., and Massart, D. L. (2000). The Mahalanobis distance. *Chemometr. Intell. Lab. Syst.* 50, 1–18. doi: 10.1016/s0169-7439(99)00047-7

Deurenberg, R. H., Bathoorn, E., Chlebowicz, M. A., Couto, N., Ferdous, M., García-Cobos, S., et al. (2017). Application of next generation sequencing in clinical microbiology and infection prevention. *J. Biotechnol.* 243, 16–24. doi: 10.1016/j.jbiotec.2016.12.022

Filzmoser, P., Maronna, R., and Werner, M. (2008). Outlier identification in high dimensions. *Comput. Stat. Data Anal.* 52, 1694–1711. doi: 10.1016/j.csda.2007.05.018

Geng, J., Zhang, W., Chen, C., Zhang, H., Zhou, A., Huang, Y., et al. (2021). Tracking the differentiation status of human neural stem cells through label-free raman spectroscopy and machine learning-based analysis. *Anal. Chem.* 93, 10453–10461. doi: 10.1021/acs.analchem.0c04941

Ho, C.-S., Jean, N., Hogan, C. A., Blackmon, L., Jeffrey, S. S., Holodniy, M., et al. (2019). Rapid identification of pathogenic bacteria using Raman spectroscopy and deep learning. *Nat. Commun.* 10:4927. doi: 10.1038/s41467-019-12898-9

Hoo, Z. H., Candlish, J., and Teare, D. (2017). What is an ROC curve? *Emerg. Med. J.* 34, 357–359. doi: 10.1136/emermed-2017-206735

Järvinen, A.-K., Laakso, S., Piiparinen, P., Aittakorpi, A., Lindfors, M., Huopaniemi, L., et al. (2009). Rapid identification of bacterial pathogens using a PCR- and microarray-based assay. *BMC Microbiol.* 9:161. doi: 10.1186/1471-2180-9-161

Kuhar, N., Sil, S., Verma, T., and Umapathy, S. (2018). Challenges in application of raman spectroscopy to biology and materials. *RSC Adv.* 8, 25888–25908. doi: 10.1039/c8ra04491k

Lussier, F., Thibault, V., Charron, B., Wallace, G. Q., and Masson, J.-F. (2020). Deep learning and artificial intelligence methods for Raman and surface-enhanced Raman scattering. *TrAC Trends Anal. Chem.* 124:115796. doi: 10.1016/j.trac.2019.115796

Orlando, A., Franceschini, F., Muscas, C., Pidkova, S., Bartoli, M., Rovere, M., et al. (2021). A comprehensive review on raman spectroscopy applications. *Chemosensors* 9:262. doi: 10.3390/chemosensors9090262

Riva, M., Sciortino, T., Secoli, R., D'Amico, E., Moccia, S., Fernandes, B., et al. (2021). Glioma biopsies classification using raman spectroscopy and machine learning models on fresh tissue samples. *Cancers* 13:1073. doi: 10.3390/cancers13051073

Singhal, N., Kumar, M., Kanaujia, P. K., and Virdi, J. S. (2015). MALDI-TOF mass spectrometry: an emerging technology for microbial identification and diagnosis. *Front. Microbiol.* 6:791. doi: 10.3389/fmicb.2015.00791

Sloan, A., Wang, G., and Cheng, K. (2017). Traditional approaches versus mass spectrometry in bacterial identification and typing. *Clin. Chim. Acta* 473, 180–185. doi: 10.1016/j.cca.2017.08.035

Tang, J. W., Liu, Q. H., Yin, X. C., Pan, Y. C., Wen, P. B., Liu, X., et al. (2021). Comparative analysis of machine learning algorithms on surface enhanced raman spectra of clinical *Staphylococcus* species. *Front. Microbiol.* 12:696921. doi: 10.3389/fmicb.2021.696921

Tuchin, V. V., Larin, K., Leahy, M., and Wang, R. (2017). *Dynamics and Fluctuations In Biomedical Photonics XIV*. San Francisco, CA: SPIE.

Wang, L., Liu, W., Tang, J. W., Wang, J. J., Liu, Q. H., Wen, P. B., et al. (2021). Applications of raman spectroscopy in bacterial infections: principles, advantages, and shortcomings. *Front. Microbiol.* 12:683580. doi: 10.3389/fmicb.2021.683580

Weng, S., Dong, R., Zhu, Z., Zhang, D., Zhao, J., Huang, L., et al. (2018). Dynamic surface-enhanced raman spectroscopy and chemometric methods for fast detection and intelligent identification of methamphetamine and 3, 4-methylenedioxy methamphetamine in human urine. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* 189, 1–7. doi: 10.1016/j.saa.2017.08.004

Xiong, M., and Ye, J. (2014). Reproducibility in surface-enhanced Raman spectroscopy. *J. Shanghai Jiaotong Univ.* 19, 681–690. doi: 10.1007/s12204-014-1566-7

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.

frontiers | Frontiers in Microbiology

ORIGINAL RESEARCH
published: 11 April 2022
doi: 10.3389/fmicb.2022.846915

NNAN: Nearest Neighbor Attention Network to Predict Drug–Microbe Associations

Bei Zhu[1†], Yi Xu[1†], Pengcheng Zhao[1], Siu-Ming Yiu[2], Hui Yu[3] and Jian-Yu Shi[1]**

[1] School of Life Sciences, Northwestern Polytechnical University, Xi'an, China, [2] Department of Computer Science, The University of Hong Kong, Hong Kong, China, [3] School of Computer Science, Northwestern Polytechnical University, Xi'an, China

OPEN ACCESS

Edited by:
Qi Zhao,
University of Science and Technology Liaoning, China

Reviewed by:
Wen Zhang,
Huazhong Agricultural University, China
Lihong Peng,
Hunan University of Technology, China

***Correspondence:**
Hui Yu
huiyu@nwpu.edu.cn
Jian-Yu Shi
jianyushi@nwpu.edu.cn

† These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Systems Microbiology, a section of the journal Frontiers in Microbiology

Received: 31 December 2021
Accepted: 14 February 2022
Published: 11 April 2022

Citation:
Zhu B, Xu Y, Zhao P, Yiu S-M, Yu H and Shi J-Y (2022) NNAN: Nearest Neighbor Attention Network to Predict Drug–Microbe Associations. Front. Microbiol. 13:846915. doi: 10.3389/fmicb.2022.846915

Many drugs can be metabolized by human microbes; the drug metabolites would significantly alter pharmacological effects and result in low therapeutic efficacy for patients. Hence, it is crucial to identify potential drug–microbe associations (DMAs) before the drug administrations. Nevertheless, traditional DMA determination cannot be applied in a wide range due to the tremendous number of microbe species, high costs, and the fact that it is time-consuming. Thus, predicting possible DMAs in computer technology is an essential topic. Inspired by other issues addressed by deep learning, we designed a deep learning-based model named Nearest Neighbor Attention Network (NNAN). The proposed model consists of four components, namely, a similarity network constructor, a nearest-neighbor aggregator, a feature attention block, and a predictor. In brief, the similarity block contains a microbe similarity network and a drug similarity network. The nearest-neighbor aggregator generates the embedding representations of drug–microbe pairs by integrating drug neighbors and microbe neighbors of each drug–microbe pair in the network. The feature attention block evaluates the importance of each dimension of drug–microbe pair embedding by a set of ordinary multi-layer neural networks. The predictor is an ordinary fully-connected deep neural network that functions as a binary classifier to distinguish potential DMAs among unlabeled drug–microbe pairs. Several experiments on two benchmark databases are performed to evaluate the performance of NNAN. First, the comparison with state-of-the-art baseline approaches demonstrates the superiority of NNAN under cross-validation in terms of predicting performance. Moreover, the interpretability inspection reveals that a drug tends to associate with a microbe if it finds its top-*l* most similar neighbors that associate with the microbe.

Keywords: deep learning, bipartite graph network, link prediction, drug–microbe association, attention matrix

INTRODUCTION

The human microbiome refers to all the microbes associated with a human body, including bacteriophages, archaea, bacteria, eukaryotes, and fungi (Lynch and Pedersen, 2016). To assess the diversity and functions of the human microbiome, the Human Microbiome Project (HMP) was supported by the National Institutes of Health (NIH) from 2007 to 2016 (Turnbaugh et al., 2007).

HMP provided a complete description of the microbiome in five tissues of the human body, including skin, gut, nostrils, vagina, and mouth (Aagaard et al., 2013). Human microbes have been verified for their close associations with human health by cell experiments, animal experiments, epidemiological studies, clinical case studies (Schwabe and Jobin, 2013; Lynch and Pedersen, 2016), etc. Previous works have revealed that abnormal microbe communities lead to metabolic disorders [e.g., non-alcoholic fatty liver disease (Younossi et al., 2016), obesity, and diabetes mellitus (Jaacks et al., 2019; Zheng et al., 2018)]. Oral drug administration is a typical treatment. Many drugs, however, can be metabolized by human microbes, and the drug metabolites would significantly alter pharmacological effects and result in low therapeutic efficacy for patients. For example, after being modified by gut microbes, the compounds can lead to their activation [e.g., *salicylazosulfapyridine* (Sousa et al., 2014)] or inactivation [e.g., inactivation of the cardiac drug digoxin by the intestinal actinomycete *Eggerthella lenta* (Haiser et al., 2013)], or induce toxicity [e.g., 70% toxicity of Brivudine may be attributed to intestinal microorganisms (Zimmermann et al., 2019b)]. The persistent findings of microbiome-induced individual pathogenesis, phenotypes, and treatment responses boost the microbiome to be an integral part of precision medicine (Kashyap et al., 2017). Therefore, drug–microbe association (DMA) prediction is of great significance for therapy and medicine development. However, the acquisition of DMAs needs a large scale of assays with high costs, low efficiency, and culturing limitations, and that are time-consuming. To identify DMAs rapidly and effectively, machine learning methods, especially deep learning-based methods, have attracted many scientists due to their inspiring applications in other areas [e.g., predicting microbe–disease associations (He et al., 2018; Peng et al., 2018), drug–drug interactions (Yu et al., 2021a), lncRNA–miRNA interactions (Zhang L. et al., 2021), and lncRNA–protein interactions (Lihong et al., 2021; Zhou et al., 2021)].

In recent years, researchers have applied Graph Attention Network [GAT (Velickovic et al., 2018)] to bioinformatics with remarkable results. For instance, Zhang Z. et al. (2021) used fragments containing functional groups to represent molecular maps for molecular property prediction through a fragment-oriented multi-scale graph attention model. Bang et al. (2021) made the prediction of polypharmacy side effects with enhanced interpretability based on graph feature attention network. Constructing a bipartite network is the most popular approach to represent associations between two types of nodes. The prediction problem of DMA can then be transformed into a link prediction problem in a bipartite graph network. However, few models predict DMAs through bipartite graph networks. For example, EGATMDA (Long et al., 2020b) used the drug–disease–microbe perspective to predict the DMAs, which does not show a direct relationship between drugs and microbes and may contain noise. HMDAKATZ (Zhu et al., 2019) predicted the interactions between drugs and microbes based on the Katz (1953); the disadvantage of this method in the node's information transmission (i.e., a node with a high central value transmits its high influence to all its neighbors) may not be appropriate in real life. GCNMDA (Long et al., 2020a) used GCN, random walk with restart, and GAT to learn node features, which relies on the parameter "step size" when using the restart random walk algorithm. HNERMDA (Long and Luo, 2020) learned the drug–microbe heterogeneous network information by metapath2vec measure, which considered the type of nodes in the meta-path-based random walk but the skip-gram does not treat them differently during training.

In the field of drug–target interaction prediction, there is a widely accepted assumption that structurally similar drugs tend to interact with the same target (Khalili et al., 2012). Analogously, we anticipate that if a drug (d_x) can associate with a microbe (b_p), the other drugs associated with the same microbe (b_p) are usually the first l nearest neighbors of the drug (d_x). Therefore, we propose a new model, Nearest Neighbor Attention Network (NNAN), which aggregates the information from nodes' neighbors according to their entity types and maps them into a unified embedding space for further predicting potential DMAs. The comparison with state-of-the-art methods on two different databases demonstrates the superiority of our NNAN. Moreover, its interpretability is illustrated and validates our assumption. Finally, the case study assesses its ability to find potential associations between drugs and microbes. In general, our contribution is as follows:

- We make use of three networks: drug–drug similarity network, microbe–microbe similarity network, and a drug–microbe bipartite graph network. Imitate the idea of KNN [K-Nearest-Neighbor (Cover and Hart, 1967)] to learn the substructures of the bipartite graph network, which can promote the accuracy of link prediction.
- We follow the idea of GAT and use multiple DNNs to learn the weights of embedding features to improve the screening efficiency of potential associations.
- In a quantitative way, we verify the hypothesis that "If a drug can associate with a microbe, the other drugs that associate with the microbe are usually the first l nearest neighbors to the drug."

MATERIALS AND METHODS

In this section, we describe a model for predicting DMAs in a bipartite graph network, named NNAN as shown in **Figure 1**. It consists of four components: a similarity network constructor, a nearest-neighbor aggregator, a feature attention block, and a predictor. Firstly, the similarity network constructor is mainly used to build a drug similarity network and a microbe similarity network (section "Similarity Networks" for details). Secondly, the nearest-neighbor aggregator generates the embedding representations of drug–microbe pairs by integrating drug neighbors and microbe neighbors of each drug–microbe pair in the network (section "Nearest-Neighbor Aggregator for Drug–Microbe Pair Embeddings" for details). Thirdly, the feature attention block evaluates the importance of each dimension of drug–microbe pair embedding by a set of ordinary multi-layer neural networks (section "Feature Attention Block" for details).

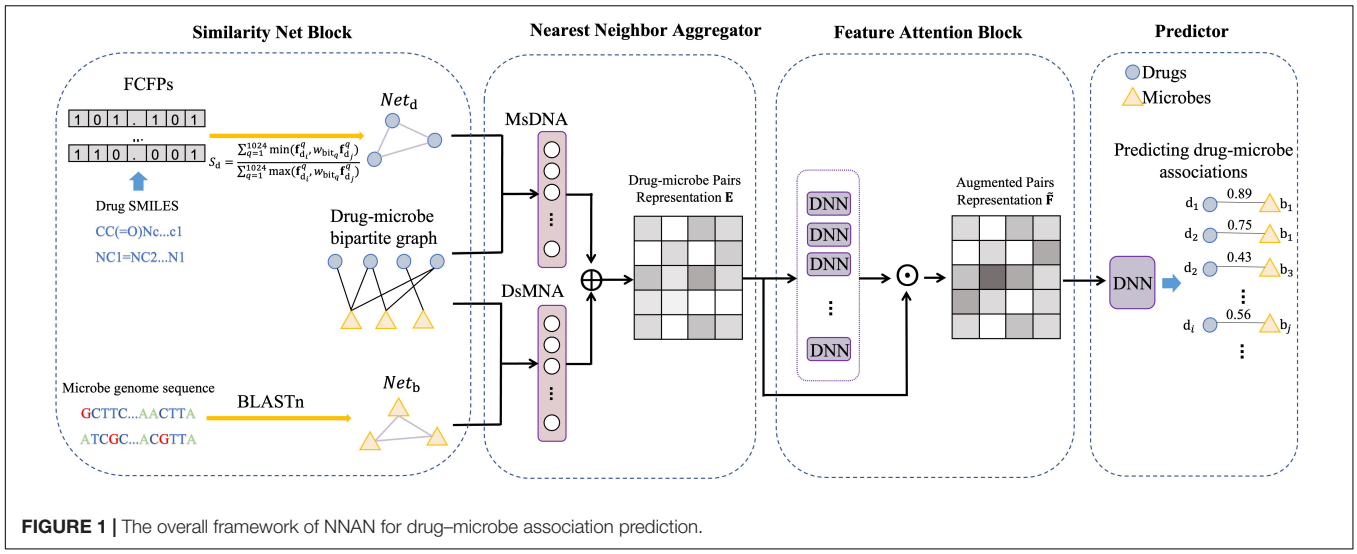


FIGURE 1 | The overall framework of NNAN for drug–microbe association prediction.

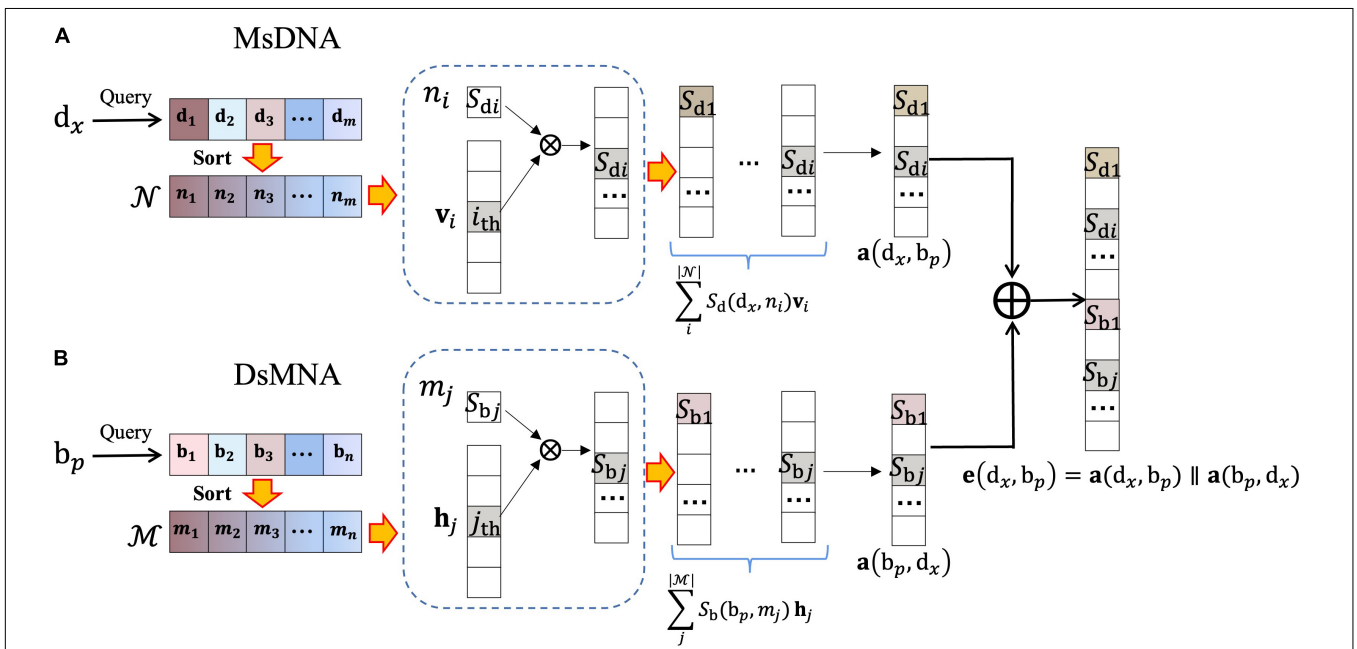


FIGURE 2 | Nearest-Neighbor Aggregator block. **(A)** Microbe-specific drug neighbor aggregator (MsDNA); the embedding representation of the unidirectional edge, which is from the drug d_x to the microbe b_p. **(B)** Drug-specific microbe neighbor aggregator (DsMNA); the embedding representation of the unidirectional edge, which is from the microbe b_p to the drug d_x, where $\mathcal{M}^x \subseteq \mathcal{M}$ is a set of instantiated keywords, $\mathcal{M}^x$ denotes the neighbors of microbe b_p in the Net_b. $S_b(b_p, m_j)$ denotes the similarity of b_p and m_j. $\mathbf{h}_j$ is the corresponding one-hot encoding vector of m_j.

Finally, we make use of a fully-connected deep neural network as a binary classifier to predict potential DMAs.

Similarity Networks

Drug Similarity Network

We calculate drug similarities by the following steps. First, drugs are represented by Functional-Class Fingerprints [FCFPs (Rogers and Hahn, 2010)], which is the generalized version of Extended-Connectivity Fingerprints [ECFPs (Rogers and Hahn, 2010)] with more attention to atom functions. The FCFPs is implemented by RDKit (Landrum, 2010). Second, the similarity between drug d_i and drug d_j is calculated by the Tanimoto coefficient (Rogers and Tanimoto, 1960) as follows:

$$S(d_i, d_j) = \frac{\mathbf{f}_{d_i} \cdot \mathbf{f}_{d_j}}{||\mathbf{f}_{d_i}|| + ||\mathbf{f}_{d_j}|| - \mathbf{f}_{d_i} \cdot \mathbf{f}_{d_j}} \quad (1)$$

where $\mathbf{f}_{d_i}$ and $\mathbf{f}_{d_j}$ represent the FCFPs vector of drug d_i and drug d_j, respectively, $||\cdot||$ indicates the norm of the vector.

Fingerprint similarity provides intuitive results: why the two molecules have been determined to be similar, but

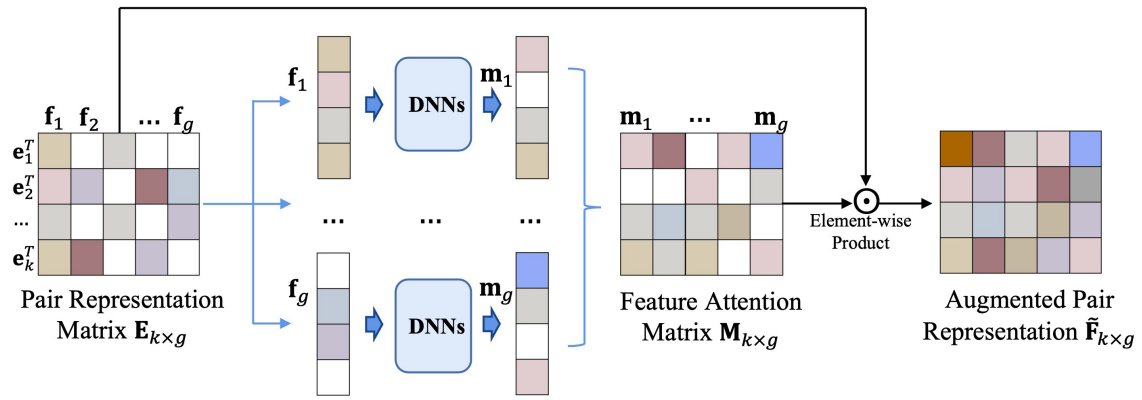


FIGURE 3 | Feature attention block. Input the representation matrix $\mathbf{E}_{k \times g}$ into a set of DNNs, then we obtain an attention matrix $\mathbf{M}_{k \times g}$of drug–microbe embedding features. After the element-wise product operation of $\mathbf{M}_{k \times g}$ and $\mathbf{E}_{k \times g}$, the final feature matrix $\widetilde{\mathbf{F}}_{k \times g}$ of the drug–microbe pairs is obtained.

TABLE 1 | The statistics of two databases.

	Drugs	Microbes	Associations
Database 1	999	133	1,708
Database 2	176	76	4,194

this transparency tends to vanish completely when molecular fingerprints are used as input to machine learning models. Inspired by the similarity maps (Riniker and Landrum, 2013), we calculate the contribution of each atom to the similarity between two molecules. To make it easier to distinguish the drugs, we regard d_i as a reference drug, d_j as a comparison drug, and $S\left(\mathrm{d}_i, \mathrm{d}_j\right)$ as the base similarity of this drug pair. The RDKit will automatically number each atom of the comparison drug d_j ($K = \{0, 1, \ldots, t-1\}$). Then, we remove the atoms of the comparison drug one by one in the order of the atomic numbers to form multiple new comparing drugs ($\mathrm{d}_j^k, k \in \mathrm{K}, \mathrm{K} = \{0, 1, \ldots, t-1\}$). We calculate the new similarity between the reference drug (d_i) and the new comparison drug (d_j^k), and regard the difference between the new similarity and the base similarity as the weight (w_j^k) of each removed atom. The weight w_j^k is formulated as:

$$w_j^k = \left| S\left(\mathrm{d}_i, \mathrm{d}_j\right) - S\left(\mathrm{d}_i, \mathrm{d}_j^k\right) \right| \tag{2}$$

We set the dimension of the FCFPs vector to 1,024 bits, of which the non-zero bits indicate the occurrences of drug feature substructures. To obtain the weight of each non-zero bit, we add up the weights of all the atoms contained in the feature substructure:

$$w_{\mathrm{bit}_q} = SUM_q(w_j^k) \tag{3}$$

where w_{bit_q} denotes the weight of the q_{th} dimensional bit of the FCFPs vector, and the function $SUM_q(\cdot)$ denotes the sum of all the atomic weights contained in the feature substructure represented by the q_{th} dimensional bit of the FCFPs.

Then, the weighted Tanimoto similarity (Ioffe, 2010) between the reference drug and the comparison drug can be calculated as follows:

$$S_{\mathrm{d}}\left(\mathrm{d}_i, \mathrm{d}_j\right) = \frac{\sum_{q=1}^{1024} \min\left(\mathbf{f}_{\mathrm{d}_i}^q, w_{\mathrm{bit}_q} \mathbf{f}_{\mathrm{d}_j}^q\right)}{\sum_{q=1}^{1024} \max\left(\mathbf{f}_{\mathrm{d}_i}^q, w_{\mathrm{bit}_q} \mathbf{f}_{\mathrm{d}_j}^q\right)} \tag{4}$$

where $\mathbf{f}_{\mathrm{d}_i}^q$ and $\mathbf{f}_{\mathrm{d}_j}^q$ denote the q_{th} dimension of the FCFPs vectors for the reference drug and the comparison drug.

Based on drug similarities, we can build a drug similarity network Net_{d}, where nodes are drugs. There are edges between the drugs if these drugs associate with the same microbe; the edges are weighted by drug similarities.

Microbe Similarity Network

To calculate microbe similarities, we use BLAST (Altschul et al., 1990) to make pairwise alignments of microbial genomes. Specifically, the main function of BLAST is to discover local similarity regions between sequences and then use the local sequence alignment algorithm (Smith and Waterman, 1981) to calculate the similarity. For example, $\mathrm{G_A} = \mathrm{g_A^1 g_A^2 \ldots g_A^n}$ and $\mathrm{G_B} = \mathrm{g_B^1 g_B^2 \ldots g_B^m}$ are the genome sequences of microbe A and microbe B, where n and m are the lengths of sequences $\mathrm{G_A}$ and $\mathrm{G_B}$, respectively. BLAST creates the scoring matrix $\mathbf{H}_{(\mathbf{n}+\mathbf{1}) \times (\mathbf{m}+\mathbf{1})}$ and makes the first row and column elements zero. The formula for the element $\mathbf{H}_{ij}$ ($\mathbf{H}_{ij} \in \mathbf{H}_{(\mathbf{n}+\mathbf{1}) \times (\mathbf{m}+\mathbf{1})}, i = 1, 2, \ldots, n; j = 1, 2, \ldots, m$) in this scoring matrix is:

$$\mathbf{H}_{ij} = \max \begin{cases} \mathbf{H}_{i-1,j-1} + \mathrm{Score} \\ \mathbf{H}_{i-k,j} - 2 \\ \mathbf{H}_{i,j-k} - 2 \\ 0 \end{cases} \quad \left(\mathrm{g}_{\mathrm{A}}^{i} = \mathrm{g}_{\mathrm{B}}^{j}, \mathrm{Score} = 1; \mathrm{g}_{\mathrm{A}}^{i} \neq \mathrm{g}_{\mathrm{B}}^{j}, \mathrm{Score} = -1\right) \tag{5}$$

the highest value in the matrix $\mathbf{H}_{(n+1) \times (m+1)}$ is chosen as $sw\left(\mathrm{G_A}, \mathrm{G_B}\right)$. The similarity between microbes A and B is adopted

by the same definition as Yamanishi et al. (2008), as follows:

$$S_{\mathrm{b}}(\mathrm{A},\ \mathrm{B}) \ = \ \frac{sw(\mathrm{G_A},\mathrm{G_B})}{\sqrt{sw(\mathrm{G_A},\mathrm{G_A}) \ \times \ sw(\mathrm{G_B},\mathrm{G_B})}} \tag{6}$$

Based on microbe similarities, we can build a microbe similarity network Net_{b}, where nodes are microbes. There are edges between the microbes if these microbes associate with the same drug; the edges are weighted by microbe similarities.

Nearest-Neighbor Aggregator for Drug–Microbe Pair Embeddings

In this section, inspired by the idea of KNN [K-Nearest-Neighbor (Cover and Hart, 1967)], we learn the substructures of the bipartite graph network to obtain the embedding representations of drug–microbe pairs.

First, we construct the drug–microbe bipartite graph network, $\mathrm{G} = (\mathrm{D}, \mathrm{B}, \mathrm{E})$, where $\mathrm{D} = \{\mathrm{d}_1, \mathrm{d}_2, ..., \mathrm{d}_m\}$ represents m drugs, $\mathrm{B} = \{\mathrm{b}_1, \mathrm{b}_2, ..., \mathrm{b}_n\}$ represents n microbes, and each edge (e_{ij}) in edge set E connects two nodes that belong to two different sets of vertexes (i.e., i in D, j in B). We regard the DMAs as bidirectional links. That is, $\mathrm{e}_{\mathrm{d}_x \to \mathrm{b}_p}$ denotes the edge pointing from the drug d_x to the microbe b_p, and $e_{\mathrm{b}_p \to \mathrm{d}_x}$ denotes the edge pointing from the microbe b_p to the drug d_x. Correspondingly, the nearest-neighbor aggregator contains two blocks (**Figure 2**), the microbe-specific drug neighbor aggregator (MsDNA), and the drug-specific microbe neighbor aggregator (DsMNA). Due to their architectures being similar, we only illustrate the MsDNA block in this section.

Microbe-specific drug neighbor aggregator (**Figure 2A**) contains a virtual key dictionary; $\mathcal{N} = \{n_1, n_2, ..., n_m\}$ indicates all the drugs. In the dictionary, we imitate the idea of KNN to learn the substructures of the bipartite graph network, where virtual keys are sorted by their semantic nearest neighbors. In simple terms, n_1 denotes d_x itself, its nearest neighbor is the second key, and the farthest neighbor is the last key. The embedding representation of the edge, which is from drug d_x to microbe b_p, is formulated as follows:

$$\mathbf{a}\left(\mathrm{d}_x, \mathrm{b}_p\right) \ = \ \sum_{i}^{|\mathcal{N}|} S_{\mathrm{d}}\left(\mathrm{d}_x, n_i\right) \mathbf{v}_i \ (if \ n_i \notin \mathcal{N}^p, S_{\mathrm{d}}\left(\mathrm{d}_x, n_i\right) \ = \ 0) \tag{7}$$

where $\mathcal{N}^p \subseteq \mathcal{N}$ is a set of instantiated keywords, and $\mathcal{N}^p$ denotes the neighbors of d_x in the Net_{d}. $S_{\mathrm{d}}\left(\mathrm{d}_x, n_i\right)$ denotes the similarity of d_x and n_i, and $\mathbf{v}_i$ is the corresponding one-hot encoding vector of n_i (i.e., the one-hot encoding has a non-zero value only in the i_{th} element, and all other position elements are zero).

Similarly, DsMNA (**Figure 2B**) makes the single directional embedding representation from b_p to d_x as $\mathbf{a}(\mathrm{b}_p, \mathrm{d}_x)$. Then, the representation of drug–microbe pair could be encoded as

$$\mathbf{e}\left(\mathrm{d}_x, \mathrm{b}_p\right) \ = \ \left[\mathbf{a}(\mathrm{d}_x, \mathrm{b}_p) \parallel \mathbf{a}(\mathrm{b}_p, \mathrm{d}_x)\right] \tag{8}$$

where $\mathbf{e}\left(\mathrm{d}_x, \mathrm{b}_p\right)$ is generated *via* the concatenation of bidirectional embedding, and $\parallel$ is the concatenation operation. All the embedding representations of drug–microbe pairs could stack as a matrix $\mathbf{E}_{k \times g}$, where k is the number of all the drug–microbe pairs and g is the dimension of each embedding. The nearest-neighbor aggregator effectively learns the bipartite graph substructures, and $\mathbf{E}_{k \times g}$ will be input into a feature attention block to select crucial features for achieving a better DMA prediction.

Feature Attention Block

To improve the performance of the prediction, we build the feature attention block (**Figure 3**) for updating the embedding of drug–microbe pairs.

Recall the equation of output feature representation in GAT (Velickovic et al., 2018):

$$\overrightarrow{h_i'} \ = \ \sigma(\sum_{j \in \mathcal{K}_i} \alpha_{ij} \mathbf{W} \overrightarrow{h_j}) \tag{9}$$

where σ is a nonlinear activation function, $\mathcal{K}_i$ is the first-order neighbors of node i (including i), α_{ij} is the coefficients computed by the attention mechanism, and $\mathbf{W}$ is a weight matrix. To make equation (9) easier to understand. We compute the coefficients as:

$$\sum_{j \in \mathcal{K}_i} \alpha_{ij} \ = \ \widetilde{\mathbf{A}} \odot \mathbf{M} \tag{10}$$

where $\widetilde{\mathbf{A}} \ = \ \mathbf{A} + \mathbf{I}$ is the adjacency matrix of the undirected graph G with added self-connections (Kipf and Welling, 2017), $\odot$ is the element-wise product operation, and $\mathbf{M}$ is the attention matrix. Then, the layer-wise propagation rules in GAT can be formulated as:

$$\mathbf{H}^{(l+1)} \ = \ \sigma(\left(\widetilde{\mathbf{A}} \odot \mathbf{M}\right) \mathbf{H}^{(l)} \mathbf{W}^{(l)}) \tag{11}$$

where σ is a nonlinear activation function, and $\mathbf{W}^{(l)}$ is the weight matrix of the l_{th} neural network layer.

Inspired by the conception of the layer-wise propagation rules in GAT, we calculate the augmented representation matrix $\widetilde{\mathbf{F}}_{k \times g}$ by

$$\widetilde{\mathbf{F}}_{k \times g} \ = \ \mathbf{E}_{k \times g} \odot \mathbf{M}_{k \times g} \tag{12}$$

where $\mathbf{E}_{k \times g}$ is the representation matrix of the drug–microbe pairs obtained from the nearest-neighbor aggregator, $\mathbf{M}_{k \times g}$ is an attention matrix of $\mathbf{E}_{k \times g}$, and $\odot$ is the element-wise product operation. We take the representation matrix $\mathbf{E}_{k \times g}$ as a feature matrix $\mathbf{F}$ ($\mathbf{F} \ = \ \{\mathbf{f}_1, \ \mathbf{f}_2, ..., \ \mathbf{f}_g\}$), which is composed of g column vectors ($\mathbf{f}_i (i \ = \ 1, 2, ..., g)$). The feature attention block mainly uses $\mathbf{M}_{k \times g}$ to indicate the importance of features in the $\mathbf{E}_{k \times g}$. Each feature dimension $\mathbf{f}_i$ can be labeled as "selected" or "discarded" in a hard way, or be associated with a probability to be selected in a soft way; we employ DNNs to model the mapping by

$$\mathbf{m}_i \ = \ \mathrm{DNNs}[\mathbf{f}_i] \tag{13}$$

the DNN contains an input layer for each element of the feature dimension $\mathbf{f}_i$ and an output layer with sigmoid as its activation function.

In total, we build $k \ \times \ g$ DNNs to obtain $\mathbf{M}_{k \times g}$. The final feature matrix $\widetilde{\mathbf{F}}_{k \times g}$ of the drug–microbe pairs is obtained after the element-wise product operation of $\mathbf{M}_{k \times g}$ and $\mathbf{E}_{k \times g}$. $\widetilde{\mathbf{F}}_{k \times g}$ is further fed into a predictor to achieve better predictive performance.

TABLE 2 | The performance comparison of DMA prediction.

Method	Database 1			Database 2		
	AUROC	AUPRC	Time (s/epoch)	AUROC	AUPRC	Time (s/epoch)
LAGCN	0.861	0.323	**0.201**	**0.944**	0.721	0.021
NIMCGCN	0.778	0.156	19.076	0.815	0.720	0.721
GCNMDA	0.894	0.042	0.341	0.821	0.177	0.127
NNAN	**0.911**	**0.502**	0.649	0.902	**0.840**	**0.019**

The highest value is indicated in bold, and the next highest value is underlined.

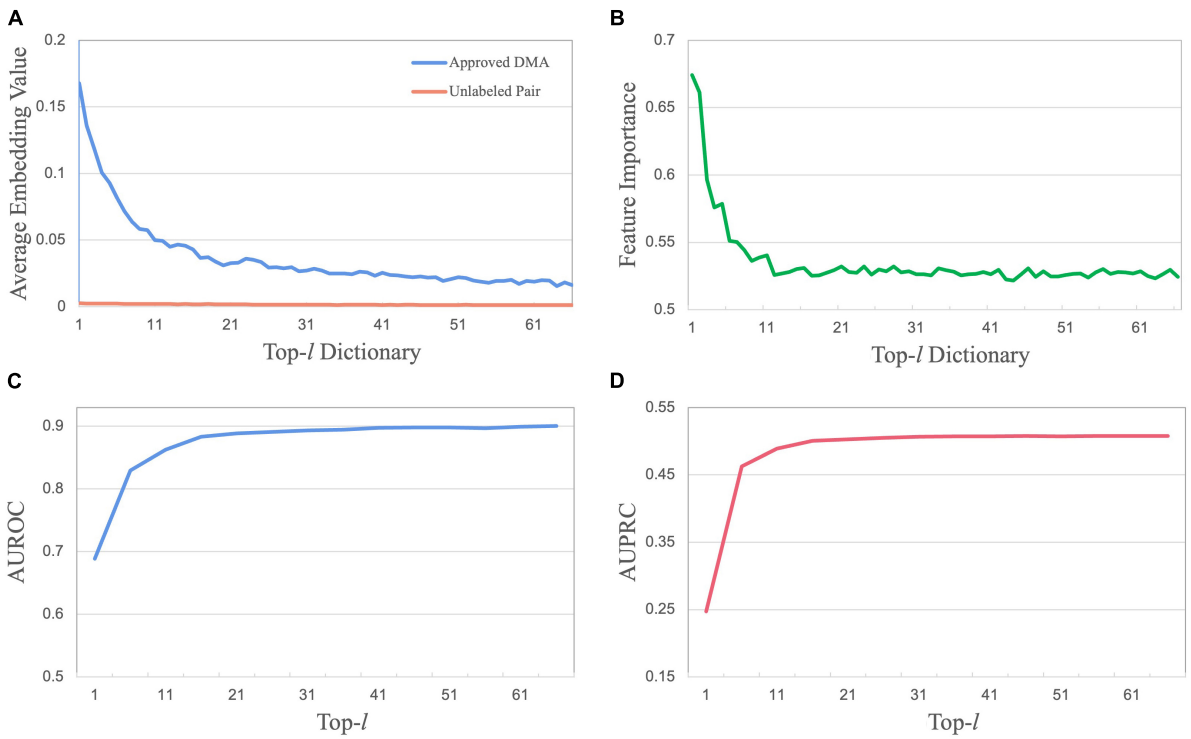


FIGURE 4 | Mensurable clues of embedding features to the association outcome. **(A)** The distribution of embedding features along with the sorted drug neighbor keys. **(B)** The distribution of feature importance along with sorted node neighbor keys. **(C)** The predictive performance with top-*l* features concerning *l* in terms of AUROC. **(D)** The predictive performance in terms of AUPRC.

Predictor

To implement the link prediction in the drug–microbe bipartite graph network, an ordinary DNN is utilized as the binary predictor that contains an input layer for the embedding representation of drug–microbe pairs, a hidden layer with ReLU as its activation function, and the two-neuron output layer with Sigmoid as its activation function. The output layer generates a probability that indicates the association likelihood of the drug and the microbe. The probability is formulated as:

$$P = \varphi(\mathcal{F}\left(\mathrm{ReLU}[\mathcal{F}\left(\widetilde{\mathbf{F}}\right)]\right)) \quad (14)$$

where φ is the sigmoid activation function, and $\mathcal{F}(\cdot)$ is the fully-connected layer.

The entire network of NNAN with the nearest-neighbor aggregator, feature attention weights, and DNN weights can be jointly optimized through the binary cross-entropy loss as follows:

$$loss = \mathbf{Y}\log\left(\mathcal{D}(\widetilde{\mathbf{F}})\right) + (1-\mathbf{Y})\log\left(1-\mathcal{D}(\widetilde{\mathbf{F}})\right) + \lambda\mathcal{R}(\theta) \quad (15)$$

where $\mathbf{Y}$ is the truth labels of drug–microbe pairs, $\mathcal{D}(\cdot)$ is the DNN, θ denotes the weight parameters in the entire network, $\mathcal{R}(\cdot)$ is an L_2-norm, and λ is coefficient of the regularization item.

EXPERIMENTS AND RESULTS

Data

In our experiments, two databases are collected from MDAD (Sun et al., 2018) and Zimmermann et al. (2019a), respectively. The former work MDAD (Sun et al., 2018) investigated 5,505 clinically or experimentally DMAs between 1,388 drugs and

TABLE 3 | The associations among *Staphylococcus aureus* and ten drugs.

Drug name	Rank	Association	Drug name	Rank	Association
Octyl gallate	1	Yes	*Tannic acid*	6	No
Butyl gallate	2	Yes	*Tea tree oil*	7	Yes
Octadecyl gallate	3	Yes	*Pentagalloylglucose*	8	Yes
Ethyl gallate	4	Yes	*4-Ethylcatechol*	9	No
Methyl gallate	5	Yes	*Hamamelitannin*	10	Yes

These ten drugs are ranked in order of their similarity to Hexyl gallate.

TABLE 4 | Top 20 predicted drugs associated with *Bacteroides fragilis*.

Drug name	Evidence	Drug name	Evidence
NATEGLINIDE	PMID: 17253883	*RAMIPRIL*	PMID: 31158845
BENAZEPRIL	PMID: 20445573	*DILTIAZEM*	unconfirmed
VORICONAZOLE	PMID: 18034666	*CLEMASTINE FUMARATE*	PMID: 31158845
FEBUXOSTAT	PMID: 18421623	*NAPROXEN (+)*	PMID: 15058617
LOPERAMIDE	PMID: 18192961	*ERGONOVINE MALEATE*	PMID: 17948937
DIGITOXIN	PMID: 1944247	*DROSPIRENONE*	PMID: 28986954
SOTALOL	PMID: 27836712	*DICYCLOMINE*	unconfirmed
EZETIMIBE	PMID: 15871634	*PROCARBAZINE*	PMID: 1316811
IRBESARTAN	PMID: 12800253	*RIZATRIPTAN BENZOATE*	unconfirmed
SUMATRIPTAN SUCCINATE	PMID:19925626	*SULPIRIDE*	PMID: 31158845

The first column records the top 10 drugs, while the third column records the top 10–20 drugs.

180 microbes. After removing redundant information, these association entries are grouped into Database 1, which contains 999 drugs, 133 microbes, and 1,708 DMAs.

The latter work (Zimmermann et al., 2019a) originally studied how 76 kinds of human gut bacteria metabolize 271 oral drugs, and found that 176 out of 217 drugs are significantly consumed by at least one bacteria strain. These associations are grouped into Database 2, which includes 176 drugs, 76 bacteria, and 4,194 associations (These two databases are shown in **Table 1**).

Comparison

Since there are few existing approaches for predicting DMAs, we compare NNAN with three state-of-the-art methods, which were raised for bipartite link prediction.

- LAGCN (Yu et al., 2021b): A layer attention graph convolutional network for the drug–disease association prediction.
- NIMCGCN (Li et al., 2020): A neural inductive matrix completion with graph convolutional networks for miRNA–disease association prediction.
- GCNMDA (Long et al., 2020a): Predicting human microbe–drug associations *via* graph convolutional network with conditional random field.

To evaluate the performance of these methods, we regard the known DMA pairs as positive samples and unlabeled DMA pairs as negative samples (Peng et al., 2020; Li et al., 2022). We set up a 5-fold cross-validation scenario in which we randomly divide positive samples and negative samples into five groups, respectively. One group of positive samples and one group of negative samples are treated as test samples in turn for each round. The remaining groups are used for training purposes. Our model is trained by Gradient Descent Optimizer (Cauchy, 2009), with batch size 3,000 for 2,000 epochs, the initial learning rate is set to 0.9, and the regularization rate is set to 2e-4. We use AUROC (area under the receiver operating characteristic curve) and AUPRC (area under the precision-recall curve) as metrics to measure the DMA prediction performance. Moreover, we investigate the running time in terms of per epoch.

The comparison (**Table 2**) shows that NNAN obtains the best AUROC value (0.911) and the best AUPRC value (0.502) in Database 1. NNAN attains the next-highest AUROC value (0.902) and the best AUPRC value (0.840) in Database 2. To further present the performance of NNAN, we calculate the running time for one epoch of the baselines and NNAN, respectively. As presented, with the same computing equipment, NNAN takes the third-shortest running time in Database 1 and the shortest running time in Database 2. In general, we can see that NNAN are comparable in terms of AUROC, AUPRC, and computation time. It demonstrates that NNAN is superior to other methods on the databases we collected.

Interpretability of Nearest Neighbor Attention Network

How does the NNAN interpret the hypothesis that "If a drug can associate with a microbe, the other drugs that associate with the microbe are usually the first l nearest neighbors to the drug."

The model has two significant advantages to enhance interpretability. First, each column vector $\mathbf{m_i}$ of $\mathbf{M}_{k \times g}$ indicates

the global importance of each feature dimension $\mathbf{f}_i$. Moreover, the element-wise product between $\mathbf{E}_{k\times g}$ and $\mathbf{M}_{k\times g}$ generates the importance map of embedding features.

We first use the MsDNA in the nearest-neighbor aggregator block to show how the representation of drug–microbe pairs can provide intuitive hints, on which embedding features lead to the association. For the queried drug d_x to the microbe b_p of associated, non-zero cells in the embedding representation of $\mathbf{a}(d_x, \mathrm{b}_p)$ stand for its attention values derived from the drugs commonly linking b_p. Since the keys are sorted in descending order from the drug itself (n_1) to the farthest neighbor (n_m), the positions of non-zero cells are crucial to the final association.

Take Database 1 as an example. By calculating two average embedding vectors for approved DMAs and unlabeled drug–microbe pairs, we obtained a distribution along with the drug key dictionary from n_1 to n_{66}(**Figure 4A**). As illustrated, the significantly high values of embedding features occurring among the first l nearest neighbors reveal that a drug (d_x) associated with a specific microbe (b_p) can always find its top-l nearest neighbors among other drugs that associate with the same microbe. This observation demonstrates that a drug is possibly associated with the microbe if it has more non-zero value cells on the positions of the first l feature dimensions. This phenomenon could be caused by the fact that over 80% of approved drugs are of "follow-on" or "me-too" drugs. Due to high cost and high risk, the design of novel drugs, except for pioneer drugs, always starts from the structures of one or several existing drugs and then slightly modify them until meeting pharmacological needs (DiMasi and Faden, 2011). Analogously, the results of the DsMNA block along the microbe neighbor aggregator keys reveal that a microbe associated with a specific drug usually finds its near neighbors associated with the same drug.

Moreover, we illustrate how the feature attention matrix $\mathbf{M}_{k\times g}$ can provide data-driven hints on which embedding features lead to the association. Since a high-value cell in $\mathbf{M}_{k\times g}$ stands for a crucial feature dimension contributing to determine the association between a queried drug and a microbe, the importance $\mathbf{m}(i,:)$ of each feature $\mathbf{f}_i$ can be measured by the average of value entries in the i_{th} column of $\mathbf{M}_{k\times g}$ (**Figure 4B**). The importance distribution along with the sorted drug neighbor keys illustrates that highly important features are usually located among the first l nearest neighbors. In addition, the predictive performance with top-l features concerning l is investigated (**Figures 4C,D**). The number of top features is tuned in the list {1, 6, 11, 16,..., 66}. As l is increasing to 16, the performance increases sharply in the top-l features. When l keeps increasing, the performance increases slowly, then even decreases at the greater value of l. Again, this illustration demonstrates that the selection of crucial features is significantly better than the set of all features.

In summary, both embedding feature matrix $\mathbf{E}_{k\times g}$, which is generated by the nearest-neighbor aggregator, and its feature attention matrix $\mathbf{M}_{k\times g}$ provide mensurable clues to the association outcome.

To complement the verification of the interpretability of NNAN, we selected one microbe (i.e., *Staphylococcus aureus*, which is a common causative agent of food poisoning) and one drug (i.e., *Hexyl gallate*, which has strong antimalarial activity against *Plasmodium falciparum*) from Database 1, and there was an association between them (de Lima Pimenta et al., 2013). We calculated the similarities between drugs using *Hexyl gallate* as the reference molecule and sorted the drugs in order of their similarity to *Hexyl gallate*. Then, we picked the top 10 drugs and checked whether these drugs were associated with *S. aureus* in Database 1. Finally, we found out that 8 out of the top 10 ranked drugs for *Hexyl gallate* are associated with *S. aureus* (**Table 3**).

From **Table 3**, it is clear that a drug tends to associate with a microbe if it finds its top-l near neighbors associate with the same microbe. Moreover, the higher the ranks of its top-l near neighbors are, the more possible it is to associate with the microbe. This conclusion would be helpful to screen drug-like molecules.

CASE STUDY OF NOVEL PREDICTION

To further confirm the effectiveness of NNAN, we apply our model on one microbe (i.e., *Bacteroides fragilis*) in Database 2 as a case study. Bacteroides are the major human colonic commensal microbes (Kuwahara et al., 2004). Although *B. fragilis* is rare in comparison to other Bacteroides species, it is the most prevalent clinical isolation of the genus (Salyers, 1984). Thus, we select *B. fragilis* for the case study experiment.

Nearest neighbor attention network predicts potential associations between drugs and *B. fragilis* by scoring drug–microbe pairs (probability). The higher the score, the more likely the association between the drugs and *B. fragilis* exists. In the case study, we verified whether NNAN could find out potential linkages between *B. fragilis* and drugs. According to the ranking of potential DMAs, we validated the top 10, 20, and 50 predicted candidate drugs by a literature search. Eventually, the validation indicates that 10, 17, and 38 out of the top 10, 20, and 50 predicted drugs associated with *B. fragilis* were found by previously published literature. For example, 85% out of the top 20 predicted candidate drugs for *B. fragilis* are validated (**Table 4**); more details can be found in the **Supplementary Material**. These results of prediction demonstrate the ability of NNAN for predicting potential DMAs in practice.

CONCLUSION

This work has introduced NNAN, a deep learning-based bipartite graph network model to predict potential associations between drugs and microbes. NNAN calculates drug similarities using the weights of feature substructures. It provides an embedding representation based on the near neighbor aggregation for drug–microbe pairs, to enhance the explanation of DMAs. In addition, the model provides a crucial feature selection attention matrix for achieving more accurate predictions. These three components of NNAN jointly reveal that a drug associated with a specific

microbe can always find its top-*l* near neighbors among other drugs that associate with the same microbe. Moreover, they uncover that the higher the ranks of its top-*l* near neighbors are, the more possible it is to associate with the microbe. Under both a cross-validation setting and a realistic potential linkage discovery setting, the empirical comparison of the proposed framework with three state-of-the-art baselines demonstrates that NNAN has significant competitive performance in predicting DMA. In addition, the framework of our model can also be evaluated in more similar biological issues (e.g., miRNA–disease, drug–target, and compound–protein associations prediction). Furthermore, there is still room to improve the model. We can set new experimental scenarios, which identify the DMAs for new drugs or new microbes, and can also integrate more biological databases to enrich the information of DMAs to improve the predictive ability.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

J-YS and HY designed and supervised the study. BZ engaged in study design, drafted the manuscript, performed experiments, and analyzed data. YX coded and implemented the model, performed experiments. PZ assisted with performing experiments. S-MY assisted with supervising the study. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by Shaanxi Provincial Key R&D Program, China (No. 2020KW-063, PI: J-YS) and National Natural Science Foundation of China (No. 61872297, PI: J-YS).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.846915/full#supplementary-material

REFERENCES

Aagaard, K., Petrosino, J., Keitel, W., Watson, M., Katancik, J., Garcia, N., et al. (2013). The human microbiome project strategy for comprehensive sampling of the human microbiome and why it matters. *FASEB J.* 27, 1012–1022. doi: 10.1096/fj.12-220806

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410.

Bang, S., Ho Jhee, J., and Shin, H. (2021). Polypharmacy side effect prediction with enhanced interpretability based on graph feature attention network. *Bioinformatics*. 37, 2955–2962 doi: 10.1093/bioinformatics/btab174

Cauchy, A.-L. (2009). *ANALYSE MATHMATIQUE. MÈthodc gÈnÈrale pour la rÈsolution des SystËmes d'Èquations SimultanÈes*. Cambridge: Cambridge University Press.

Cover, T. M., and Hart, P. E. (1967). Nearest neighbor pattern classification. *IEEE Trans. Inf. Theory* 13, 21–27.

de Lima Pimenta, A., Chiaradia-Delatorre, L. D., Mascarello, A., de Oliveira, K. A., Leal, P. C., Yunes, R. A., et al. (2013). Synthetic organic compounds with potential for bacterial biofilm inhibition, a path for the identification of compounds interfering with quorum sensing. *Int. J. Antimicrob. Agents* 42, 519–523. doi: 10.1016/j.ijantimicag.2013.07.006

DiMasi, J. A., and Faden, L. B. (2011). Competitiveness in follow-on drug R&D: a race or imitation? *Nat. Rev. Drug Discov.* 10, 23–27. doi: 10.1038/nrd3296

Haiser, H. J., Gootenberg, D. B., Chatman, K., Sirasani, G., Balskus, E. P., and Turnbaugh, P. J. (2013). Predicting and manipulating cardiac drug inactivation by the human gut bacterium *Eggerthella lenta*. *Science* 341, 295–298. doi: 10.1126/science.1235872

He, B. S., Peng, L. H., and Li, Z. (2018). Human microbe-disease association prediction with graph regularized non-negative matrix factorization. *Front. Microbiol.* 9:2560. doi: 10.3389/fmicb.2018.02560

Ioffe, S. (2010). "Improved consistent sampling, weighted minhash and L1 sketching," in *Proceedings of the 2010 IEEE International Conference on Data Mining*, Sydney, 246–255.

Jaacks, L. M., Vandevijvere, S., Pan, A., McGowan, C. J., Wallace, C., Imamura, F., et al. (2019). The obesity transition: stages of the global epidemic. *Lancet Diabetes Endocrinol.* 7, 231–240. doi: 10.1016/S2213-8587(19)30026-9

Kashyap, P. C., Chia, N., Nelson, H., Segal, E., and Elinav, E. (2017). Microbiome at the Frontier of personalized medicine. *Mayo Clin. Proc.* 92, 1855–1864. doi: 10.1016/j.mayocp.2017.10.004

Katz, L. (1953). A new status index derived from sociometric analysis. *Psychometrika* 18, 39–43. doi: 10.1007/bf02289026

Khalili, H., Godwin, A., Choi, J. W., Lever, R., and Brocchini, S. (2012). Comparative binding of disulfide-bridged PEG-Fabs. *Bioconjug. Chem.* 23, 2262–2277. doi: 10.1021/bc300372r

Kipf, T. N., and Welling, M. (2017). Semi-supervised classification with graph convolutional networks. *arXiv* [Preprint] doi: 10.48550/arXiv.1609.02907

Kuwahara, T., Yamashita, A., Hirakawa, H., Nakayama, H., Toh, H., Okada, N., et al. (2004). Genomic analysis of *Bacteroides fragilis* reveals extensive DNA inversions regulating cell surface adaptation. *Proc. Natl. Acad. Sci. U S A.* 101, 14919–14924. doi: 10.1073/pnas.0404172101

Landrum. (2010). *RDKit: Open-Source Cheminformatics. Release 2014.03.1.*

Li, F., Dong, S., Leier, A., Han, M., Xu, J., et al. (2022). Positive-unlabeled learning in bioinformatics and computational biology: a brief review. *Brief. Bioinform.* 23:bbab461. doi: 10.1093/bib/bbab461

Li, J., Zhang, S., Liu, T., Ning, C., Zhang, Z., and Zhou, W. (2020). Neural inductive matrix completion with graph convolutional networks for miRNA-disease association prediction. *Bioinformatics* 36, 2538–2546. doi: 10.1093/bioinformatics/btz965

Lihong, P., Wang, C., Tian, X., Zhou, L., and Li, K. (2021). Finding lncRNA-protein interactions based on deep learning with dual-net neural architecture. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 14:1. doi: 10.1109/TCBB.2021.3116232

Long, Y., and Luo, J. (2020). Association mining to identify microbe drug interactions based on heterogeneous network embedding representation. *IEEE J. Biomed. Health Informatics* 25, 266–275. doi: 10.1109/JBHI.2020.2998906

Long, Y., Wu, M., Kwoh, C. K., Luo, J., and Li, X. (2020a). Predicting human microbe-drug associations via graph convolutional network with conditional random field. *Bioinformatics* 36, 4918–4927. doi: 10.1093/bioinformatics/btaa598

Long, Y., Wu, M., Liu, Y., Kwoh, C. K., Luo, J., and Li, X. (2020b). Ensembling graph attention networks for human microbe-drug association prediction. *Bioinformatics* 36(Suppl_2), i779–i786. doi: 10.1093/bioinformatics/btaa891

Lynch, S. V., and Pedersen, O. (2016). The human intestinal microbiome in health and disease. *N. Engl. J. Med.* 375, 2369–2379.

Peng, L., Shen, L., Liao, L., Liu, G., and Zhou, L. (2020). RNMFMDA: a microbe-disease association identification method based on reliable negative sample selection and logistic matrix factorization with neighborhood regularization. *Front. Microbiol.* 11:592430. doi: 10.3389/fmicb.2020.592430

Peng, L. H., Yin, J., Zhou, L., Liu, M. X., and Zhao, Y. (2018). Human microbe-disease association prediction based on adaptive boosting. *Front. Microbiol.* 9:2440. doi: 10.3389/fmicb.2018.02440

Riniker, S., and Landrum, G. A. (2013). Similarity maps – a visualization strategy for molecular fingerprints and machine-learning methods. *J. Cheminform.* 5:43. doi: 10.1186/1758-2946-5-43

Rogers, D., and Hahn, M. (2010). Extended-connectivity fingerprints. *J. Chem. Inf. Model.* 50, 742–754. doi: 10.1021/ci100050t

Rogers, D. J., and Tanimoto, T. T. A. (1960). Computer program for classifying plants. *Science* 132, 1115–1118. doi: 10.1126/science.132.3434.1115

Salyers, A. A. (1984). *Bacteroides* of the human lower intestinal tract. *Annu. Rev. Microbiol.* 38, 293–313. doi: 10.1146/annurev.mi.38.100184.001453

Schwabe, R. F., and Jobin, C. (2013). The microbiome and cancer. *Nat. Rev. Cancer.* 13, 800–812.

Smith, T. F., and Waterman, M. S. (1981). Identification of common molecular subsequences. *J. Mol. Biol.* 147, 195–197. doi: 10.1016/0022-2836(81)90087-5

Sousa, T., Yadav, V., Zann, V., Borde, A., Abrahamsson, B., and Basit, A. W. (2014). On the colonic bacterial metabolism of azo-bonded prodrugsof 5-aminosalicylic acid. *J. Pharm. Sci.* 103, 3171–3175. doi: 10.1002/jps.24103

Sun, Y. Z., Zhang, D. H., Cai, S. B., Ming, Z., Li, J. Q., and Chen, X. M. D. A. D. (2018). A special resource for microbe-drug associations. *Front. Cell. Infect. Microbiol.* 8:424. doi: 10.3389/fcimb.2018.00424

Turnbaugh, P. J., Ley, R. E., Hamady, M., Fraser-Liggett, C. M., Knight, R., and Gordon, J. I. (2007). The human microbiome project. *Nature* 449, 804–810.

Velickovic, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., and Bengio, Y. (2018). Graph attention networks. *arXiv* [Preprint]. doi: 10.48550/arXiv.1710.10903

Yamanishi, Y., Araki, M., Gutteridge, A., Honda, W., and Kanehisa, M. (2008). Prediction of drug-target interaction networks from the integration of chemical and genomic spaces. *Bioinformatics* 24, i232–i240.

Younossi, Z. M., Koenig, A. B., Abdelatif, D., Fazel, Y., Henry, L., and Wymer, M. (2016). Global epidemiology of nonalcoholic fatty liver disease-meta-analytic assessment of prevalence, incidence, and outcomes. *Hepatology* 64, 73–84. doi: 10.1002/hep.28431

Yu, H., Dong, W., and Shi, J. Y. (2021a). RANEDDI: Relation-aware network embedding for prediction of drug-drug interactions. *Inf. Sci.* 582, 167–180.

Yu, Z., Huang, F., Zhao, X., Xiao, W., and Zhang, W. (2021b). Predicting drug-disease associations through layer attention graph convolutional network. *Brief. Bioinform.* 22:bbaa243. doi: 10.1093/bib/bbaa243

Zhang, L., Yang, P., Feng, H., Zhao, Q., and Liu, H. (2021). Using network distance analysis to predict lncRNA-miRNA Interactions. *Interdiscip. Sci.* 13, 535–545. doi: 10.1007/s12539-021-00458-z

Zhang, Z., Guan, J., and Zhou, S. (2021). FraGAT: a fragment-oriented multi-scale graph attention model for molecular property prediction. *Bioinformatics* 37, 2981–2987. doi: 10.1093/bioinformatics/btab195

Zheng, Y., Ley, S. H., and Hu, F. B. (2018). Global aetiology and epidemiology of type 2 diabetes mellitus and its complications. *Nat. Rev. Endocrinol.* 14, 88–98. doi: 10.1038/nrendo.2017.151

Zhou, L., Wang, Z., Tian, X., and Peng, L. (2021). LPI-deepGBDT: a multiple-layer deep framework based on gradient boosting decision trees for lncRNA-protein interaction identification. *BMC Bioinform.* 22:479. doi: 10.1186/s12859-021-04399-8

Zhu, L., Duan, G., Yan, C., and Wang, J. (2019). "Prediction of microbe-drug associations based on KATZ measure," in *Proceedings of the 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, San Diego, CA.

Zimmermann, M., Zimmermann-Kogadeeva, M., Wegmann, R., and Goodman, A. L. (2019b). Separating host and microbiome contributions to drug pharmacokinetics and toxicity. *Science* 363:eaat9931. doi: 10.1126/science.aat9931

Zimmermann, M., Zimmermann-Kogadeeva, M., Wegmann, R., and Goodman, A. L. (2019a). Mapping human microbiome drug metabolism by gut bacteria and their genes. *Nature* 570, 462–467. doi: 10.1038/s41586-019-1291-3

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.

frontiers | Frontiers in Microbiology

CORRECTION
published: 30 May 2022
doi: 10.3389/fmicb.2022.944952

Corrigendum: NNAN: Nearest Neighbor Attention Network to Predict Drug–Microbe Associations

Bei Zhu[1†], *Yi Xu*[1†], *Pengcheng Zhao*[1], *Siu-Ming Yiu*[2], *Hui Yu*[3*] and *Jian-Yu Shi*[1*]

[1] School of Life Sciences, Northwestern Polytechnical University, Xi'an, China, [2] Department of Computer Science, The University of Hong Kong, Hong Kong, China, [3] School of Computer Science, Northwestern Polytechnical University, Xi'an, China

Keywords: deep learning, bipartite graph network, link prediction, drug-microbe association, attention matrix

OPEN ACCESS

Approved by:
Frontiers Editorial Office, Frontiers Media SA, Switzerland

***Correspondence:**
Hui Yu
huiyu@nwpu.edu.cn
Jian-Yu Shi
jianyushi@nwpu.edu.cn

[†] These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Systems Microbiology, a section of the journal Frontiers in Microbiology

Received: *16 May 2022*
Accepted: *17 May 2022*
Published: *30 May 2022*

Citation:
Zhu B, Xu Y, Zhao P, Yiu S-M, Yu H and Shi J-Y (2022) Corrigendum: NNAN: Nearest Neighbor Attention Network to Predict Drug–Microbe Associations. Front. Microbiol. 13:944952. doi: 10.3389/fmicb.2022.944952

A Corrigendum on

NNAN: Nearest Neighbor Attention Network to Predict Drug–Microbe Associations
by Zhu, B., Xu, Y., Zhao, P., Yiu, S.-M., Yu, H., and Shi, J.-Y. (2022). Front. Microbiol. 13:846915. doi: 10.3389/fmicb.2022.846915

In the original article, there was an error in "affiliation 1" as published. Instead of "School of Life Sciences, Northwestern Polytechnic University, Xi'an, China," it should be "School of Life Sciences, Northwestern Polytechnical University, Xi'an, China."

In the original article, there was an error in "affiliation 3" as published. Instead of "School of Computer Science, Northwestern Polytechnic University, Xi'an, China," it should be "School of Computer Science, Northwestern Polytechnical University, Xi'an, China."

The authors apologize for this error and state that this does not change the scientific conclusions of the article in any way. The original article has been updated.

frontiers | Frontiers in Microbiology

ORIGINAL RESEARCH
published: 25 April 2022
doi: 10.3389/fmicb.2022.818358

Outline, Divergence Times, and Phylogenetic Analyses of Trechisporales (Agaricomycetes, Basidiomycota)

*Zhan-Bo Liu[1], Ying-Da Wu[1,2], Heng Zhao[1], Ya-Ping Lian[1], Ya-Rong Wang[1], Chao-Ge Wang[1], Wei-Lin Mao[1] and Yuan Yuan[1]**

[1] School of Ecology and Nature Conservation, Beijing Forestry University, Beijing, China, [2] Key Laboratory of Forest and Grassland Fire Risk Prevention, Ministry of Emergency Management, China Fire and Rescue Institute, Beijing, China

Phylogenetic analyses inferred from the nuc rDNA ITS1-5.8S-ITS2 (ITS) data set and the combined 2-locus data set [5.8S + nuc 28S rDNA (nLSU)] of taxa of Trechisporales around the world show that *Sistotremastrum* family forms a monophyletic lineage within Trechisporales. Bayesian evolutionary and divergence time analyses on two data sets of 5.8S and nLSU sequences indicate an ancient divergence of *Sistotremastrum* family from Hydnodontaceae during the Triassic period (224.25 Mya). *Sistotremastrum* family is characterized by resupinate and thin basidiomata, smooth, verruculose, or odontoid-semiporoid hymenophore, a monomitic hyphal structure, and generative hyphae bearing clamp connections, the presence of cystidia and hyphidia in some species, thin-walled, smooth, inamyloid, and acyanophilous basidiospores. In addition, four new species, namely, *Trechispora dentata*, *Trechispora dimitiella*, *Trechispora fragilis*, and *Trechispora laevispora*, are described and illustrated. In addition, three new combinations, namely, *Brevicellicium daweishanense*, *Brevicellicium xanthum*, and *Sertulicium limonadense*, are also proposed.

Keywords: Hydnodontaceae, phylogenetic analysis, *Trechispora*, taxonomy, wood-rotting fungi

OPEN ACCESS

Edited by:
Qi Zhao,
University of Science and Technology Liaoning, China

Reviewed by:
Angelina De Meiras-Ottoni, Federal University of Pernambuco, Brazil
Renata Dos Santos Chikowski, Federal University of Pernambuco, Brazil
Alexander Ordynets, University of Kassel, Germany

***Correspondence:**
Yuan Yuan
yuanyuan1018@bjfu.edu.cn

Specialty section:
This article was submitted to Systems Microbiology, a section of the journal Frontiers in Microbiology

Received: *19 November 2021*
Accepted: *28 February 2022*
Published: *25 April 2022*

Citation:
Liu Z-B, Wu Y-D, Zhao H, Lian Y-P, Wang Y-R, Wang C-G, Mao W-L and Yuan Y (2022) Outline, Divergence Times, and Phylogenetic Analyses of Trechisporales (Agaricomycetes, Basidiomycota). Front. Microbiol. 13:818358. doi: 10.3389/fmicb.2022.818358

INTRODUCTION

Trechisporales K.H. Larss. was established by Hibbett et al. (2007). Most species in this order are corticioid fungi with smooth, grandinioid, odontioid, or hydnoid hymenophores, and others are polypores. All species have a monomitic or dimitic hyphal system with generative hyphae bearing clamp connections, and many species have rhizomorphs (mycelial cords) (Larsson, 2007).

At present, there is only an acknowledged and a named family belonging to Trechisporales, i.e., Hydnodontaceae Jülich. Hydnodontaceae contains 11 genera now, namely, *Brevicellicium* K.H. Larss. and Hjortstam, *Dextrinocystis* Gilb. and M. Blackw., *Fibrodontia* Parmasto, *Pteridomyces* Jülich, *Luellia* K.H. Larss. and Hjortstam, *Porpomyces* Jülich, *Scytinopogon* Singer, *Subulicystidium* Parmasto, *Suillosporium* Pouzar, *Trechispora* P. Karst., and *Tubulicium* Oberw (Larsson, 2007; Spirin et al., 2021).

Trechispora is the genus type of Trechisporales and Hydnodontaceae. It is the largest genus in this order, with more than 50 accepted species (Meiras-Ottoni et al., 2021; Zhao and Zhao, 2021). Identification keys for *Trechispora* species recorded in China and Brazil have been provided by

some fungal taxonomists (Chikowski et al., 2020; Meiras-Ottoni et al., 2021; Zong et al., 2021). *Trechispora* was typified with *Trechispora onusta* P. Karst. [= *Trechispora hymenocystis* (Berk. and Broome) K.H. Larss.] (Karsten, 1890). It is characterized by the resupinate basidiomata (a few species have stipitate, flabellate, and effused–reflexed basidiomata) with smooth grandinioid, odontioid, hydnoid, or poroid hymenophores, a monomitic or dimitic hyphal structure with clamped generative hyphae and smooth to verrucose or aculeate basidiospores (Larsson, 1992; Larsson et al., 2004). Most species in *Trechispora* are soil-dwelling (Larsson et al., 2004). One remarkable character is the presence of ampullate septa on the subicular and especially on some hyphae of the mycelial cords. Above all, ampullate septa are only known from *Scytinopogon*, *Trechispora,* and *Porpomyces mucidus* (Pers.) Jülich within Trechisporales (Furtado et al., 2021; Meiras-Ottoni et al., 2021).

Larsson (2007) used the term "*Sistotremastrum* family" for the first time to accommodate *Sistotremastrum suecicum* Litsch. ex J. Erikss. and *Sistotremastrum niveocremeum* [= *Sertulicium niveocremeum* (Höhn. and Litsch.) Spirin and K.H. Larss.]. Since then, "*Sistotremastrum* family" has been adopted by some taxonomists (Telleria et al., 2013; Liu et al., 2019). In this work, the phylogeny of Trechisporales is carried out based on combined 5.8S + nLSU sequences. In addition, Bayesian evolutionary and divergence time analyses are also carried out to indicate the divergence time of Trechisporales, Hydnodontaceae, and *Sistotremastrum* family. We outline the *Sistotremastrum* family and discuss the difference between Hydnodontaceae and *Sistotremastrum* family.

During investigations on the diversity of wood-rotting fungi, seven resupinate specimens were collected from China and Malaysia. Their morphology corresponds to the concept of *Trechispora*. To confirm their affinity, phylogenetic analyses based on the ITS sequences are carried out. Both morphological characteristics and molecular evidence demonstrate that these seven resupinate specimens represent the four new species of Trechispora.

In addition, we downloaded the type sequences of *Trechispora daweishanensis* C.L. Zhao, *Trechispora xantha* C.L. Zhao, and *Sistotremastrum limonadense* G. Gruhn and P. Alvarado from GenBank. We also studied the type specimens of *T. daweishanensis* and *T. xantha*. In conclusion, *T. daweishanensis* and *T. xantha* were transferred to *Brevicellicium*, while *S. limonadense* was transferred to *Sertulicium*.

MATERIALS AND METHODS

Morphological Studies

Macro-morphological descriptions are based on field notes and dry herbarium specimens. Microscopic structures are photographed using a Nikon Digital Sight DS-L3 (Japan) or Leica ICC50 HD (Japan) camera. Microscopic measurements are made from slide preparations of dry tissues stained with 1% Phloxine B ($C_{20}H_4Br_4Cl_2K_2O_5$) (Fan et al., 2021). We also use other reagents, such as Cotton Blue and Melzer's reagent following Dai's (2010) study. Spore measurements include both with ornamentation and without ornamentation. The following abbreviations are used: KOH = 5% potassium hydroxide; CB = Cotton Blue; CB(+) = weakly cyanophilous; CB− = acyanophilous in Cotton Blue; IKI = Melzer's reagent; IKI− = neither amyloid nor dextrinoid in Melzer's reagent; L = mean spore length (arithmetic average of all spores including ornamentation); W = mean spore width (arithmetic average of all spores including ornamentation); Q = a variation in the L/W ratios between the specimens studied; L' = mean spore length (arithmetic average of all spores excluding ornamentation); W' = mean spore width (arithmetic average of all spores excluding ornamentation); Q' = a variation in the L'/W' ratios between the specimens studied; n (a/b) = the number of spores (a) measured from a given number of specimens (b). When presenting spore size variation, 5% of measurements are excluded from each end of the range and these values are given in parentheses. Special color terms follow Petersen (1996). Herbarium abbreviations follow Thiers (2018). The studied specimens are deposited at the herbarium of the Institute of Microbiology, Beijing Forestry University (BJFC), and the herbarium of Southwest Forestry University (SWFC).

DNA Extraction, Polymerase Chain Reaction Amplification, and Sequencing

Total genomic DNA from the dried specimens is extracted by a CTAB rapid plant genome extraction kit (Aidlab Biotechnologies Company Limited, Beijing, China) according to the manufacturer's instructions with some modifications (Liu and Yuan, 2020; Du et al., 2021). The ITS regions are amplified with the primers ITS4 and ITS5 (White et al., 1990). The nLSU regions are amplified with the primers LR0R and LR7 (Vilgalys and Hester, 1990).

The polymerase chain reaction (PCR) procedure for ITS is as follows: initial denaturation at 95°C for 3 min, followed by 35 cycles at 94°C for 40 s, 58°C for 45 s, and 72°C for 1 min, and a final extension of 72°C for 10 min. The PCR procedure for nLSU was as follows: initial denaturation at 94°C for 1 min, followed by 35 cycles at 94°C for 30 s, 48°C for 1 min, and 72°C for 1.5 min, and a final extension of 72°C for 10 min (Zhao et al., 2015; Liu and Dai, 2021). The PCR products are purified and sequenced in the Beijing Genomics Institute, China, with the same primers used in the PCR reactions.

Phylogenetic Analyses

Two combined matrices, an ITS1-5.8S-ITS2 (ITS) data set and a two-gene data set (5.8S + nLSU), are used for phylogenetic analyses. Phylogenetic analyses are performed with maximum likelihood (ML), maximum parsimony (MP), and Bayesian inference (BI) methods in the ITS data set. Phylogenetic analyses are performed with ML and BI methods in the combined two-gene data set (5.8S + nLSU). Species and strain sequences are adopted partly from 28S- and ITS-based tree topologies established by Meiras-Ottoni et al. (2021) and Spirin et al. (2021). New sequences generated in this study, along with reference sequences retrieved from GenBank (**Table 1**), are

TABLE 1 | Information of taxa used in phylogenetic analyses.

Species	Collector ID (herbarium ID)	GenBank accession no.	
		ITS	nLSU
Auricularia sp.	PBM 2295	DQ200918	AY634277
Brevicellicium atlanticum	LISU 178566 (holotype)	NR_119820	HE963774
Brevicellicium atlanticum	LISU 178590	HE963775	HE963776
Brevicellicium daweishanense	**CLZhao 18255 (SWFC)**	**MW302338**	**MW293867**
Brevicellicium daweishanense	**CLZhao 17860 (SWFC, holotype)**	**MW302337**	**MW293866**
Brevicellicium exile	MA-Fungi 26554 (holotype)	HE963777	HE963778
Brevicellicium olivascens	KHL 8571 (GB)	HE963792	HE963793
Brevicellicium olivascens	MA-Fungi 23496	HE963787	HE963788
Brevicellicium xanthum	**CLZhao 17781 (SWFC)**	**MW302340**	**MW293869**
Brevicellicium xanthum	**CLZhao 2632 (SWFC, holotype)**	**MW302339**	**MW293868**
Dextrinocystis calamicola	He 5700 (BJFC)	MK204534	MK204547
Dextrinocystis calamicola	He 5693 (BJFC)	MK204533	MK204546
Exidia recisa	EL 15-98 (GB)	AF347112	AF347112
Exidiopsis calcea	MW 331	AF291280	AF291326
Fibrodontia alba	TNM F24944 (holotype)	KC928274	KC928275
Fibrodontia gossypina	AFTOL-ID 599	DQ249274	AY646100
Hyphodontia floccosa	Berglund 150-02 (GB)	DQ873618	DQ873617
Hyphodontia subalutacea	GEL2196 (KAS)	DQ340341	DQ340362
Porpomyces mucidus	Dai 12692 (BJFC)	KT157833	KT157838
Porpomyces submucidus	Cui 5183 (BJFC)	KU509521	KT152145
Pteridomyces galzinii	GB0150230	LR694188	LR694210
Pteridomyces galzinii	Bernicchia 8122 (GB)	MN937559	MN937559
Scytinopogon angulisporus	TFB13611	–	JQ684661
Scytinopogon chartaceum	FLOR56185	MK458775	–
Scytinopogon pallescens	He 5192 (BJFC)	–	MK204553
Sertulicium chilense	MA-Fungi 86368 (holotype)	HG315521	–
Sertulicium granuliferum	He 3338	MK204552	MK204540
Sertulicium jacksonii	Spirin 10425 (H)	MN987943	MN987943
Sertulicium lateclavigerum	LY 13467	MG913225	–
Sertulicium limonadense	**LIP 0001683 (holotype)**	**MT180981**	**MT180978**
Sertulicium limonadense	**He 6276 (BJFC)**	**OK298489***	**OK298947***
Sertulicium niveocremeum	KHL13727 (GB)	MN937563	MN937563
Sertulicium vernale	Soderholm 3886 (H, holotype)	MT002311	MT664174
Sistotremastrum aculeatum	Miettinen 10380.1 (H)	MN991176	MW045423
Sistotremastrum aculeatum	Cui 8401 (BJFC)	KX081133	KX081184
Sistotremastrum aculeocrepitans	KHL 16097 (URM)	MN937564	MN937564
Sistotremastrum confusum	KHL 16004 (URM)	MN937567	MN937567
Sistotremastrum denticulatum	Motato-Vásquez 894 (SP, holotype)	MN954694	MW045424
Sistotremastrum fibrillosum	LIP 0001413 (holotype)	NR_161047	NG_075239
Sistotremastrum fibrillosum s. l.	GUY13-119 (GG)	MG913224	MG913210
Sistotremastrum fibrillosum s. l.	KHL 16988 (MG)	MN937568	MN937568
Sistotremastrum geminum	Miettinen 14333 (MAN, holotype)	MN937568	MN937568
Sistotremastrum induratum	Spirin 8598 (H, holotype)	MT002324	MT664173
Sistotremastrum mendax	KHL 12022 (O, holotype)	MN937570	MN937570
Sistotremastrum rigidum	Motato-Vásquez 833 (SP, holotype)	MN954693	MW045435
Sistotremastrum suecicum	Kunttu 5959 (H)	MT075859	MT002335
Sistotremastrum suecicum	Miettinen 14550.1 (H)	MT075860	MT002336
Sistotremastrum suecicum	KHL 11849 (GB)	MN937571	MN937571
Sistotremastrum vigilans	Fonneland 2011-78 (O, holotype)	MN937572	MN937572
Sistotremastrum vigilans	Spirin 8778 (H)	MN991182	MN991182
Subulicystidium tropicum	He 3968 (BJFC)	MK204531	MK204544
Suillosporium cystidiatum	Spirin 3830 (H)	MN937573	MN937573
Trechispora alnicola	AFTOL-ID 665	DQ411529	AY635768
Trechispora araneosa	KHL8570 (GB)	AF347084	AF347084
Trechispora bambusicola	CLZhao 3302 (SWFC)	MW544021	MW520171
Trechispora bispora	CBS 142.63 (holotype)	MH858241	MH869842
Trechispora cohaerens	TU 110332	UDB008249	–
Trechispora cohaerens	TU 115568	UDB016421	–
Trechispora confinis	KHL11064 (GB)	AF347081	AF347081

(Continued)

TABLE 1 | (Continued)

Species	Collector ID (herbarium ID)	GenBank accession no.	
		ITS	nLSU
Trechispora copiosa	AMO456	MN701019	MN687976
Trechispora copiosa	AMO422 (holotype)	MN701013	MN687971
Trechispora cyatheae	FR-0219442	UDB024014	UDB024014
Trechispora cyatheae	FR-0219443 (holotype)	UDB024015	UDB024015
Trechispora dentata	**Dai 22565 (BJFC)**	**OK298491***	**OM049408***
Trechispora dimitiella	**Dai 21931 (BJFC)**	**OK298492***	**OK298948***
Trechispora dimitiella	**Dai 21181 (BJFC)**	**OK298493***	**OK298949***
Trechispora echinocristallina	FR-0219445 (holotype)	UDB024018	UDB024019
Trechispora echinocristallina	FR-0219448	UDB024022	–
Trechispora echinospora	MA-Fungi 82485 (holotype)	JX392845	JX392846
Trechispora farinacea	KHL 8793 (GB)	AF347089	AF347089
Trechispora farinacea	KHL 8451 (GB)	AF347082	AF347082
Trechispora fimbriata	CLZhao 7969 (SWFC)	MW544024	MW520174
Trechispora fimbriata	CLZhao 4154 (SWFC, holotype)	MW544023	MW520173
Trechispora fissurata	CLZhao 4571 (SWFC, holotype)	MW544027	MW520177
Trechispora fissurata	CLZhao 995 (SWFC)	MW544026	MW520176
Trechispora fragilis	**Dai 20535 (BJFC)**	**OK298494***	**OK298950***
Trechispora gelatinosa	AMO1139 (holotype)	MN701021	MN687978
Trechispora gelatinosa	AMO824	MN701020	MN687977
Trechispora havencampii	SFSU DED8300 (holotype)	NR_154418	NG_059993
Trechispora hymenocystis	TL 11112 (holotype)	UDB000778	UDB000778
Trechispora hymenocystis	KHL 8795 (GB)	AF347090	AF347090
Trechispora incisa	GB0090648	KU747095	KU747087
Trechispora incisa	GB0090521	KU747093	–
Trechispora kavinioides	KGN 981002 (GB)	AF347086	AF347086
Trechispora laevispora	**Dai 21655 (BJFC)**	**OK298495***	**OM108710**
Trechispora minispora	MEXU 28300 (holotype)	MK328886	MK328894
Trechispora minispora	MEXU 28301	MK328886	MK328895
Trechispora mollis	URM 85884 (holotype)	MK514945	MH280003
Trechispora mollusca	DLL2011-186 (CFMR)	KJ140681	–
Trechispora mollusca	DLL2010-077 (CFMR)	JQ673209	–
Trechispora nivea	GB0102694	KU747096	AY586720
Trechispora nivea	MA-Fungi 74044	JX392832	JX392833
Trechispora papillosa	AMO713	MN701022	MN687979
Trechispora papillosa	AMO795 (holotype)	MN701023	MN687981
Trechispora regularis	KHL10881 (GB)	AF347087	AF347087
Trechispora rigida	URM 85754	MT406381	MH279999
Trechispora sp.	AMO799	MN701008	MN687969
Trechispora sp.	AMO440	MN701006	MN687967
Trechispora sp.	KHL16968 (O)	MH290763	MH290763
***Trechispora* sp.**	**Dai 22173 (BJFC)**	**OK298496***	**OK298951***
***Trechispora* sp.**	**Dai 22174 (BJFC)**	**OK298497***	**OK298952***
Trechispora stevensonii	TU 115499	UDB016467	UDB016467
Trechispora stevensonii	MA-Fungi 70669	JX392841	JX392842
Trechispora subsphaerospora	KHL 8511 (GB)	AF347080	AF347080
Trechispora termitophila	AMO396 (holotype)	MN701025	MN687983
Trechispora termitophila	AMO390	MN701024	MN687982
Trechispora torrendii	URM 85886 (holotype)	MK515148	MH280004
Tubulicium raphidisporum	He 3191 (BJFC)	MK204537	MK204545

**Newly generated sequences for this study. New species and new combinations or putatively new species are in bold.*

aligned by MAFFT 7 (Katoh et al., 2019[1]) using the "G-INS-i" strategy and manually adjusted in BioEdit (Hall, 1999). Unreliably aligned sections are removed before analyses and attempts are made to manually inspect and improve alignment. The data matrix is edited in Mesquite v3.70 software (Maddison and Maddison, 2021). The sequence alignment is deposited at TreeBase (submission ID 29141 and 29142). Sequences of *Auricularia* sp., *Exidia recisa* (Ditmar) Fr., and *Exidiopsis calcea* (Pers.) K. Wells are included in phylogenetic analyses. They belong to another order, Auriculariales Bromhead. The order is close to Trechisporales (Sulistyo et al., 2021). We add these three sequences in the combined two-gene data set (5.8S + nLSU) to demonstrate that Trechisporales forms a strongly supported sister clade to Auriculariales. Sequences of *Hyphodontia floccosa*

[1]http://mafft.cbrc.jp/alignment/server/

(Bourdot and Galzin) J. Erikss. and *Hyphodontia subalutacea* (P. Karst.) J. Erikss. in Hymenochaetales Oberw. obtained from GenBank are used as outgroups to root trees in the 5.8S + nLSU analysis. Two sequences of *Brevicellicium atlanticum* Melo, Tellería, M. Dueñas and M.P. Martín obtained from GenBank are used as outgroups to root trees in the ITS analysis.

The MP analysis is applied to the ITS data set sequences. Approaches to phylogenetic analysis follow Liu and Dai (2021), and the tree construction procedure is performed in PAUP* version 4.0 beta 10 software (Swofford, 2002). All characters are equally weighted, and gaps are treated as missing data. Trees are inferred using the heuristic search option with tree bisection and reconnection (TBR) branch swapping, and 1,000 random sequence additions maxtrees are set to 5,000, branches of zero length are collapsed, and all parsimonious trees are saved. Clade robustness is assessed using a bootstrap (BT) analysis with 1,000 replicates (Felsenstein, 1985). Descriptive tree statistics tree length (TL), consistency index (CI), retention index (RI), rescaled consistency index (RC), and homoplasy index (HI) are calculated for each maximum parsimonious tree (MPT) generated.

Maximum likelihood research is conducted with RAxML-HPC v. 8.2.3 (Stamatakis, 2014) and RAxML-HPC through the CIPRES Science Gateway (Miller et al., 2009[2]). Statistical support values (BS) are obtained using nonparametric bootstrapping with 1,000 replicates. The BI analysis is performed with MrBayes 3.2.7a (Ronquist and Huelsenbeck, 2003). Four Markov chains are run for two runs from random starting trees for 4 million generations (ITS) and 8 million generations (5.8S + nLSU) until the split deviation frequency value reaches <0.01, and trees are sampled every 1,000 generations. The first 25% of the sampled trees are discarded as burn-in, and the remaining ones are used to reconstruct a majority rule consensus tree and to calculate Bayesian posterior probabilities (BPP) of the clades.

The optimal substitution models for the combined data set are determined using the Akaike information criterion (AIC) implemented in MrModeltest 2.3 (Posada and Crandall, 1998; Nylander, 2004) after scoring 24 models of evolution by PAUP* version 4.0 beta 10 software (Swofford, 2002). The selected model applied in the BI analyses and ML analyses is the model GTR + I + G.

Branches that received BT support for ML (BS), MP (BP), and BPP greater than 65% (BS), 70% (BP), and 0.9 (BPP) are considered as significantly supported, respectively. Additionally, the ML analysis results in the best tree, and only the ML tree is presented along with the support values from the MP and BI analyses. FigTree v1.4.4 (Rambaut, 2018) is used to visualize the resulting tree.

Divergence Time Estimation

Divergence time is estimated with the BEAST v2.6.5 software package (Bouckaert et al., 2019) with 5.8S and nLSU sequences representing all main lineages in Basidiomycota (**Table 2**). Sequences of the species are adopted partly from the topology established by Wang et al. (2021). *Neurospora crassa* Shear and B.O. Dodge from Ascomycota are designated as outgroup taxon

[2]http://www.phylo.org

TABLE 2 | Information of taxa used in molecular clock analysis.

Species	Specimen no.	ITS	nLSU
Amylocorticium cebennense	HHB-2808	GU187505	GU187561
Anomoloma myceliosum	MJL-4413	GU187500	GU187559
Athelia arachnoidea	CBS 418.72	GU187504	GU187557
Auricularia heimuer	Xiaoheimao	LT716074	KY418890
Auricularia sp.	PBM 2295	DQ200918	AY634277
Australovuilleminia coccinea	MG75	HM046875	HM046931
Boletopsis leucomelaena	AFTOL-ID 1527	DQ484064	DQ154112
Bondarzewia montana	AFTOL-ID 452	DQ200923	DQ234539
Brevicellicium atlanticum	LISU 178566	NR_119820	HE963774
Brevicellicium atlanticum	LISU 178590	HE963775	HE963776
Brevicellicium daweishanense	CLZhao 17860	MW302337	MW293866
Brevicellicium daweishanense	CLZhao 18255	MW302338	MW293867
Brevicellicium exile	MA-Fungi 26554	HE963777	HE963778
Brevicellicium olivascens	KHL8571	HE963792	HE963793
Brevicellicium olivascens	MA-Fungi 23496	HE963787	HE963788
Brevicellicium xanthum	CLZhao 17781	MW302340	MW293869
Brevicellicium xanthum	CLZhao 2632	MW302339	MW293868
Bridgeoporus sinensis	Cui 10013	KY131832	KY131891
Calocera cornea	AFTOL-ID 438	AY789083	AY701526
Coltricia perennis	Cui 10319	KU360687	KU360653
Coltriciella dependens	Dai 10944	KY693737	KY693757
Corticium roseum	MG43	GU590877	AY463401
Craterocolla cerasi	TUB 020203	KF061265	KF061265
Cryptococcus humicola	AFTOL-ID 1552	DQ645516	DQ645514
Dacryopinax spathularia	AFTOL-ID 454	AY854070	AY701525
Dextrinocystis calamicola	He 5700	MK204534	MK204547
Dextrinocystis calamicola	He5693	MK204533	MK204546
Exidia recisa	EL 15-98	AF347112	AF347112
Exidiopsis calcea	MW 331	AF291280	AF291326
Fasciodontia brasiliensis	MSK-F 7245a	MK575201	MK598734
Fasciodontia bugellensis	MSK-F 5548	MK575204	MK598736
Fibrodontia alba	TNMF 24944	KC928274	KC928275
Fibrodontia gossypina	AFTOL-ID 599	DQ249274	AY646100
Fomitiporia hartigii	MUCL 53551	JX093789	JX093833
Fomitiporia mediterranea	AFTOL 688	AY854080	AY684157
Gloeophyllum sepiarium	Wilcox-3BB	HM536091	HM536061
Gloeophyllum striatum	ARIZAN 027866	HM536092	HM536063
Grifola frondosa	AFTOL-ID 701	AY854084	AY629318
Gymnopilus picreus	ZRL2015011	LT716066	KY418882
Hymenochaete rubiginosa	He1049	JQ716407	JQ279667
Hyphodontia densispora	LWZ 20170908-5	MT319426	MT319160
Hyphodontia zhixiangii	LWZ 20170818-13	MT319420	MT319151
Jaapia argillacea	CBS 252.74	GU187524	GU187581
Gomphidius roseus	MB 95-038	DQ534570	DQ534669
Kneiffiella barba-jovis	KHL 11730	DQ873609	DQ873610
Kneiffiella subalutacea	LWZ 20170816-9	MT319407	MT319139
Lepiota cristata	ZRL20151133	LT716026	KY418841
Leptosporomyces raunkiaeri	HHB-7628	GU187528	GU187588
Leucophellinus hobsonii	Cui 6468	KT203288	KT203309
Lyomyces macrosporus	LWZ20170817-2	MT319459	MT319194
Multiclavula mucida	AFTOL-ID 1130	DQ521417	AY885163
Neoantrodiella gypsea	Cui 10372	KT203290	MT319396
Neoantrodiella thujae	Dai 5065	KT203293	*MT319397*
Neurospora crassa	OR74A	HQ271348	AF286411

(Continued)

TABLE 2 | (Continued)

Species	Specimen no.	ITS	nLSU
Nigrofomes melanoporus	JV 1704/39	MF629835	MF629831
Nigrofomes sinomelanoporus	Cui 5277	MF629836	MT319398
Porodaedalea chinensis	Cui 10252	KX673606	MH152358
Porpomyces mucidus	Dai 12692	KT157833	KT157838
Porpomyces submucidus	Cui 5183	KU509521	KT152145
Pteridomyces galzinii	GB0150230	LR694188	LR694210
Pteridomyces galzinii	Bernicchia8122	MN937559	MN937559
Ramaria rubella	AFTOL-ID 724	AY854078	AY645057
Rigidoporus corticola	ZRL20151459	LT716075	KY418899
Rigidoporus ginkgonis	Cui 5555	KT203295	KT203316
Scytinopogon angulisporus	TFB13611	–	JQ684661
Scytinopogon pallescens	He 5192	–	MK204553
Sertulicium chilense	MA-Fungi 86368	HG315521	–
Sertulicium granuliferum	He 3338	MK204552	MK204540
Sertulicium jacksonii	Spirin 10425	MN987943	MN987943
Sertulicium lateclavigerum	LY 13467	MG913225	–
Sertulicium limonadense	LIP 0001683	MT180981	MT180978
Sertulicium niveocremeum	KHL13727	MN937563	MN937563
Sertulicium vernale	Soderholm 3886	MT002311	MT664174
Sistotremastrum aculeatum	Cui 8401	KX081133	KX081184
Sistotremastrum aculeatum	Miettinen 10380.1	MN991176	MW045423
Sistotremastrum aculeocrepitans	KHL 16097	MN937564	MN937564
Sistotremastrum confusum	KHL 16004	MN937567	MN937567
Sistotremastrum denticulatum	MV894	MN954694	MW045424
Sistotremastrum fibrillosum	LIP 0001413	NR_161047	NG_075239
Sistotremastrum fibrillosum s. l.	GUY13-119	MG913224	MG913210
Sistotremastrum fibrillosum s. l.	KHL 16988	MN937568	MN937568
Sistotremastrum geminum	Miettinen 14333	MN937568	MN937568
Sistotremastrum induratum	Spirin 8598	MT002324	MT664173
Sistotremastrum mendax	KHL12022	MN937570	MN937570
Sistotremastrum rigidum	MV833	MN954693	MW045435
Sistotremastrum suecicum	Kunttu 5959	MT075859	MT002335
Sistotremastrum suecicum	Miettinen 14550.1	MT075860	MT002336
Sistotremastrum suecicum	KHL 11849 (GB)	MN937571	MN937571
Sistotremastrum vigilans	Fonneland 2011-78	MN937572	MN937572
Sistotremastrum vigilans	Spirin 8778	MN991182	MN991182
Subulicystidium tropicum	He3968	MK204531	MK204544
Suillosporium cystidiatum	VS3830	MN937573	MN937573
Suillus pictus	AFTOL 717	AY854069	AY684154
Thelephora ganbajun	ZRL20151295	LT716082	KY418908
Trametes versicolor	ZRL20151477	LT716079	KY418903
Trechispora hymenocystis	KHL8795	AF347090	AF347090
Tremellodendron sp.	PBM2324	DQ411526	–
Tubulicium raphidisporum	He 3191	MK204537	MK204545
Ustilago maydis	AFTOL 505	AY854090	AF453938
Xylodon heterocystidiatus	LWZ 20171015-33	MT319518	MT319264

(Wang et al., 2021). A BEAST XML input file is generated with BEATUti v2. The estimation of rates of evolutionary changes at nuclear acids is using ModelTest 3.7 with the GTR substitution model (Posada and Crandall, 1998). A log-normal distribution is employed for molecular clock analysis (Drummond and Rambaut, 2007). A Yule speciation model is selected as prior assuming a constant speciation rate per lineage. Three fossil fungi, *viz. Paleopyrenomycites devonicus* (Taylor et al., 1999, 2005), *Archaeomarasmius leggetti* (Hibbett et al., 1995, 1997), and *Quatsinoporites cranhamii* (Smith et al., 2004; Berbee and Taylor, 2010) are taken from Wang et al.'s (2021) study. An XML file is conducted for 10 billion generations, producing log files and trees files. The log file is analyzed in Tracer 1,[3] and a maximum clade credibility (MCC) tree is interpreted in TreeAnnotator by trees file, removing the first 10% of the sampled trees as burn-in, and viewed in FigTree v1.4.2.

RESULTS

Phylogenetic Analyses

The concatenated 5.8S + nLSU data set contains 50 5.8S and 50 nLSU sequences from 52 fungal specimens representing 35 taxa in Trechisporales. The data set has an aligned length of 1,528 characters, of which 1,126 are constant, 89 are variable but parsimony-uninformative, and 313 are parsimony-informative. The average standard deviation (SD) of split frequencies is 0.005271 (BI). Three new combinations, namely, *Brevicellicium daweishanense*, *Brevicellicium xanthum*, and *Sertulicium limonadense*, are proposed based on the examination of type materials and phylogenetic analyses of type sequences (**Figure 1**).

The ITS data set contains sequences from 58 fungal specimens representing 36 *Trechispora* taxa (4 new species and another 32 *Trechispora* taxa). The data set has an aligned length of 753 characters, of which 284 are constant, 72 are variable but parsimony-uninformative, and 397 are parsimony-informative. MP analysis yields 13 equally parsimonious trees (TL = 2,318, CI = 0.398, RI = 0.638, RC = 0.254, and HI = 0.602). The average SD of split frequencies in BI analyses is 0.006959 (BI). The phylogenetic tree (**Figure 2**) reveals four new and independent lineages represented by our specimens, indicating that they are phylogenetically distinct from the species currently known in the genus. In addition, another taxon (Dai 22173 and Dai 22174) is treated as *Trechispora* sp.

The combined data set for the molecular clock analysis includes 100 collections, of which 47 belonged to Trechisporales. This data set results in a concatenated alignment of 1,588 characters with GTR as the best-fit evolutionary model. The MCC tree is used to study divergence time. The tree shows that Trechisporales occurs in a mean stem age of 270.85 Mya with a 95% highest posterior density (HPD) of 234.1–307.93 Mya (**Figure 3**). The tree also shows that the *Sistotremastrum* family and Hydnodontaceae occur in a mean stem age of 224.25 Mya [posterior probabilities (PP) = 0.8] with a 95% HPD of 182.47–266.75 Mya.

Taxonomy

***Sistotremastrum* family**

"Type genus": *Sistotremastrum* J. Erikss.

Habitat: It grows on rotten angiosperm and gymnosperm wood.

Basidioma are resupinate, thin, pruinose, or waxy. Hymenophores are smooth, verruculose, or odontioid-semiporoid. The hyphal structure is monomitic; generative hyphae bear clamp connections, CB(+). Cystidia and hyphidia

[3]http://beast.community/tracer

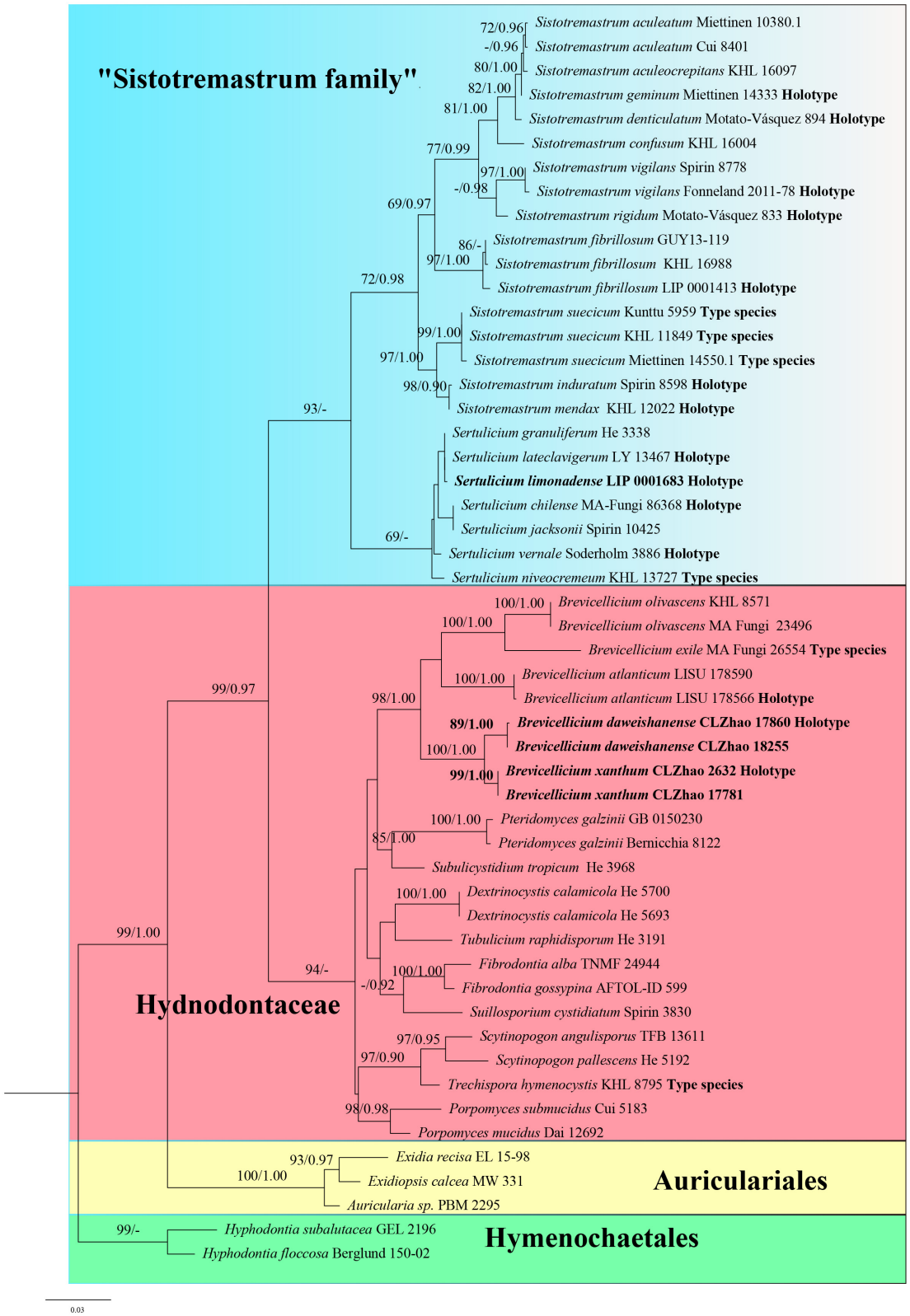


FIGURE 1 | Phylogeny of Trechisporales generated by maximum likelihood (ML) analyses based on combined 5.8S + nLSU sequences. Branches are labelled with ML bootstrap (BT) >65%, and Bayesian posterior probabilities (BPP) >0.90, respectively. New combinations, the sequence origin from holotype and the type status of the species in the genus are indicated in bold.

FIGURE 2 | Phylogeny of *Trechispora* generated by ML analyses based on combined ITS sequences. Branches are labelled with ML BT >65%, Parsimony Bootstrap Proportions >70%, and BPP >0.90, respectively. Putatively new species, the sequence origin from the holotype, and the type status of the species in the genus are indicated in bold.

are present in some species. Basidia are clavate or cylindrical, often with a median constriction, mostly with 2–4 or 4–6 sterigmata, and rarely with 6–8 sterigmata. Basidiospores are narrowly ellipsoid, ovoid, or cylindrical, thin-walled (but the wall is distinct), smooth, inamyloid, and acyanophilous.

Notes: *Sistotremastrum* family accommodates the genera *Sistotremastrum* and *Sertulicium* in the order Trechisporales based on its distinct lineage in the phylogenetic analysis. The combined phylogeny of two-gene data (**Figure 1**) demonstrates that *Sistotremastrum* family forms a supported sister clade to Hydnodontaceae. Basidia of most species in the *Sistotremastrum* family have more than four sterigmata, and basidiospores are smooth, while basidia of species in Hydnodontaceae have four sterigmata and their basidiospores are smooth to verrucose or aculeate. In addition, ampullate septa are only present in *Scytinopogon*, *Trechispora*, and *P. mucidus* in Hydnodontaceae.

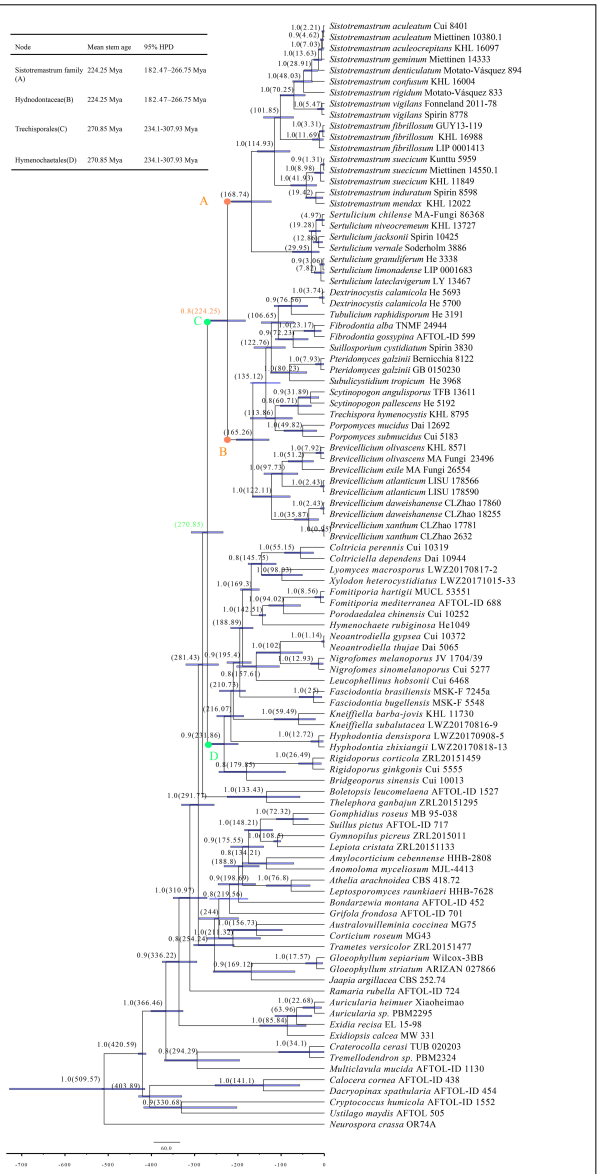


FIGURE 3 | Maximum clade credibility (MCC) chronogram and estimated divergence times of all main lineages in Basidiomycota inferred from the combined data set of 5.8S and LSU regions. The estimated divergence times of 95% highest posterior density (HPD) for all clades are indicated as node bars. The colored dots refer to the positions of the mean stem age of *Sistotremastrum* family, Hydnodontaceae, Trechisporales, and Hymenochaetales. The BPP above 0.8 and the mean divergence times of clades are labelled above and below the branches, respectively, at the nodes.

Trechispora dentata Z.B. Liu and Yuan Yuan, sp. November **Figure 4**

MycoBank number: MB 842865.

Type: China, Yunnan province, Sipsongpanna, Mengla County, XiShuangBanNa Tropical Botanical Garden, on soil, in southwestern China, ca. E 101° 25′, N 21° 41′, alt. 570 m. The vegetation is a natural tropical forest. 4 July 2021, Y.C. Dai 22565 (holotype BJFC 037139).

Etymology: *Dentata* (Lat.): It refers to the species having a dentate hymenophore.

Basidioma: They are annual, resupinate, soft when fresh, fragile when dry, easily separable from the substratum, up to 2.5-cm long, 2-cm wide, and less than 1-mm thick at the center; hymenial surface irpicoid, white when fresh, becoming cream (4A2/3) when dry; margin indistinct and fimbriate, mycelial cords absent; pores or aculei 3–4/mm; hymenophore lacerate to dentate; subiculum very thin to almost absent; tubes or aculei concolorous with a hymenial surface, less than 1 mm long.

Hyphal structure: Hyphal system is monomitic; generative hyphae bear clamp connections; ampullate septa occasionally present in subiculum and trama, up to 5-μm wide; all hyphae IKI−, CB− are unchanged in KOH; rhomboidal calcium oxalate crystals are scattered.

Subiculum: Generative hyphae hyaline, thin- to thick-walled, frequently branched, loosely interwoven, 2–4 μm in diameter.

Tubes or aculei: Generative hyphae in trama hyaline, thin- to thick-walled, frequently branched, loosely interwoven, 2–3 μm in diameter; cystidia and cystidioles are absent; basidia are clavate or barrel-shaped, hyaline, bearing four sterigmata and a basal clamp connection, 10–15 × 4–5 μm; basidioles are similar to basidia in shape but slightly shorter.

Basidiospores: They are ellipsoid, hyaline, thick-walled, aculeate, occasionally with one guttule, IKI−, CB−, (4−)4.1–5 × (3−)3.2–4(−4.1) μm (including ornamentation), L = 4.46 μm, W = 3.66 μm, Q = 1.22 (n = 60/1); (2.2−)2.6–3.7(−3.8) × 2–2.5 μm (excluding ornamentation), L' = 3.17 μm, W' = 2.23 μm, and Q' = 1.42 (n = 60/1).

Notes: *T. dentata* was discovered in the Yunnan Province of China. Phylogenetically, *T. dentata* is close to *Trechispora regularis* (Murrill) Liberta with strong support (96% BS, 96% BP, 1.00 BPP; **Figure 2**). However, *T. regularis* is strictly poroid (Liberta, 1973), and basidiospores of *T. dentata* are smaller than that of *T. regularis* [4.1–5 × 3.2–4 μm vs. 4–5.5 × 3.5–5 μm in *T. regularis* (including ornamentation); Liberta, 1973].

Trechispora dimitiella Z.B. Liu and Yuan, sp. November **Figure 5**

MycoBank number: MB 842866.

Type: China, Hainan Province, Haikou, Jinniuling Park, on a rotten leaf, in southwestern China, ca. E 110° 19′, N 20° 1′, alt. 17 m. The vegetation is a plantation in tropical China. 7 November 2020, Y.C. Dai 21931 (holotype BJFC 035830).

Etymology: *Dimitiella* (Lat.): It refers to the species having a dimitic hyphal system.

Basidioma: They are annual, resupinate, soft when fresh, fragile when dry, easily separable from the substratum, up to 6-cm long, 4-cm wide, and approximately 3-mm thick at the center; the hymenial surface is poroid, pore surface white to cream (4A2/3) when fresh, becoming white to buff-yellow (4A4) when dry; margin indistinct, often with emerging mycelial cords; pores angular, 5–6/mm; dissepiments thin, lacerate; subiculum up to 1 mm thick; tubes concolorous with a poroid surface, up to 2 mm long.

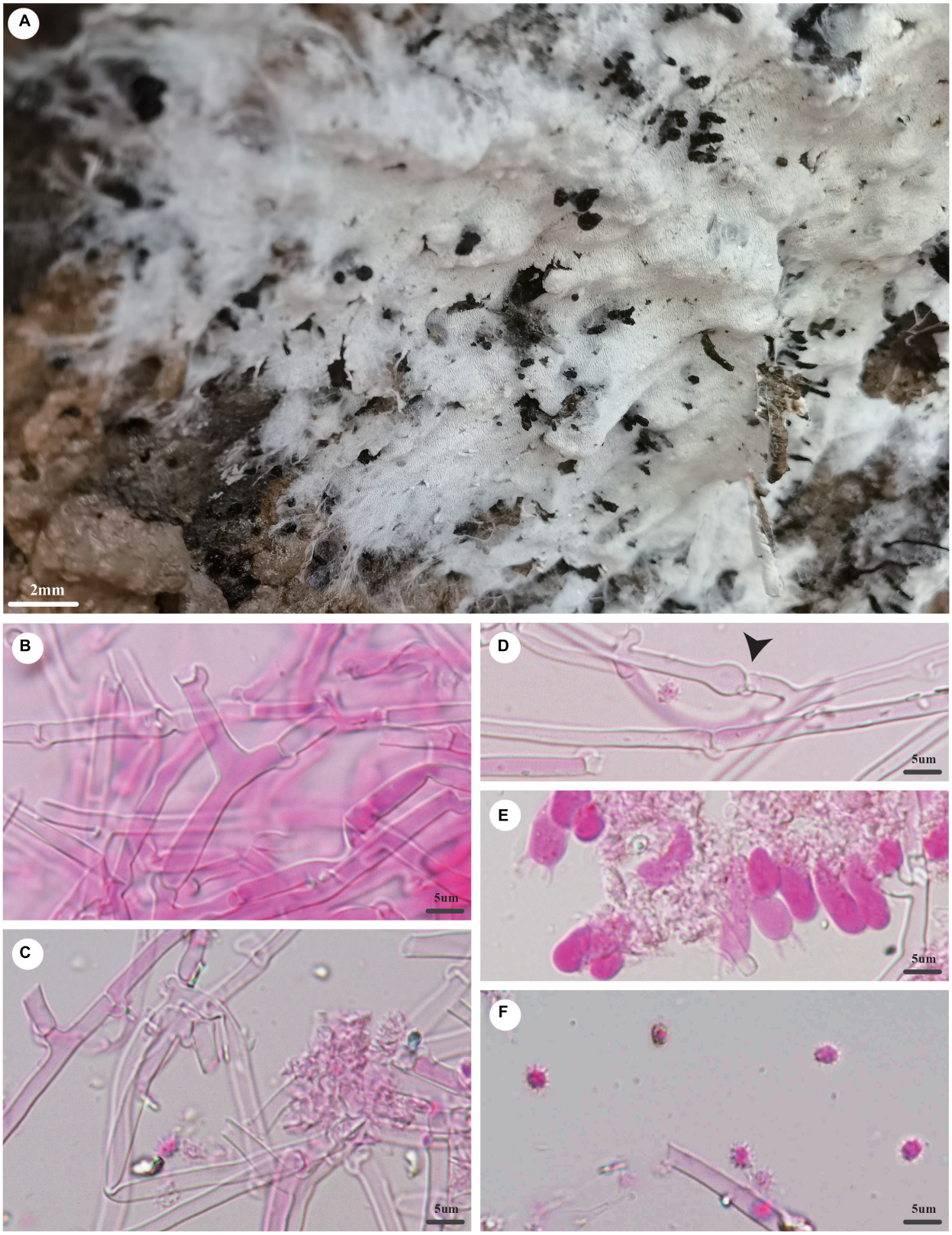


FIGURE 4 | *Trechispora dentata* (holotype, Dai 22565). **(A)** A basidioma, **(B)** hyphae from subiculum, **(C)** hyphae from trama, **(D)** hyphae with ampullate septa (black arrow), **(E)** basidia and basidioles, and **(F)** basidiospores. Photo by Ya-Ping Lian and Zhan-Bo Liu.

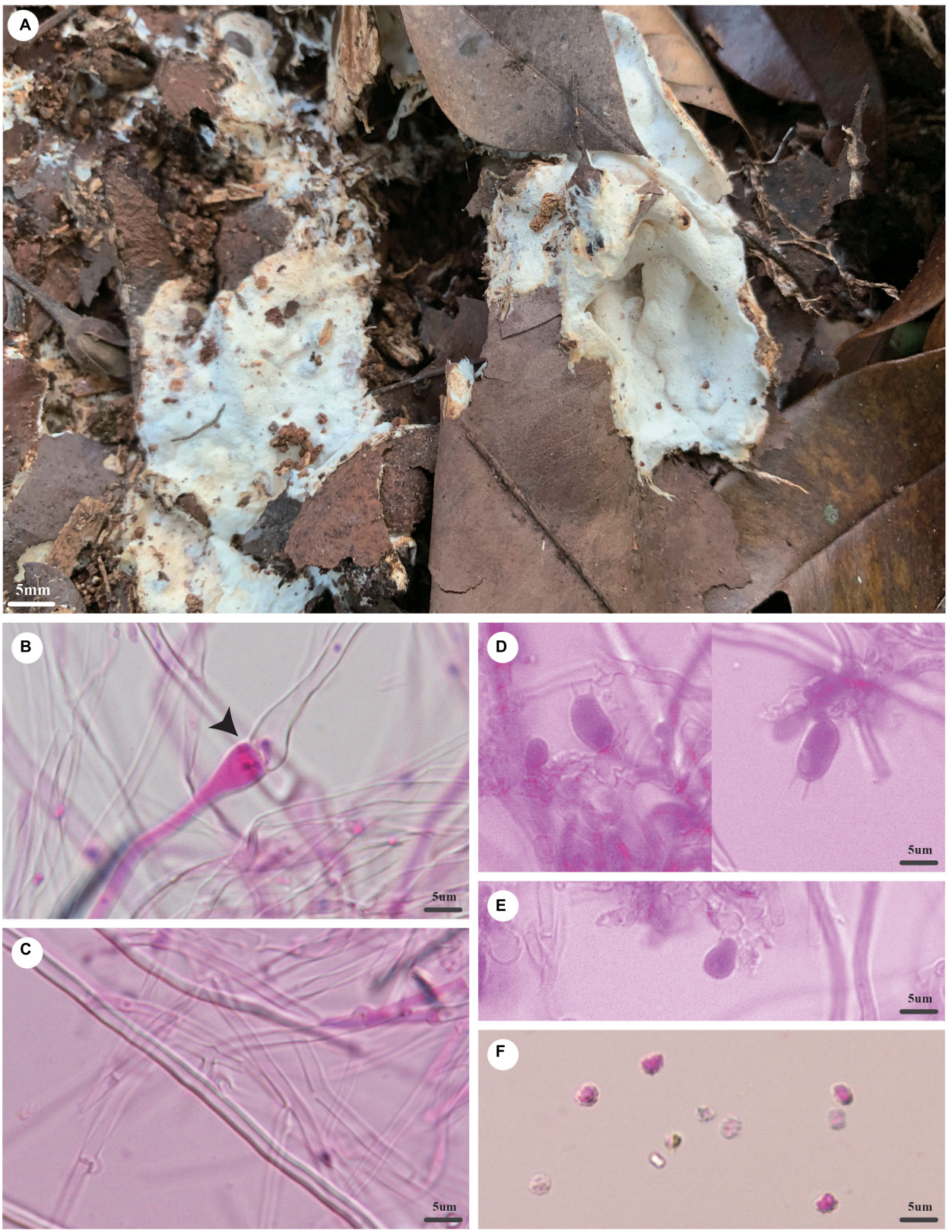


FIGURE 5 | *Trechispora dimitiella* (holotype, Dai 21931). **(A)** A basidioma, **(B)** hyphae with ampullate septa from subiculum (black arrow), **(C)** hyphae from tubes, **(D)** basidia, **(E)** basidioles, and **(F)** basidiospores. Photo by Ya-Ping Lian and Zhan-Bo Liu.

Hyphal structure: Hyphal system is dimitic; generative hyphae bear clamp connections; ampullate septa occasionally present in subiculum and trama, up to 4.5 μm wide; all hyphae IKI−, CB− are unchanged in KOH; rhomboidal calcium oxalate crystals are scattered.

Subiculum: Generative hyphae hyaline, thin-walled, rarely branched, 2–3 μm in diameter; skeletal hyphae thick-walled with a wide lumen, unbranched, loosely interwoven, 2–4 μm diameter.

Tubes: Generative hyphae hyaline, thin-walled, rarely branched, 1.5–2.5 μm in diameter; skeletal hyphae thick-walled with a wide lumen, unbranched, loosely interwoven, 2–3 μm in diameter; cystidia and cystidioles are absent; basidia are barrel-shaped, hyaline, bearing four sterigmata and a basal clamp connection, 9.5–12 × 4–5 μm; basidioles are similar to basidia in shape but slightly shorter.

Basidiospores: They are ellipsoid, hyaline, thick-walled, aculeate, IKI−, CB−, (3.5−)3.6–4(−4.2) × (2.5–)2.7–3.1(−3.2) μm (including ornamentation), L = 3.84 μm, W = 2.92 μm, Q = 1.31–1.33 (n = 60/2); (2.6−)2.7–3.4(−3.7) × 2–2.6(−2.9) μm (excluding ornamentation), L' = 3.04 μm, W' = 2.18 μm, and Q' = 1.38–1.4 (n = 60/2).

Additional specimen examined (paratypes): China, Yunnan Province, Jinghong, Primeval Forest Park, on soil, 7 July 2021, Y.C. Dai 22601 (BJFC), Dai 22602 (BJFC). Malaysia, Selangor, Kota Damansara, Community Forest Reserve, on rotten angiosperm wood, 7 December 2019, Y.C. Dai 21181 (BJFC 032835).

Notes: *T. dimitiella* was discovered in China and Malaysia. Most species in *Trechispora* are corticioid fungi with a monomitic hyphal structure, but *T. dimitiella* is different. Morphologically, *T. dimitiella* and *Trechispora brasiliensis* (Corner) K.H. Larss. share the poroid hymenophore with a dimitic hyphal system and aculeate basidiospores. However, the basidiospores of *T. dimitiella* are smaller than that of *T. brasiliensis* [3.6–4 × 2.7–3.1 μm vs. 4–4.5 × 3–4 μm in *T. brasiliensis* (including ornamentation), Larsson, 1992]. Phylogenetically, *T. dimitiella* is close to *Trechispora incisa* K.H Larss. (80% BS, 0.99 BPP; **Figure 2**), but *T. dimitiella* can be easily distinguished from *T. incisa* due to its poroid hymenophore with a dimitic hyphal system because *T. incisa* has arachnoid to farinose or minutely granulose hymenophore with a monomitic hyphal system (Larsson, 1996).

Trechispora fragilis Z.B. Liu and Yuan Yuan, sp. November **Figure 6**

MycoBank number: MB 842867.

Type: China, Yunnan Province, Sipsongpanna, Mengla County, XiShuangBanNa Tropical Botanical Garden, on the ground of the forest, in southwestern China, ca. E 101° 25′, N 21° 41′, alt. 570 m. The vegetation is a natural tropical forest. 18 August 2019, Y.C. Dai 20535 (holotype BJFC 032203).

Etymology: *Fragilis* (Lat.): It refers to the species having fragile basidiocarps.

Basidioma: They are annual, resupinate, soft when fresh, fragile when dry, easily separable from the substratum, up to 3 cm long, 2 cm wide, and less than 1 mm thick at the center; the hymenial surface is odontoid, white when fresh, becoming cream (4A2/3) to buff-yellow (4A4) when dry; margin is indistinct and fimbriate, often with emerging mycelial cords; aculei sparse, 4–6/mm; subiculum very thin to almost absent; aculei concolorous with a hymenial surface, less than 1 mm long.

Hyphal structure: Hyphal system monomitic; generative hyphae bear clamp connections; ampullate septa occasionally present in subiculum and aculei, up to 7 μm wide; all hyphae IKI−, CB− are unchanged in KOH; rhomboidal calcium oxalate crystals are scattered.

Subiculum: Generative hyphae hyaline, thin- to thick-walled, frequently branched, loosely interwoven, 1.5–4 μm in diameter.

Aculei: Generative hyphae in trama hyaline, thin- to thick-walled, frequently branched, loosely interwoven, 1.5–3 μm in diameter; cystidia and cystidioles are absent; basidia are clavate shaped, hyaline, bearing four sterigmata, and a basal clamp connection, 12–14 × 3.5–4 μm; basidioles are similar to basidia in shape but slightly shorter.

Basidiospores: Ellipsoid, hyaline, thick-walled, aculeate, IKI−, CB−, (3.2−)3.8–4(−4.2) × (2.4−)2.5–3 μm (including ornamentation), L = 3.53 μm, W = 2.79 μm, Q = 1.27 (n = 60/1); (2.6−)2.8–3.7(−4) × (1.9−)2–2.7(−3.1) μm (excluding ornamentation), L' = 3.16 μm, W' = 2.26 μm, and Q' = 1.40 (n = 60/1).

Notes: *T. fragilis* was discovered in the Yunnan Province of China. Phylogenetically, *T. fragilis* groups with *Trechispora termitophila* Meiras-Ottoni and Gibertoni and *Trechispora havencampii* (Desjardin and B.A. Perry) Meiras-Ottoni and Gibertoni (69% BS, 0.92 BPP; **Figure 2**). *T. termitophila* can be easily distinguished from *T. fragilis* due to its coralloid basidioma. In addition, the basidiospores of *T. fragilis* are smaller than that of *T. termitophila* [6.5–7.5 μm vs. 4.5–5 μm in *T. termitophila* (including ornamentation), Meiras-Ottoni et al., 2021]. *T. havencampii* can also be easily distinguished from *T. fragilis* due to its coralloid basidioma. In addition, basidiospores of *T. fragilis* are smaller than that of *T. havencampii* [3.8–4 × 2.5–3 μm vs. 5.2–6.5 × 3.5–4.2 μm in *T. havencampii* (including ornamentation), Desjardin and Perry, 2015].

Trechispora laevispora Z.B. Liu, Y.D. Wu and Yuan Yuan, sp. November **Figure 7**

MycoBank number: MB 842868.

Type: China, Inner Mongolia Autonomous Region, Arxan, Bailang Feng Scenic Spot, on the charred trunk of *Larix*, in southwestern China, ca. E 119° 56′, N 47° 10′, alt. 1,511 m. The vegetation is a natural boreal forest. 25 August 2020, Y.C. Dai 21655 (holotype BJFC 035556).

Etymology: *Laevispora* (Lat.): It refers to the species having smooth basidiospores.

Basidioma: They are annual, resupinate, soft when fresh and dry, up to 8 cm long, 3 cm wide, and less than 1 mm thick at the center; the hymenial surface is smooth, white when fresh and dry; margin is indistinct and fimbriate, often with emerging mycelial cords; subiculum very thin to almost absent.

Hyphal structure: Hyphal system monomitic; generative hyphae bear clamp connections; ampullate septa frequently present in subiculum and hymenium, up to 7 μm wide; all hyphae

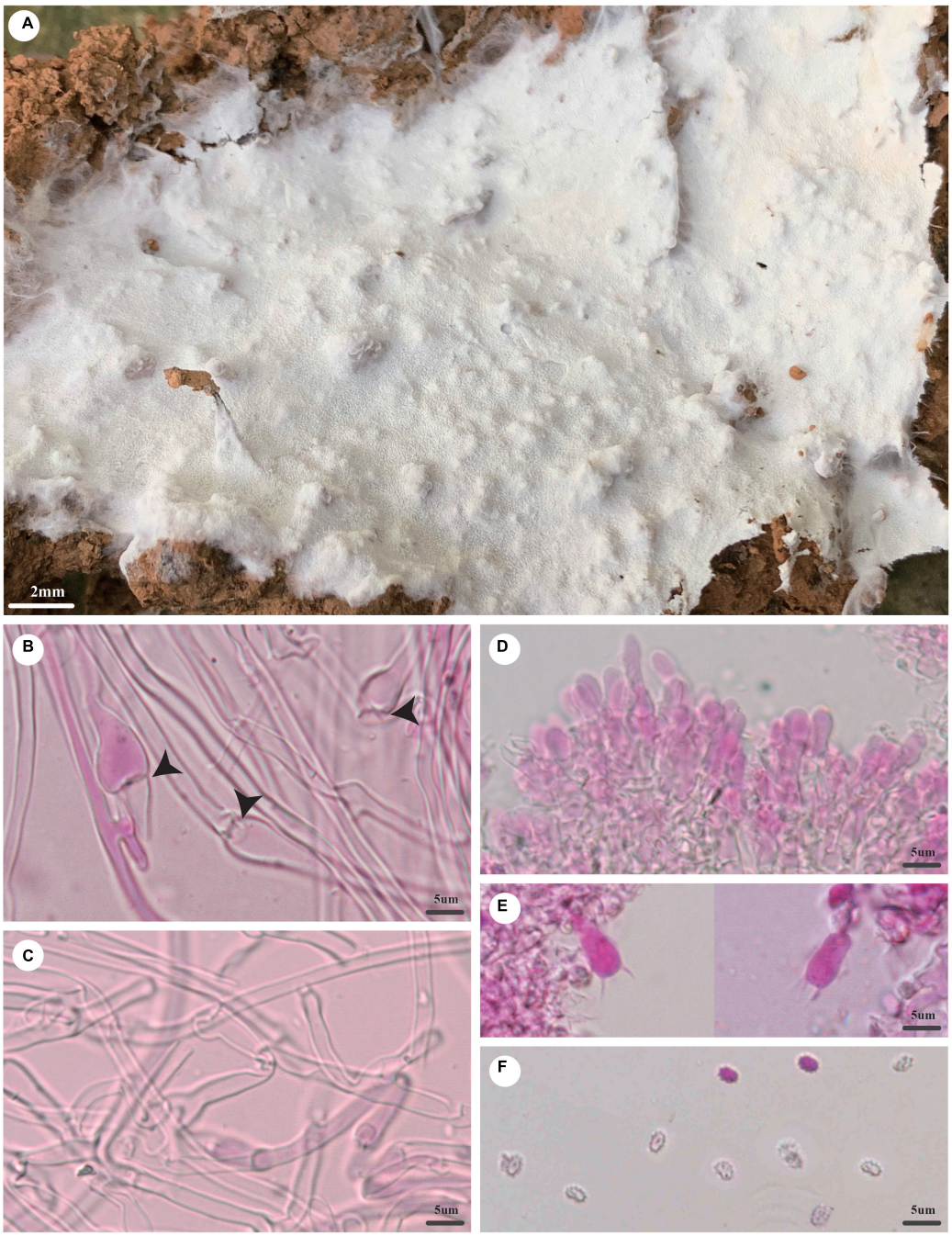


FIGURE 6 | *Trechispora fragilis* (holotype, Dai 20535). **(A)** A basidioma, **(B)** hyphae with ampullate septa from subiculum (black arrows), **(C)** hyphae from trama, **(D)** hymenium with basidioles, **(E)** basidia, and **(F)** basidiospores. Photo by Ya-Ping Lian and Zhan-Bo Liu.

IKI−, CB− are unchanged in KOH; rhomboidal calcium oxalate crystals are abundant.

Subiculum: Generative hyphae hyaline, thin-walled, frequently branched, loosely interwoven, 1.5–3 μm in diameter.

Hymenium: Generative hyphae in subhymenium hyaline, thin-walled, frequently branched, 1.5–3 μm in diameter; cystidia and cystidioles are absent; basidia are clavate shaped, hyaline, bearing four sterigmata and a basal clamp connection, 11.5–15 × 4–5 μm; basidioles are similar to basidia in shape but slightly shorter.

Basidiospores: Ellipsoid, hyaline, thin-walled, smooth, IKI−, CB−, (2.5−) 2.6–3.2(−3.3) × (1.8−)1.9–2.2(−2.5) μm, $L = 2.92$ μm, $W = 2.04$ μm, and $Q = 1.43$ ($n = 60/1$).

Notes: *T. laevispora* was discovered in the Inner Mongolia Autonomous Region of China. Phylogenetically, *T. laevispora* groups with *Trechispora cohaerens* (Schwein.) Jülich and Stalpers

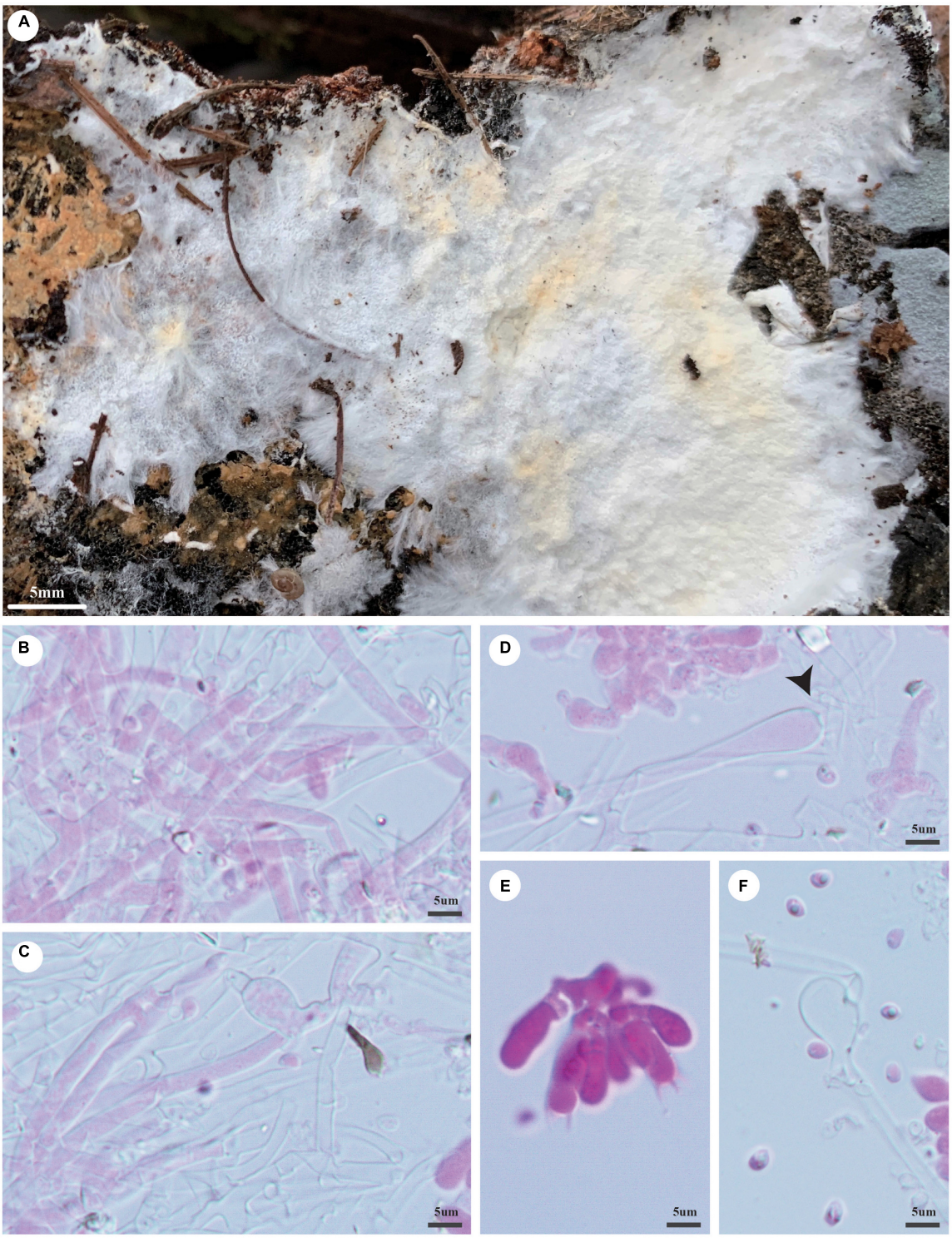


FIGURE 7 | *Trechispora laevispora* (holotype, Dai 21655). **(A)** A basidioma, **(B,C)** hyphae from subiculum, **(D)** subicular hyphae with ampullate septa (black arrow) and a piece of hymenium, **(E)** basidia and basidioles, and **(F)** basidiospores. Photo by Ya-Ping Lian and Zhan-Bo Liu.

with strong support (94% BS, 96% BP, 1.00 BPP; **Figure 2**). Both species share a smooth hymenophore, a monomitic hyphal system with smooth basidiospores. However, basidiospores of *T. cohaerens* are thick-walled and larger than that of *T. laevispora* (3.5–4 × 2.2–2.5 μm in *T. cohaerens*; Larsson, 1992).

B. daweishanense (C.L. Zhao) Z.B. Liu and Yuan Yuan, comb. November

MycoBank number: MB 842869.

Basionym: *T. daweishanensis* C.L. Zhao, Phytotaxa 479(2): 153 (2021).

Type: China. Yunnan Province, Honghe, Pingbian County, Daweishan National Nature Reserve, on the fallen branch of angiosperms, 1 August 2019, CLZhao 17860 (holotype SWFC).

Description: See Zong et al. (2021, as *T. daweishanensis*).

B. xanthum (C.L. Zhao) Z.B. Liu and Yuan Yuan, comb. November

MycoBank number: MB 842870.

Basionym: *T. xantha* C.L. Zhao, Phytotaxa 479(2): 155 (2021).

Type: China. Yunnan Province, Yuxi, Xinping County, Mopanshan National Forestry Park, on the trunk of *Albizia julibrissin*, 20 August 2017, CLZhao 2632 (holotype SWFC).

Description: See Zong et al. (2021, as *T. xantha*).

Notes: Zong et al. (2021) described *T. daweishanensis* and *T. xantha* as new species. However, in our phylogeny, they belong to the genus *Brevicellicium* (98% BS, 1.00 BPP; **Figure 1**). The type specimens of abovementioned species are studied [CLZhao 17860 (SWFC); CLZhao 2632 (SWFC)]. We do not observe ampullate hyphae from type materials as mentioned by Zong et al. (2021). We suppose that Zong et al. (2021) confused basidioles with ampullate hyphae (ampullate septa on some generative hyphae), which are remarkable characters of *Trechispora*. In fact, *T. daweishanensis* and *T. xantha* have a smooth hymenophore, a monomitic hyphal structure with clamped generative hyphae, and the absence of ampullate septa. They fit *Brevicellicium* well. Herein, we combine these two species in *Brevicellicium* based on morphological and phylogenetic evidence (**Figure 1**).

S. limonadense (G. Gruhn and P. Alvarado) Z.B. Liu and Yuan Yuan, comb. November

MycoBank number: MB 842871.

Basionym: *S. limonadense* G. Gruhn and P. Alvarado, Phytotaxa 498(1): 36 (2021).

Type: French Guiana. On the bark of an unidentified dead trunk lying on the ground, October 22, 2013, LIP 0001683 (holotype).

Description: See Gruhn and Alvarado (2021, as *S. limonadense*).

Notes: Gruhn and Alvarado (2021) described *S. limonadense* as a new species. However, at the same time, Spirin et al. (2021) segregated the species around *S. niveocremeum* (Höhn. and Litsch.) J. Erikss. into the new genus *Sertulicium*. In our phylogeny, *S. limonadense* groups with *Sertulicium granuliferum* (Hallenb.) Spirin and Volobuev *Sertulicium lateclavigerum* (Boidin and Gilles) Spirin and Viner (**Figure 1**). We did not study specimens, but *S. limonadense* is characterized by smooth to tuberculate hymenophore and basidia have 6–8 sterigmata (Gruhn and Alvarado, 2021) and fits *Sertulicium* better. Hence, we transfer *S. limonadense* to *Sertulicium*.

DISCUSSION

Larsson (2007) showed that *S. suecicum* and *S. niveocremeum* (= *S. niveocremeum*) formed a strongly supported sister clade (94% BS, 1.00 BPP) to Hydnodontaceae within Trechisporales. However, in his phylogenetic analysis of 5.8S + nLSU, there were a few species in Hydnodontaceae and *Sistotremastrum* to establish a new family for *S. suecicum* and *S. niveocremeum*. Hence, Larsson (2007) named this clade *Sistotremastrum* family. The same strongly supported topology was recovered by Telleria et al. (2013); Gruhn et al. (2018), and Meiras-Ottoni et al. (2021) by the nLSU phylogenetic analysis. Spirin et al. (2021) presented a comprehensive study of *Sistotremastrum* and *Sertulicium* with 17 species. They used the nLSU region to perform phylogenetic analyses of 16 species in the two genera (**Figure 1** in Spirin et al., 2021), except for *Sertulicium chilense* (Telleria, M. Dueñas and M.P. Martín) Spirin and Volobuev because the nLSU sequences of *S. chilense* were absent. However, they were not able to generate high support values for the node connecting *Sistotremastrum* and *Sertulicium* (87% BS, 0.87 BPP, **Figure 1** in Spirin et al., 2021). As a result, they gave up establishing a new family too.

ITS1-5.8S-ITS2 is an important marker used for the barcoding of fungal species (Liu et al., 2021; Wangsawat et al., 2021). However, the difficulty in aligning ITS sequences for fungi in Trechisporales is evident because it is a data set covering taxa in distinct taxonomic levels (Larsson, 2007). Therefore, it is not a good idea to run combined analyses of ITS + nLSU, so we use the most stable and conservative portion of ITS (5.8S) and nLSU to our phylogenetic analyses of *Sistotremastrum* and *Sertulicium* (5.8S + nLSU) (**Figure 1**). We add *S. chilense* and *S. limonadense* to phylogenetic analyses. Our results of the *Sistotremastrum* are the same as phylogenetic analyses by Spirin et al. (2021, **Figure 1**). However, our phylogenetic analyses of *Sertulicium* are a bit different from that by Spirin et al. (2021, **Figure 1**) because the data sets used in both studies are different. Above all, we generate high support values for the node connecting *Sistotremastrum* and *Sertulicium* from ML analysis (93% BS) based on 5.8S and nLSU sequences; however, BI fails to provide support for the node (0.76 BPP).

Divergence time is estimated with 5.8S and nLSU sequences representing all main lineages in Basidiomycota (**Figure 3**). The MCC tree shows that Basidiomycota occurs in a mean stem age of 509.57 Mya. Trechisporales occurs in a mean stem age of 270.85 Mya. The tree also shows that the *Sistotremastrum* family and Hydnodontaceae occur in a mean stem age of 224.25 Mya (PP = 0.8). Zhao et al. (2017) indicate that the divergence times of Basidiomycota are 530 Mya (the mean stem age). He et al. (2019) indicate that the divergence times of Trechisporales and Hydnodontaceae are 259 Mya (the mean stem age). Our experimental results agree with them. In this paper, we update the divergence times of Trechisporales and Hydnodontaceae and define the divergence time of the *Sistotremastrum* family.

Bayesian phylogenetic inference fails to provide support for the node of *Sistotremastrum* and *Sertulicium*, so we use the term "*Sistotremastrum* family" for the two genera without a formal description of the new family. In the future, we will sequence

additional DNA regions or whole genomes, for a more robust phylogenetic analysis.

At present, there are only two species in the *Sistotremastrum* family ever been recorded from China, i.e., *Sistotremastrum aculeatum* Miettinen and Viner (Cui 8401) and *S. granuliferum* (He 3338; CLZhao 5531, 9771). Recently, we collected a specimen from the Yunnan Province of China (He 6276), and its morphological and DNA data demonstrated the specimen is S. limonadense. The species is a new record in China, and we have uploaded ITS and nLSU sequences of the specimen (He 6276) to GenBank. Above all, we study all the Chinese specimens of species in the Sistotremastrum family seriously, and their morphology fits the descriptions of Gruhn and Alvarado (2021) and Spirin et al. (2021). We also collected a specimen from the Hainan Province of China (Dai 17696). The ITS (OK298490) region is different from *Sistotremastrum fibrillosum* G. Gruhn and P. Alvarado by 6%, and morphologically it is similar to *S. fibrillosum*. However, we only have a single specimen, so for the time being we regard Dai 17696 as *Sistotremastrum* sp.

In this article, we use the whole ITS region in analyses of *Trechispora* to visualize the genetic distances among new taxa and those already described. *T. dentata*, *T. dimitiella*, *T. fragilis*, and *T. laevispora* are described as new to science based on morphological characteristics and molecular evidence (**Figure 2**). Most of these new species are found in subtropical or tropical Asia and conform to the phenomenon that subtropical or tropical Asia harbors high taxonomic diversity for all wood-decaying fungi (Dai, 2012; Cui et al., 2019). We also collected two resupinate specimens (Dai 22173 and Dai 22174) from the Hainan Province of China. The morphology of the two specimens corresponds to the concept of *Trechispora* and forms a distinct lineage within the *Trechispora* clade (100% BS, 1.00 BPP; **Figure 2**). However, these specimens are sterile, so we regard Dai 22173 and Dai 22174 as *Trechispora* spp. temporarily here.

Molecular phylogenetic analyses in the present study show that *Brevicellicium* forms a monophyletic clade in which all *Brevicellicium* species are included (98% BS, 1.00 BPP; **Figure 1**). However, when we add sequences of *T. xantha* and *T. daweishanensis*, we find sequences of a two-species cluster with *Brevicellicium* with high support (100% BS, 1.00 BPP; **Figure 1**). We request and examine type specimens from Zhao and find *T. xantha* and *T. daweishanensis* corresponding to the concept of *Brevicellicium* and they should be transferred to the genus *Brevicellicium* (see the notes of *B. daweishanense*).

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/supplementary material.

AUTHOR CONTRIBUTIONS

Z-BL: design of the research, performance of the research, and writing and revising this manuscript. Z-BL, HZ, Y-PL, Y-RW, C-GW, and W-LM: data analysis and interpretation. Z-BL, YY, and Y-DW: a collection of the materials. All authors contributed to the article and approved the submitted version.

FUNDING

The research is supported by the National Natural Science Foundation of China (Project Nos. 31870007 and 32011540380).

ACKNOWLEDGMENTS

We thank Prof. Dr. Chang-Lin Zhao (SWFC, China) and Prof. Yu-Cheng Dai for allowing us to study their specimens.

REFERENCES

Berbee, M. L., and Taylor, J. W. (2010). Dating the molecular clock in fungi – how close are we? *Fungal Biol. Rev.* 24, 1–16. doi: 10.1016/j.fbr.2010.03.001

Bouckaert, R., Vaughan, T. G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina, A., et al. (2019). BEAST 2.5: an advanced software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* 15:e1006650. doi: 10.1371/journal.pcbi.1006650

Chikowski, R., Larsson, K. H., and Gibertoni, T. B. (2020). Taxonomic novelties in *Trechispora* (*Trechisporales*, Basidiomycota) from Brazil. *Mycol. Prog.* 19, 1403–1414. doi: 10.1007/s11557-020-01635-y

Cui, B. K., Li, H. J., Ji, X., Zhou, J. L., Song, J., Si, J., et al. (2019). Species diversity, taxonomy and phylogeny of *Polyporaceae* (Basidiomycota) in China. *Fungal Divers.* 97, 137–392. doi: 10.1007/s13225-019-00427-4

Dai, Y. C. (2010). Hymenochaetaceae (Basidiomycota) in China. *Fungal Divers.* 45, 131–343. doi: 10.1007/s13225-010-0066-9

Dai, Y. C. (2012). Polypore diversity in China with an annotated checklist of Chinese polypores. *Mycoscience* 53, 49–80. doi: 10.1007/s10267-011-0134-3

Desjardin, D. E., and Perry, B. A. (2015). A new species of Scytinopogon from the island of príncipe, republic of são tomé and príncipe, West Africa. *Mycosphere* 6, 434–441. doi: 10.5943/mycosphere/6/4/5

Drummond, A. J., and Rambaut, A. (2007). BEAST: bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* 7:214–221. doi: 10.1186/1471-2148-7-214

Du, P., Cao, T. X., Wu, Y. D., Zhou, M., and Liu, Z. B. (2021). Two new species of *Hymenochaetaceae* on *Dracaena cambodiana* from tropical China. *MycoKeys* 80, 1–17. doi: 10.3897/mycokeys.80.63997

Fan, L. F., Alvarenga, R. L. M., Gibertoni, T. B., Wu, F., and Dai, Y. C. (2021). Four new species in the *Tremella fibulifera* complex (Tremellales, Basidiomycota). *MycoKeys* 82, 33–56. doi: 10.3897/mycokeys.82.63241

Felsenstein, J. (1985). Confidence intervals on phylogenetics: an approach using bootstrap. *Evolution* 39, 783–791. doi: 10.2307/2408678

Furtado, A. N. M., Danils, P. P., Reck, M. A., and Neves, M. A. (2021). Scytinopogon caulocystidiatus and *S. foetidus* spp. nov. and five other species recorded from Brazil. *Mycotaxon* 136, 107–130. doi: 10.5248/136.107

Gruhn, G., Alvarado, P., Hallenberg, N., Roy, M., and Courtecuisse, R. (2018). Contribution to the taxonomy of *Sistotremastrum* (Trechisporales, Basidiomycota) and the description of two new species, *S. fibrillosum* and *S. aculeocrepitans*. *Phytotaxa* 379:1. doi: 10.11646/phytotaxa.379.1.2

Gruhn, G., and Alvarado, P. (2021). Sistotremastrum limonadense sp. nov. from French Guiana. *Phytotaxa* 498, 35–43. doi: 10.11646/phytotaxa.498.1.4

Hall, T. A. (1999). Bioedit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* 41, 95–98.

He, M. Q., Zhao, R. L., Hyde, K. D., Dominik, B., Martin, K., Andrey, Y., et al. (2019). Notes, outline and divergence times of Basidiomycota. *Fungal Divers.* 99, 105–367. doi: 10.1007/s13225-019-00435-4

Hibbett, D. S., Binder, M., Bischoff, J. F., Blackwell, M., Cannon, P. F., Eriksson, O. E., et al. (2007). A higher-level phylogenetic classification of the Fungi. *Mycol. Res.* 111, 509–547. doi: 10.1016/j.mycres.2007.03.004

Hibbett, D. S., Grimaldi, D., and Donoghue, M. J. (1995). Cretaceous mushrooms in amber. *Nature* 377:487. doi: 10.1038/377487a0

Hibbett, D. S., Grimaldi, D., and Donoghue, M. J. (1997). Fossil mushrooms from Miocene and *Cretaceous ambers* and the evolution of Homobasidiomycetes. *Am. J. Bot* 1997, 981–991. doi: 10.2307/2446289

Karsten, P. A. (1890). Fragmenta mycologica XXIX. *Hedwigia* 29, 147–149.

Katoh, K., Rozewicki, J., and Yamada, K. D. (2019). MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief. Bioinformatics* 20, 1160–1166. doi: 10.1093/bib/bbx108

Larsson, K. H. (1992). *The Genus Trechispora. (Corticiaceae, Basidiomycetes).* Ph.D. thesis. Gothenburg: University of Göteborg.

Larsson, K. H. (1996). New species and combination in *Trechispora* (*Corticiaceae*, Basidiomycotina). *Nordic J. Bot.* 16, 83–98. doi: 10.1111/j.1756-1051.1996.tb00218.x

Larsson, K. H. (2007). Re-thinking the classification of corticioid fungi. *Mycol. Res.* 111, 1040–1063. doi: 10.1016/j.mycres.2007.08.001

Larsson, K. H., Larsson, E., and Kõljalg, U. (2004). High phylogenetic diversity among corticioid Homobasidiomycetes. *Mycol. Res.* 108, 983–1002.

Liberta, A. E. (1973). The genus *Trechispora* (Basidiomycetes, *Corticiaceae*). *Botany* 51, 1871–1892. doi: 10.1139/b73-240

Liu, S. L., Ma, H. X., He, S. H., and Dai, Y. C. (2019). Four new corticioid species in Trechisporales (Basidiomycota) from East Asia and notes on phylogeny of the order. *MycoKeys* 48, 97–113. doi: 10.3897/mycokeys.48.31956

Liu, Z. B., and Dai, Y. C. (2021). *Steccherinum fragile* sp. nov. and *S. subcollabens* comb. nov. (*Steccherinaceae*, Polyporales), evidenced by morphological characters and phylogenetic analysis. *Phytotaxa* 483, 106–116. doi: 10.11646/phytotaxa.483.2.3

Liu, Z. B., and Yuan, Y. (2020). Luteoporia citriniporia sp. nov. (Polyporales, Basidiomycota), evidenced by morphological characters and phylogenetic analysis. *Phytotaxa* 461, 31–39. doi: 10.11646/phytotaxa.461.1.4

Liu, Z. B., Zhou, M., Yuan, Y., and Dai, Y. C. (2021). Global diversity and taxonomy of Sidera (Hymenochaetales, Basidiomycota): four new species and keys to species of the genus. *J. Fungi* 7:251. doi: 10.3390/jof7040251

Maddison, W. P., and Maddison, D. R. (2021). *Mesquite: A Modular System For Evolutionary Analysis. Version 3.70.* Available online at: https://www.mesquiteproject.org/ (Accessed March, 2022).

Meiras-Ottoni, A. D., Larsson, K. H., and Gibertoni, T. B. (2021). Additions to *Trechispora* and the status of *Scytinopogon* (*Trechisporales*, Basidiomycota). *Mycol. Prog.* 20, 203–222. doi: 10.1007/s11557-021-01667-y

Miller, M. A., Holder, M. T., Vos, R., Midford, P. E., Liebowitz, T., Chan, L., et al. (2009). *The CIPRES Portals. CIPRES.* Available online at: http://www.phylo.org/sub_sections/portal. (Accessed Aug 4, 2009). (Archived by WebCite® at: http://www.webcitation.org/5 imQlJeQa)

Nylander, J. A. A. (2004). *MrModeltest v2: Program Distributed By The Author*. Uppsala: Evolutionary Biology Centre, Uppsala University.

Petersen, J. H. (1996). *The Danish Mycological Society's colour-chart*. Greven: Foreningen til Svampekundskabens Fremme, 1–6.

Posada, D., and Crandall, K. A. (1998). Modeltest: testing the model of DNA substitution. *Bioinformatics* 14, 817–818. doi: 10.1093/bioinformatics/14.9.817

Rambaut, A. (2018). *Molecular Evolution, Phylogenetics and Epidemiology. FigTree ver. 1.4.4 Software.* Available online at: http://tree.bio.ed.ac.uk/software/figtree/ (accessed March, 2022).

Ronquist, F., and Huelsenbeck, J. P. (2003). Mrbayes 3: bayesian phylogenetic inference under mixed models. *Bioinformatics* 19, 1572–1574. doi: 10.1093/bioinformatics/btg180

Smith, S. Y., Currah, R. S., and Stockey, R. A. (2004). Cretaceous and Eocene poroid hymenophores from Vancouver Island, British Columbia. *Mycologia* 96, 180–186. doi: 10.2307/3762001

Spirin, V., Volobuev, S., Viner, I., Miettinen, O., Vlasák, J., Schoutteten, N., et al. (2021). On Sistotremastrum and similar-looking taxa (*Trechisporales*, Basidiomycota). *Mycol. Prog.* 20, 453–476. doi: 10.1007/s11557-021-01682-z

Stamatakis, A. (2014). RAxML Version 8: a tool for phylogenetic analyses and post analyses of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033

Sulistyo, B. P., Larsson, K. H., Haelewaters, D., and Ryberg, M. (2021). Multigene phylogeny and taxonomic revision of Atheliales s.l.: reinstatement of three families and one new family, *Lobuliciaceae* fam. nov. *Fungal Biol.* 125, 239–255. doi: 10.1016/j.funbio.2020.11.007

Swofford, D. L. (2002). *PAUP*: Phylogenetic Analysis Using Parsimony (*And Other Methods). Version 4.0b10*. Sunderland, MA: Sinauer Associates.

Taylor, T. N., Hass, H., and Kerp, H. (1999). The oldest fossil ascomycetes. *Nature* 399:648. doi: 10.1038/21349

Taylor, T. N., Hass, H., Kerp, H., Krings, M., and Hanlin, R. T. (2005). Perithecial ascomycetes from the 400 million year old Rhynie chert: an example of ancestral polymorphism. *Mycologia* 97, 269–285. doi: 10.1080/15572536.2006.11832862

Telleria, M. T., Melo, I., Dueñas, M., Larsson, K. H., and Martín, P. (2013). Molecular analyses confirm *Brevicellicium* in *Trechisporales*. *IMA Fungus* 4, 21–28. doi: 10.5598/imafungus.2013.04.01.03

Thiers, B. (2018). *Index Herbariorum: A Global Directory Of Public Herbaria And Associated Staff*. New York, NY: New York Botanical Garden's Virtual Herbarium.

Vilgalys, R., and Hester, M. (1990). Rapid genetic identification and mapping of enzymatically amplified ribosomal DNA from several *Cryptococcus* species. *J. Bacteriol.* 172, 4238–4246. doi: 10.1128jb.172.8.4238-4246.1990

Wang, X. W., Tom, W. M., Liu, X. L., and Zhou, L. W. (2021). Towards a natural classification of hyphodontia sensu lato and the trait evolution of basidiocarps within hymenochaetales (Basidiomycota). *J. Fungi* 7:478. doi: 10.3390/jof7060478

Wangsawat, N., Ju, Y. M., Phosri, C., Whalley, A. J. S., and Suwannasai, N. (2021). Twelve new taxa of xylaria associated with termite nests and soil from northeast thailand. *Biology* 10:575. doi: 10.3390/biology10070575

White, T. J., Bruns, T., Lee, S., and Taylor, J. (1990). "Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics," in *PCR Protocols, A Guide To Methods And Applications*, eds M. A. Innis, D. H. Gelfand, J. J. Sninsky, and T. J. White (New York, NY: Academic Press), 315–322. doi: 10.1016/B978-0-12-372180-8.50042-1

Zhao, C. L., Cui, B. K., Song, J., and Dai, Y. C. (2015). Fragiliporiaceae, a new family of Polyporales (Basidiomycota). *Fungal Divers.* 70, 115–126. doi: 10.1007/s13225-014-0299-0

Zhao, R. L., Li, G. J., Sanchez-Ramirez, S., Stata, M., Yang, Z. L., Wu, G., et al. (2017). A six-gene phylogenetic overview of Basidiomycota and allied phyla with estimated divergence times of higher taxa and a phyloproteomics perspective. *Fungal Divers.* 84, 43–74. doi: 10.1007/s13225-017-0381-5

Zhao, W., and Zhao, C. L. (2021). The phylogenetic relationship revealed three new wood-inhabiting fungal species from genus *Trechispora*. *Front. Microbiol.* 12:650195. doi: 10.3389/fmicb.2021.650195

Zong, T. K., Liu, C. M., Wu, J. R., and Zhao, C. L. (2021). Trechispora daweishanensis and *T. xantha* spp. nov. (*Hydnodontaceae*, *Trechisporales*) found in Yunnan Province of China. *Phytotaxa* 479, 147–159. doi: 10.11646/phytotaxa.479.2.1

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.

frontiers | Frontiers in Microbiology

ORIGINAL RESEARCH
published: 28 April 2022
doi: 10.3389/fmicb.2022.860308

Streptomyces marincola sp. nov., a Novel Marine Actinomycete, and Its Biosynthetic Potential of Bioactive Natural Products

Songbiao Shi[1,3], Linqing Cui[1,3], Kun Zhang[1,3], Qi Zeng[1,3], Qinglian Li[1,2], Liang Ma[1,2], Lijuan Long[1,2] and Xinpeng Tian[1,2]**

[1] Key Laboratory of Tropical Marine Bio-Resources and Ecology, Chinese Academy of Sciences, Guangdong Key Laboratory of Marine Materia Medica, RNAM Center for Marine Microbiology, Sanya Institute of Oceanology, South China Sea Institute of Oceanology, Chinese Academy of Sciences, Guangzhou, China, [2] Southern Marine Science and Engineering Guangdong Laboratory, Guangzhou, China, [3] University of Chinese Academy of Sciences, Beijing, China

OPEN ACCESS

Edited by:
Qi Zhao,
University of Science and Technology Liaoning, China

Reviewed by:
Chung Thanh Nguyen,
Hanoi Open University, Vietnam
Dipesh Dhakal,
University of Florida, United States

***Correspondence:**
Lijuan Long
longlj@scsio.ac.cn
Xinpeng Tian
xinpengtian@scsio.ac.cn

Specialty section:
This article was submitted to Systems Microbiology, a section of the journal Frontiers in Microbiology

Received: 22 January 2022
Accepted: 24 February 2022
Published: 28 April 2022

Citation:
Shi S, Cui L, Zhang K, Zeng Q, Li Q, Ma L, Long L and Tian X (2022) Streptomyces marincola sp. nov., a Novel Marine Actinomycete, and Its Biosynthetic Potential of Bioactive Natural Products. Front. Microbiol. 13:860308. doi: 10.3389/fmicb.2022.860308

Marine actinomycetes are an important source of antibiotics, but many of them are yet to be explored in terms of taxonomy, ecology, and functional activity. In this study, two marine actinobacterial strains, designated SCSIO 64649^T and SCSIO 03032, were isolated, and the potential for bioactive natural product discovery was evaluated based on genome mining, compound detection, and antimicrobial activity. Phylogenetic analysis of the 16S rRNA gene sequences showed that strain SCSIO 64649^T formed a single clade with SCSIO 03032 (similarity 99.5%) and sister clades with the species *Streptomyces specialis* DSM 41924^T (97.1%) and *Streptomyces manganisoli* MK44^T (96.8%). The whole genome size of strain SCSIO 64649^T was 6.63 Mbp with a 73.6% G + C content. The average nucleotide identity and digital DNA–DNA hybridization between strain SCSIO 64649^T and its closest related species were well below the thresholds recommended for species delineation. Therefore, according to the results of polyphasic taxonomy analysis, the strains SCSIO 64649^T and SCSIO 03032 are proposed to represent a novel species named *Streptomyces marincola* sp. nov. Furthermore, strains SCSIO 64649^T and 03032 encode 37 putative biosynthetic gene clusters, and *in silico* analysis revealed that this new species has a high potential to produce unique natural products, such as a novel polyene polyketide compounds, two mayamycin analogs, and a series of post-translationally modified peptides. In addition, other important bioactive natural products, such as heronamide F, piericidin A1, and spiroindimicin A, were also detected in strain SCSIO 64649^T. Finally, this new species' metabolic crude extract showed a strong antimicrobial activity. Thanks to the integration of all these analyses, this study demonstrates that the novel species *Streptomyces marincola* has a unique and novel secondary metabolite biosynthetic potential that not only is beneficial to possible marine hosts but that could also be exploited for industrial applications.

Keywords: *Strepomyces marincola* sp. nov., biosynthetic potential, polyphasic taxonomy, new bioactive secondary metabolites, marine actinobacteria

INTRODUCTION

In recent years, the search and discovery of novel microbes producing new active secondary metabolites have been urgently needed to counter and reverse the spread of new and emerging diseases and antibiotic-resistant pathogens in recent years (Payne et al., 2007). Actinobacteria are prolific producers of antibiotics and important suppliers to the pharmaceutical industry as they can produce a wide variety of secondary metabolites (van der Meij et al., 2017). Actinobacteria are also widely distributed throughout marine habitats, which, different from terrestrial habitats, are characterized by highly dynamic pressure, salinity, pH, dissolved oxygen, and light intensity. Marine actinobacteria have been attracting particular attention as new producers of novel antibiotics such as salinosporamide (Feling et al., 2003), ilamycin (Ma J. et al., 2017), and anticancer agents with unusual properties (Manivasagan et al., 2014a; Hassan and Shaikh, 2017; Davies-Bolorunduro et al., 2021). The genus *Streptomyces* alone accounts for a remarkable 80% of the actinobacterial natural products and, therefore, has unrivaled biosynthetic capacity in the microbial world (Manivasagan et al., 2014b; Hu et al., 2015). In particular, marine *Streptomyces* are known to produce broad-range-active natural products with immunosuppressant, antifungal, anticancer, antiparasitic, or antithrombotic activities (Ser et al., 2017; Wang et al., 2021), such as pactamides (Saha et al., 2017) and streptoseomycin (Zhang et al., 2019). Thus, the isolation and characterization of novel marine streptomycetes species are important for identifying new potential bioactive compounds.

The strain *Streptomyces* sp. SCSIO 03032 was isolated from a deep-sea sediment sample of the Indian Ocean in 2012. This strain was reported to produce α-pyridone antibiotics (piericidins A1/E1) (Chen et al., 2014), new polycyclic macrolactams (heronamides D–F) (Zhang et al., 2014a; Zhu et al., 2015), and unusual bisindole alkaloids (spiroindimicins A–D, G/H, indimicins A–G, and lynamicins A/D/F/G) (Zhang et al., 2012; Saurav et al., 2014; Zhang et al., 2014b; Ma L. et al., 2017; Liu et al., 2019) with excellent cytotoxic activity and antimicrobial activity. The intact biosynthetic gene clusters (BGCs) of piericidins (Chen et al., 2014) and heronamides (Zhu et al., 2015) and partial BGCs of spiroindimicins/indimicins/lynamicins (Ma L. et al., 2017; Liu et al., 2019) have also been reported from this strain. Interestingly, while investigating coral symbiotic microbial diversity, we isolated strain SCSIO 64649^T, which was revealed to have 99.5% 16S rRNA gene sequence similarity with strain SCSIO 03032. This discovery was key as, while strain SCSIO 03032 is a producer of highly active compounds, its taxonomic status is undetermined. Therefore, this study was designed to establish the taxonomic status of strains SCSIO 64649^T and SCSIO 03032 using a polyphasic taxonomic approach. We identify the strains as one, same new species named *Streptomyces marincola* sp. nov. and evaluate its biosynthesis potential for novel natural product discovery through genome mining, compound detection, and antimicrobial activity evaluation.

Abbreviations: DPG, diphosphatidylglycerol; PG, phosphatidylglycerol; PE, phosphatidylethanolamine; GL, glycerol lipid; PIM, phosphatidylinositol mannoside; PI, phosphatidylinositol; PL, unidentified phospholipid; L, unidentified lipid. R2A, Reasoner's 2 agar; TSA, tryptic soy agar; NA, nutrient agar; ISP, International *Streptomyces* Project medium; ANI, average nucleotide identity; dDDH, digital DNA–DNA hybridization; high-performance liquid chromatography (HPLC); BGC, biosynthetic gene cluster; PKS, polyketide synthase; NRPS, non-ribosomal peptides synthase; RiPPs, ribosomally synthesized and post-translationally modified peptides.

MATERIALS AND METHODS

Isolation and Maintenance

Strain SCSIO 64649^T was isolated from colonies of *Favites* sp. scleractinian corals collected at a depth of 2 m from the South China Sea off the Luhuitou Peninsula, Hainan Province, China (18.50°N, 109.46°E). The coral samples were washed with sterile natural seawater and then processed according to Zhou et al. (2020). The samples were diluted 100 times and plated onto 1/10 tryptic soy agar (TSA) prepared with natural seawater. After inverted culturing at 28°C for 15 days, strain SCSIO 64649^T was selected and purified by routine cultivation on 2216E medium at 28°C.

Strain SCSIO 03032 had been isolated using a modified ISP 2 medium from a deep-sea sediment sample collected at a depth of –3,412 m from the Bay of Bengal in the Indian Ocean (9.988°N, 87.995°E) (Zhang et al., 2012). Strains SCSIO 64649^T and SCSIO 03032 were preserved in glycerol suspensions (30%, v/v) at –80°C. The strain *S. specialis* DSM 41924^T was obtained from the Marine Culture Collection of China (MCCC) and cultured under the same conditions as the reference strain.

Phylogenetic Analyses

Genomic DNA was extracted using a genomic DNA extraction kit (QIAGEN, Germany), and the amplification of the 16S rRNA sequence was carried out as previously described in Li et al. (2007). The identification and calculation of pairwise 16S rRNA gene sequence similarity were determined using EzBioCloud.[1] Phylogenetic relationships were investigated using the neighbor-joining (Saitou and Nei, 1987), maximum-likelihood (Felsenstein, 1981), and maximum-parsimony methods (Fitch, 1971) on the MEGA 11 program with a bootstrap value of 1,000 resampling replicates (Felsenstein, 1985).

Whole-Genome Sequencing, Analysis, and Biosynthetic Evaluation

The complete genome of strain SCSIO 64649^T was sequenced on a PacBio RS II platform by the Tianjin Biochip Corporation (Tianjin, China). *De novo* genome assembly was carried out following a hierarchical genome-assembly process (HGAP) (Chin et al., 2013), using HGAP4 (Pacific Biosciences, SMRT Link V6.0). The phylogenomic tree was reconstructed using 120 marker genes with the GTDB-Tk software toolkit (Chaumeil et al., 2019). The genomes were annotated by the Rapid Annotation using Subsystem Technology (RAST version 2.0) (Overbeek et al., 2014). Barrnap was used to predict rRNA information, and tRNAscan was used to predict the tRNAs (Chan and Lowe, 2019). CRISPR arrays and their associated

[1] https://www.ezbiocloud.net

proteins were searched by CRISPRCasFinder (Couvin et al., 2018). Prophage Hunter was used for finding active prophages from the whole genome (Song et al., 2019). Metabolic pathways in a single bacterium were reconstructed using the online tool KEGG Mapper (Kanehisa and Sato, 2020). The whole genome and orthologous genes among *Streptomyces* species were compared using OrthoVenn2 (Xu et al., 2019). Genomic island prediction was performed by using IslandViewer 4 (Bertelli et al., 2017). BGCs of secondary metabolisms were predicted in antiSMASH web service (version 6.0) (Blin et al., 2021). The compound structures were predicted based on genome sequence using PRISM4 (Skinnider et al., 2020). The average nucleotide identity (ANI) values were calculated using ChunLab's online ANI calculator (Yoon et al., 2017). The estimated digital DNA–DNA hybridization (dDDH) values were calculated using the Genome-to-Genome Distance Calculator (GGDC 2.1), and Formula 2 was used as recommended for the calculation of dDDH (Meier-Kolthoff et al., 2013). The estimation of average amino acid identity (AAI) was determined using the tool AAI calculator.[2]

Cultural and Phenotypic Properties

After incubation on ISP 2 at 28°C for 14 days, cell morphology was observed using a scanning electron microscope (Hitachi s-3400N). Cultural characteristics were tested on ISP 1, ISP 2, ISP 3, ISP 4, ISP 5, ISP 6, ISP 7 agar, R2A agar, Czapek's agar, TSA, 2216E, and nutrient agar (NA) for 2 weeks at 28°C. The color of aerial and substrate mycelium and soluble pigments was determined using the ISCC-NBS color charts. Anaerobic growth was determined after 4 weeks of incubation at 28°C using the GasPak EZ Anaerobe Pouch Systems (BD). Growth at different temperatures, salinities, and pH was tested in ISP 2 broth as in Wang et al. (2018). Catalase activity was determined as the production of bubbles after the addition of 3% (v/v) hydrogen peroxide (H_2O_2). Tests for hydrolysis of starch, cellulose, gelatin, and Tweens (20, 40, 60, and 80) and H_2S production, coagulation, and peptonization of milk were performed using the methods previously described (Gonzalez et al., 1978). Biochemical properties and enzyme activities were tested using API 20NE and API ZYM kits (bioMérieux, France) according to the manufacturer's instructions. The ability to metabolize sole sources of carbon and nitrogen was tested with Biolog GEN III microplates. The susceptibility to antimicrobial agent was determined by the disk diffusion method (Bauer et al., 1966) with the following antibiotics (microgram per disk, Oxoid, United Kingdom): amikacin (30), amoxicillin (10), ampicillin (10), chloramphenicol (30), ciprofloxacin (5), erythromycin (15), gentamicin (10), lincomycin (2), neomycin (30), norfloxacin (10), novobiocin (5), penicillin G (10), rifampicin (5), streptomycin (10), kanamycin (30), tetracycline (30), tobramycin (10), and vancomycin (30).

Chemotaxonomy

The cell biomass was collected for chemotaxonomic analyses after growing the strains on ISP 2 at 28°C for 1 week. Fatty acids from strains SCSIO 64649^T and SCSIO 03032 and the reference strain were extracted and analyzed using the standard protocol of the MIDI system (Sherlock version 6.1; database TSBA6). Polar lipids were examined and identified by two-dimensional TLC using silica gel 60 plates (Merck) with four dye agents (Minnikin et al., 1984). Menaquinones were extracted from freeze-dried biomass, purified, and analyzed by high-performance liquid chromatography (HPLC) (Collins et al., 1977) using an Agilent TC-C18 column (250 × 4.6 mm, 5 μm). The cell-wall diamino acid was analyzed from whole-cell hydrolyzates as previously described (Tang et al., 2009). For sugar analysis, cell walls were hydrolyzed in 0.5 M H_2SO_4 at 100°C for 2 h and analyzed by TLC on cellulose plates (Whiton et al., 1985).

Identification of Bioactive Compounds

To identify the compounds produced by strains SCSIO 64649^T and SCSIO 03032, both strains were fermented in ISP 3 and ISP 4 media with 3% sea salt, in a 250 ml Erlenmeyer flask, and cultivated on a rotary shaker (200 rpm/min) at 28°C for 7, 9, and 11 days. For each fermentation sample, 5 ml was extracted with 10 ml butanone, and the crude extracts were dissolved with DMSO after evaporation. The extracts were analyzed by HPLC and LC-HR-MS after filtration through 0.2 μm syringe filters. Ten microliter per samples was injected on an Agilent 1260 HPLC equipped with a diode array detector (DAD) and Agilent TC-C18 Column (250 × 4.6 mm, 5 μm). The HPLC gradient was as follows: UV detection at 254 nm; solvent A, acetonitrile (10%) in water with formic acid (0.1%); solvent B, acetonitrile (90%) in water; 5–100% B (0–20 min); 100% B (20–21 min); 100%–5% B (21–22 min); and 5% B (22–30 min) with a flow rate of 1.0 ml/min. ESI-MS data were measured with an LCQ Deca XP HPLC/MS spectrometer (Bruker).

In vitro Antimicrobial Activity Assay

The antimicrobial activity of crude extracts was evaluated by the agar well diffusion method against seven indicator microorganisms: *Staphylococcus aureus* ssp. *aureus* CGMCC 1.2386, *Bacillus subtilis* ssp. *spizizenii* CGMCC 1.1849, *Escherichia coli* CGMCC 1.2385, *Pseudomonas aeruginosa* DSM 50071, *Candida* albicans CGMCC 2.2086, *Aspergillus niger* CICC 2487, and *Micrococcus luteus* CGMCC 1.2299. The butanone extracts were dissolved in DMSO for antimicrobial activity detection. After incubating the plate at 28°C for 48 h, the antimicrobial activity was determined by the inhibition zone around the samples.

RESULTS AND DISCUSSION

Phylogenetic Analyses

The nearly complete 16S rRNA gene sequence of strain SCSIO 64649^T was obtained (1,432 bp; GenBank accession number MZ889118). Based on 16S rRNA gene sequence comparison, strain SCSIO 64649^T showed a high sequence similarity with SCSIO 03032 (99.5%), the published species *S. specialis* DSM 41924^T (97.1%), *S. manganisoli* MK44^T (96.8%), and *S. sediminis* MKSP12^T (96.6%) and less than 96.5% similarity with other

[2]http://enve-omics.ce.gatech.edu/

species. Likewise, strain SCSIO 03032 was most similar to the same type species with minor differences in the percentage similarity: *S. specialis* DSM 41924^T (97.7%), *S. sediminis* MKSP12^T (97.1%), and *S. manganisoli* MK44^T (97.0%). The 16S rRNA gene similarity between SCSIO 64649^T and SCSIO 03032 is higher than the threshold of 98.65% for differentiating two species (Kim et al., 2014), while the similarities between these two strains with other known species are below this threshold, therefore supporting the idea that the strains represent the same species. Phylogenetic analysis showed that the strains SCSIO 64649^T and SCSIO 03032 clustered together with *S. sediminis* MKSP12^T and formed a separate phylogenetic branch parallel with *S. specialis* DSM 41924^T and *S. manganisoli* MK44^T. This topology was supported by all three algorithms employed (**Figure 1**). The phylogenomic tree showed that strain SCSIO 64649^T stably clustered with SCSIO 03032 and their nearest-neighbor, *S. specialis* DSM 41924^T (**Figure 2**). Strain SCSIO 64649^T showed an ANI value of 80.2%, a dDDH value of 23.8%, and an AAI value of 74.5% with *S. specialis* DSM 41924^T and was followed by *S. hoynatensis* KCTC 29097^T (79.0, 22.4, and 73.3%, respectively) (**Table 1**). Nearly identical values were observed between strain SCSIO 03032 and *S. specialis* DSM 41924^T. The lower values of ANI, dDDH, and AAI between these two strains and other closely related *Streptomyces* species are also shown in **Table 1**. These values are all far from the recommended similarity thresholds (ANI < 95–96%, dDDH < 70%, and AAI < 95%) (Richter and Rossello-Mora, 2009; Meier-Kolthoff et al., 2013; Konstantinidis et al., 2017). However, strains SCSIO 64649^T and SCSIO 03032 revealed high ANI, dDDH, and AAI similarity (96.6, 84.9, and 96.4%, respectively) and therefore fall within the recommended thresholds. Given all this evidence, we suggest that strains SCSIO 64649^T and SCSIO 03032 represent the same novel species in the genus *Streptomyces*.

Genome Characteristics

The complete genome of strain SCSIO 64649^T was composed of one linear chromosome (6,629,020 bp, GenBank accession number CP084541) with G + C content of 73.6%, 5,774 genes, 5,567 protein-coding genes, 133 pseudo genes, 60 tRNA genes, and 15 rRNA genes (5 23S, 5 5S, and 5 16S) (**Supplementary Table 1**). The genome of strain SCSIO 03032 has been previously described (Ma et al., 2021). The 16S rRNA gene sequences of SCSIO 64649^T obtained from its genome and Sanger sequencing method showed 99.3% sequence similarity. Eleven and seven genomic islands were identified in the genome of strains SCSIO 64649^T and SCSIO 03032, respectively (**Supplementary Table 2**). Evidence of genomic islands linking secondary metabolism to functional adaptation has been provided in marine actinobacteria *Salinispora* (Penn et al., 2009), which may explain the cosmopolitan distribution of SCSIO 64649^T and SCSIO 03032. Three prophage-like sequences were identified in genomes of strains SCSIO 64649^T and SCSIO 03032, and one out of three was active. The closest elements related to *Mycobacterium* phage Hammy were found in SCSIO 64649^T, while the *Xanthomonas* phage Xoo-sp2 was found in SCSIO 03032 (**Supplementary Table 3**). A total of 37 secondary metabolite BGCs with eight new clusters were discovered in strains SCSIO 64649^T and SCSIO 03032 genome sequences. These secondary metabolite BGCs mainly covered polyketide synthase (PKS), terpene, siderophore, non-ribosomal peptide synthase (NRPS), thiopeptide, lanthipeptide, lasso peptide, indole, and 15 unknown gene clusters. The BGC maps of strains SCSIO 64649^T and SCSIO 03032 were drawn (**Supplementary Figure 1**). Strain SCSIO 64649^T encoded for 10 CRISPR arrays and 15 Cas-proteins, including Cas1-6, Csh2, and the Cse3-5 family. A novel knock-in CRISPR-based approach introducing the *kasO**p promoter cassette to drive expression of putative BGCs was successfully used for *S. viridochromogenes* (Zhang M. M. et al., 2017), leading to expression and production of novel secondary metabolites. Since the presence of putative CRISPR arrays and all known cascade proteins in two genomes implicates the activity of the CRISPR/Cas immune system in the two strains, in the future, the activation of this species' silent and unusual BGCs could be possible using the above-mentioned CRISPR-based tools.

Phenotypic Characteristics

Strains SCSIO 64649^T and SCSIO 03032 are Gram-stain-positive and aerobic actinomycetes with extensively branched substrate mycelia and aerial hyphae, which differentiate into spiral spore chains consisting of elliptical or short-rod spores (∼1.0–1.3 × 0.7–0.9 μm) with smooth surfaces (**Figure 3**). The strains grow well on ISP 2, ISP 4, ISP 7, NA, and 2216E and moderately well on ISP 1, ISP 3, ISP 5, ISP 6, TSA, and CA media (**Supplementary Table 4** and **Supplementary Figure 2**). The colors of the aerial and substrate mycelia are media dependent. The diffusible melanin is observed only on the ISP 2 medium. Growth of strain SCSIO 64649^T occurs at 15–40°C (optimal 28°C), pH 6–9 (optimal 7–8), and up to 9% NaCl (optimal 4%), different from *S. specialis* DSM 41924^T in growth conditions (**Table 2**). Strain SCSIO 64649^T can be easily distinguished from the type strains of its closest neighbors through its phenotypic properties (**Table 2** and **Supplementary Table 5**).

Chemotaxonomic Analyses

The cellular fatty acids of strain SCSIO 64649^T detected (>10%) were iso-$C_{16:0}$ (37.7%) and $C_{12:0}$ (13.1%). While the major fatty acids of *S. specialis* DSM 41924^T are iso-$C_{16:0}$ (33.2%) and anteiso-$C_{17:0}$ (19.3%). And also the difference between them is in the amounts of $C_{18:3}$ ω6*c*, iso-$C_{16:1}$ H, etc. (**Table 3**). The polar lipid profiles of strain SCSIO 64649^T comprised diphosphatidylglycerol, phosphatidylglycerol, phosphatidylethanolamine, phosphatidylinositol mannoside, phosphatidylinositol, glycerol lipid, and six unidentified phospholipids. Strain SCSIO 64649^T differed from *S. specialis* DSM 41924^T in its polar lipid composition by having four unidentified phospholipids (PL3-6) and missing unidentified lipids (L1-2) (**Supplementary Figure 3**). *LL*-2,6-Diaminopimelic acid was identified as the cell-wall diamino acid of strains SCSIO 64649^T and DSM 41924^T. The cell sugars identified in both strains were galactose, glucose, xylose, and ribose. The predominant menaquinone of strain SCSIO 64649^T was MK-10(H_4) (75.5%) and MK-10(H_6) (16.1%), and minor amounts

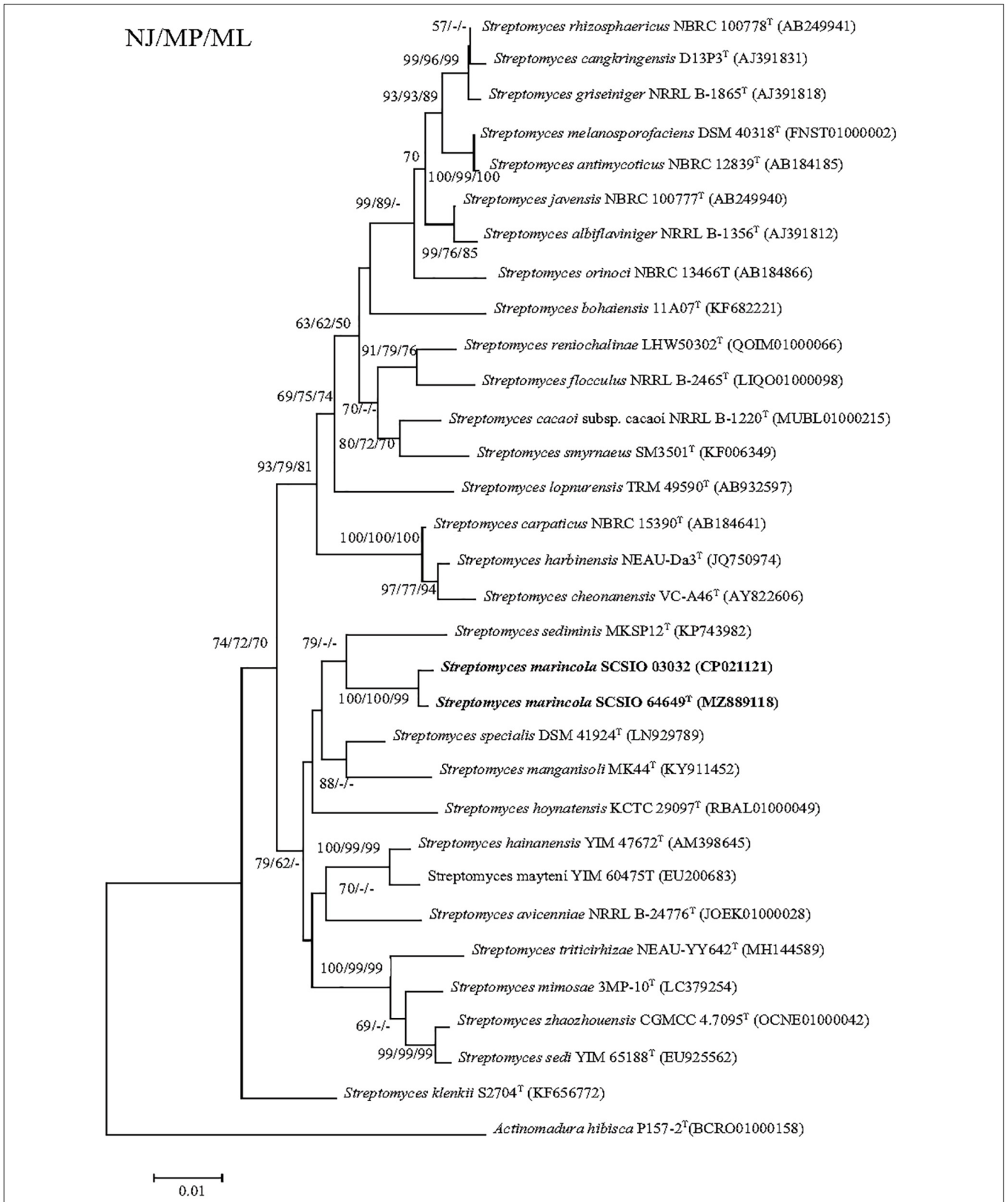


FIGURE 1 | Neighbor-joining tree showing phylogenetic relationships between strains SCSIO 64649^T, SCSIO 03032, and related *Streptomyces* species, based on 16S rRNA gene sequences. *Allostreptomyces psammosilenae* YIM DR4008^T (KX689228) was added as an outgroup. Bootstrap values are shown from left to right for neighbor-joining, maximum-likelihood, and maximum-parsimony trees based on 1,000 replications. Bar, 0.01 sequence divergence.

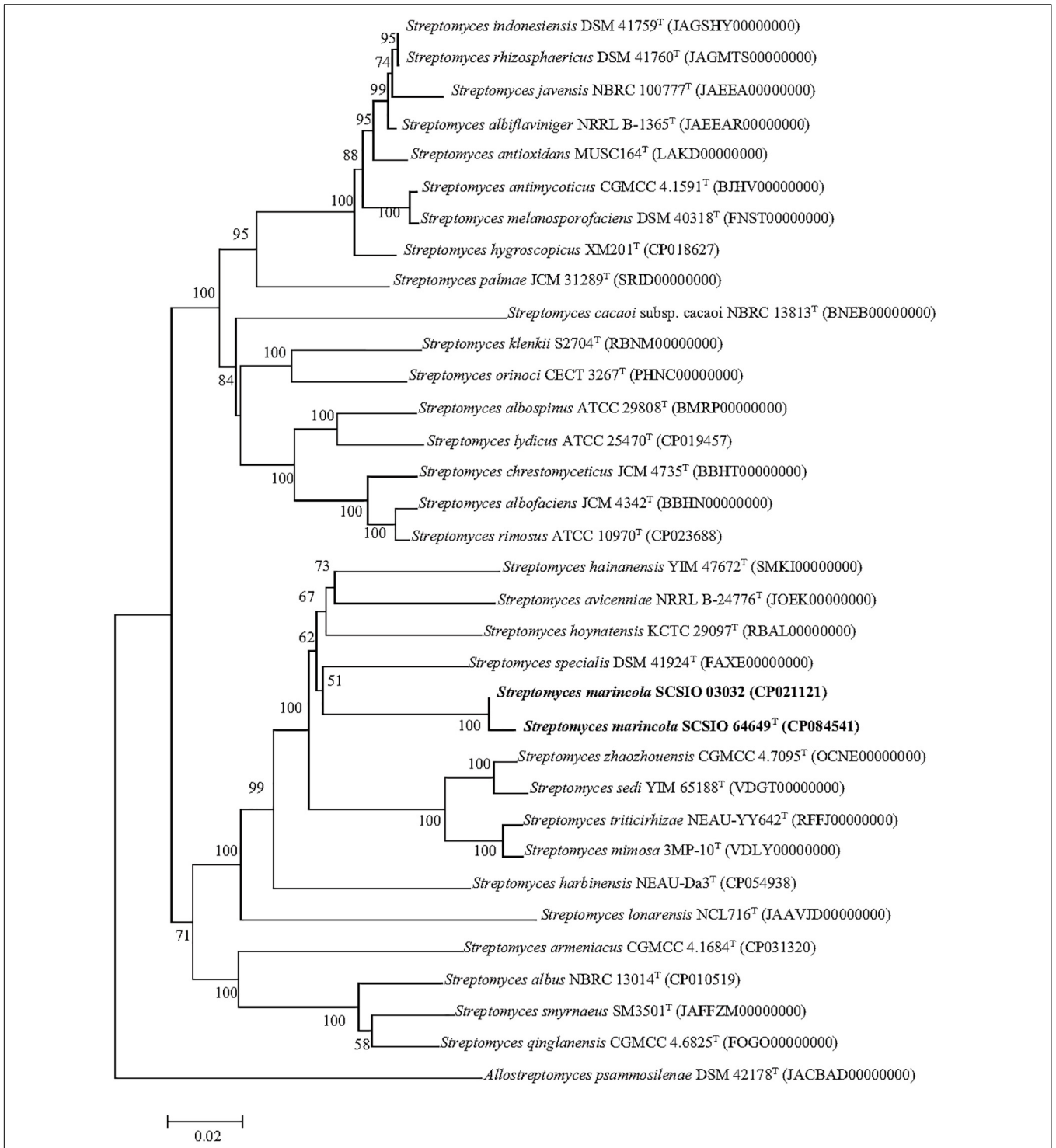


FIGURE 2 | Phylogenetic analysis based on genome sequences of strains SCSIO 64649^T, SCSIO 03032, and related *Streptomyces* species. The RAxML tree was calculated with the PhyloPhlAn software. *Allostreptomyces psammosilenae* YIM DR4008^T was added as an outgroup. Bar, 0.02 sequence divergence.

(<3%) of MK-9(H_4), MK-9(H_6), and MK-10 (H_8) were also detected. All the above menaquinones were also detected in *S. specialis* DSM 41924^T; however, the proportions of MK-10(H_4) and MK-10(H_6) were different (79.6 and 16.0%, respectively).

Comparative Genome Analysis

The core genes and specific genes of the two isolates and the species *S. specialis* DSM 41924^T and *S. hoynatensis* KCTC 29097^T were determined. With OrthoVenn2, a total of 2,676

TABLE 1 | ANI, dDDH values, and AAI found between isolates and their closest species.

Strains	ANI (%)		dDDH (%)		AAI (%)	
	SCSIO 64649^T	SCSIO 03032	SCSIO 64649^T	SCSIO 03032	SCSIO 64649^T	SCSIO 03032
SCSIO 64649^T	100	96.6	100	84.9	100	96.4
SCSIO 03032	96.6	100	84.9	100	96.4	100
S. specialis DSM 41924^T	80.2	80.3	23.8	23.7	74.5	74.7
S. hoynatensis KCTC 29097^T	79.0	79.0	22.4	22.5	73.3	73.3
S. hainanensis YIM 47672^T	79.2	79.2	22.8	22.9	72.8	73.1
S. klenkii KCTC 29202^T	76.5	79.0	21.2	21.2	65.2	65.2

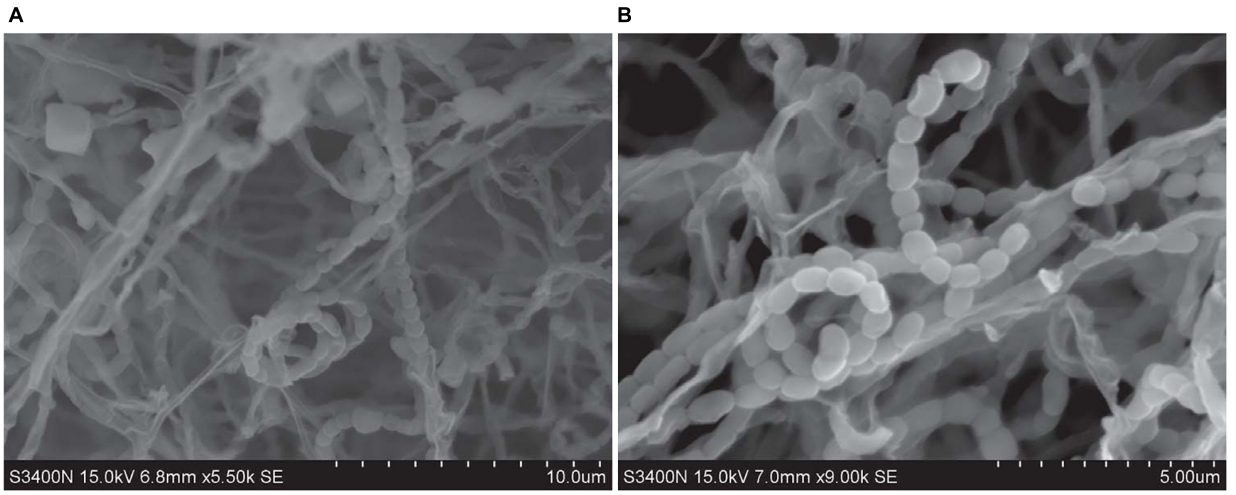


FIGURE 3 | Scanning transmission electron microscope images show the cell morphology of strains SCSIO 64649^T **(A)** and SCSIO 03032 **(B)** after incubation on ISP 2 medium for 14 days at 28°C.

core genes were found in the four strains (**Figure 4**), and 10 genes were unique to SCSIO 64649^T. These genes encoded functional minor molecules, such as oxidoreductase, transferase, and hydrolase activity. A comparison of the orthologous gene numbers revealed that SCSIO 64649^T shares 90.5% (4,748/5,249) sequence similarity with SCSIO 03032 and a lower proportion with *S. specialis* DSM 41924^T (63.0%) and *S. hoynatensis* KCTC 29097^T (60.4%). These results help to separate strains SCSIO 64649^T and SCSIO 03032 from closely related type strains and identify them as the same, novel species. For a comprehensive genome comparison, a synteny block analysis was performed on highly conserved large segment sequences in strains SCSIO 64649^T, SCSIO 03032, *S. specialis* DSM 41924^T, and *S. hoynatensis* KCTC 29097^T by a progressive mauve tool. Although the four strains shared many locally collinear blocks with their reference genomes, all exhibited large-scale genome rearrangement (**Figure 5**). The positions of the locally collinear blocks highlight the complex evolutionary history of these strains. A total of 37 putative secondary metabolite secondary metabolite BGCs were detected in strains SCSIO 64649^T and SCSIO 03032 (**Supplementary Table 6**). The BGCs identified share homology to 22 known gene clusters with known metabolic products, such as the compatible solute ectoine, siderophore desferrioxamine B, the carotenoid light-harvesting pigment isorenieratene, and terpene hopene. In addition, SCSIO 64649^T contained some unique gene clusters, including one encoding for ketomemicin B3/B4 (Kawata et al., 2017). The BGCs of DSM 41924^T and KCTC 29097^T were also analyzed, and the type and number of coding gene clusters were markedly different from strains SCSIO 64649^T and SCSIO 03032 (**Supplementary Figure 4**). These results further illustrate the diverging metabolic potential of this new *Streptomyces* species.

Biosynthetic Potential of the New *Streptomyces* Species

To evaluate the secondary metabolite biosynthetic potential of the new species, the genomes of strains SCSIO 64649^T and SCSIO 03032 were analyzed with antiSMASH. A total of 32 and 29 putative secondary metabolite BGCs were detected and accounted for 19.7 and 14.4% of their genomes, respectively. This analysis revealed that there is a higher percentage of the genome dedicated to secondary metabolites in this species compared with the representative actinomycetes strain *S. coelicolor* (4.5%) (Cimermancic et al., 2014). Strain SCSIO 64649^T comprises 10 different types of BGCs, including those coding for ribosomally synthesized and post-translationally modified peptides (RiPPs) (thiopeptide, lanthipeptide, lasso

TABLE 2 | Phenotypic properties that distinguish strains SCSIO 64649^T and SCSIO 03032 from their closest phylogenomic relatives.

Characteristics	SCSIO 64649^T	SCSIO 03032	*S. specialis* DSM 41924^T	*S. manganisoli* MK44^{T*}
Sources	**Stony coral**	**Sediment**	**Soil**	**Soil**
pH range	6.0–9.0	6.0–9.0	6.0–10.0	5.0–9.0
Temperature range (°C)	15–40	15–40	25–40	10–40
NaCl range (%)	0–9	0–5	0–7	0–5
Gelatin liquefaction	+	+	+	−
Hydrolysis of aesculin	+	+	−	−
Carbon source utilization				
Myo-Inositol	−	+	+	−
D-Arabinose	−	+	+	+
Cellobiose	−	+	−	+
D-Fructose	−	+	−	−
D-Galactose	+	−	−	−
D-Mannose	−	−	−	+
D-Ribose	−	−	+	−
D-Rhamnose	−	+	−	+
Lactose	−	−	+	+
Maltose	+	+	−	+
Xylitol	−	−	+	−
G+C content (%)	73.6	73.5	72.8	75.7

*All strains are positive for hydrolysis of starch and Tweens 40 and 80 but negative for nitrate reduction, H_2S production, milk coagulation, and peptonization. All data come from this study (except data marked with *). +, positive; –, negative.*
**Data from Mo et al. (2018).*

TABLE 3 | Cellular fatty acid profiles of strain SCSIO 64649^T and its neighbors in genus *Streptomyces*.

Fatty acid	SCSIO 64649^T	SCSIO 03032	*S. specialis* DSM 41924^T	*S. manganisoli* MK44$^{T\&}$
$C_{9:0}$	2.0	ND	TR	ND
$C_{12:0}$	**13.1**	ND	2.8	1.0
Iso-$C_{15:0}$	2.3	1.0	2.8	2.0
Anteiso-$C_{15:0}$	2.2	TR	1.9	1.0
$C_{15:1}$ ω6*c*	1.1	TR	TR	0.5
Iso-$C_{16:1}$ G	ND	**25.2**	ND	**14.4**
Iso-$C_{16:1}$ H	TR	ND	6.0	ND
Iso-$C_{16:0}$	**37.7**	**43.2**	**33.2**	**51.3**
$C_{16:0}$	6.2	1.5	2.6	ND
Anteiso-$C_{17:1}$ ω9*c*	5.2	5.5	9.2	4.4
Iso-$C_{17:0}$	2.0	1.9	3.6	1.6
Anteiso-$C_{17:0}$	7.3	6.2	**19.3**	**7.0**
$C_{17:0}$ cyclo	5.1	1.0	1.7	0.4
Iso-$C_{18:0}$	TR	2.0	3.6	0.2
Iso-$C_{18:1}$ H	1.2	1.9	TR	1.2
$C_{18:3}$ ω6*c*	5.6	ND	TR	ND
Sum in feature 3	2.1	1.9	4.3	3.4
Sum in feature 9	2.7	2.5	5.4	8.6

Strains: 1, SCSIO 64649^T; 2, SCSIO 03032; 3, S. specialis DSM 41924^T; 4, S. manganisoli MK44^T. All data from this study except for S. manganisoli MK44^T; cells were collected after incubation on 2216E at 28°C for 7 days. The major fatty acids (greater than 10%) are shown in bold. TR, less than 1%; ND, not detected. Summed feature 3 comprises $C_{16:1}$ ω6c and/or $C_{16:1}$ ω7c. Summed feature 9 comprises iso-$C_{17:1}$ ω9c and 10-methyl $C_{16:0}$.
&Data from Mo et al. (2018).

peptide, RRE-containing, and RiPP-like), PKS, NRPS, indole, terpene, siderophore, guanidinotides, melanin, ectoine, and phenazine (**Supplementary Table 6**). In addition to the BGCs encoding for common secondary metabolites produced by genus *Streptomyces*, such as desferrioxamine E, melanin, and hopene, 11 BGCs showed no similarity to any reference BGCs, and nine BGCs had 60% of genes with high similarity to homologs from known BGCs. This highlights the potential of SCSIO 64649^T to produce novel secondary metabolites.

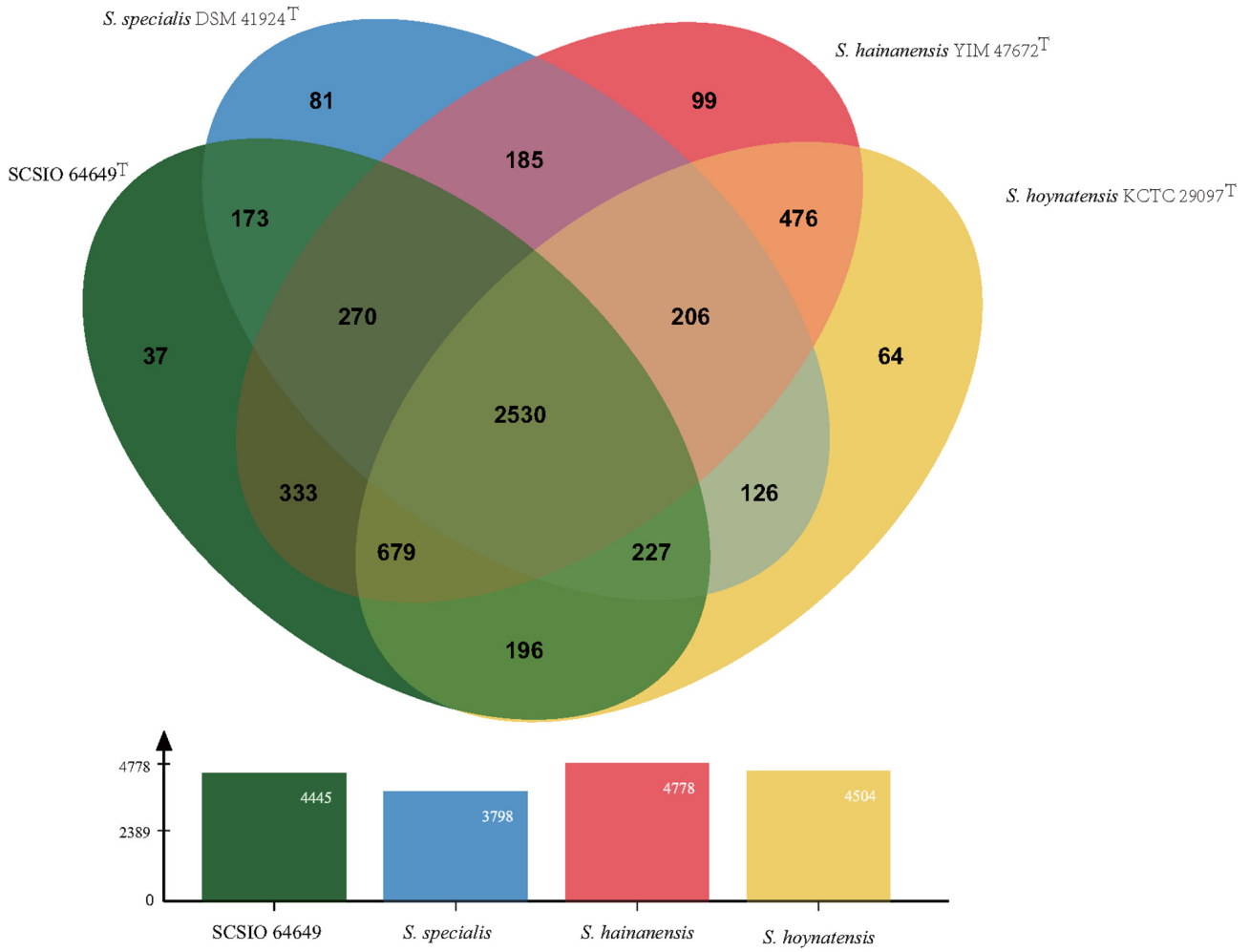


FIGURE 4 | Venn diagram representing the core orthologs and unique genes for strain SCSIO 64649^T and closely related type strains.

A remarkable feature of the strain SCSIO 64649^T genome is the presence of nine PKS-coding BGCs, encoding three different kinds of PKS (six PKS I, two PKS II, and one PKS III). The six PKS I clusters (#3, #4, #9, #13, #15, and #30) showed variable similarity with reported BGCs (6–100%). Clusters #3 and #13 showed 100% similarity with BGCs encoding for piericidin A1 and heronamide F, which have been reported in SCSIO 03032 by Chen et al. (2014) and Zhu et al. (2015), respectively. SCSIO 64649^T presented intact BGCs of piericidin A1 and heronamide F. However, the predicted constructs of gene clusters 3# and 13# are quite different from piericidin A1 and heronamide F (**Supplementary Figure 5**) clarified in strain SCSIO 03032. *In silico* analysis of cluster #30 revealed low similarity with candicidin (23%, MIBiG accession number BGC0000034), which is a member of polyene polyketides possessing a series of conjugated unsaturated double bonds and exhibiting potent activities against fungal pathogens (Zhang L. et al., 2017; Sun et al., 2018). The polyketide backbone of candicidin is characterized by seven conjugated double bonds and is constructed by 21 PKS modules through condensing a *p*-aminobenzoic acid starter unit, 4 propionate molecules, and 14 acetate units (Chen et al., 2003). Detailed bioinformatics analysis revealed that the five core PKS genes (ctg1_5542-5546) in cluster 30# encode 18 modules, which were predicted to be responsible for condensation of seven propionates and 11 acetate units to form a polyketide backbone with three conjugated double bonds. In addition to the core PKSs, the post-PKS tailoring enzymes also showed much lower homologous similarity compared with candicidin coding sequences. Collectively, the difference in the core PKSs and the low sequence similarity of the post-PKS tailoring enzymes strongly indicated that cluster #30 encodes a new polyketide product. The predicted structure of the compound is shown in **Supplementary Figure 5**. Clusters #4 and #15 displayed low similarity with known gene clusters salinomycin (6%, MIBiG accession number BGC0000144) and stambomycins (16%, MIBiG accession number BGC0000151). This potential to produce novel metabolites, which cannot be predicted with bioinformatics analyses, needs to be confirmed by further secondary metabolite separation.

The type II PKS genes clusters #5 and #20 showed, respectively, 68 and 54% similarities with the mayamycin BGC

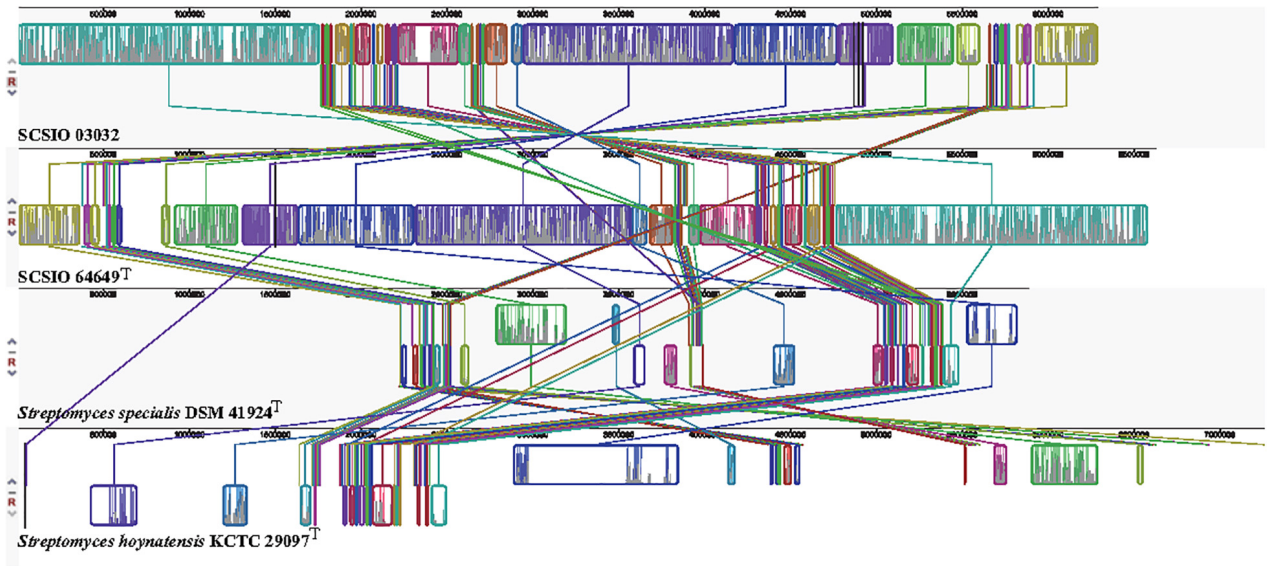


FIGURE 5 | Genome alignment between strains SCSIO 64649^T, SCSIO 03032, *S. specialis* DSM 41924^T, and *S. hoynatensis* KCTC 29097^T.

(MIBiG accession number BGC0001661) from *Streptomyces* sp. 120454 (Bo et al., 2018). Mayamycin is a member of the angucycline-type polycyclic compounds that predominantly display anticancer and antibacterial activity and feature a tetracyclic benz[a]anthracene scaffold, which is derived via successive decarboxylative Claisen condensations of an acetyl-CoA starter unit and nine malonyl-CoA extender units (Kharel et al., 2012). Six PKS genes in BGC #5, encoding ketoacyl synthase, chain length factor, acyl carrier protein, two cyclases, and one ketoreductase, showed high similarity to the corresponding PKS enzymes (May 12–17) from the mayamycin gene cluster (Bo et al., 2018). This suggests the formation of angular tetracyclic rings in its biosynthetic pathway. Although the core PKS enzymes in BGC #5 showed high similarity to the homologs from angucycline-type BGC, the genes responsible for sugar biosynthesis are different from those involved in the amino sugar biosynthetic pathway of mayamycin. The mayamycin BGC contains six genes (*may5*, *6*, *7*, *9*, *10*, and *22*) encoding NDP-glucose phosphate nucleotidyltransferase, NDP-hexose 4,6-dehydratase, NDP-deoxyglucose-2,3-dehydratase, NDP-deoxyhexose 3-aminotransferase, NDP-4-keto-6-deoxyhexose reductase, and *N*-methyl transferase for the construction of amino sugar (Bo et al., 2018). The sequence analysis of BGC #5 revealed five open reading frames (ORFs) that could potentially be involved in the amino sugar biosynthetic pathway; these five ORFs are LC193_03460, 03465, 03470, 03480, and 03485, encoding for NDP-glucose phosphate nucleotidyltransferase, NDP-hexose 4,6-dehydratase, NDP-deoxyglucose-2,3-dehydratase, NDP-deoxyhexose aminotransferase, and dTDP-4-dehydrorhamnose-3,5-epimerase, respectively. This suggests the biosynthesis of novel mayamycin analogs with different amino sugars (**Supplementary Table 7**), and its predicted structure is shown in **Supplementary Figure 5**. Cluster #20 showed 54% similarity with the mayamycin BGC, indicating that this cluster also produces angucycline-type polycyclic compounds. However, the lower amino sequence similarity and different organization (**Supplementary Table 8**) indicate that cluster #20 may synthesize a novel mayamycin analog (**Supplementary Figure 5**), which is different from the product of cluster #5.

Nine BGCs are involved in the biosynthesis of RiPPs (thiopeptide, lanthipeptide, lasso peptide, RRE-containing, and RiPP-like). Only cluster #10 of nine RiPPs BGCs showed a high similarity (80%) to class III lanthipeptide of AmfS, which comprises biological surfactants that positively regulate the formation of aerial mycelia (Ueda et al., 2002). In the remaining eight RiPP BGCs, two BGCs (#8, #28) showed low similarities (<50%) to the known BGCs, and six BGCs (#10, #14, #17, #23, #24, and #29) did not match with known gene clusters. These findings revealed that strain SCSIO 64649^T has the potential to produce the novel RiPPs.

The remaining clusters in SCSIO 64649^T, #16, #19, #21, and #26, are terpene BGCs, assumed to be similar to the BGCs of isorenieratene, geosmin, carotenoid, and hopene, respectively. Except for cluster #19, which shows 100% BGC similarity with geosmin, the other three clusters, #16 (37%), #21 (27%), and #26 (30%), showed low similarities with known BGCs, indicating that the strain also has the potential to produce novel terpene compounds.

Phage-encoded serine integrases are powerful tools for molecular genetics, because they can catalyze site-specific integration of DNA into bacterial host chromosomes in a highly controllable and predictable way (Gao et al., 2020). Twenty-one serine and subtilisin-like serine integrase genes were also annotated in the SCSIO 64649^T genome (**Supplementary Table 9**), and they belong to a family of proteins known to play several different biological roles (Karlsson et al., 2007). Further research on these specific proteases may also be relevant

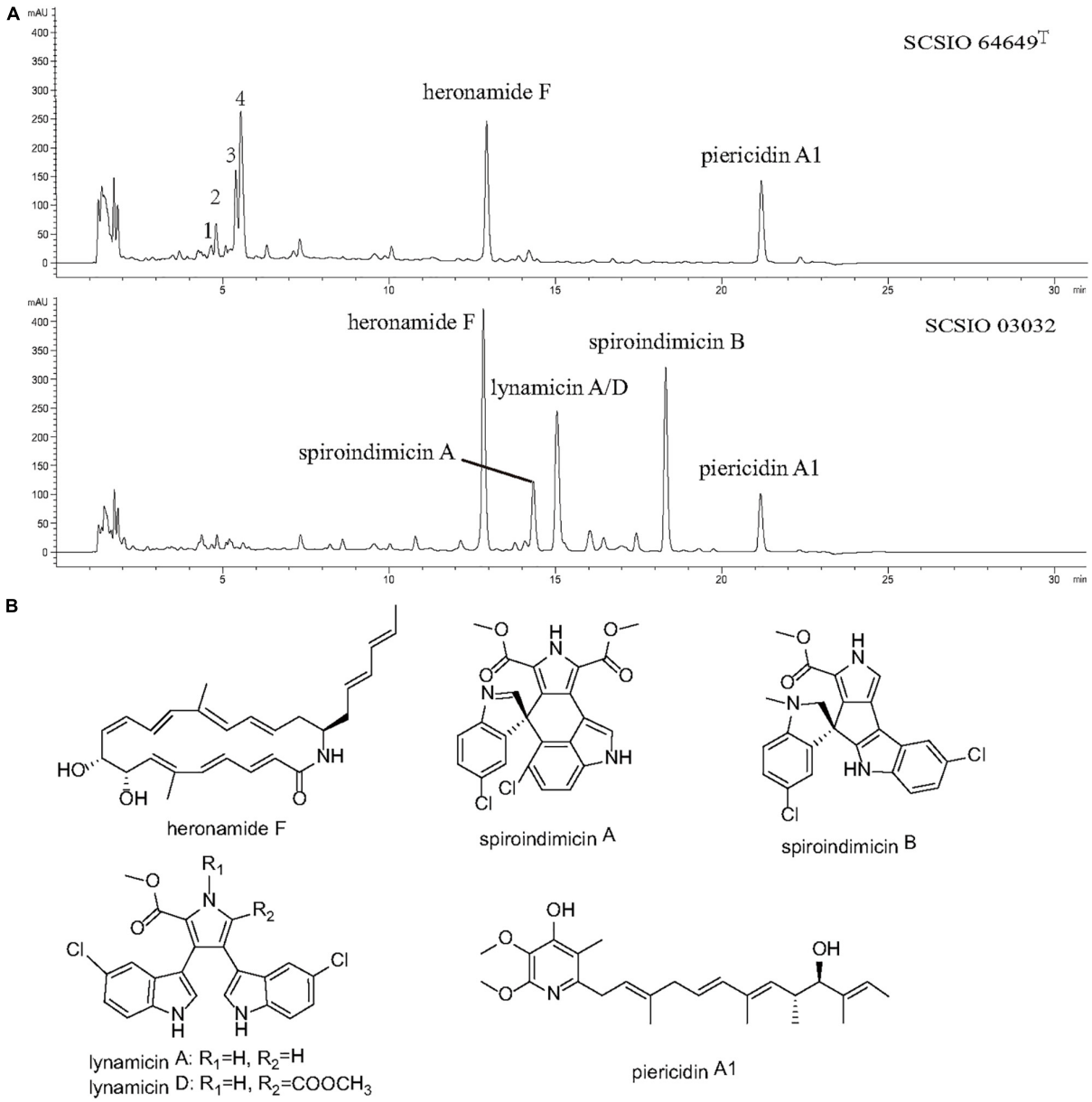


FIGURE 6 | (A) LC-DAD isoplot of the analyzed extract showing compounds produced. **(B)** Structural formula of compounds isolated from strains SCSIO 64649^T and SCSIO 03032.

from an industrial perspective. Through the analysis of synthetic pathways by KEGG, we also found that, in addition to natural products, this new species has the potential to produce a variety of cofactors and vitamins, such as riboflavin, biotin, and VB12 (cobalamin). These results indicate that this new species offers the potential to discover novel natural products.

Identification of Bioactive Compounds and Antimicrobial Activity Assay

Strain SCSIO 03032 has been demonstrated to be able to produce five categories of bioactive compounds (piericidins, heronamides, spiroindimicins, indimicins, and lynamicins) under laboratory culture conditions (Zhang et al., 2012; Chen et al., 2014; Zhang et al., 2014a,b; Zhu et al., 2015; Ma L. et al., 2017; Liu et al., 2019). To identify the bioactive compounds produced by strain SCSIO 64649^T, both strains were fermented using ISP 3 and ISP 4 media with the same conditions. The fermentation extracts were analyzed using LC-HR-MS and then compared. When fermented with ISP 4 medium, strain SCSIO 03032 produced four types of compounds (heronamide F, piericidin A1, spiroindimicins A and B, and lynamicins A/D); however, strain SCSIO 64649^T was found to only produce heronamide F and piericidin A1 as evidenced by ESI-MS data (422.2706

$[M + H]^+$ and 416.2722 $[M + H]^+$, respectively) (**Figure 6A** and **Supplementary Figure 6**). A yield of spiroindimicin B was detected in SCSIO 03032 but not detected in strain SCSIO 64649^T under the same fermentation conditions. The gene cluster coding for spiroindimicins in SCSIO 64649^T showed similar genetic organization, and the amino acid sequence of eight genes had a high degree of homology with a similarity greater than 90% with SCSIO 03032 (**Supplementary Table 10**). The inactivation of the spiroindimicin BGC in SCSIO 64649^T may be attributed to the unsuitability of the ISP 4 medium to produce spiroindimicin B. Indeed, when the strain SCSIO 64649^T was fermented in ISP 3 medium, spiroindimicin A was also readily detected (**Supplementary Figure 7**). Although the two strains belong to the same species, their secondary metabolites could be different because of their adaptations to different natural environments.

Piericidin A1 is a member of the piericidin family and features a 4-pyridinol core linked to a methylated polyketide side chain (Liu et al., 2012; Zhou and Fenical, 2016). A recent study revealed that piericidin A1 shows anti-renal cell carcinoma activity, which has laid the foundation for the development of new anti-kidney cancer drugs (Zhou et al., 2019). Heronamides D–F were isolated from strain SCSIO 03032 (Zhang et al., 2014a). Heronamides are a rare class of polyketide 20-membered macrolactams featuring unprecedented carbon/heteroatom skeletons that show potent antifungal metabolites and are produced by marine-derived actinomycetes (Raju et al., 2010). The spiroindimicins are a unique class of chlorinated indole alkaloids characterized by featuring unique [5,5] or [5,6] spiro-rings, which displayed moderate cytotoxicity against several cancer cell lines. The discovery of these bioactive compounds in strains SCSIO 64649^T and SCSIO 03032 suggests that this new species is an important source of piericidins, heronamides, and spiroindimicins.

Interestingly, four new peaks were found in extracts of strain SCSIO 64649^T fermented with ISP 4 medium. These appeared within 4–6 min with a molecular weight of 190.1226, 179.1171, 196.1691, and 208.1335 (**Figure 6**), but they were not detected in the fermentation extracts of strain 03032. Future research should focus on large-scale fermentation and separation to obtain more metabolites and extend the detection range of biological activity.

The fermentation extract of strain SCSIO 64649^T was also subjected to an antimicrobial activity assay to assess its potential to produce antimicrobial compounds. The analysis revealed that the fermentation extracts showed strong antimicrobial activity against *Aspergillus niger* and weak activity against *Micrococcus luteus* from two different fermentation media, but no antimicrobial activities to the other five indicative microorganisms (**Supplementary Figure 8**). It was also revealed that the fermentation extract on the ninth day showed better antimicrobial activity than that on the 7th and 11th days, which indicated that the production of active natural products is related to fermentation time. Although heronamide F has been reported to possess antifungal activity against *Candida albicans*, we detected only antifungal activity against *Aspergillus niger*. Therefore, it was deduced that strains SCSIO 64649^T and SCSIO 03032 produced different antifungal compounds. Results from genome mining and antimicrobial activity detection indicate that this new species has the great potential to produce novel natural products with potent antimicrobial activities.

Ecological Role

Genome annotation and analysis revealed that this new species encodes many Eukaryotic-like proteins (ELPs), such as tetratricopeptide repeats, ankyrin repeats, and WD-40 proteins. These molecules may mediate bacterial–host interactions and modulate the host's behavior (Reynolds and Thomas, 2016; Robbins et al., 2019). Four WD-40 repeats and one tetratricopeptide repeat were discovered in SCSIO 64649^T, which indicates that it may have the potential to form symbiotic relationships with coral hosts. In fact, bacteria in sponges express ankyrin genes to avoid phagocytosis becoming residents (Jahn et al., 2019). Although strain SCSIO 03032 was isolated from deep-sea sediment, 18 tetratricopeptide repeats, 2 ankyrin repeats, and 1 WD-40 repeat were also discovered in its genome, suggesting a possible role for these motifs in maintaining symbiotic associations. ATP-binding cassette (ABC) transporters are ATP-dependent protein complexes, which are vital in mediating the transport of both organic and inorganic molecules across cell membranes (Denecke et al., 2021). A total of 240 ABC transporters and 144 ABC transporters were annotated in the genomes of SCSIO 64649^T and SCSIO 03032, respectively. These are thought to be involved in nutrient acquisition and to help maintain osmotic balance in the cell. In addition, strain SCSIO 64649^T has genes encoding for cobalt and zinc resistance as well as genes for copper oxidase, which are stress genes associated with osmotic and oxidative processes. VB12 plays an important role as an essential co-factor in various biochemical processes, is only produced by some bacteria and archaea, and requires more than 30 enzymes for the *de novo* synthesis (Banerjee and Ragsdale, 2003). Strains SCSIO 64649^T and SCSIO 03032 possess complete synthetic pathway genes and synthesis ability of VB12 by aerobic routes; therefore, they have the potential to supply VB12 for their hosts, which cannot produce it themselves. The antimicrobial activity of this new species may also protect the host against pathogens and predators. In return, the host may provide shelter from predators and adverse conditions to these microorganisms residing in its tissues and supply the essential nutrients for their growth and metabolism. All these features indicate that strain SCSIO 64649^T may be a beneficial microorganism for the coral holobiont thanks to its versatility and high adaptability to harsh environmental conditions.

DISCUSSION

There is a recognized positive correlation between the isolation and discovery of new actinomycetes and novel bioactive compound discovery. The present study was designed to establish the taxonomic status of this novel species and, further, to describe its biosynthesis potential to produce novel natural products through genome mining, compound detection, and antimicrobial activity analysis.

Morphological, phylogenetic, chemotaxonomic, and genomic analyses indicated that strains SCSIO 64649^T and SCSIO 03032

belong to the genus *Streptomyces*. Extensive analyses revealed that they are very different from their closest relatives *S. specialis* DSM 41924^T and *S. manganisoli* MK44^T in physiological, biochemical, and chemotaxonomic properties (**Tables 2**, **3** and **Supplementary Table 5**). Based upon these results and the ANI, dDDH, and AAI values, strains SCSIO 64649^T and SCSIO 03032 were found to represent the same novel species in genus *Streptomyces*, for which the name *Streptomyces marincola* sp. nov. is here proposed, with the type strain SCSIO 64649^T.

A high percentage of this new species genome is dedicated to secondary metabolite production, as indicated by the length of the BGC-related sequences. Thirty-two secondary metabolite BGCs in strain SCSIO 64649^T were distributed across 10 different types. Eleven of them show no similarity to any reference BGCs, while nine BGCs representing 60% of related genes showed high similarity to homologs from known BGCs. This indicates that this novel species has the potential to produce new secondary metabolites. Detailed analysis of the PKS-coding BGCs and RiPPs revealed that the new species has potential for the biosynthesis of a novel polyene polyketide compound, two mayamycin analogs, and a series of RiPPs.

Three remarkable bioactive secondary metabolites were detected from SCSIO 64649^T as compared with 22 compounds from strain SCSIO 03032 found in a previous study. Most of these compounds showed multiple biological activities, which suggests that this new species could be an important source of useful compounds for the medical and agricultural industries. Extending the detection range of biological activity can improve the application of these compounds, such as piericidin A1. However, in this study, the putatively novel compounds were not further characterized, either because of unsuitable fermentation conditions or lack of expression. In addition, this new species showed strong antimicrobial activity, which was inconsistent with the activity of known compounds, indicating that different, unknown antifungal compounds were produced.

In conclusion, a "new species owns the novel genes, which relate the novel natural products"; this new *Streptomyces* species shows great potential to produce novel natural products that could be used by the medical and agricultural industries. Since, in recent years, several approaches for activating the BGCs have been developed (Liu et al., 2020, 2021; Nguyen et al., 2020), future studies should focus on not only the discovery of the uncultured or new microorganisms but also how to isolate more bioactive products with new methods and also to use the metabolic profiling in species-level systematics research.

Descriptions of *Streptomyces marincola* sp. nov.

Streptomyces marincola (ma.rin'co.la. L. n. *mare* the sea; L. n. *incola* inhabitant; N.L. n. *marincola* inhabitant of the sea).

Gram-stain-positive and aerobic actinomycete that forms an extensively branched substrate mycelium and aerial hyphae that differentiate into spiral spore chains consisting of elliptical or short rod spores with smooth surfaces and grows well on ISP 2, ISP 4, ISP 7, NA, and 2216E media. The colors of the aerial and substrate mycelium are media dependent. The diffusible melanin is only observed on ISP 2 medium. Growth occurs at 15–40°C (optimal 28°C), at pH 6–9 (optimal pH 7–8), and up to 9% NaCl (optimal 4%). It is catalase and oxidase negative and positive for hydrolysis of gelatin; aesculin, Tweens 20, 40, 60, and 80; and starch, but negative for milk coagulation and peptonization, nitrate reduction, hydrolysis of cellulose, and H_2S production. It is positive for lipase (C14), leucine arylamidase, naphthol-AS-BI-phosphohydrolase, α-glucosidase, β-glucosidase, *N*-acetyl-β-glucosaminidase, α-mannosidase, and β-fucosidase and weakly positive for alkaline phosphatase, esterase (C4), valine arylamidase, cystine arylamidase, and acid phosphatase. It utilizes sole carbon sources D-maltose, sucrose, β-methyl-D-glucoside, *N*-acetyl-β-D-mannosamine, a-D-glucose, D-galactose, L-arginine, D-glucuronic acid, quinic acid, L-lactic acid, and L-malic acid. The predominant fatty acids (>10%) were iso-$C_{16:0}$ and $C_{12:0}$. The major polar lipid comprised diphosphatidylglycerol, phosphatidylglycerol, phosphatidylethanolamine, phosphatidylinositol mannoside, phosphatidylinositol, glycerol lipid, and six unidentified phospholipids. The cell-wall peptidoglycan contains *LL*-2,6-diaminopimelic acid, and the whole-cell sugars were galactose, glucose, xylose, and ribose. The predominant menaquinones were MK-10(H_4) and MK-10(H_6).

The type strain, SCSIO 64649^T (MCCC 1K06255^T = VKM Ac-2908^T), was isolated from the stony coral *Favites* sp. collected from the South China Sea off the Luhuitou peninsula, Sanya, Hainan province, China. The complete genome of SCSIO 64649^T was composed of one linear chromosome 6,629,020 bp long with a G + C content of 73.6%, a total of 5,774 genes, and 32 biosynthetic gene clusters. The genome sequences for strains SCSIO 64649^T and SCSIO 03032 have been deposited to GenBank under accession numbers CP084541 and CP021121, respectively. The 16S rRNA gene sequences of strains SCSIO 64649^T and SCSIO 03032 have been deposited to GenBank under accession numbers MZ889118 and JN798514, respectively.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

SS and XT conceived and designed the study. SS and LC carried out all the experiments. LM and QL performed the LC/MS-based identification. KZ and QZ ran the bioinformatics analysis. SS prepared the manuscript. XT and LL revised the manuscript. All authors reviewed and approved the manuscript.

FUNDING

This work was supported by the grants from Finance Science and Technology Project of Hainan Province

(no. ZDKJ202018), Local Innovative and Research Teams Project of Guangdong Pearl River Talents Program. (no. 2019BT02Y262), and Academy of South China Sea Ecology and Environmental Engineering, Chinese Academy of Sciences (no. ISEE2018ZD02).

ACKNOWLEDGMENTS

We are grateful to Xuan Ma, Shi-kun Dai, Zhihui Xiao, Aijun Sun, and Yun Zhang in the analytical facilities of SCSIO. We thank the research South China Sea Open Cruise by R/V Shiyan 1 for sample collection supported by NSFC Shiptime Sharing Project.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.860308/full#supplementary-material

REFERENCES

Banerjee, R., and Ragsdale, S. W. (2003). The many faces of vitamin B12: catalysis by cobalamin-dependent enzymes. *Annu. Rev. Biochem*. 72, 209–247. doi: 10.1146/annurev.biochem.72.121801.161828

Bauer, A. W., Kirby, W. M. M., Sherris, J. C., and Turch, M. (1966). Antibiotic susceptibility testing by a standard single disk method. *Am. J. Clin. Pathol*. 45, 493–496.

Bertelli, C., Laird, M. R., Williams, K. P., Simon Fraser University Research Computing, G., Lau, B. Y., Hoad, G., et al. (2017). IslandViewer 4: expanded prediction of genomic islands for larger-scale datasets. *Nucleic Acids Res*. 45, W30–W35. doi: 10.1093/nar/gkx343

Blin, K., Shaw, S., Kloosterman, A. M., Charlop-Powers, Z., van Wezel, G. P., Medema, M. H., et al. (2021). antiSMASH 6.0: improving cluster detection and comparison capabilities. *Nucleic Acids Res*. 49, W29–W35. doi: 10.1093/nar/gkab335

Bo, S. T., Xu, Z. F., Yang, L., Cheng, P., Tan, R. X., Jiao, R. H., et al. (2018). Structure and biosynthesis of mayamycin B, a new polyketide with antibacterial activity from *Streptomyces* sp. 120454. *J. Antibiot.* 71, 601–605. doi: 10.1038/s41429-018-0039-x

Chan, P. P., and Lowe, T. M. (2019). tRNAscan-SE: searching for tRNA genes in genomic sequences. *Methods Mol. Biol*. 1962, 1–14. doi: 10.1007/978-1-4939-9173-0_1

Chaumeil, P. A., Mussig, A. J., Hugenholtz, P., and Parks, D. H. (2019). GTDB-Tk: a toolkit to classify genomes with the genome taxonomy database. *Bioinformatics* 36, 1925–1927. doi: 10.1093/bioinformatics/btz848

Chen, S., Huang, X., Zhou, X., Bai, L., He, J., Jeong, K. J., et al. (2003). Organizational and mutational analysis of a complete FR-008/candicidin gene cluster encoding a structurally related polyene complex. *Chem. Biol*. 10, 1065–1076. doi: 10.1016/j.chembiol.2003.10.007

Chen, Y., Zhang, W., Zhu, Y., Zhang, Q., Tian, X., Zhang, S., et al. (2014). Elucidating hydroxylation and methylation steps tailoring piericidin A1 biosynthesis. *Org. Lett*. 16, 736–739. doi: 10.1021/ol4034176

Chin, C. S., Alexander, D. H., Marks, P., Klammer, A. A., Drake, J., Heiner, C., et al. (2013). Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat. Methods* 10, 563–569. doi: 10.1038/nmeth.2474

Cimermancic, P., Medema, M. H., Claesen, J., Kurita, K., Wieland Brown, L. C., Mavrommatis, K., et al. (2014). Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. *Cell* 158, 412–421. doi: 10.1016/j.cell.2014.06.034

Collins, M. D., Pirouz, T., Goodfellow, M., and Minnikin, D. E. (1977). Distribution of menaquinones in actinomycetes and corynebacteria. *J. Gen. Microbiol*. 100, 221–230. doi: 10.1099/00221287-100-2-221

Couvin, D., Bernheim, A., Toffano-Nioche, C., Touchon, M., Michalik, J., Neron, B., et al. (2018). CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. *Nucleic Acids Res*. 46, W246–W251. doi: 10.1093/nar/gky425

Davies-Bolorunduro, O. F., Osuolale, O., Saibu, S., Adeleye, I. A., and Aminah, N. S. (2021). Bioprospecting marine actinomycetes for antileishmanial drugs: current perspectives and future prospects. *Heliyon* 7:e07710. doi: 10.1016/j.heliyon.2021.e07710

Denecke, S., Rankic, I., Driva, O., Kalsi, M., Luong, N. B. H., Buer, B., et al. (2021). Comparative and functional genomics of the ABC transporter superfamily across arthropods. *BMC Genomics* 22:553. doi: 10.1186/s12864-021-07861-2

Feling, R. H., Buchanan, G. O., Mincer, T. J., Kauffman, C. A., Jensen, P. R., and Fenical, W. (2003). Salinosporamide A: a highly cytotoxic proteasome inhibitor from a novel microbial source, a marine bacterium of the new genus *Salinospora*. *Angew. Chem. Int. Ed. Engl.* 42, 355–357. doi: 10.1002/anie.200390115

Felsenstein, J. (1981). Evolutionary trees from DNA sequences a maximum likelihood. *J. Mol. Evol*. 17, 368–376. doi: 10.1007/BF01734359

Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39, 783–791. doi: 10.1111/j.1558-5646.1985.tb00420.x

Fitch, W. M. (1971). Toward defining the course of evolution: minimum change for a specific tree topology. *Syst. Biol*. 20, 406–416. doi: 10.1093/sysbio/20.4.406

Gao, H., Taylor, G., Evans, S. K., Fogg, P. C. M., and Smith, M. C. M. (2020). Application of serine integrases for secondary metabolite pathway assembly in *Streptomyces*. *Synth. Syst. Biotechnol.* 5, 111–119. doi: 10.1016/j.synbio.2020.05.006

Gonzalez, C., Gutierrez, C., and Ramirez, C. (1978). *Halobacterium vallismortis* sp. nov. an amylolytic and carbohydrate-metabolizing, extremely halophilic bacterium. *Can. J. Microbiol*. 24, 710–715. doi: 10.1139/m78-119

Hassan, S. S. U., and Shaikh, A. L. (2017). Marine actinobacteria as a drug treasure house. *Biomed. Pharmacother*. 87, 46–57. doi: 10.1016/j.biopha.2016.12.086

Hu, Y., Chen, J., Hu, G., Yu, J., Zhu, X., Lin, Y., et al. (2015). Statistical research on the bioactivity of new marine natural products discovered during the 28 years from 1985 to 2012. *Mar. Drugs* 13, 202–221. doi: 10.3390/md13010202

Jahn, M. T., Arkhipova, K., Markert, S. M., Stigloher, C., Lachnit, T., Pita, L., et al. (2019). A phage protein aids bacterial symbionts in eukaryote immune wvasion. *Cell. Host Microbe* 26:e545. doi: 10.1016/j.chom.2019.08.019

Kanehisa, M., and Sato, Y. (2020). KEGG Mapper for inferring cellular functions from protein sequences. *Protein Sci.* 29, 28–35. doi: 10.1002/pro.3711

Karlsson, C., Andersson, M. L., Collin, M., Schmidtchen, A., Bjorck, L., and Frick, I. M. (2007). SufA–a novel subtilisin-like serine proteinase of *Finegoldia magna*. *Microbiology* 153, 4208–4218. doi: 10.1099/mic.0.2007/010322-0

Kawata, J., Naoe, T., Ogasawara, Y., and Dairi, T. (2017). Biosynthesis of the carbonylmethylene structure found in the Ketomemicin Class of *Pseudotripeptides*. *Angew. Chem. Int. Ed. Engl*. 56, 2026–2029. doi: 10.1002/anie.201611005

Kharel, M. K., Pahari, P., Shepherd, M. D., Tibrewal, N., Nybo, S. E., Shaaban, K. A., et al. (2012). Angucyclines: biosynthesis, mode-of-action, new natural products, and synthesis. *Nat. Prod. Rep*. 29, 264–325. doi: 10.1039/c1np00068c

Kim, M., Oh, H. S., Park, S. C., and Chun, J. (2014). Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. *Int. J. Syst. Evol. Microbiol*. 64, 346–351. doi: 10.1099/ijs.0.059774-0

Konstantinidis, K. T., Rossello-Mora, R., and Amann, R. (2017). Uncultivated microbes in need of their own taxonomy. *ISME J*. 11, 2399–2406. doi: 10.1038/ismej.2017.113

Li, W. J., Xu, P., Schumann, P., Zhang, Y. Q., Pukall, R., Xu, L. H., et al. (2007). *Georgenia ruanii* sp. nov., a novel actinobacterium isolated from forest soil in Yunnan (China), and emended description of the genus *Georgenia*. Int. J. Syst. Evol. Microbiol. 57, 1424–1428. doi: 10.1099/ijs.0.64749-0

Liu, J., Xie, X., and Li, S. M. (2020). Increasing cytochrome P450 enzyme diversity by identification of two distinct cyclodipeptide dimerases. *Chem. Commun.* 56, 11042–11045. doi: 10.1039/d0cc04772d

Liu, Q., Yao, F., Chooi, Y. H., Kang, Q., Xu, W., Li, Y., et al. (2012). Elucidation of Piericidin A1 biosynthetic locus revealed a thioesterase-dependent mechanism

of alpha-pyridone ring formation. *Chem. Biol*. 19, 243–253. doi: 10.1016/j.chembiol.2011.12.018

Liu, Z., Ma, L., Zhang, L., Zhang, W., Zhu, Y., Chen, Y., et al. (2019). Functional characterization of the halogenase *SpmH* and discovery of new deschloro-tryptophan dimers. *Org. Biomol. Chem*. 17, 1053–1057. doi: 10.1039/c8ob02775g

Liu, Z., Zhao, Y., Huang, C., and Luo, Y. (2021). Recent advances in silent gene cluster activation in *Streptomyces*. *Front. Bioeng. Biotechnol*. 9:632230. doi: 10.3389/fbioe.2021.632230

Ma, J., Huang, H., Xie, Y., Liu, Z., Zhao, J., Zhang, C., et al. (2017). Biosynthesis of ilamycins featuring unusual building blocks and engineered production of enhanced anti-tuberculosis agents. *Nat. Commun*. 8:391. doi: 10.1038/s41467-017-00419-5

Ma, L., Zhang, W., Liu, Z., Huang, Y., Zhang, Q., Tian, X., et al. (2021). Complete genome sequence of *Streptomyces* sp. SCSIO 03032 isolated from Indian Ocean sediment, producing diverse bioactive natural products. *Mar. Genomics* 55:100803. doi: 10.1016/j.margen.2020.100803

Ma, L., Zhang, W., Zhu, Y., Zhang, G., Zhang, H., Zhang, Q., et al. (2017). Identification and characterization of a biosynthetic gene cluster for tryptophan dimers in deep sea-derived *Streptomyces* sp. SCSIO 03032. *Appl. Microbiol. Biotechnol*. 101, 6123–6136. doi: 10.1007/s00253-017-8375-5

Manivasagan, P., Kang, K. H., Sivakumar, K., Li-Chan, E. C., Oh, H. M., and Kim, S. K. (2014a). Marine actinobacteria: an important source of bioactive natural products. *Environ. Toxicol. Pharmacol*. 38, 172–188. doi: 10.1016/j.etap.2014.05.014

Manivasagan, P., Venkatesan, J., Sivakumar, K., and Kim, S. K. (2014b). Pharmaceutically active secondary metabolites of marine actinobacteria. *Microbiol. Res*. 169, 262–278. doi: 10.1016/j.micres.2013.07.014

Meier-Kolthoff, J. P., Auch, A. F., Klenk, H. P., and Göker, M. (2013). Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics* 14:60. doi: 10.1186/1471-2105-14-60

Minnikin, D. E., O'Donnell, A. G., Goodfellow, M., Alderson, G., Athalye, M., Schaal, A., et al. (1984). An integrated procedure for the extraction of bacterial isoprenoid quinones and polar lipids. *J. Microbiol. Methods* 2, 233–241. doi: 10.1016/0167-7012(84)90018-6

Mo, P., Zhao, J., Li, K., Tang, X., and Gao, J. (2018). *Streptomyces manganisoli* sp. nov., a novel actinomycete isolated from manganese-contaminated soil. *Int. J. Syst. Evol. Microbiol*. 68, 1890–1895. doi: 10.1099/ijsem.0.002762

Nguyen, C. T., Dhakal, D., Pham, V. T. T., Nguyen, H. T., and Sohng, J. K. (2020). Recent advances in strategies for activation and discovery/characterization of cryptic biosynthetic gene clusters in *Streptomyces*. *Microorganisms* 8:616. doi: 10.3390/microorganisms8040616

Overbeek, R., Olson, R., Pusch, G. D., Olsen, G. J., Davis, J. J., Disz, T., et al. (2014). The SEED and the rapid annotation of microbial genomes using subsystems technology (RAST). *Nucleic Acids Res*. 42, D206–D214. doi: 10.1093/nar/gkt1226

Payne, D. J., Gwynn, M. N., Holmes, D. J., and Pompliano, D. L. (2007). Drugs for bad bugs: confronting the challenges of antibacterial discovery. *Nat. Rev. Drug Discov*. 6, 29–40. doi: 10.1038/nrd2201

Penn, K., Jenkins, C., Nett, M., Udwary, D. W., Gontang, E. A., McGlinchey, R. P., et al. (2009). Genomic islands link secondary metabolism to functional adaptation in marine Actinobacteria. *ISME J*. 3, 1193–1203. doi: 10.1038/ismej.2009.58

Raju, R., Piggott, A. M., Conte, M. M., and Capon, R. J. (2010). Heronamides A-C, new polyketide macrolactams from an Australian marine-derived *Streptomyces* sp. A biosynthetic case for synchronized tandem electrocyclization. *Org. Biomol. Chem*. 8, 4682–4689. doi: 10.1039/c0ob00267d

Reynolds, D., and Thomas, T. (2016). Evolution and function of eukaryotic-like proteins from sponge symbionts. *Mol. Ecol*. 25, 5242–5253. doi: 10.1111/mec.13812

Richter, M., and Rossello-Mora, R. (2009). Shifting the genomic gold standard for the prokaryotic species definition. *Proc. Natl. Acad. Sci. U. S. A*. 106, 19126–19131. doi: 10.1073/pnas.0906412106

Robbins, S. J., Singleton, C. M., Chan, C. X., Messer, L. F., Geers, A. U., Ying, H., et al. (2019). A genomic view of the reef-building coral *Porites lutea* and its microbial symbionts. *Nat. Microbiol*. 4, 2090–2100. doi: 10.1038/s41564-019-0532-4

Saha, S., Zhang, W., Zhang, G., Zhu, Y., Chen, Y., Liu, W., et al. (2017). Activation and characterization of a cryptic gene cluster reveals a cyclization cascade for polycyclic tetramate macrolactams. *Chem. Sci*. 8, 1607–1612. doi: 10.1039/c6sc03875a

Saitou, N., and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol*. 4, 406–425. doi: 10.1093/oxfordjournals.molbev.a040454

Saurav, K., Zhang, W., Saha, S., Zhang, H., Li, S., Zhang, Q., et al. (2014). In silico molecular docking, preclinical evaluation of spiroindimicins A-D, lynamicin A and D isolated from deep marine sea derived *Streptomyces* sp. SCSIO 03032. *Interdiscip. Sci*. 6, 187–196. doi: 10.1007/s12539-013-0200-y

Ser, H. L., Tan, L. T., Law, J. W., Chan, K. G., Duangjai, A., Saokaew, S., et al. (2017). Focused review: cytotoxic and antioxidant potentials of mangrove-derived *Streptomyces*. *Front. Microbiol*. 8:2065. doi: 10.3389/fmicb.2017.02065

Skinnider, M. A., Johnston, C. W., Gunabalasingam, M., Merwin, N. J., Kieliszek, A. M., MacLellan, R. J., et al. (2020). Comprehensive prediction of secondary metabolite structure and biological activity from microbial genome sequences. *Nat. Commun*. 11:6058. doi: 10.1038/s41467-020-19986-1

Song, W., Sun, H. X., Zhang, C., Cheng, L., Peng, Y., Deng, Z., et al. (2019). Prophage Hunter: an integrative hunting tool for active prophages. *Nucleic Acids Res*. 47, W74–W80. doi: 10.1093/nar/gkz380

Sun, F., Xu, S., Jiang, F., and Liu, W. (2018). Genomic-driven discovery of an amidinohydrolase involved in the biosynthesis of mediomycin A. *Appl. Microbiol. Biotechnol*. 102, 2225–2234. doi: 10.1007/s00253-017-8729-z

Tang, S. K., Wang, Y., Chen, Y., Lou, K., Cao, L. L., Xu, L. H., et al. (2009). *Zhihengliuella alba* sp. nov., and emended description of the genus *Zhihengliuella*. Int. J. Syst. Evol. Microbiol. 59, 2025–2031. doi: 10.1099/ijs.0.007344-0

Ueda, K., Oinuma, K., Ikeda, G., Hosono, K., Ohnishi, Y., Horinouchi, S., et al. (2002). AmfS, an extracellular peptidic morphogen in *Streptomyces griseus*. *J. Bacteriol*. 184, 1488–1492. doi: 10.1128/JB.184.5.1488-1492.2002

van der Meij, A., Worsley, S. F., Hutchings, M. I., and van Wezel, G. P. (2017). Chemical ecology of antibiotic production by actinomycetes. *FEMS Microbiol. Rev*. 41, 392–416. doi: 10.1093/femsre/fux005

Wang, C., Du, W., Lu, H., Lan, J., Liang, K., and Cao, S. (2021). A Review: halogenated compounds from marine actinomycetes. *Molecules* 26:2754. doi: 10.3390/molecules26092754

Wang, Y., Xia, Z., Liu, Z., Wan, C., Luo, X., and Zhang, L. (2018). *Streptomyces carminius* sp. nov., a novel actinomycete isolated from *Sophora alopecuroides* in Xinjiang, China. *Antonie Van Leeuwenhoek* 111, 1807–1814. doi: 10.1007/s10482-018-1069-x

Whiton, R. S., Lau, P., Morgan, S. L., Gilbart, J., and Fox, A. (1985). Modifications in the alditol acetate method for analysis of muramic acid and other neutral and amino sugars by capillary gas chromatography-mass spectrometry with selected ion monitoring. *J. Chromatogr*. 347, 109–120. doi: 10.1016/s0021-9673(01)95474-3

Xu, L., Dong, Z., Fang, L., Luo, Y., Wei, Z., Guo, H., et al. (2019). OrthoVenn2: a web server for whole-genome comparison and annotation of orthologous clusters across multiple species. *Nucleic Acids Res*. 47, W52–W58. doi: 10.1093/nar/gkz333

Yoon, S. H., Ha, S. M., Lim, J., Kwon, S., and Chun, J. (2017). A large-scale evaluation of algorithms to calculate average nucleotide identity. *Antonie Van Leeuwenhoek* 110, 1281–1286. doi: 10.1007/s10482-017-0844-4

Zhang, B., Wang, K. B., Wang, W., Wang, X., Liu, F., Zhu, J., et al. (2019). Enzyme-catalysed [6+4] cycloadditions in the biosynthesis of natural products. *Nature* 568, 122–126. doi: 10.1038/s41586-019-1021-x

Zhang, L., Hashimoto, T., Qin, B., Hashimoto, J., Kozone, I., Kawahara, T., et al. (2017). Characterization of giant modular PKSs provides insight into genetic mechanism for structural diversification of aminopolyol polyketides. *Angew. Chem. Int. Ed. Engl*. 56, 1740–1745. doi: 10.1002/anie.201611371

Zhang, M. M., Wong, F. T., Wang, Y., Luo, S., Lim, Y. H., Heng, E., et al. (2017). CRISPR-Cas9 strategy for activation of silent *Streptomyces* biosynthetic gene clusters. *Nat. Chem. Biol*. 13, 607–609. doi: 10.1038/nchembio.2341

Zhang, W., Li, S., Zhu, Y., Chen, Y., Chen, Y., Zhang, H., et al. (2014a). Heronamides D-F, polyketide macrolactams from the deep-sea-derived *Streptomyces* sp. SCSIO 03032. *J. Nat. Prod*. 77, 388–391. doi: 10.1021/np400665a

Zhang, W., Ma, L., Li, S., Liu, Z., Chen, Y., Zhang, H., et al. (2014b). Indimicins A-E, bisindole alkaloids from the deep-sea-derived *Streptomyces* sp. SCSIO 03032. *J. Nat. Prod*. 77, 1887–1892. doi: 10.1021/np500362p

Zhang, W., Liu, Z., Li, S., Yang, T., Zhang, Q., and Ma, L. (2012). Spiroindimicins A-D: new bisindole alkaloids from a deep-sea-derived actinomycete. *Org. Lett.* 14, 3364–3367. doi: 10.1021/ol301343n

Zhou, X., and Fenical, W. (2016). The unique chemistry and biology of the piericidins. *J. Antibiot*. 69, 582–593. doi: 10.1038/ja.2016.71

Zhou, X., Liang, Z., Li, K., Fang, W., Tian, Y., Luo, X., et al. (2019). Exploring the natural piericidins as anti-renal cell carcinoma agents targeting Peroxiredoxin 1. *J. Med. Chem*. 62, 7058–7069. doi: 10.1021/acs.jmedchem.9b00598

Zhou, Y., Pei, S., Xie, F., Gu, L., and Zhang, G. (2020). *Saccharopolyspora coralli* sp. nov. a novel actinobacterium isolated from the stony coral *Porites*. Int. J. Syst. Evol. Microbiol. 70, 3241–3246. doi: 10.1099/ijsem.0.004162

Zhu, Y., Zhang, W., Chen, Y., Yuan, C., Zhang, H., Zhang, G., et al. (2015). Characterization of Heronamide biosynthesis reveals a tailoring hydroxylase and indicates migrated double bonds. *Chembiochem* 16, 2086–2093. doi: 10.1002/cbic.201500281

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.

frontiers | Frontiers in Microbiology

PERSPECTIVE
published: 09 June 2022
doi: 10.3389/fmicb.2022.888414

Improving the Diagnostic Potential of Extracellular miRNAs Coupled to Multiomics Data by Exploiting the Power of Artificial Intelligence

Alessandro Paolini[1†], Antonella Baldassarre[1†], Stefania Paola Bruno[1,2†], Cristina Felli[1†], Chantal Muzi[1], Sara Ahmadi Badi[3,4], Seyed Davar Siadat[3,4], Meysam Sarshar[1‡] and Andrea Masotti[1‡]*

[1] Research Laboratories, Bambino Gesù Children's Hospital-IRCCS, Rome, Italy, [2] Department of Science, University Roma Tre, Rome, Italy, [3] Microbiology Research Center (MRC), Pasteur Institute of Iran, Tehran, Iran, [4] Mycobacteriology and Pulmonary Research Department, Pasteur Institute of Iran, Tehran, Iran

OPEN ACCESS

Edited by:
George Tsiamis,
University of Patras, Greece

Reviewed by:
Francesco Pampaloni,
Goethe University Frankfurt, Germany
Elvan Ciftci,
Üsküdar University, Turkey

***Correspondence:**
Andrea Masotti
andrea.masotti@opbg.net

[†] These authors have contributed equally to this work and share first authorship

[‡] These authors have contributed equally to this work and share senior authorship

Specialty section:
This article was submitted to Systems Microbiology, a section of the journal Frontiers in Microbiology

Received: 02 March 2022
Accepted: 11 May 2022
Published: 09 June 2022

Citation:
Paolini A, Baldassarre A, Bruno SP, Felli C, Muzi C, Ahmadi Badi S, Siadat SD, Sarshar M and Masotti A (2022) Improving the Diagnostic Potential of Extracellular miRNAs Coupled to Multiomics Data by Exploiting the Power of Artificial Intelligence. Front. Microbiol. 13:888414. doi: 10.3389/fmicb.2022.888414

In recent years, the clinical use of extracellular miRNAs as potential biomarkers of disease has increasingly emerged as a new and powerful tool. Serum, urine, saliva and stool contain miRNAs that can exert regulatory effects not only in surrounding epithelial cells but can also modulate bacterial gene expression, thus acting as a "master regulator" of many biological processes. We think that in order to have a holistic picture of the health status of an individual, we have to consider comprehensively many "omics" data, such as miRNAs profiling form different parts of the body and their interactions with cells and bacteria. Moreover, Artificial Intelligence (AI) and Machine Learning (ML) algorithms coupled to other multiomics data (i.e., big data) could help researchers to classify better the patient's molecular characteristics and drive clinicians to identify personalized therapeutic strategies. Here, we highlight how the integration of "multiomic" data (i.e., miRNAs profiling and microbiota signature) with other omics (i.e., metabolomics, exposomics) analyzed by AI algorithms could improve the diagnostic and prognostic potential of specific biomarkers of disease.

Keywords: circulating miRNAs, fecal miRNAs biomarkers, urinary miRNAs detection, oral diagnostics, Machine Learning (ML), Artificial Intelligence (AI)

INTRODUCTION

Over the past decade, the Human Microbiome Project has started many interconnected activities and projects, which allowed us to understand that we have "another" genome (i.e., the microbiome) (Human Microbiome Jumpstart Reference Strains Consortium et al., 2010; Qin et al., 2010). This topic boosted an incredible number of ever increasing interdisciplinary studies applied to medicine [Integrative HMP (iHMP) Research Network Consortium, 2019]. Therefore, the importance of gut microbiota is enormous, since it has been unraveled in these years that an imbalance of the microbial composition may lead to shift from the physiological state to dysbiosis, hence, from health to disease status (Althani et al., 2016). The microbial composition in the gut may be modulated also by a different diet (Del Chierico et al., 2014). In fact, different feeding modalities during the first months of life is among the factors that explain the high variability of the microbiota

after birth (Putignani et al., 2014). Manipulation of intestinal microbes due to mode of delivery (cesarean section versus vaginal delivery), breast feeding/formula/mixed feeding, overuse/misuse of antibiotics, prebiotics and probiotics, dietary modification, regional lifestyle, and ultimately fecal microbiota transplantation are among the external factors influencing gut microbiota composition (Arrieta et al., 2014; Zhuang et al., 2019; Niu et al., 2020).

However, many other biologically relevant molecules (i.e., microRNAs or miRNAs) can modulate gut microbiota homeostasis and composition. In the last few years, the discovery of miRNAs in biological fluids has generated a great interest for their potential use as biomarkers. Moreover, many miRNAs are stably expressed in various body fluids and a recent paper reported a good correlation of circulating miRNAs in body fluids with that of tissue miRNAs, opening the way to use miRNAs as biomarkers to monitor corresponding specific human diseases (Cui and Cui, 2020).

Circulating biomarkers play a significant role in clinical applications especially for the diagnosis of specific diseases, to monitor the therapeutic effect of a drug or to predict the tumor recurrence in chemotherapy-treated patients (Kosaka et al., 2010). As an example, Chim and coworkers were the first to identify circulating placental miRNAs in the plasma of pregnant women (Chim et al., 2008). Circulating miRNAs have many of the essential characteristics of good biomarkers: they are stable in the circulation and resistant to digestion by RNAses, extreme pH, high temperatures, extended storage and multiple freeze-thaw cycles (Mitchell et al., 2008; Chen et al., 2009). In many cases, changes in circulating miRNA expression levels have been associated with different diseases or certain biological/pathological stages. Circulating miRNAs are released in the bloodstream into many forms, although their origin and the mechanism of their release have not been completely elucidated. The reasons of the high stability of circulating miRNAs remain largely unknown as well, although several hypothesis have been suggested (Cortez et al., 2011).

Despite the few information about the biogenesis and stability of circulating miRNAs, we know more about their functional role in health and disease, as briefly discussed in the following paragraphs.

SERUM/PLASMA miRNAs

Circulating miRNAs are small RNA molecules that can be detected, free or encapsulated in vesicles (i.e., exosomes), in many biological fluids (i.e., blood, serum, plasma, saliva, urine, etc.) (Weber et al., 2010). Since many years, circulating miRNAs are considered powerful diagnostic biomarkers for many diseases and the novel frontier of liquid biopsies (Pardini et al., 2019). It is quite impossible to cite all the papers dealing with circulating miRNAs in health and disease conditions, as many papers appeared in the literature but we will cite only a few of them related to pediatric diseases.

In neonatology, miRNAs from cord blood showed a very low correlation with maternal blood miRNAs expression but instead they appeared to be powerful early biomarkers of child health (Cretoiu et al., 2016; Dypas et al., 2020). Circulating miRNAs are also considered good diagnostic tools for example in celiac disease and its treatment (Felli et al., 2022) or to detect the evolution of liver diseases (Calkins et al., 2020; Resaz et al., 2021; Nagura et al., 2022).

Circulating miRNAs are also widely studied also in cancer disease (Colletti et al., 2019; Galardi et al., 2019). In cardiology, specific sets of miRNAs are able to distinguish patients with supraventricular (SVa) and ventricular (Va) arrhythmias from the controls (Moric-Janiszewska et al., 2021).

In neurological and neurodevelopmental child disorders, serum miRNAs have been used as biomarker for autism spectrum disorder (ASD) (Vasu et al., 2014), and temporal lobe epilepsy (TLE) in children (Niu et al., 2021).

Finally, the identification of specific circulating miRNAs in children with type-1 diabetes (T1D) allowed to discriminate early and late stages of diabetes (Erener et al., 2017; Margaritis et al., 2021), the glycemic status and the ongoing islet autoimmunity in high-risk individuals (Akerman et al., 2018). Circulating miRNAs can also predict metabolic diseases such as insulin resistance in obese pre-schoolers (Masotti et al., 2017).

Circulating miRNAs can act not only as "passive" biomarkers but to have also an "active" communication role in distant organs (Valadi et al., 2007). However, one of their most interesting role is to be part of a more complex network of communication mediated by exosomes. In particular, miRNAs have both a regulatory effect on pathogen infections (i.e., by mediating further infection by transmitting pathogen-related molecules, participating in the immune escape of pathogens and inhibiting immune responses by favoring immune cell apoptosis) and an anti-infection role (i.e., by inhibiting pathogen proliferation and infection directly or inducing immune responses such as those related to the function of monocyte-macrophages, NK cells, T cells, and B cells) (Zhang et al., 2018).

We have reported that HIV-1 Nef protein is secreted also within exosomes and contributes to regulate the intercellular communication exploiting the vesicular trafficking machinery of the host (Felli et al., 2017b). This can be considered as a potential inter-kingdom communication pathway between virus and humans, where viral Nef contributes to modulate and post-transcriptionally regulate the host gene expression and immune response. Therefore, we cannot exclude that these exosomes contain also miRNAs and that these circulating molecules might have a modulatory role, not exclusively in infections.

FECAL miRNAs

Recently, the novel concept of intestinal epithelial cells able to release luminal regulatory miRNAs (i.e., fecal miRNAs) was described (Masotti, 2012; Liu and Weiner, 2016; Felli et al., 2017a; O'Brien et al., 2018; Li et al., 2020; Sarshar et al., 2020b). Noticeably, less is known about the effect of bacterial pathogens on host miRNA expression, as well as the reciprocal effect (i.e., the role of host cell miRNAs on modulating bacterial infections). Interestingly, the expression of miRNAs have been recently

reported to be correlated with the richness and diversity of the microbial community (Liu et al., 2016; Ragusa et al., 2020; Del Pozo-Acebo et al., 2021). Thus, miRNAs may enter bacteria and, in turn, affect their biological processes by regulating bacterial gene expression and growth, therefore, giving to pathogenic bacteria the opportunity to expand, leading to dysbiosis (Liu and Weiner, 2016; Liu et al., 2016; Wu et al., 2016). Liu and colleagues showed that fecal samples contained small RNA species (size ranging between ∼20–100 nt) similar to extracellular/exosomal RNA. They further suggested that fecal miRNAs, mainly released by intestinal epithelial cells (IEC) and Hopx-positive cells, regulate bacterial gene transcripts and affect bacterial growth such as *Fusobacterium nucleatum* and *Escherichia coli* (*E. coli*) (Liu et al., 2016). In addition, they found that gut dysbiosis in deficient IEC-fecal miRNA could be restored through WT fecal miRNA transplantation. This group later proposed the term of "miRNA-microbiome axis" as a potential therapeutic approach.

A recent study reported that IECs release miRNAs packed in exosomes and these IEC-exosomal miRNAs in turn influence the composition of gut bacterial populations (Kumar et al., 2021). Specifically, IEC-exosomes could be taken up by gut bacteria and inhibit the expression of the *E. coli tnaA* gene and its indole production, an intracellular signal in microbial communities mainly in Gram negative bacteria.

SALIVARY miRNAs AND ORAL MICROBIOTA

For many years until now, blood has been considered the biological source of choice for the diagnosis of many diseases and for clinical monitoring. So far, no alternatives were considered valid and reliable. The concept of using less invasive methods, such as saliva (i.e., oral diagnostics) has been increasingly adopted in the clinical field owing to the presence of specific molecules or microorganisms that can be used as valuable biomarkers. In fact, these components are not only altered in oral diseases but also correlated with damaged tissues or distal organs. Therefore, oral fluids have been studied in details for both diagnostic and potential therapeutic applications. Five major constituents can be considered for these purposes: proteins, mRNAs, miRNAs, metabolites and microbes that can be altered alone or in combination according to the analyzed disease.

Some authors focused on salivary miRNAs as diagnostic biomarkers of hand, foot and mouth disease (HFMD) in pediatric patients, and they found that miR-221 was consistently and significantly downregulated in the patients cohort (Min et al., 2018) supporting the method of using salivary miRNAs to diagnose this infection. Other authors identified the salivary miR-4668 as a novel potential biomarker of eosinophilic esophagitis (Bhardwaj et al., 2020).

Recent evidences have demonstrated that salivary molecules, as well as bacterial populations, can be dysregulated by several pathological conditions including neuro-psychiatric diseases such as the elusive Autistic Spectrum Disorder (ASD) (Ragusa et al., 2020). The authors performed a combined approach of miRNA expression profiling and 16S rRNA microbiome analysis on saliva from 53 ASD and 27 neurologically unaffected control children. The authors found that miRNAs and microbes were statistically associated to different neuropsychological scores related to anomalies in social interaction and communication and among the prevalent miRNA/bacteria associations, the most relevant was the negative correlation between salivary miR-141-3p expression and *Tannerella* spp. abundance. Furthermore, the authors suggest that a potential cross-talk between circulating miRNAs and resident bacteria could occur in saliva of ASD.

To add further complexity to these relationships, the recent evidences that circadian changes may be in relation to, or affect, the host microbiota (i.e., gut microbiota) (Thaiss et al., 2016; Dickson, 2017) solicited Hicks et al. (2018) to investigate the daily oscillations in salivary miRNAs and microbial RNAs. They explored relationships between circadian oscillations in human salivary miRNAs and microbial RNAs that may have distinct implications in human health and disease (Hicks et al., 2018). However, all of these studies suggest the need of a reliable method to study this complex relationship.

In the last few years, the Salivaomics Knowledge Base (SKB) consisting in data repositories, management systems and dedicated web resources has been established to support human salivary research in these fields (i.e., proteomics, transcriptomics, miRomics, metabolomics and microbiota analysis) (Shah, 2018).

URINARY miRNAs AND MICROBIOTA

Urine is the metabolic fluid excreted through the urinary system and contains not only bacteria, but also exosomes and miRNAs, among the others (Weber et al., 2010). Urinary microbiota refers to the microbial community present in urines, and generally contains a different and less abundant microbial population compared to other sites of the body (i.e., the gut). Nevertheless, dysbiosis in the urinary tract leads to urinary tract infections (UTIs) such as cystitis, although the interaction between different microbiota ecosystems has been poorly investigated (Ceprnja et al., 2021; Martinez et al., 2021; Perez-Carrasco et al., 2021). However, we know that different species of *Lactobacillus* (i.e., *L. crispatus*, *L. gasseri*, *L. iners*, and *L. jensenii*) are dominant in the urinary tract and protect the host, and further maintain a good microbial balance. In women, a part of this microbiota is represented by *Gardnerella*, *Sneathia*, *Staphylococcus*, and *Enterobacteriacae* members (Brubaker and Wolfe, 2017), whereas in men, the urobioma includes mainly *Firmicutes*, *Actinobacteria*, *Proteobacteria* and *Bacteroidetes* (Nelson et al., 2010). Similarly to what already observed for gut microbiota, the urinary microbiota can be affected by external environmental factors that may alter its composition leading to dysbiosis. Interestingly, the discovery of exosomes present also in urine, referred as urinary exosomes, outlined some interesting properties of these vesicles.

UTIs, which include cystitis and pyelonephritis, are mainly caused by uropathogenic *E. coli* (UPEC) strains that originate from the intestinal microbiota through the migration to the perianal region and to the urinary tract (the fecal-perineal-urethral route) (Scribano et al., 2021) and the establishment of adaptive strategies during host cell adhesion (Sarshar et al., 2022).

Several strategies (i.e., anti-adhesive molecules) have been suggested as good antibiotic alternatives to prevent UTIs (Sarshar et al., 2020a; Scribano et al., 2020). Interestingly, also urinary exosomes have the ability to inhibit the growth of, and kill, bacteria such as UPEC in the urinary tract (Hiemstra et al., 2014). Exosomes can also have an inhibitory role in viral infections while they may also facilitate the spread of viral (Caobi et al., 2020) or parasitic (Wu et al., 2019) infections. Since exosomes contain miRNAs, potent gene expression modulators, it is not surprising to observe these effects and expect to find circulating miRNAs in urine. In fact, Zhao and collaborators showed that there were a large number of differentially expressed miRNAs in urinary exosomes of type 2 diabetes mellitus (T2DM) as well as Diabetic kidney disease (DKD) patients, and that some of these urinary miRNAs are also able to predict early microalbuminuria (Zhao et al., 2020). A study reported by Scian and coworkers showed that urinary miRNAs are correlated with renal dysfunction and to pathophysiological processes. In fact, the authors investigated the expression profiles of miRNAs in patients with chronic renal transplant dysfunction both in kidney biopsies and in paired urine samples (Scian et al., 2011) and detected a differential expression for several miRNAs in renal biopsies, confirmed by the same dysregulated miRNAs in urines.

Another study by Yang and collaborators evaluated the expression changes of urinary miRNAs in the progression of diabetic nephropathy (DN) and observed a distinct profile of dysregulated miRNAs in these patients. This findings further suggested the potentiality of urine-specific miRNAs as novel biomarkers for the diagnosis of early stages of diabetic nephropathy (Yang et al., 2013).

In the oncology field, other studies validated the reliability of urinary miRNAs as biomarkers of diseases, such as in the case of esophageal cancers, where the urinary levels of five miRNAs were reported significantly higher in the adenocarcinoma and esophageal squamous cell carcinoma patient group compared to the control group (Okuda et al., 2021). In pancreatic adenocarcinomas, three urinary miRNAs were significantly overexpressed in patients with stage-I cancer compared to healthy individuals (Debernardi et al., 2015), whereas nine urinary miRNAs were able to distinguish early-stages of renal cell carcinomas (RCC) or progressive and non-progressive tumors (Di Meo et al., 2020).

ARTIFICIAL INTELLIGENCE AND SYSTEMS BIOLOGY: FROM DIAGNOSIS TO PERSONALIZED MEDICINE

In the previous paragraphs, we have discussed the recent findings reported in the literature about the use of miRNAs present in different biological fluids and specimens (i.e., serum, urine, saliva, stool) for diagnostic purposes. However, to have a more complete picture of the disease and to unravel novel biological and pathological links among different interactors, we think that we should analyze data coming from multiple "omics" techniques (i.e., circulating miRNAs, microbiome, metabolome, exposome, etc.) with a more integrated approach that takes into consideration the big-data produced by all of these techniques. The approach that we think feasible is to employ Artificial Intelligence (AI) and Machine Learning (ML) algorithms (**Figure 1**).

ML is defined as an ensemble of AI algorithms that enable a robust interrogation of multiple datasets aimed at identifying previously undiscovered patterns and relationships in the data. This approach has been applied successfully to the comprehension of neurodegenerative diseases and to the early diagnosis, prognosis and discovery of potential therapies (Myszczynska et al., 2020) or to predict the associations between miRNAs and diseases (Chen et al., 2017, 2018; He et al., 2018; Yao et al., 2019). Therefore, many different ML algorithms are available and depending on the data to analyze, it is essential to choose the most suitable one. In general, all of the approaches involve the division of the overall sample size in two subgroups: discovery and validation sets. ML algorithms have the potential to acquire information about the prognosis or to facilitate patient stratification and classification. The authors conclude by saying that large sets of curated datasets are needed to run ML algorithms and that one should rely on robust methods to obtain reliable diagnostic and prognostic models.

According to Lin and Lane, there is a huge demand for establishing algorithms in machine learning and systems genomics (MLSG) to establish genotype-phenotype relationships and find useful protocols for the discovery of diagnostics and therapeutics (Lin and Lane, 2017). Without going into details of the various algorithms employed in ML methods (i.e., naive Bayes, random forest, C4.5 decision tree, artificial neural networks, support vector machine, k-Means, k-nearest neighbors, and regression), we would like to present our personal view on the ideal approach that could be followed with miRNAs profiling and microbiota data by discussing some recent applications.

In serum/plasma miRNAs discovery, four distinct ML methods (i.e., the random forest wrapper method "Boruta," the regression partition trees "Rpart," LASSO regression, and the extreme gradient boosting "XGBoost") and one univariate analysis have been employed to identify two (out of 20) circulating miRNAs (miR-636 and miR-187-5p) that have been found at the same time by all of the methods and that were able to diagnose with high accuracy patients with pulmonary arterial hypertension (PAH) (Errington et al., 2021). Moreover, the target genes of these two miRNAs were validated in two independent sets of human pulmonary artery smooth muscle cells, affording two genes differentially expressed *in vitro*, giving also some insights into the pathogenesis of PAH.

For urinary miRNAs, the majority of papers are related to cancer biomarker discovery (i.e., prostate cancers) and to the identification of small RNAs or miRNAs able to discriminate, for example, prostate cancers from benign prostate hyperplasia (BPH). In fact, Markert and collaborators used the random forest method to identify candidate small RNAs (Markert et al., 2021), and in other studies the same group focused on miRNAs (Holdmann et al., 2022).

Salivary miRNAs obtained by small RNA sequencing have been validated using ML with random forests as accessible

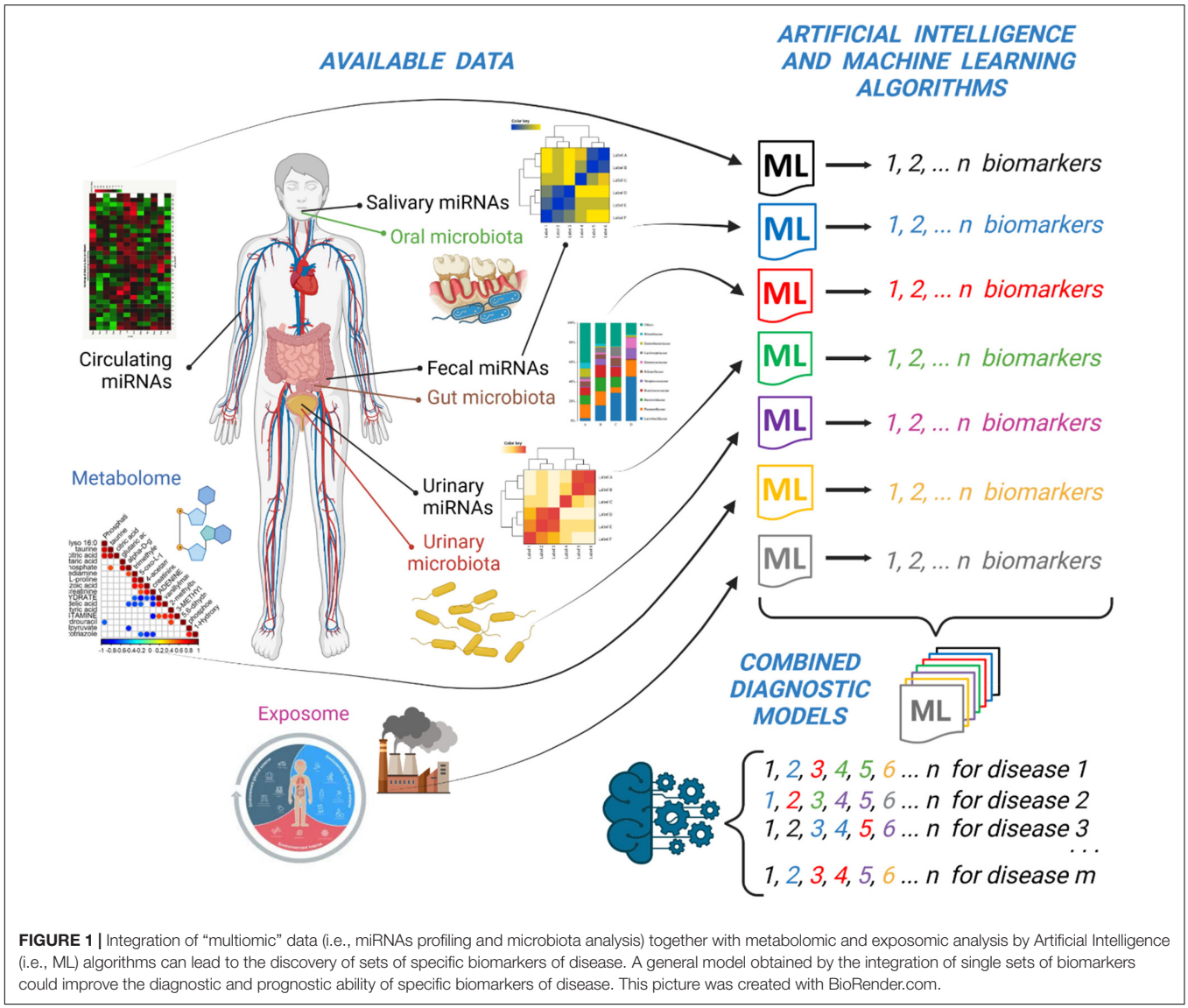


FIGURE 1 | Integration of "multiomic" data (i.e., miRNAs profiling and microbiota analysis) together with metabolomic and exposomic analysis by Artificial Intelligence (i.e., ML) algorithms can lead to the discovery of sets of specific biomarkers of disease. A general model obtained by the integration of single sets of biomarkers could improve the diagnostic and prognostic ability of specific biomarkers of disease. This picture was created with BioRender.com.

and affordable biomarkers of alcohol dependence (AD) in a large group of 56 African-Americans and 64 European-Americans (Rosato et al., 2019). In another study, by applying support vector machine (SVM) models and leave-one-out cross-validation, the authors reported that salivary miRNAs were effective non-invasive biomarkers of hepatocellular carcinoma (Mariam et al., 2022).

Surprisingly, no studies dealing with fecal miRNAs and disease biomarker discovery have applied ML methods, but this can be easily explained by considering that fecal miRNAs represent a novel and emerging topic in the clinical diagnostic field.

On the contrary, both stool and salivary microbiota have been analyzed by ML algorithms. In a recent study, Saboo and collaborators aimed at determining the ability of stool vs. salivary microbiota to differentiate between cirrhosis and hepatic encephalopathy groups based on disease severity using ML. In particular, they classified stool and salivary microbiota using four different ML techniques (i.e., random forest, support vector machine, logistic regression, and gradient boosting) leading to the conclusion that stool microbiota are superior to saliva in distinguishing these two groups (Saboo et al., 2022).

Interestingly, in order to evaluate whether SARS-CoV-2 infected individuals should undergo hospitalization and reach the emergency department according to their symptoms, salivary metabolome have been correlated to serum biomarkers and salivary microbiota and analyzed by ML algorithms (Pozzi et al., 2022). Saliva and blood samples were collected over the course of two observational cohort studies. The results outlined that nine salivary metabolites are enough to assess the need of hospitalization. However, when these metabolites were combined with serum biomarkers and correlated with modulated microbiota taxa, just two salivary metabolites and one serum protein were sufficient to identify the patient to hospitalize.

Therefore, this is a clear example that a combined omic analysis coupled to ML algorithms is advantageous to stratify patients affected by several diseases.

Therefore, we suggest to consider an AI method that could integrate the results of many ML algorithms applied to different profiles of miRNAs, microbiota or metabolites under investigation (**Figure 1**). This AI model should be ideally made by a combination of these single models obtained by single experiments. In this way, the distinctive circulating miRNAs obtained by one ML method can be listed together with fecal or urinary miRNAs or to other metabolites obtained by different ML methods. The overall "matrix" of data will be made by the *ensemble* of all of the potential biomarkers that will constitute a more robust model to predict a disease. It could be possible in the future to diagnose a disease (i.e., celiac disease) by simply looking at a couple of circulating miRNAs and a selection of specific fecal or urinary biomarkers (Paolini et al., 2022) or to the level of some metabolites obtained by analytical laboratory techniques (**Figure 1**).

OTHER POTENTIAL INTERACTORS TO COMPLETE THE OVERALL PICTURE

As reported by Liu, miRNAs and bacteria are functionally linked and a mutual host-guest regulation mediated by miRNAs and bacterial extracellular vesicles cannot be excluded. We also know that miRNAs produced in one tissue or circulating in serum/plasma is a direct consequence of what happens in distant organs, or they are a representation of a healthy/disease status. Therefore, we cannot exclude that circulating miRNAs, salivary miRNAs, urinary or fecal miRNAs can have a common denominator in describing the healthy/disease status of an individual and that a direct or indirect correlation may exist in different "districts" analyzed (**Figure 2**). Moreover, owing to the direct effects of miRNAs on bacterial gene regulation and therefore their function, as demonstrated by Liu, we can have multiple "regulatory pathways" coming from different parts of the human body that can have an overall beneficial or detrimental effect for human health (**Figure 2**).

Finally, it has been recently reported that also prokaryote cells are able to release outer membrane vesicles (OMVs) or extracellular vesicles (EVs) into the extracellular environments (Ahmadi Badi et al., 2020) and that specific isolation protocols can be used to characterize their structure (Sarra et al., 2020) and their small RNA content (Ahmadi Badi et al., 2020). This adds further complexity to the overall picture and another layer of possible interactions that should be considered by ML algorithms and investigated further.

For these reasons, AI approaches are needed not only to describe better the overall picture of a throughout screening of the patient, but also to unravel possible interactions among different districts, that are characteristics of the beginning or the progression of a certain disease. Moreover, AI and ML approach can help clinicians to classify the patient in a holistic way, by taking different analytical data and analyzing them in a coordinated and comprehensive way, and to define the best "category" where the patient belong to and opening the way to a laboratory-based personalized medicine approach. Finally, we foresee that the proposed strategy can be applied not only in the diagnostic field but also for the therapy of different adult and pediatric diseases and the follow-up of patients through the accurate identification and validation of the best miRNA targets (Major et al., 2020).

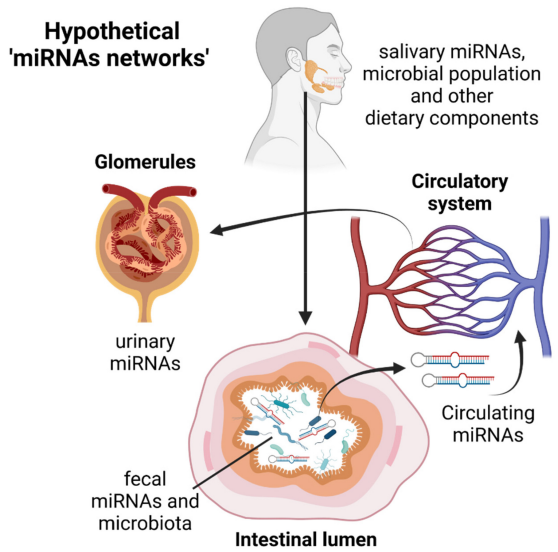


FIGURE 2 | The miRNAs journey through the human body: from the mouth to the intestine through the digestive tract, migration from the gut to the circulatory systems and excretion by urine. The possible miRNA-host-pathogen cross talks and network of interactions during this journey could better classify the patient's molecular characteristics (and the healthy/disease status) and drive clinicians to identify personalized therapeutic treatments. This picture was created with BioRender.com.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

AP, AB, SPB, and CF drafted the manuscript. CM prepared the pictures. SAB and SS revised the manuscript. MS and AM drafted the manuscript and edited the final version of the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

We would like to thank the Italian Ministry of Health for funding (Ricerca 5 × 1,000 to AM and Ricerca Finalizzata Starting Grant SG-2018-12365432 to MS).

REFERENCES

Ahmadi Badi, S., Bruno, S. P., Moshiri, A., Tarashi, S., Siadat, S. D., and Masotti, A. (2020). Small RNAs in Outer Membrane Vesicles and Their Function in Host-Microbe Interactions. *Front. Microbiol.* 11:1209. doi: 10.3389/fmicb.2020.01209

Akerman, L., Casas, R., Ludvigsson, J., Tavira, B., and Skoglund, C. (2018). Serum miRNA levels are related to glucose homeostasis and islet autoantibodies in children with high risk for type 1 diabetes. *PLoS One* 13:e0191067. doi: 10.1371/journal.pone.0191067

Althani, A., Marei, H., Hamdi, W. S., Nasrallah, G. K., El Zowalaty, M. E., Al Khdor, S., et al. (2016). Human Microbiome and Its Association With Health and Diseases. *J. Cell. Physiol.* 231, 1688–1694. doi: 10.1002/jcp.25284

Arrieta, M. C., Stiemsma, L. T., Amenyogbe, N., Brown, E. M., and Finlay, B. (2014). The intestinal microbiome in early life: health and disease. *Front. Immunol.* 5:427. doi: 10.3389/fimmu.2014.00427

Bhardwaj, N., Sena, M., Ghaffari, G., and Ishmael, F. (2020). MiR-4668 as a Novel Potential Biomarker for Eosinophilic Esophagitis. *Allergy Rhinol.* 11:2152656720953378. doi: 10.1177/2152656720953378

Brubaker, L., and Wolfe, A. J. (2017). The Female Urinary Microbiota/Microbiome: Clinical and Research Implications. *Rambam Maimonides Med. J.* 8:e0015. doi: 10.5041/RMMJ.10292

Calkins, K. L., Thamotharan, S., Ghosh, S., Dai, Y., and Devaskar, S. U. (2020). MicroRNA 122 Reflects Liver Injury in Children with Intestinal Failure-Associated Liver Disease Treated with Intravenous Fish Oil. *J. Nutr.* 150, 1144–1150. doi: 10.1093/jn/nxaa001

Caobi, A., Nair, M., and Raymond, A. D. (2020). Extracellular Vesicles in the Pathogenesis of Viral Infections in Humans. *Viruses* 12:1200. doi: 10.3390/v12101200

Ceprnja, M., Oros, D., Melvan, E., Svetlicic, E., Skrlin, J., Barisic, K., et al. (2021). Modeling of Urinary Microbiota Associated With Cystitis. *Front. Cell. Infect. Microbiol.* 11:643638. doi: 10.3389/fcimb.2021.643638

Chen, X., Wu, Q. F., and Yan, G. Y. (2017). RKNNMDA: Ranking-based KNN for MiRNA-Disease Association prediction. *RNA Biol.* 14, 952–962. doi: 10.1080/15476286.2017.1312226

Chen, X., Xie, D., Wang, L., Zhao, Q., You, Z. H., and Liu, H. (2018). BNPMDA: Bipartite Network Projection for MiRNA-Disease Association prediction. *Bioinformatics* 34, 3178–3186. doi: 10.1093/bioinformatics/bty333

Chen, Y., Gelfond, J. A., McManus, L. M., and Shireman, P. K. (2009). Reproducibility of quantitative RT-PCR array in miRNA expression profiling and comparison with microarray analysis. *BMC Genomics* 10:407. doi: 10.1186/1471-2164-10-407

Chim, S. S., Shing, T. K., Hung, E. C., Leung, T. Y., Lau, T. K., Chiu, R. W., et al. (2008). Detection and characterization of placental microRNAs in maternal plasma. *Clin. Chem.* 54, 482–490. doi: 10.1373/clinchem.2007.097972

Colletti, M., Paolini, A., Galardi, A., Di Paolo, V., Pascucci, L., Russo, I., et al. (2019). Expression profiles of exosomal miRNAs isolated from plasma of patients with desmoplastic small round cell tumor. *Epigenomics* 11, 489–500. doi: 10.2217/epi-2018-0179

Cortez, M. A., Bueso-Ramos, C., Ferdin, J., Lopez-Berestein, G., Sood, A. K., and Calin, G. A. (2011). MicroRNAs in body fluids–the mix of hormones and biomarkers. *Nat. Rev. Clin. Oncol.* 8, 467–477. doi: 10.1038/nrclinonc.2011.76

Cretoiu, D., Xu, J., Xiao, J., Suciu, N., and Cretoiu, S. M. (2016). Circulating MicroRNAs as Potential Molecular Biomarkers in Pathophysiological Evolution of Pregnancy. *Dis. Markers* 2016:3851054. doi: 10.1155/2016/3851054

Cui, C., and Cui, Q. (2020). The relationship of human tissue microRNAs with those from body fluids. *Sci. Rep.* 10, 5644–5620. doi: 10.1038/s41598-020-62534-6

Debernardi, S., Massat, N. J., Radon, T. P., Sangaralingam, A., Banissi, A., Ennis, D. P., et al. (2015). Noninvasive urinary miRNA biomarkers for early detection of pancreatic adenocarcinoma. *Am. J. Cancer. Res.* 5, 3455–3466.

Del Chierico, F., Vernocchi, P., Dallapiccola, B., and Putignani, L. (2014). Mediterranean diet and health: food effects on gut microbiota and disease control. *Int. J. Mol. Sci.* 15, 11678–11699. doi: 10.3390/ijms150711678

Del Pozo-Acebo, L., Lopez, de Las Hazas, M. C., Margolles, A., Davalos, A., and Garcia-Ruiz, A. (2021). Eating microRNAs: pharmacological opportunities for cross-kingdom regulation and implications in host gene and gut microbiota modulation. *Br. J. Pharmacol.* 178, 2218–2245. doi: 10.1111/bph.15421

Di Meo, A., Brown, M. D., Finelli, A., Jewett, M. A. S., Diamandis, E. P., and Yousef, G. M. (2020). Prognostic urinary miRNAs for the assessment of small renal masses. *Clin. Biochem.* 75, 15–22. doi: 10.1016/j.clinbiochem.2019.10.002

Dickson, I. (2017). Gut microbiota: Intestinal microbiota oscillations regulate host circadian physiology. *Nat. Rev. Gastroenterol. Hepatol.* 14:67. doi: 10.1038/nrgastro.2016.205

Dypas, L. B., Gutzkow, K. B., Olsen, A. K., and Duale, N. (2020). MiRNA profiles in blood plasma from mother-child duos in human biobanks and the implication of sample quality: Circulating miRNAs as potential early markers of child health. *PLoS One* 15:e0231040. doi: 10.1371/journal.pone.0231040

Erener, S., Marwaha, A., Tan, R., Panagiotopoulos, C., and Kieffer, T. J. (2017). Profiling of circulating microRNAs in children with recent onset of type 1 diabetes. *JCI Insight* 2:e89656. doi: 10.1172/jci.insight.89656

Errington, N., Iremonger, J., Pickworth, J. A., Kariotis, S., Rhodes, C. J., Rothman, A. M., et al. (2021). A diagnostic miRNA signature for pulmonary arterial hypertension using a consensus machine learning approach. *EBioMedicine* 69:103444. doi: 10.1016/j.ebiom.2021.103444

Felli, C., Baldassarre, A., and Masotti, A. (2017a). Intestinal and Circulating MicroRNAs in Coeliac Disease. *Int. J. Mol. Sci.* 18:1907. doi: 10.3390/ijms18091907

Felli, C., Baldassarre, A., Uva, P., Alisi, A., Cangelosi, D., Ancinelli, M., et al. (2022). Circulating microRNAs as novel non-invasive biomarkers of paediatric celiac disease and adherence to gluten-free diet. *EBioMedicine* 76:103851. doi: 10.1016/j.ebiom.2022.103851

Felli, C., Vincentini, O., Silano, M., and Masotti, A. (2017b). HIV-1 Nef Signaling in Intestinal Mucosa Epithelium Suggests the Existence of an Active Inter-kingdom Crosstalk Mediated by Exosomes. *Front. Microbiol.* 8:1022. doi: 10.3389/fmicb.2017.01022

Galardi, A., Colletti, M., Di Paolo, V., Vitullo, P., Antonetti, L., Russo, I., et al. (2019). Exosomal MiRNAs in Pediatric Cancers. *Int. J. Mol. Sci.* 20:4600. doi: 10.3390/ijms20184600

He, B. S., Qu, J., and Chen, M. (2018). Prediction of potential disease-associated microRNAs by composite network based inference. *Sci. Rep.* 8, 15813–15818. doi: 10.1038/s41598-018-34180-6

Hicks, S. D., Khurana, N., Williams, J., Dowd Greene, C., Uhlig, R., and Middleton, F. A. (2018). Diurnal oscillations in human salivary microRNA and microbial transcription: Implications for human health and disease. *PLoS One* 13:e0198288. doi: 10.1371/journal.pone.0198288

Hiemstra, T. F., Charles, P. D., Gracia, T., Hester, S. S., Gatto, L., Al-Lamki, R., et al. (2014). Human urinary exosomes as innate immune effectors. *J. Am. Soc. Nephrol.* 25, 2017–2027. doi: 10.1681/ASN.2013101066

Holdmann, J., Markert, L., Klinger, C., Kaufmann, M., Schork, K., Turewicz, M., et al. (2022). MicroRNAs from urinary exosomes as alternative biomarkers in the differentiation of benign and malignant prostate diseases. *J. Circ. Biomark* 11, 5–13. doi: 10.33393/jcb.2022.2317

Human Microbiome, Jumpstart Reference, Strains Consortium, Nelson, K. E., Weinstock, G. M., Highlander, S. K., et al. (2010). A catalog of reference genomes from the human microbiome. *Science* 328, 994–999.

Integrative HMP (iHMP) Research Network Consortium. (2019). The Integrative Human Microbiome Project. *Nature* 569, 641–648.

Kosaka, N., Iguchi, H., and Ochiya, T. (2010). Circulating microRNA in body fluid: a new potential biomarker for cancer diagnosis and prognosis. *Cancer. Sci.* 101, 2087–2092. doi: 10.1111/j.1349-7006.2010.01650.x

Kumar, A., Ren, Y., Sundaram, K., Mu, J., Sriwastva, M. K., Dryden, G. W., et al. (2021). miR-375 prevents high-fat diet-induced insulin resistance and obesity by targeting the aryl hydrocarbon receptor and bacterial tryptophanase (tnaA) gene. *Theranostics* 11, 4061–4077. doi: 10.7150/thno.52558

Li, M., Chen, W. D., and Wang, Y. D. (2020). The roles of the gut microbiota-miRNA interaction in the host pathophysiology. *Mol. Med.* 26, 101–120. doi: 10.1186/s10020-020-00234-7

Lin, E., and Lane, H. Y. (2017). Machine learning and systems genomics approaches for multi-omics data. *Biomark Res* 5:2. doi: 10.1186/s40364-017-0082-y

Liu, S., da Cunha, A. P., Rezende, R. M., Cialic, R., Wei, Z., Bry, L., et al. (2016). The Host Shapes the Gut Microbiota via Fecal MicroRNA. *Cell. Host Microbe* 19, 32–43. doi: 10.1016/j.chom.2015.12.005

Liu, S., and Weiner, H. L. (2016). Control of the gut microbiome by fecal microRNA. *Microb. Cell.* 3, 176–177. doi: 10.15698/mic2016.04.492

Major, J. L., Bagchi, R. A., Pires, and da Silva, J. (2020). Application of microRNA Database Mining in Biomarker Discovery and Identification of Therapeutic Targets for Complex Disease. *Methods Protoc.* 4:5. doi: 10.3390/mps4010005

Margaritis, K., Margioula-Siarkou, G., Margioula-Siarkou, C., Petousis, S., Kotanidou, E. P., Christoforidis, A., et al. (2021). Circulating serum and plasma levels of micro-RNA in type-1 diabetes in children and adolescents: A systematic review and meta-analysis. *Eur. J. Clin. Invest.* 51:e13510. doi: 10.1111/eci.13510

Mariam, A., Miller-Atkins, G., Moro, A., Rodarte, A. I., Siddiqi, S., Acevedo-Moreno, L. A., et al. (2022). Salivary miRNAs as non-invasive biomarkers of hepatocellular carcinoma: a pilot study. *PeerJ* 10:e12715. doi: 10.7717/peerj.12715

Markert, L., Holdmann, J., Klinger, C., Kaufmann, M., Schork, K., Turewicz, M., et al. (2021). Small RNAs as biomarkers to differentiate benign and malign prostate diseases: An alternative for transrectal punch biopsy of the prostate? *PLoS One* 16:e0247930. doi: 10.1371/journal.pone.0247930

Martinez, J. E., Vargas, A., Perez-Sanchez, T., Encio, I. J., Cabello-Olmo, M., and Barajas, M. (2021). Human Microbiota Network: Unveiling Potential Crosstalk between the Different Microbiota Ecosystems and Their Role in Health and Disease. *Nutrients* 13:2905. doi: 10.3390/nu13092905

Masotti, A. (2012). Interplays between gut microbiota and gene expression regulation by miRNAs. *Front. Cell. Infect. Microbiol.* 2:137. doi: 10.3389/fcimb.2012.00137

Masotti, A., Baldassarre, A., Fabrizi, M., Olivero, G., Loreti, M. C., Giammaria, P., et al. (2017). Oral glucose tolerance test unravels circulating miRNAs associated with insulin resistance in obese preschoolers. *Pediatr. Obes.* 12, 229–238. doi: 10.1111/ijpo.12133

Min, N., Sakthi Vale, P. D., Wong, A. A., Tan, N. W. H., Chong, C. Y., Chen, C. J., et al. (2018). Circulating Salivary miRNA hsa-miR-221 as Clinically Validated Diagnostic Marker for Hand, Foot, and Mouth Disease in Pediatric Patients. *EBioMedicine* 31, 299–306. doi: 10.1016/j.ebiom.2018.05.006

Mitchell, P. S., Parkin, R. K., Kroh, E. M., Fritz, B. R., Wyman, S. K., Pogosova-Agadjanyan, E. L., et al. (2008). Circulating microRNAs as stable blood-based markers for cancer detection. *Proc. Natl. Acad. Sci. U.S.A.* 105, 10513–10518. doi: 10.1073/pnas.0804549105

Moric-Janiszewska, E., Smolik, S., Morka, A., Szydlowski, L., and Kapral, M. (2021). Expression levels of serum circulating microRNAs in pediatric patients with ventricular and supraventricular arrhythmias. *Adv. Med. Sci.* 66, 411–417. doi: 10.1016/j.advms.2021.08.003

Myszczynska, M. A., Ojamies, P. N., Lacoste, A. M. B., Neil, D., Saffari, A., Mead, R., et al. (2020). Applications of machine learning to diagnosis and treatment of neurodegenerative diseases. *Nat. Rev. Neurol.* 16, 440–456. doi: 10.1038/s41582-020-0377-8

Nagura, Y., Matsuura, K., Iio, E., Fujita, K., Inoue, T., Matsumoto, A., et al. (2022). Serum miR-192-5p levels predict the efficacy of pegylated interferon therapy for chronic hepatitis B. *PLoS One* 17:e0263844. doi: 10.1371/journal.pone.0263844

Nelson, D. E., Van Der Pol, B., Dong, Q., Revanna, K. V., Fan, B., Easwaran, S., et al. (2010). Characteristic male urine microbiomes associate with asymptomatic sexually transmitted infection. *PLoS One* 5:e14116. doi: 10.1371/journal.pone.0014116

Niu, J., Xu, L., Qian, Y., Sun, Z., Yu, D., Huang, J., et al. (2020). Evolution of the Gut Microbiome in Early Childhood: A Cross-Sectional Study of Chinese Children. *Front. Microbiol.* 11:439. doi: 10.3389/fmicb.2020.00439

Niu, X., Zhu, H. L., Liu, Q., Yan, J. F., and Li, M. L. (2021). MiR-194-5p serves as a potential biomarker and regulates the proliferation and apoptosis of hippocampus neuron in children with temporal lobe epilepsy. *J. Chin. Med. Assoc.* 84, 510–516. doi: 10.1097/JCMA.0000000000000518

O'Brien, J., Hayder, H., Zayed, Y., and Peng, C. (2018). Overview of MicroRNA Biogenesis. Mechanisms of Actions, and Circulation. *Front. Endocrinol.* 9:402. doi: 10.3389/fendo.2018.00402

Okuda, Y., Shimura, T., Iwasaki, H., Fukusada, S., Nishigaki, R., Kitagawa, M., et al. (2021). Urinary microRNA biomarkers for detecting the presence of esophageal cancer. *Sci. Rep* 11, 8508–8521. doi: 10.1038/s41598-021-87925-1

Paolini, A., Sarshar, M., Felli, C., Bruno, S. P., Rostami-Nejad, M., Ferretti, F., et al. (2022). Biomarkers to Monitor the Adherence to Gluten-free Diet by Celiac Disease Patients: Gluten Immunogenic Peptides and Urinary miRNAs. *Foods* 11:1380.

Pardini, B., Sabo, A. A., Birolo, G., and Calin, G. A. (2019). Noncoding RNAs in Extracellular Fluids as Cancer Biomarkers: The New Frontier of Liquid Biopsies. *Cancers* 11:1170. doi: 10.3390/cancers11081170

Perez-Carrasco, V., Soriano-Lerma, A., Soriano, M., Gutierrez-Fernandez, J., and Garcia-Salcedo, J. A. (2021). Urinary Microbiome: Yin and Yang of the Urinary Tract. *Front. Cell. Infect. Microbiol.* 11:617002. doi: 10.3389/fcimb.2021.617002

Pozzi, C., Levi, R., Braga, D., Carli, F., Darwich, A., Spadoni, I., et al. (2022). A 'Multiomic' Approach of Saliva Metabolomics, Microbiota, and Serum Biomarkers to Assess the Need of Hospitalization in Coronavirus Disease 2019. *Gastro Hep. Adv.* 1, 194–209. doi: 10.1016/j.gastha.2021.12.006

Putignani, L., Del Chierico, F., Petrucca, A., Vernocchi, P., and Dallapiccola, B. (2014). The human gut microbiota: a dynamic interplay with the host from birth to senescence settled during childhood. *Pediatr. Res.* 76, 2–10. doi: 10.1038/pr.2014.49

Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K. S., Manichanh, C., et al. (2010). A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* 464, 59–65. doi: 10.1038/nature08821

Ragusa, M., Santagati, M., Mirabella, F., Lauretta, G., Cirnigliaro, M., Brex, D., et al. (2020). Potential Associations Among Alteration of Salivary miRNAs, Saliva Microbiome Structure, and Cognitive Impairments in Autistic Children. *Int. J. Mol. Sci.* 21:6203. doi: 10.3390/ijms21176203

Resaz, R., Cangelosi, D., Segalerba, D., Morini, M., Uva, P., Bosco, M. C., et al. (2021). Exosomal MicroRNAs as Potential Biomarkers of Hepatic Injury and Kidney Disease in Glycogen Storage Disease Type Ia Patients. *Int. J. Mol. Sci.* 23:328. doi: 10.3390/ijms23010328

Rosato, A. J., Chen, X., Tanaka, Y., Farrer, L. A., Kranzler, H. R., Nunez, Y. Z., et al. (2019). Salivary microRNAs identified by small RNA sequencing and machine learning as potential biomarkers of alcohol dependence. *Epigenomics* 11, 739–749. doi: 10.2217/epi-2018-0177

Saboo, K., Petrakov, N. V., Shamsaddini, A., Fagan, A., Gavis, E. A., Sikaroodi, M., et al. (2022). Stool microbiota are superior to saliva in distinguishing cirrhosis and hepatic encephalopathy using machine learning. *J. Hepatol.* 76, 600–607. doi: 10.1016/j.jhep.2021.11.011

Sarra, A., Celluzzi, A., Bruno, S. P., Ricci, C., Sennato, S., Ortore, M. G., et al. (2020). Biophysical Characterization of Membrane Phase Transition Profiles for the Discrimination of Outer Membrane Vesicles (OMVs) From *Escherichia coli* Grown at Different Temperatures. *Front. Microbiol.* 11:290. doi: 10.3389/fmicb.2020.00290

Sarshar, M., Behzadi, P., Ambrosi, C., Zagaglia, C., Palamara, A. T., and Scribano, D. (2020a). FimH and Anti-Adhesive Therapeutics: A Disarming Strategy Against Uropathogens. *Antibiotics* 9:397. doi: 10.3390/antibiotics9070397

Sarshar, M., Scribano, D., Ambrosi, C., Palamara, A. T., and Masotti, A. (2020b). Fecal microRNAs as Innovative Biomarkers of Intestinal Diseases and Effective Players in Host-Microbiome Interactions. *Cancers* 12:2174. doi: 10.3390/cancers12082174

Sarshar, M., Scribano, D., Limongi, D., Zagaglia, C., Palamara, A. T., and Ambrosi, C. (2022). Adaptive strategies of uropathogenic *Escherichia coli* CFT073: from growth in lab media to virulence during host cell adhesion. *Int. Microbiol.* [Epub ahead of print]. doi: 10.1007/s10123-022-00235-y

Scian, M. J., Maluf, D. G., David, K. G., Archer, K. J., Suh, J. L., Wolen, A. R., et al. (2011). MicroRNA profiles in allograft tissues and paired urines associate with chronic allograft dysfunction with IF/TA. *Am. J. Transplant.* 11, 2110–2122. doi: 10.1111/j.1600-6143.2011.03666.x

Scribano, D., Sarshar, M., Fettucciari, L., and Ambrosi, C. (2021). Urinary tract infections: Can we prevent uropathogenic *Escherichia coli* infection with dietary intervention? *Int. J. Vitam. Nutr. Res.* 91, 391–395. doi: 10.1024/0300-9831/a000704

Scribano, D., Sarshar, M., Prezioso, C., Lucarelli, M., Angeloni, A., Zagaglia, C., et al. (2020). d-Mannose Treatment neither Affects Uropathogenic *Escherichia coli* Properties nor Induces Stable FimH Modifications. *Molecules* 25:316. doi: 10.3390/molecules25020316

Shah, S. (2018). Salivaomics: The current scenario. *J. Oral. Maxillofac. Pathol.* 22, 375–381. doi: 10.4103/jomfp.JOMFP_171_18

Thaiss, C. A., Levy, M., Korem, T., Dohnalová, L., Shapiro, H., Jaitin, D. A., et al. (2016). Microbiota Diurnal Rhythmicity Programs Host Transcriptome Oscillations. *Cell* 167, 1495.e–1510.e. doi: 10.1016/j.cell.2016.11.003

Valadi, H., Ekstrom, K., Bossios, A., Sjostrand, M., Lee, J. J., and Lotvall, J. O. (2007). Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells. *Nat. Cell Biol.* 9, 654–659. doi: 10.1038/ncb1596

Vasu, M. M., Anitha, A., Thanseem, I., Suzuki, K., Yamada, K., Takahashi, T., et al. (2014). Serum microRNA profiles in children with autism. *Mol. Autism* 5, 40–2392–5–40. eoeton 2014. doi: 10.1186/2040-2392-5-40

Weber, J. A., Baxter, D. H., Zhang, S., Huang, D. Y., Huang, K. H., Lee, M. J., et al. (2010). The microRNA spectrum in 12 body fluids. *Clin. Chem.* 56, 1733–1741. doi: 10.1373/clinchem.2010.147405

Wu, Z., Qin, W., Wu, S., Zhu, G., Bao, W., and Wu, S. (2016). Identification of microRNAs regulating *Escherichia coli* F18 infection in Meishan weaned piglets. *Biol. Direct* 11, 59–16. doi: 10.1186/s13062-016-0160-3

Wu, Z., Wang, L., Li, J., Wang, L., Wu, Z., and Sun, X. (2019). Extracellular Vesicle-Mediated Communication Within Host-Parasite Interactions. *Front. Immunol.* 9:3066. doi: 10.3389/fimmu.2018.03066

Yang, Y., Xiao, L., Li, J., Kanwar, Y. S., Liu, F., and Sun, L. (2013). Urine miRNAs: potential biomarkers for monitoring progression of early stages of diabetic nephropathy. *Med. Hypotheses* 81, 274–278. doi: 10.1016/j.mehy.2013.04.031

Yao, D., Zhan, X., and Kwoh, C. K. (2019). An improved random forest-based computational model for predicting novel miRNA-disease associations. *BMC Bioinformatics* 20, 624–619. doi: 10.1186/s12859-019-3290-7

Zhang, W., Jiang, X., Bao, J., Wang, Y., Liu, H., and Tang, L. (2018). Exosomes in Pathogen Infections: A Bridge to Deliver Molecules and Link Functions. *Front. Immunol.* 9:90. doi: 10.3389/fimmu.2018.00090

Zhao, Y., Shen, A., Guo, F., Song, Y., Jing, N., Ding, X., et al. (2020). Urinary Exosomal MiRNA-4534 as a Novel Diagnostic Biomarker for Diabetic Kidney Disease. *Front. Endocrinol.* 11:590. doi: 10.3389/fendo.2020.00590

Zhuang, L., Chen, H., Zhang, S., Zhuang, J., Li, Q., and Feng, Z. (2019). Intestinal Microbiota in Early Life and Its Implications on Childhood Health. *Genomics Proteomics Bioinform.* 17, 13–25. doi: 10.1016/j.gpb.2018.10.002

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.

frontiers | Frontiers in Microbiology

MINI REVIEW
published: 17 June 2022
doi: 10.3389/fmicb.2022.883734

Metaomics in Clinical Laboratory: Potential Driving Force for Innovative Disease Diagnosis

Liang Wang[1†], Fen Li[2†], Bin Gu[1†], Pengfei Qu[3], Qinghua Liu[4], Junjiao Wang[1], Jiawei Tang[1], Shubin Cai[5], Qi Zhao[6] and Zhong Ming[5*]*

[1] Department of Bioinformatics, School of Medical Informatics and Engineering, Xuzhou Medical University, Xuzhou, China, [2] Department of Laboratory Medicine, Huaiyin Hospital, Huai'an, China, [3] The First School of Clinical Medicine, Xuzhou Medical University, Xuzhou, China, [4] State Key Laboratory of Quality Research in Chinese Medicines, Macau University of Science and Technology, Taipa, Macao SAR, China, [5] College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, China, [6] School of Computer Science and Software Engineering, University of Science and Technology Liaoning, Anshan, China

OPEN ACCESS

Edited by:
Huanzi Zhong, Beijing Genomics Institute (BGI), China

Reviewed by:
Arianna Basile, University of Padua, Italy
Yu-Wei Wu, Taipei Medical University, Taiwan

***Correspondence:**
Qi Zhao
zhaoqi@lnu.edu.cn
Zhong Ming
mingz@szu.edu.cn

[†] These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Systems Microbiology, a section of the journal Frontiers in Microbiology

Received: *25 February 2022*
Accepted: *18 May 2022*
Published: *17 June 2022*

Citation:
Wang L, Li F, Gu B, Qu P, Liu Q, Wang J, Tang J, Cai S, Zhao Q and Ming Z (2022) Metaomics in Clinical Laboratory: Potential Driving Force for Innovative Disease Diagnosis. Front. Microbiol. 13:883734. doi: 10.3389/fmicb.2022.883734

Currently, more and more studies suggested that reductionism was lack of holistic and integrative view of biological processes, leading to limited understanding of complex systems like microbiota and the associated diseases. In fact, microbes are rarely present in individuals but normally live in complex multispecies communities. With the recent development of a variety of metaomics techniques, microbes could be dissected dynamically in both temporal and spatial scales. Therefore, in-depth understanding of human microbiome from different aspects such as genomes, transcriptomes, proteomes, and metabolomes could provide novel insights into their functional roles, which also holds the potential in making them diagnostic biomarkers in many human diseases, though there is still a huge gap to fill for the purpose. In this mini-review, we went through the frontlines of the metaomics techniques and explored their potential applications in clinical diagnoses of human diseases, e.g., infectious diseases, through which we concluded that novel diagnostic methods based on human microbiomes shall be achieved in the near future, while the limitations of these techniques such as standard procedures and computational challenges for rapid and accurate analysis of metaomics data in clinical settings were also examined.

Keywords: microbiology, microbiome, omics, biomarker, diseases, rapid diagnosis

INTRODUCTION

In recent years, human microbiome studies revealed that dysbiosis of microbial communities could lead to dysfunction of host machineries, causing a broad spectrum of diseases (Wang et al., 2017; Kho and Lal, 2018). Thus, understanding the associations of particular bacterial species with diseases could hold the potential of providing new treatment targets and therapeutic approaches in clinical settings (Almeida et al., 2019). Until the last two decades, conventional methods such as bacterial culture and biochemical tests were normally considered as gold standards of bacterial diagnosis and widely employed in clinical laboratory (Wang et al., 2021). Driven by the technological developments and economic benefits, molecular methods such as PCR and immunoassay are gradually becoming available and popular for bacterial diagnosis.

However, both conventional microbiology and novel molecular techniques only satisfy with the simplicity and controllability of the reductionism framework by focusing on limited number of genes and bacterial species. Although the reductionist approach could reveal the individual genetics and physiology, contributing to the understanding of complex microbial behaviors in nature (Tecon et al., 2019), these observations and conclusions are difficult to be directly applied to the physiology of whole ecological systems like human-microbiota interactions (Fang et al., 2011). Around a decade ago, microbiome was merely a word that was mainly heard of by fellow scientists and the public was rarely familiar with the concept. With the recent rise of microbiome research, more and more studies acknowledge that microbes work together as a community to achieve key functions related with various aspects of human health ranging from metabolic disease to gastrointestinal disorders to emotional disturbance. A variety of techniques have been developed so far to dissect the human microbial communities in common niches such as mouth, gut, vagina, etc. both spatially and temporally, which include metagenomics, metatranscriptomics, metaproteomics, and metabolomics (**Figure 1**). These techniques are also known as metaomics when combined for integrated analysis. In addition, both the public and popular press show more and more interests in this novel field (Marcon et al., 2021), which lies the ground for metaomics to be developed and accepted as innovative bacterial diagnostics tools.

Metaomics is an innovative integration approach that is based on the in-depth analysis of human microbiomes, which has spurred a paradigm shift in understanding human health and detecting infectious diseases (Xu and Yang, 2021). Apparent advantages have been reported that makes these techniques with promising potentials in clinical diagnosis of bacterial infections, such as quantification of bacterial compositions, detection of unculturable bacterial pathogens, profiling of bacterial antibiotics-resistant genes, identification of virulence factors in large scale, and establishment of associations between bacteria and diseases, etc., all of which could be realized through metagenomic analysis (Wang et al., 2022). In addition, the dynamics of microbe–microbe interplays, host–microbe interactions, energy metabolism, and chemical cycling during bacterial infection could be elucidated through metatranscriptomic studies, which could not only improve the understanding of bacterial pathogenicity, but also facilitate biomarker discovery and development of microbial therapeutics (Zhang et al., 2021). Moreover, metatranscriptomics is also able to identify active bacteria and temporal variability of bacterial gene expressions during infection. Metaproteomics focuses on the dynamic changes of whole proteins in specified microbial communities, which could not only obtain functional information of bacterial communities, but be also able to link genes (proteins) with underlying phenotypes, which could also contribute to the development of biomarkers for therapies and diagnosis. As for metabolomics, it is a community overview of individual microbial metabolism, which focuses on global profiles of metabolites (small molecules), aiming to reveal biomarkers for bacterial infection diagnosis and also unravel metabolites concerning human health. Due to the complex interplays between metabolites during microbe-microbe and microbe-host interactions, metabolic networks based on constraint-based reconstruction and analysis (COBRA) and genomic-scale metabolic models (GEMs) are frequently constructed so as to understand microbiome–metabolome links and facilitate the translation of the findings into effective and novel therapeutics (Heinken et al., 2021; Jansma and El Aidy, 2021).

Although metaomics studies are increasingly wide-spreading and are presumed to be novel diagnostic tools in clinical laboratory in future, these techniques are mainly confined to the research field at current stage due to the disadvantages that are so far hard to overcome, such as high costs for experimental procedures and lack of gold standard for sample collection and data analysis, etc. (Chiu and Miller, 2019; Shakya et al., 2019; Dias et al., 2020). In addition, for metabolomics, techniques with acceptable sensitivity are only just being developed, while computational analysis and integration of metaomics data are other challenges that hinder the potential application of metaomics techniques in clinical settings, though data management and comparative analysis system are actively explored at current stage (Chen et al., 2019; Wang et al., 2022). In this mini-review, we will not look into the technical details of metaomics approaches; in contrast, we endeavor to focus on the application potentials of metaomics techniques for their rapid and accurate diagnosis of bacterial pathogens and infections. However, it should be noted that, in most studies, the presence of certain species, the altered levels of microbes, and the changed abundances of microbial transcripts, proteins or metabolites, have not been proven as causes for diseases but only associations. Therefore, the perspectives for advancing the functional and translational microbiome research in clinical settings, which may also facilitate the implementation of metaomics-based precision medicines, will be discussed in general manner in this mini-review.

POTENTIAL APPLICATIONS OF METAOMICS IN CLINICAL DIAGNOSIS

Metagenomics

In clinical laboratories, many pathogens are unamenable to be cultured or sometimes exist in a viable but non-culturable (VBNC) state, which makes them very difficult to be detected *via* conventional microbiological approaches such as microscopy and biological tests, leading to a great risk to public health (Li et al., 2014). Metagenomic next generation sequencing (mNGS) is the analysis of a collection of genomes from a mixed community of microbial organisms, which can rapidly quantify the organism diversity and microbial composition of a specific microbiota in a timely manner, showing attractive features for clinical diagnosis. It should be emphasized that mNGS (whole shotgun metagenomic sequencing) is not the same as 16S rRNA gene sequencing (16S sequencing) because the single amplicon sequencing cannot be analyzed together with other omics datasets. In particular, 16S sequencing only amplifies portions of the hypervariable regions (V1–V9) of the bacterial 16S rRNA gene, which could lead to potential

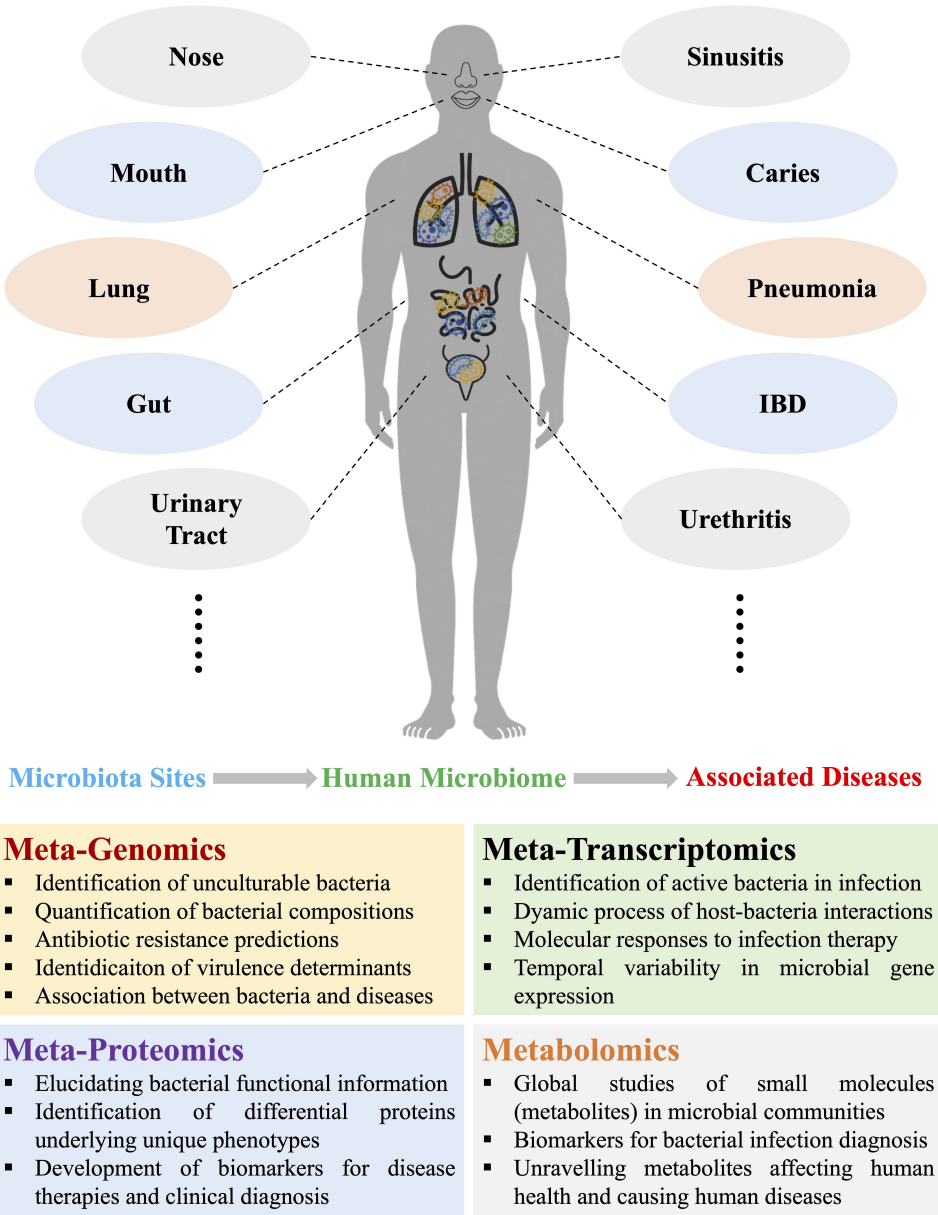


FIGURE 1 | Schematic illustration of the four omics approaches used in current and potential studies of human microbiomes and the associated diseases due to microbiota dysbiosis, which mainly involves metagenomics, metatranscriptomics, metaproteomics, and metabolomics. Representative functions of each of the four metaomics techniques were also listed. IBD, inflammatory bowel disease.

biases in the representation of the taxonomic units due to the choice of primers (Laudadio et al., 2018). In addition, studies also showed that 16S sequencing only detects a part of gut microbiota community revealed by mNGS (Durazzi et al., 2021). In contrast, the mNGS approach sequences all the DNA materials (viruses, bacteria, fungi, and micro-eukaryotes) in the microbiome rather than just bacteria as found through 16S sequencing, which generated more sequenced reads per sample, hence, higher resolutions in taxonomic assignments at species level and also higher sequencing costs (Peterson et al., 2021).

In a systematic review, Quince et al. (2017) summarized in details of the metagenomic analysis procedures from sample preparations to computational pipelines, which offers a biotechnological promise in therapeutic discovery of human health. In a recent perspective, Segata (2018) emphasized the importance of accurately elucidating human-associated microbial communities at strain level through developing new computational tools, which can link strain variants to host phenotypes and holds the potential of understanding the personalized host–microbiome interactions. In fact, with the fast development of metagenomic techniques, this

culture-independent approach has been applied in detecting microbial pathogens in public health (Miller et al., 2013; Chiu and Miller, 2019), identifying genes or genetic mutations conferring resistance to antimicrobial drugs (De, 2019; De Abreu et al., 2021), and enabling genotyping analysis for molecular epidemiology and so on (Robert and Filkins, 2019), which makes the method gradually transiting from research fields to clinical laboratories (Chiu and Miller, 2019) and slowly integrating into clinician's toolbox to identify infectious diseases (De Vries, 2021), though it functions as a diagnostic tool yet to be widely established due to a variety of issues such as costs, turnaround time, sensitivity, specificity, validation, and reproducibility, etc. in clinical microbiology laboratories.

In specificity, the clinical applications of mNGS involves dissecting healthy microbial compositions in various body parts such as mouth, respiratory tract, gut, central nervous system (CNS), urinary tract, vagina, etc. (Gu et al., 2019) and revealing the aberrant bacterial compositions in various clinical samples such as saliva, bronchoalveolar lavage fluid, cerebrospinal fluid (CSF), urine, vaginal secretion, and other body fluids or infected tissues (Chiu and Miller, 2019), through which abnormal bacterial genera and species could be identified and might be used to serve for potential clinical diagnosis of human infectious diseases such as periodontitis (Curtis et al., 2020), pneumoniae (Thibeault et al., 2021), meningitis (Moir, 2015), urethritis (Srinivasan et al., 2021), vaginosis (Onderdonk et al., 2016), etc. In addition, non-infectious human diseases were also reported to be associated with microbiota dysbiosis. For example, it was identified that during diabetes and cardiovascular disease (CVD), the microbial diversity of blood microbiota is vastly transformed, in which the two bacterial genera *Staphylococcus* and *Klebsiella* were predominant in the blood of patients with type 2 diabetes mellitus (T2DM), while high *Actinobacteria/Proteobacteria* ratio was consistently associated with CVD (Velmurugan et al., 2020). Therefore, these alterations in bacterial compositions hold the promise to be translated into potential indictors for the clinical diagnosis of the two diseases. As for the CNS, it was suggested that no detectable microbial community existed in healthy CSF because blood–brain barrier (BBB) is able to protect against microbial invasions, though such a claim is still controversial due to the difficulties in the identification of contamination (Kang et al., 2021). Recently, under pathological conditions, studies revealed that bacterial pathogen *Porphyromonas gingivalis* was found in the brains of Alzheimer's disease (AD) patients and CSF of patients with probable AD (Dominy et al., 2019); however, the presence of *P. gingivalis* DNA in CSF serving as a diagnostic marker for AD promising but debatable, which required further explorations. In addition, there are many other cases involving metagenomic analysis confirmed the application potential of the metaomics techniques in clinical diagnosis due to the associations between human diseases and microbiota dysbiosis. In addition, microbiome research also holds the potential to identify microbial species that are causally associated with cancer phenotypes and unravel the underlying mechanisms behind these associations, which could facilitate cancer diagnosis and transform the treatment strategies for patients with cancer (Banerjee et al., 2015). For a brief summary of the representative studies on the associations between diseases and aberrant microbiota, please refer to **Table 1**. Taken together, metagenomics can serve as a potential driving force for clinical diagnosis of microbial infections and microbiota-dysbiosis-related diseases with personalized patient cares in future, though there is still a huge gap to fill between basic researches and clinical translations. Therefore, different from microbial culture and biochemical testing, there is still a long way for mNGS to go before the technique could become a vital tool in any clinical testing algorithms.

Metatranscriptomics

Different from metagenomic analysis that focused on the study of taxonomical profiles and microbial compositions in human samples, metatranscriptomics aims to elucidate the functional profiles of metagenomes that inform of the genes that are expressed by the community as a whole under specific conditions, leading to the dynamic understanding of functional ecology of human microbiome (Franzosa et al., 2014; Aguiar-Pulido et al., 2016; Shakya et al., 2019). In addition, during certain circumstances, no linkages between microbiome and diseases could be found at metagenomic level, while correlations at metatranscriptome level could be established. For example, Feng et al. (2019) recruited both metagenomic and metatranscriptomic analyses to dissect the human prostate microbiota from patients with prostate cancer, through which the study revealed that the bacterial composition was not significantly changed between tumor and adjacent benign tissue while gene expression profiles of *Pseudomonas* may be related with metastasis. In fact, with the emergence of the novel notion that microbial associations with certain diseases like oral cancer are actually at functional level of microbial communities rather than at microbial compositional level (Banavar et al., 2021), more and more studies implemented metatranscriptomics or combined metagenomics with metatranscriptomics to determine gene expressions and regulations when the microbiota responded to certain conditions or in certain abnormal states in order to gain comprehensive and functional understandings of human microbiomes (Shakya et al., 2019). Interestingly, metatranscriptomic profiles were more individualized than metagenomic profiles, which had less variable when compared with microbial compositions (Franzosa et al., 2014; Abu-Ali et al., 2018). Currently, many studies have taken the advantages of metatranscriptomics and aimed to elucidate the dynamic gene expressions in the study of human microbiota. For example, Banavar et al. (2021) used both metagenomics and metatranscriptomics to characterize salivary microbiota, which discovered relative abundance of specific bacterial species and gene expressions associated with periodontitis and dental caries. Thus, theses bacterial species and active genes might be possible for evaluating saliva for potential periodontitis and dental caries at pre-clinical stages. Another example using metatranscriptome to study lung cancer patients found that the active presence of two bacteria, *Bacillus megaterium* and *Mycobacterium franklinii*, might play an important role in the occurrence of lung cancer tumors (Chang et al., 2021), which confirmed the potential of metatranscriptomics in identifying the dynamic interactions between microbes and human host

TABLE 1 | Comparison of healthy and disturbed microbiota that might contribute to the understanding of certain diseases from microbial perspectives.

Organ, tissues, fluids	Healthy microbiota (predominant bacterial genera)	Disturbed microbiota (abundant bacterial genera/species)	Representative human diseases associated with disturbed microbiota	References
Blood	*Achromobacter, Pseudomonas, Serratia, Sphingomonas, Staphylococcus, Corynebacterium, Acinetobacter*	*Staphylococcus, Klebsiella*	Type 2 diabetes mellitus (T2DM)	Velmurugan et al., 2020
		High *Actinobacteria/ Proteobacteria* ratio	Cardiovascular disease (CVD)	
Central nervous system (cerebrospinal fluid)	No detectable microbial community	*Porphyromonas gingivalis*	Alzheimer's disease (AD)	Roos, 2015; Velmurugan et al., 2020; Kang et al., 2021
		Streptococcus pneumoniae, Neisseria meningitidis	Meningitis	
		Staphylococcus aureus	Spinal epidural abscess	
Gut (feces)	*Ruminococcus, Clostridium, Lactobacillus, Enterococcus, Bacteroides, Prevotella, Bifidobacterium, Escherichia, Akkermansia*	*Enterobacteriaceae*	Inflammatory bowel disease (IBD)	Durack and Lynch, 2019; Velmurugan et al., 2020
		Bacteroides spp.	Type 2 diabetes mellitus	
		Collinsella, Corynebacterium, Lactobacillus	Behavioral disorders	
		Faecalibacterium, Akkermansia, Lachnospira	Atopic asthma	
Lung (bronchoalveolar lavage fluid)	*Prevotella, Streptococcus, Veillonella, Neisseria, Haemophilus, Fusobacterium*	*Staphylococcus, Haemophilus*	Asthma	Faner et al., 2017; Mathieu et al., 2018
		Staphylococcus aureus, Burkholderia cepacia	Cystic fibrosis	
Milk	*Staphylococcus, Streptococcus, Corynebacterium, Cutibacterium, Lactobacillus, Lactococcus, Bifidobacterium*	*Lactobacillus iners, Neisseria subflava, Streptococcus lactarius, Streptococcus cristatus, Staphylococcus aureus*	Sub-acute lactational mastitis	Fernández et al., 2020
Mouth (saliva)	*Streptococcus, Veillonella, Granulicatella, Gemella, Actinomyces, Corynebacterium, Rothia, Fusobacterium, Porphyromonas, Prevotella, Capnocytophaga, Neisseria, Haemophilus, Treponema, Lactobacterium, Eikenella, Leptotrichia, Peptostreptococcus, Staphylococcus, Eubacteria, Propionibacterium*	*Prevotella, Fusobacterium*	Periodontitis	Zarco et al., 2012; Willis and Gabaldón, 2020
		Neisseria, Selenomonas, Propionibacterium	Dental caries	
		Veillonella, Atopobium, Prevotella, Leptotrichia	Rheumatoid arthritis	
Stomach (gastric juice)	*Streptococcus, Prevotella*	*Firmicutes, Fusobacteria*	Gastroesophageal reflux disease (due to the use of proton pump inhibitor)	Ohno and Satoh-Takayama, 2020
Urinary tract (urine)	*Prevotella, Escherichia, Enterococcus, Streptococcus, Citrobacter*	*Herbaspirillum, Porphyrobacter, Bacteroides*	Urothelial carcinoma	Perez-Carrasco et al., 2021
Vagina (vaginal secretion)	*Lactobacillus* spp., *Actinobacteria, Prevotella, Veillonellaceae, Streptococcus, Proteobacteria, Bifidobacteriaceae, Bacteroides, Burkholderiales*	*Gardnerella, Prevotella, Atopobium, Mobiluncus, Bifidobacterium, Sneathia, Leptotrichia*	Bacterial vaginosis	Chen et al., 2021

It should be emphasized that the presence of certain species has not been proven as causes for diseases but only associations in most studies. Therefore, we only discuss the possibilities for mNGS method in clinical diagnosis of human diseases through the composition of bacteria in disturbed microbiota, rather than confirming the real applications of the mNGS methods in clinical settings.

in terms of disease progression and severity. It was also recently reported that metatranscriptomics was able to assess the clearance of burn wound infection through differentiating between live and dead organisms and understanding rapid microbial alterations in complex host-microbe samples (Ojala et al., 2021). A variety of other human diseases were also investigated through metatranscriptomics like bacterial vaginosis (Ravel et al., 2013), which could facilitate the identification of the most metabolically active species present in the patients with particular diseases. Therefore, metatranscriptomics is an integral part of the metaomics toward a system level understanding the dynamics of human microbiome in responses to diverse factors.

Metaproteomics

All the proteins in a microbial community are termed as a metaproteome while the study of taxonomic and functional composition of a microbiota through overall identification of proteins using mass spectrometry is terms as metaproteomics, which is a crucial approach to understand microbial functions in communities (Heyer et al., 2019). Due to its direct insights into microbial phenotypes on large-scale molecular levels, metaproteomics is also a promising tool for clinical diagnostics of human diseases. For example, Long et al. investigated the pathogenesis of colorectal cancer (CRC) through the quantitative comparisons of microbial protein abundances between the CRC patients and the healthy volunteers, which identified 341 significantly different proteins that may serve as biomarkers for distinguishing pathological states and showed that metaproteomics had great value for guiding clinical diagnosis in the future (Long et al., 2020). In addition, a recent in-depth investigation studied the functional compositions of gut microbiota and proteins in a set of fecal samples (treatment-naïve type 2 diabetic, $n = 77$; pre-diabetic, $n = 80$; and normal glucose tolerant, $n = 97$); through a combination of metagenomics and metaproteomics, distinct gut metagenomics and metaproteomics signatures in prediabetics and treatment-naïve type 2 diabetics were discovered, leading to the potential translation of microbiota features into clinical diagnosis biomarkers (Zhong et al., 2019). Previously, Lassek et al. (2015) also used metaproteomics approach to explore the interactions between host and pathogens during catheter-associated urinary tract infections (UTI), which revealed that the asymptomatic phase of catheter-associated UTI could be due to the well-maintained balance of protein levels between bacterial virulence factors and human immune system. Therefore, metaproteomics is also able to elucidate the potential molecular mechanisms of clinical problems. However, so far, the clinical application of quantitative metaproteomics is still in its infancy because of methodological limitations in sample preparations and computational analyses, etc. For example, the great heterogenicity of microbial proteins in any clinical sample significantly hinders the analysis and interpretation of the metaproteome result; in addition, it is also computationally challenging to integrate metaproteomic data with clinical data sets in order to gain clinically meaningful explanations (Blackburn and Martens, 2016). To sum up, further technical developments and innovations are required to facilitate the progress of this promising field.

Metabolomics

Metabolomics is an analytical technique for the study of metabolic networks by examining the overall changes of small metabolites in biological systems (Wang et al., 2019). As for metabolomics in the study of microbiota, it is a recently emerged application for determining all the metabolites released by microbiomes. Thus, it is a community-based version of single microbial metabolomics in a particular physiological state, which is also known as community metabolomics or environmental metabolomics. In clinical settings, metabolomics could solve the questions like what metabolites are produced under different conditions by the microbiome. In addition, metabolites released by microbial communities normally have responsibilities for the human health that they inhabit, which makes them eligible to serve as biomarkers for clinical diagnosis. In fact, the molecular mechanisms behind how human microbiomes in different body parts correlated with the dynamic alterations of metabolites and causing diseases are starting to be elucidated, which could contribute to the development of preventive and treatment strategies for complex human diseases (Lee-Sarwar et al., 2020). For example, Jansson et al. recruited Ion Cyclotron Resonance Fourier Transform Mass Spectrometry (ICR-FT/MS) to study the causes and etiology of Crohn's disease (CD) *via* fecal samples from 17 identical twins that were with and without CD, respectively (Jansson et al., 2009). According to the study, the non-targeted metabolic profiling revealed metabolic biomarkers of CD that might serve as diagnostic aims or monitoring tools for CD therapy and prevention (Jansson et al., 2009). In addition, a comprehensive study conducted by Walker et al. (2014) revealed obesity-related metabolite profiles in two different C57BL/6 mouse strains, C57J and C57N, which identified new factors that might be responsible for high-fat diet induced obesity, providing potential new strategy for obesity diagnosis and treatment. In addition, Han et al. (2021) recently developed a novel metabolomics pipeline, which provided a powerful tool for characterizing microorganisms and deciphering the interactions between microorganisms and their host in terms of small metabolites. Although metabolomics is powerful technique and is sensitive enough in profiling metabolites in batch, due to the complex interplays between metabolites, there are still many limitations for its robust applications in clinical settings, which should be addressed and solved during the continuing development and in-depth application of the technique.

Integration of Metaomics Techniques

Microbial community is a complex but integral part of our body (human ecosystem) that is tightly associated with our health and disease (Segata et al., 2013). In order to comprehensively and accurately understand the microbial communities and their interactions with the hosts, an integrated approach that combines multi-omics data is starting to be under active develop, rather than relying on any single omics method. However, it is inherently difficult to integrate multi-omics data, e.g., metagenomes, metatranscriptomes, metaproteomes,

and metabolomes, for systematic analysis (Knight et al., 2018) because these data are largely heterogeneous and are sourced from different time scales. Due to the importance of metaomics in comprehensive understanding of microbiomes, the studies and tools for the integration of different multi-omics data sets are becoming increasingly available, which greatly facilitates the development and translational potential of the metaomics approach in the field of human microbiota. So far, many pilot studies, preliminary analyses and comprehensive researches have innovatively explored the metaomics approach in the dissection of microbial communities and its interplays with hosts (Darzi et al., 2015; Aguiar-Pulido et al., 2016; Valles-Colomer et al., 2016; Boeri et al., 2022). For example, Valles-Colomer et al. (2016) systematically reviewed the application of metaomics in the complex and multifactorial disease IBD, which revealed that the approach held great promise in providing insights into IDB, though the interpretation of the metaomics data at multiple levels were very challenging. Boeri et al. (2022) summarized the current advantages of using metaomics approach to study microbiota–host interactions in the understanding of epilepsy with focuses on sample collection, extraction, and data processing, which could help in recognizing molecular pathways and biomarkers for microbiota–epilepsy connection, leading to development of novel clinical diagnostic methods. In addition, since the metaomics approach is data-intensive, many computational tools have also been developed and pipelines constructed for the comparative metaomics analysis so as to decipher the adaptations of microbial communities and microbiota–host interactions (Segata et al., 2013; Zhai et al., 2017; Sequeira et al., 2019). However, more computational tools are needed in this field in order to overcome the challenges of diversity and heterogeneity during the integration of the metaomics data. The phenotypes of complex microbial communities are constantly shaped by the dynamic interactions between hosts and their associated microbiota. In order to explore the full extent of microbial functions during the process, optimal, and efficient integration of multi-omics data derived from metagenomics, metatranscriptomics, metaproteomics, and metabolomics is essential, which significantly improves our knowledge of the human microbiome and its specific roles in the health and disease states of human beings. This is the reason why it is necessary to provide a timely and updated perspective overview of this exciting field. For an illustrative summary of the integration of the four metaomics approaches, refer to **Figure 2** below.

CHALLENGES OF METAOMICS APPROACHES

Microbiota has been extensively studied for the past two decades from environment like water and soil to human body sites such as gut and skin, etc. In clinical settings, many diseases that are directly linked with microbial infections such as pneumonia, gastritis and vaginosis have been known to be caused by the disturbance of normal microflora. However, many well-known diseases that were previously unexpected to be microbe-relevant were also shown to have tight associations with the dysbiosis of human microbiomes such as mental disorders, CVD and cancer, etc. (Elinav et al., 2019; Velmurugan et al., 2020; Xu and Yang, 2021). Therefore, thorough understanding of the dynamic changes of human microbiome at both pathological and healthy states will greatly facilitate the understanding of disease mechanisms and promote the discovery of novel biomarkers at different levels (DNA, RNA, protein, metabolite, and species) *via* metaomics approaches, which could significantly improve the diagnostic efficiency and accuracy of multiple diseases in clinical laboratories. However, both the standalone omics techniques and combined metaomics approaches still face many challenges for their routine uses and real-world applications.

Drawbacks of Standalone Omics Methods

The routine deployment of mNGS in clinical settings involves sample collection, nucleic acid extraction, library preparation, sequencing, computational analysis, and clinical interpretation of the data (Chiu and Miller, 2019). During the implementation of the mNGS pipeline, multiple factors should be considered for increasing the accuracy of the clinical diagnosis, such as sample stability during collection and transportation, diagnostic cost, turnaround time, computational complexity of datasets, and patient privacy, etc. (Chiu and Miller, 2019). In addition, sequencing and data extraction biases should also be considered since next generation sequencing is very well known to be biases toward certain GC range (Browne et al., 2020), which could be solved with methodological optimizations. During clinical diagnosis, the final and desired result is unbiased detection and reporting of all pathogens in a clinical sample, which involves targeted sequence capture, specialized computational tools, and explicative result reports and so on (Dekker and Dulanto Chiang, 2020). As for the metatranscriptomics, although it is complementary to metagenomics through dynamic characterization of microbiomes, some important restrictions should be pointed out in order to enhance the reproducibility and applicability of the approach, which may enable the integration of metatranscriptomic data into clinical settings (Bashiardes et al., 2016). Among these technical challenges, potential host RNA contamination and the short half-life of mRNA in the sample have been proven to be problematic (Bashiardes et al., 2016), which should be carefully handled during sample collection and RNA extraction. In fact, the procedures of RNA isolation, processing, sequencing, and analysis should be standardized so as to integrate the data into microbiome research. In addition, the metatranscriptomic data involves large-scale expression of genes, the discovery of which should be also validated *via* conventional diagnostic methods such as quantitative polymerase chain reaction (qPCR).

However, it is well known that the presence of DNA (metagenome) and mRNA (metatranscriptome) does not guarantee the presence of proteins and protein activities, not even mentioning the bioactive metabolites. In fact, different from metagenomics and metatranscriptomics, both metaproteomics and metabolomics are considered as functional tools to characterize microbial activities involving

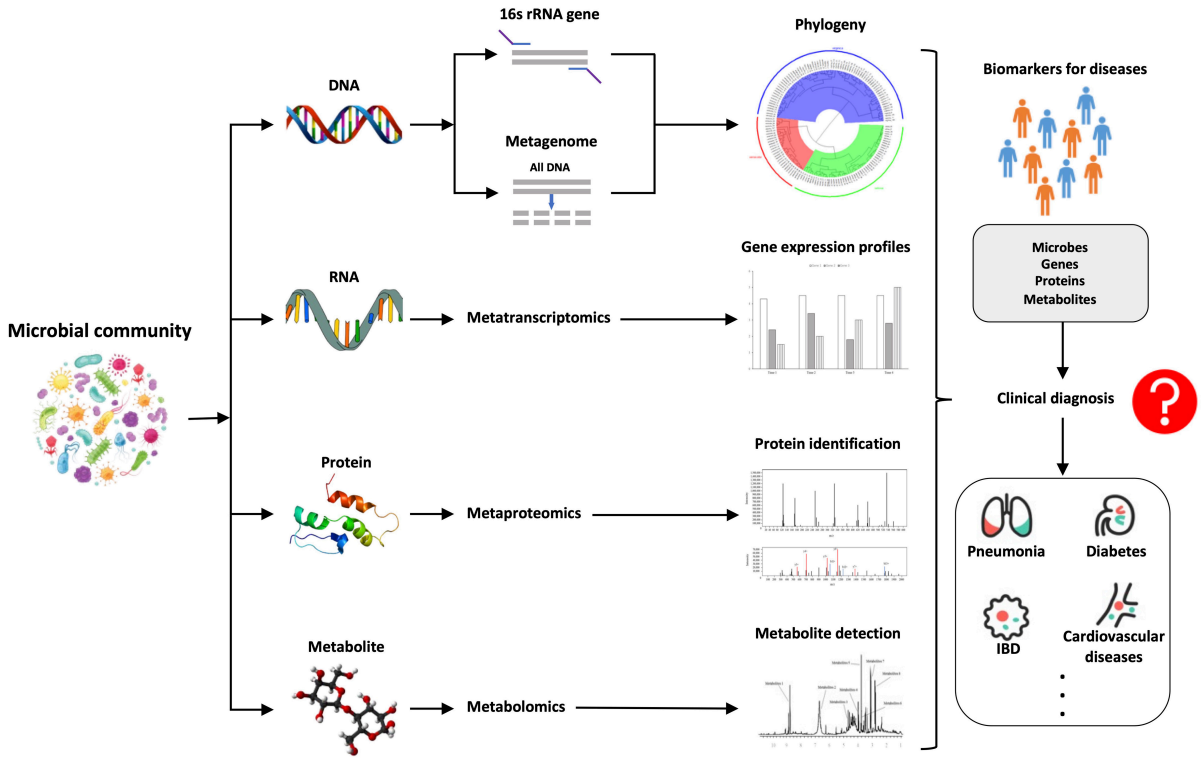


FIGURE 2 | A brief summary of the comparative illustration of the integration of the four metaomics approaches, that is, metagenomics (DNA), metatranscriptomics (RNA), metaproteomics (proteins), and metabolomics (metabolites), through which novel biomarkers such as microbes, genes, proteins, and metabolites could be identified, which might have the potential to be used for the rapid and accurate diagnosis of human diseases caused by microbiota dysbiosis.

healthy and pathological states in human beings (Zhu et al., 2021). Therefore, metaomics pipeline integrating different omics approaches is necessary to generate a holistic view of clinical samples, which is also why metaproteomics and metabolomics are needed for sample analysis. Although sample preparation protocols for liquid chromatography-mass spectrometry (LC-MS) that were used for metaproteomics and metabolomics are becoming standardized, real-world analyses of these data are still facing many difficulties. For example, both metaproteomics and metabolomics experience lack of standardized protocols for sample preparations, inaccuracy of MS to measure low-concentration molecules, high costs of data generation and sophisticated downstream data analyses (Nyholm et al., 2020). In addition, the LC-MS approach is also limited to both insufficiency of reference database and inadequacy of normalization procedures (Ejigu et al., 2013; Vinaixa et al., 2016). Therefore, during the analysis of human microbiomes, the procedures should be scrutinized for best practice and the results should be carefully interpreted for accuracy.

Limitations of Integrated Metaomics

It is true that metaomics has many advantages for microbial studies and holds the promise to revolutionize clinical diagnosis in foreseeable future. In fact, some of metaomics approaches have already been implemented in clinical diagnosis for certain circumstances such as precision medicine for drug-resistant tuberculosis (Leong et al., 2018) and identification of bacterial pathogens directly from clinical urine samples (Schmidt et al., 2017), etc. In addition, metaomics approach has also been applied to study complex disease such as epilepsy and IBD in order to understand the functions of microbiota in these diseases (Valles-Colomer et al., 2016; Boeri et al., 2022). Since metaomics is intrinsically a data-intensive field, well-trained personnel should also be a part of clinical diagnosis team during the coming metaomics era. Previously, restrictions to the use of metaomics such as low standardization of sample preparations and high costs of experiments have gradually been overcome, though there is still a gap that needs to be filled before the approach could be applied in real-world clinical settings. In addition to all the experimental procedures, novel and efficient computational tools are also essential for the application of metaomics, especially for data heterogenicity between and data integration across metagenomic, metatranscriptomic, metaproteomic, and metabolomic data sets, not even mentioning other more specialized omics techniques such as glycomics and lipidomics, etc. (Blackburn and Martens, 2016; Wang, 2022). Moreover, the development of pipelines to integrate standalone omics data, together with the equipment of sufficient computational storage space are also necessary for fast and efficient analysis during integration of metaomics data (Segata et al., 2013). Considering the complexity of the metaomics dataset, machine learning algorithms also provided a promising strategy to explore the microbiota-host interactions (Yuan et al., 2021). Taken together, in order to achieve a

holistic analysis of microbiome and facilitate its diagnostic application in clinical settings, both experimental procedures and computational approaches should be enhanced and integrated to form a network-based approach in order to find true and reliable biomarkers for human diseases during clinical diagnosis.

CONCLUSION AND PERSPECTIVES

Although each omics approach provides valuable information separately for human microbiome analysis, it has been shown by various studies that these techniques could generate a more comprehensive picture for clinical diagnosis of diseases when combined together as metaomics. In fact, with the advancement of metaomics techniques in microbiome studies, many limitations for conventional clinical diagnosis could be overcome such as rapid recognition of unculturable pathogens, profiling of antibiotic resistance, causing pathogens of diseases, and harmful bioactive molecules, etc., which will greatly facilitate the efficient treatment and rapid management of microbial infections. In addition, supported by cumulative evidence of metaomics studies, it is gradually revealed that microbiota is indispensably involved in the basic biological activities of human beings through host-microbe interactions and the modulation of important human metabolic processes, while many studies have established the associations between human microbiomes and a variety of diseases such as obesity, diabetes mellitus, CVD and cancer, etc., though causative relationships between these associations still need further in-depth explorations. However, novel biomarkers from microbial perspectives, e.g., microbial compositions, gene levels, protein types, and metabolite concentrations, are still promising and hold the application potential in clinical settings. In this mini-review, we went through recent applications of standalone omics techniques and integrated metaomics in clinical setting, together with their current challenges, which reinforced the future of these novel methods in rapid and accurate disease diagnosis of human diseases.

AUTHOR CONTRIBUTIONS

LW, QZ, and ZM conceived and designed the framework of the study and provided the platform and resources. LW and QZ were responsible for project administration. LW, FL, BG, PQ, QL, JW, and SC carried out the literature review and performed the data analysis and visualization. All authors wrote and approved the final manuscript.

FUNDING

LW appreciated the financial support from National Natural Science Foundation of China (Nos. 31900022 and 32171281), Young Science and Technology Innovation Team of Xuzhou Medical University (No. TD202001), and Jiangsu Qinglan Project (2020). QZ appreciated the financial support from Foundation of Education Department of Liaoning Province (Grant No. LJKZ0280).

ACKNOWLEDGMENTS

We appreciated the reviewers for their genuine and constructive comments and suggestions.

REFERENCES

Abu-Ali, G. S., Mehta, R. S., Lloyd-Price, J., Mallick, H., Branck, T., Ivey, K. L., et al. (2018). Metatranscriptome of human faecal microbial communities in a cohort of adult men. *Nat. Microbiol.* 3, 356–366. doi: 10.1038/s41564-017-0084-4

Aguiar-Pulido, V., Huang, W., Suarez-Ulloa, V., Cickovski, T., Mathee, K., and Narasimhan, G. (2016). Metagenomics, metatranscriptomics, and metabolomics approaches for microbiome analysis. *Evol. Bioinform.* 12(Suppl. 1), 5–16. doi: 10.4137/EBO.S36436

Almeida, A., Mitchell, A. L., Boland, M., Forster, S. C., Gloor, G. B., Tarkowska, A., et al. (2019). A new genomic blueprint of the human gut microbiota. *Nature* 568, 499–504. doi: 10.1038/s41586-019-0965-1

Banavar, G., Ogundijo, O., Toma, R., Rajagopal, S., Lim, Y. K., Tang, K., et al. (2021). The salivary metatranscriptome as an accurate diagnostic indicator of oral cancer. *NPJ Genom. Med.* 6:105. doi: 10.1038/s41525-021-00257-x

Banerjee, J., Mishra, N., and Dhas, Y. (2015). Metagenomics: a new horizon in cancer research. *Meta Gene* 5, 84–89. doi: 10.1016/j.mgene.2015.05.005

Bashiardes, S., Zilberman-Schapira, G., and Elinav, E. (2016). Use of metatranscriptomics in microbiome research. *Bioinform. Biol. Insights* 10, 19–25. doi: 10.4137/BBI.S34610

Blackburn, J. M., and Martens, L. (2016). The challenge of metaproteomic analysis in human samples. *Expert Rev. Proteomics* 13, 135–138. doi: 10.1586/14789450.2016.1135058

Boeri, L., Donnaloja, F., Campanile, M., Sardelli, L., Tunesi, M., Fusco, F., et al. (2022). Using integrated meta-omics to appreciate the role of the gut microbiota in epilepsy. *Neurobiol. Dis.* 164:105614. doi: 10.1016/j.nbd.2022.105614

Browne, P. D., Nielsen, T. K., Kot, W., Aggerholm, A., Gilbert, M. T. P., Puetz, L., et al. (2020). GC bias affects genomic and metagenomic reconstructions, underrepresenting GC-poor organisms. *GigaScience* 9:giaa008. doi: 10.1093/gigascience/giaa008

Chang, Y.-S., Hsu, M.-H., Tu, S.-J., Yen, J.-C., Lee, Y.-T., Fang, H.-Y., et al. (2021). Metatranscriptomic analysis of human lung metagenomes from patients with lung cancer. *Genes* 12:1458. doi: 10.3390/genes12091458

Chen, I. M. A., Chu, K., Palaniappan, K., Pillay, M., Ratner, A., Huang, J., et al. (2019). IMG/M v.5.0: an integrated data management and comparative analysis system for microbial genomes and microbiomes. *Nucleic Acids Res.* 47, D666–D677. doi: 10.1093/nar/gky901

Chen, X., Lu, Y., Chen, T., and Li, R. (2021). The female vaginal microbiome in health and bacterial vaginosis. *Front. Cell. Infect. Microbiol.* 11:631972. doi: 10.3389/fcimb.2021.631972

Chiu, C. Y., and Miller, S. A. (2019). Clinical metagenomics. *Nat. Rev. Genet.* 20, 341–355.

Curtis, M. A., Diaz, P. I., Van Dyke, T. E., and Mariano, R. J. (2020). The role of the microbiota in periodontal disease. *Periodontology* 83, 14–25.

Darzi, Y., Falony, G., Vieira-Silva, S., and Raes, J. (2015). Towards biome-specific analysis of meta-omics data. *ISME J.* 10, 1025–1028. doi: 10.1038/ismej.2015.188

De, R. (2019). Metagenomics: aid to combat antimicrobial resistance in diarrhea. *Gut Pathog.* 11:47. doi: 10.1186/s13099-019-0331-8

De Abreu, V. A. C., Perdigão, J., and Almeida, S. (2021). Metagenomic approaches to analyze antimicrobial resistance: an overview. *Front. Genet.* 11:575592. doi: 10.3389/fgene.2020.575592

De Vries, J. J. C. (2021). The multidimensional nature of metagenomics drives interdisciplinary diagnostics. *EBioMedicine* 74:103694. doi: 10.1016/j.ebiom.2021.103694

Dekker, J. P., and Dulanto Chiang, A. (2020). From the pipeline to the bedside: advances and challenges in clinical metagenomics. *J. Infect. Dis.* 221(Suppl. 3), S331–S340. doi: 10.1093/infdis/jiz151

Dias, C. K., Starke, R., Pylro, V. S., and Morais, D. K. (2020). Database limitations for studying the human gut microbiome. *PeerJ Comput. Sci.* 6:e289. doi: 10.7717/peerj-cs.289

Dominy, S. S., Lynch, C., Ermini, F., Benedyk, M., Marczyk, A., Konradi, A., et al. (2019). Porphyromonas gingivalis in Alzheimer's disease brains: evidence for disease causation and treatment with small-molecule inhibitors. *Sci. Adv.* 5:eaau3333. doi: 10.1126/sciadv.aau3333

Durack, J., and Lynch, S. V. (2019). The gut microbiome: relationships with disease and opportunities for therapy. *J. Exp. Med.* 216, 20–40. doi: 10.1084/jem.20180448

Durazzi, F., Sala, C., Castellani, G., Manfreda, G., Remondini, D., and De Cesare, A. (2021). Comparison between 16S rRNA and shotgun sequencing data for the taxonomic characterization of the gut microbiota. *Sci. Rep.* 11:3030. doi: 10.1038/s41598-021-82726-y

Ejigu, B. A., Valkenborg, D., Baggerman, G., Vanaerschot, M., Witters, E., Dujardin, J.-C., et al. (2013). Evaluation of normalization methods to pave the way towards large-scale LC-MS-based metabolomics profiling experiments. *OMICS* 17, 473–485. doi: 10.1089/omi.2013.0010

Elinav, E., Garrett, W. S., Trinchieri, G., and Wargo, J. (2019). The cancer microbiome. *Nat. Rev. Cancer* 19, 371–376.

Faner, R., Sibila, O., Agustí, A., Bernasconi, E., Chalmers, J. D., Huffnagle, G. B., et al. (2017). The microbiome in respiratory medicine: current challenges and future perspectives. *Eur. Respir. J.* 49:1602086. doi: 10.1183/13993003.02086-2016

Fang, F. C., Casadevall, A., and Payne, S. M. (2011). Reductionistic and holistic science. *Infect. Immun.* 79, 1401–1404. doi: 10.1128/IAI.01343-10

Feng, Y., Ramnarine, V. R., Bell, R., Volik, S., Davicioni, E., Hayes, V. M., et al. (2019). Metagenomic and metatranscriptomic analysis of human prostate microbiota from patients with prostate cancer. *BMC Genomics* 20:146. doi: 10.1186/s12864-019-5457-z

Fernández, L., Pannaraj, P. S., Rautava, S., and Rodríguez, J. M. (2020). The microbiota of the human mammary ecosystem. *Front. Cell. Infect. Microbiol.* 10:586667. doi: 10.3389/fcimb.2020.586667

Franzosa, E. A., Morgan, X. C., Segata, N., Waldron, L., Reyes, J., Earl, A. M., et al. (2014). Relating the metatranscriptome and metagenome of the human gut. *Proc. Natl. Acad. Sci. U.S.A.* 111, E2329–E2338. doi: 10.1073/pnas.1319284111

Gu, W., Miller, S., and Chiu, C. Y. (2019). Clinical metagenomic next-generation sequencing for pathogen detection. *Annu. Rev. Pathol. Mech. Dis.* 14, 319–338.

Han, S., Van Treuren, W., Fischer, C. R., Merrill, B. D., Defelice, B. C., Sanchez, J. M., et al. (2021). A metabolomics pipeline for the mechanistic interrogation of the gut microbiome. *Nature* 595, 415–420. doi: 10.1038/s41586-021-03707-9

Heinken, A., Basile, A., Hertel, J., Thinnes, C., and Thiele, I. (2021). Genome-scale metabolic modeling of the human microbiome in the era of personalized medicine. *Annu. Rev. Microbiol.* 75, 199–222. doi: 10.1146/annurev-micro-060221-012134

Heyer, R., Schallert, K., Büdel, A., Zoun, R., Dorl, S., Behne, A., et al. (2019). A robust and universal metaproteomics workflow for research studies and routine diagnostics within 24 h using phenol extraction, FASP digest, and the metaproteomeanalyzer. *Front. Microbiol.* 10:1883. doi: 10.3389/fmicb.2019.01883

Jansma, J., and El Aidy, S. (2021). Understanding the host-microbe interactions using metabolic modeling. *Microbiome* 9:16. doi: 10.1186/s40168-020-00955-1

Jansson, J., Willing, B., Lucio, M., Fekete, A., Dicksved, J., et al. (2009). Metabolomics reveals metabolic biomarkers of crohn's disease. *PLoS One* 4:e6386. doi: 10.1371/journal.pone.0006386

Kang, Y., Ji, X., Guo, L., Xia, H., Yang, X., Xie, Z., et al. (2021). Cerebrospinal fluid from healthy pregnant women does not harbor a detectable microbial community. *Microbiol. Spectr.* 9:e0076921. doi: 10.1128/Spectrum.00769-21

Kho, Z. Y., and Lal, S. K. (2018). The human gut microbiome – a potential controller of wellness and disease. *Front. Microbiol.* 9:1835. doi: 10.3389/fmicb.2018.01835

Knight, R., Vrbanac, A., Taylor, B. C., Aksenov, A., Callewaert, C., Debelius, J., et al. (2018). Best practices for analysing microbiomes. *Nat. Rev. Microbiol.* 16, 410–422. doi: 10.1038/s41579-018-0029-9

Lassek, C., Burghartz, M., Chaves-Moreno, D., Otto, A., Hentschker, C., Fuchs, S., et al. (2015). A metaproteomics approach to elucidate host and pathogen protein expression during catheter-associated urinary tract infections (CAUTIs). *Mol. Cell. Proteom.* 14, 989–1008. doi: 10.1074/mcp.M114.043463

Laudadio, I., Fulci, V., Palone, F., Stronati, L., Cucchiara, S., and Carissimi, C. (2018). Quantitative assessment of shotgun metagenomics and 16S rDNA amplicon sequencing in the study of human gut microbiome. *OMICS* 22, 248–254. doi: 10.1089/omi.2018.0013

Lee-Sarwar, K. A., Lasky-Su, J., Kelly, R. S., Litonjua, A. A., and Weiss, S. T. (2020). Metabolome–microbiome crosstalk and human disease. *Metabolites* 10:181. doi: 10.3390/metabo10050181

Leong, J. M., Gröschel, M. I., Walker, T. M., Van Der Werf, T. S., Lange, C., Niemann, S., et al. (2018). Pathogen-based precision medicine for drug-resistant tuberculosis. *PLoS Pathog.* 14:e1007297. doi: 10.1371/journal.ppat.1007297

Li, L., Mendis, N., Trigui, H., Oliver, J. D., and Faucher, S. P. (2014). The importance of the viable but non-culturable state in human bacterial pathogens. *Front. Microbiol.* 5:258. doi: 10.3389/fmicb.2014.00258

Long, S., Yang, Y., Shen, C., Wang, Y., Deng, A., Qin, Q., et al. (2020). Metaproteomics characterizes human gut microbiome function in colorectal cancer. *NPJ Biofilms Microbiomes* 6:14. doi: 10.1038/s41522-020-0123-4

Marcon, A. R., Turvey, S., and Caulfield, T. (2021). 'Gut health' and the microbiome in the popular press: a content analysis. *BMJ Open* 11:e052446. doi: 10.1136/bmjopen-2021-052446

Mathieu, E., Escribano-Vazquez, U., Descamps, D., Cherbuy, C., Langella, P., Riffault, S., et al. (2018). Paradigms of lung microbiota functions in health and disease, particularly, in asthma. *Front. Physiol.* 9:1168. doi: 10.3389/fphys.2018.01168

Miller, R. R., Montoya, V., Gardy, J. L., Patrick, D. M., and Tang, P. (2013). Metagenomics for pathogen detection in public health. *Genome Med.* 5:81. doi: 10.1186/gm485

Moir, J. W. B. (2015). Meningitis in adolescents: the role of commensal microbiota. *Trends Microbiol.* 23, 181–182. doi: 10.1016/j.tim.2015.02.004

Nyholm, L., Koziol, A., Marcos, S., Botnen, A. B., Aizpurua, O., Gopalakrishnan, S., et al. (2020). Holo-omics: integrated host-microbiota multi-omics for basic and applied biological research. *iScience* 23:101414. doi: 10.1016/j.isci.2020.101414

Ohno, H., and Satoh-Takayama, N. (2020). Stomach microbiota, *Helicobacter pylori*, and group 2 innate lymphoid cells. *Exp. Mol. Med.* 52, 1377–1382. doi: 10.1038/s12276-020-00485-8

Ojala, T., Lindford, A., Savijoki, K., Lagus, H., Tommila, J., Medlar, A., et al. (2021). Metatranscriptomic assessment of burn wound infection clearance. *Clin. Microbiol. Infect.* 27, 144–146. doi: 10.1016/j.cmi.2020.07.021

Onderdonk, A. B., Delaney, M. L., and Fichorova, R. N. (2016). The human microbiome during bacterial vaginosis. *Clin. Microbiol. Rev.* 29, 223–238. doi: 10.1128/CMR.00075-15

Perez-Carrasco, V., Soriano-Lerma, A., Soriano, M., Gutiérrez-Fernández, J., and Garcia-Salcedo, J. A. (2021). Urinary microbiome: yin and yang of the urinary tract. *Front. Cell. Infect. Microbiol.* 11:617002. doi: 10.3389/fcimb.2021.617002

Peterson, D., Bonham, K. S., Rowland, S., Pattanayak, C. W., and Klepac-Ceraj, V. (2021). Comparative analysis of 16S rRNA gene and metagenome sequencing in pediatric gut microbiomes. *Front. Microbiol.* 12:670336. doi: 10.3389/fmicb.2021.670336

Quince, C., Walker, A. W., Simpson, J. T., Loman, N. J., and Segata, N. (2017). Shotgun metagenomics, from sampling to analysis. *Nat. Biotechnol.* 35, 833–844.

Ravel, J., Twin, J., Bradshaw, C. S., Garland, S. M., Fairley, C. K., Fethers, K., et al. (2013). The potential of metatranscriptomics for identifying screening targets for bacterial vaginosis. *PLoS One* 8:e76892. doi: 10.1371/journal.pone.0076892

Robert, S., and Filkins, L. (2019). "Clinical metagenomics for infection diagnosis," in *Genomic and Precision Medicine: Infectious and Inflammatory Disease*, eds G. Ginsburg, H. Willard, C. Woods, and E. Tsalik (Amsterdam: Elsevier), 35–60.

Roos, K. L. (2015). Bacterial Infections of the Central Nervous System. *CONTINUUM* 21, 1679–1691.

Schmidt, K., Mwaigwisya, S., Crossman, L. C., Doumith, M., Munroe, D., Pires, C., et al. (2017). Identification of bacterial pathogens and antimicrobial resistance

directly from clinical urines by nanopore-based metagenomic sequencing. *J. Antimicrob. Chemother.* 72, 104–114. doi: 10.1093/jac/dkw397

Segata, N. (2018). On the road to strain-resolved comparative metagenomics. *mSystems* 3:e00190-17. doi: 10.1128/mSystems.00190-17

Segata, N., Boernigen, D., Tickle, T. L., Morgan, X. C., Garrett, W. S., and Huttenhower, C. (2013). Computational meta'omics for microbial community studies. *Mol. Syst. Biol.* 9:666. doi: 10.1038/msb.2013.22

Sequeira, J. C., Rocha, M., Madalena Alves, M., and Salvador, A. F. (2019). "MOSCA: an automated pipeline for integrated metagenomics and metatranscriptomics data analysis," in *Proceedings of the 12th International Conference Practical Applications of Computational Biology and Bioinformatics*, Vol. 803, eds F. Fdez-Riverola, M. Mohamad, M. Rocha, J. De Paz, and P. González (Cham: Springer), 183–191.

Shakya, M., Lo, C.-C., and Chain, P. S. G. (2019). Advances and challenges in metatranscriptomic analysis. *Front. Genet.* 10:904. doi: 10.3389/fgene.2019.00904

Srinivasan, S., Chambers, L. C., Tapia, K. A., Hoffman, N. G., Munch, M. M., Morgan, J. L., et al. (2021). Urethral microbiota in men: association of *Haemophilus influenzae* and mycoplasma penetrans with nongonococcal urethritis. *Clin*. Infect. Dis. 73, e1684–e1693. doi: 10.1093/cid/ciaa1123

Tecon, R., Mitri, S., Ciccarese, D., Or, D., Van Der Meer, J. R., and Johnson, D. R. (2019). Bridging the holistic-reductionist divide in microbial ecology. *mSystems* 4:e00265-18. doi: 10.1128/mSystems.00265-18

Thibeault, C., Suttorp, N., and Opitz, B. (2021). The microbiota in pneumonia: from protection to predisposition. *Sci. Transl. Med.* 13:eaba0501. doi: 10.1126/scitranslmed.aba0501

Valles-Colomer, M., Darzi, Y., Vieira-Silva, S., Falony, G., Raes, J., and Joossens, M. (2016). Meta-omics in inflammatory bowel disease research: applications, challenges, and guidelines. *J. Crohns Colitis* 10, 735–746. doi: 10.1093/ecco-jcc/jjw024

Velmurugan, G., Dinakaran, V., Rajendhran, J., and Swaminathan, K. (2020). Blood microbiota and circulating microbial metabolites in diabetes and cardiovascular disease. *Trends Endocrinol. Metab.* 31, 835–847. doi: 10.1016/j.tem.2020.01.013

Vinaixa, M., Schymanski, E. L., Neumann, S., Navarro, M., Salek, R. M., and Yanes, O. (2016). Mass spectral databases for LC/MS- and GC/MS-based metabolomics: state of the field and future prospects. *TrAC Trends Anal. Chem.* 78, 23–35.

Walker, A., Pfitzner, B., Neschen, S., Kahle, M., Harir, M., Lucio, M., et al. (2014). Distinct signatures of host–microbial meta-metabolome and gut microbiome in two C57BL/6 strains under high-fat diet. *ISME J.* 8, 2380–2396. doi: 10.1038/ismej.2014.79

Wang, B., Yao, M., Lv, L., Ling, Z., and Li, L. (2017). The human microbiota in health and disease. *Engineering* 3, 71–82.

Wang, L., Du, Y., Xu, B.-J., Deng, X., Liu, Q.-H., Zhong, Q.-Q., et al. (2019). Metabolomics study of metabolic changes in renal cells in response to high-glucose exposure based on liquid or gas chromatography coupled with mass spectrometry. *Front. Pharmacol.* 10:928. doi: 10.3389/fphar.2019.00928

Wang, L., Liu, W., Tang, J.-W., Wang, J.-J., Liu, Q.-H., Wen, P.-B., et al. (2021). Applications of raman spectroscopy in bacterial infections: principles, advantages, and shortcomings. *Front. Microbiol.* 12:683580. doi: 10.3389/fmicb.2021.683580

Wang, L., Tay, A. C. Y., Li, J., and Zhao, Q. (2022). Editorial: computational predictions, dynamic tracking, and evolutionary analysis of antibiotic resistance through the mining of microbial genomes and metagenomic data. *Front. Microbiol.* 13:880967. doi: 10.3389/fmicb.2022.880967

Wang, W. (2022). Glycomedicine: the current state of the art. *Engineering* 1–4. doi: 10.1016/j.eng.2022.03.009

Willis, J. R., and Gabaldón, T. (2020). The human oral microbiome in health and disease: from sequences to ecosystems. *Microorganisms* 8:308. doi: 10.3390/microorganisms8020308

Xu, J., and Yang, Y. (2021). Gut microbiome and its meta-omics perspectives: profound implications for cardiovascular diseases. *Gut Microbes* 13:1936379. doi: 10.1080/19490976.2021.1936379

Yuan, L., Zhao, J., Sun, T., and Shen, Z. (2021). A machine learning framework that integrates multi-omics data predicts cancer-related LncRNAs. *BMC Bioinformatics* 22:332. doi: 10.1186/s12859-021-04256-8

Zarco, M. F., Vess, T. J., and Ginsburg, G. S. (2012). The oral microbiome in health and disease and the potential impact on personalized dental medicine. *Oral Dis.* 18, 109–120. doi: 10.1111/j.1601-0825.2011.01851.x

Zhai, P., Yang, L., Guo, X., Wang, Z., Guo, J., Wang, X., et al. (2017). MetaComp: comprehensive analysis software for comparative meta-omics including comparative metagenomics. *BMC Bioinformatics* 18:434. doi: 10.1186/s12859-017-1849-8

Zhang, Y., Thompson, K. N., Branck, T., Yan, Y., Nguyen, L. H., Franzosa, E. A., et al. (2021). Metatranscriptomics for the human microbiome and microbial community functional profiling. *Annu. Rev. Biomed. Data Sci.* 4, 279–311. doi: 10.1146/annurev-biodatasci-031121-103035

Zhong, H., Ren, H., Lu, Y., Fang, C., Hou, G., Yang, Z., et al. (2019). Distinct gut metagenomics and metaproteomics signatures in prediabetics and treatment-naïve type 2 diabetics. *EBioMedicine* 47, 373–383. doi: 10.1016/j.ebiom.2019.08.048

Zhu, X., Li, B., Lou, P., Dai, T., Chen, Y., Zhuge, A., et al. (2021). The relationship between the gut microbiome and neurodegenerative diseases. *Neurosci. Bull.* 37, 1510–1522. doi: 10.1007/s12264-021-00730-8

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.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership