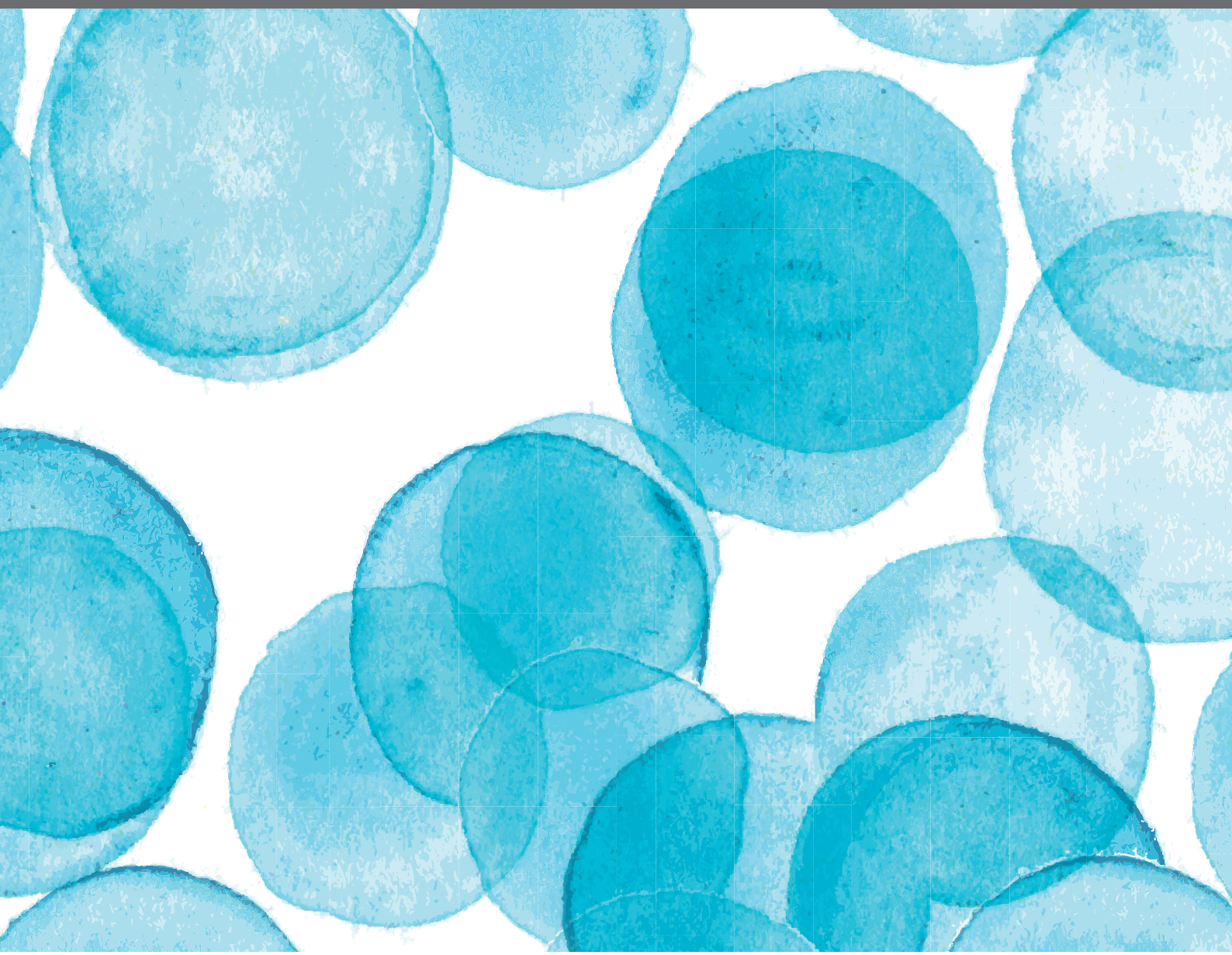
MULTI-OMICS PROFILING OF UNIQUE NICHES TO REVEAL THE MICROBIAL AND METABOLITE COMPOSITION

EDITED BY: Roshan Kumar, Vasvi Chaudhry and Om Prakash
PUBLISHED IN: Frontiers in Microbiology and Frontiers in Genetics



frontiers Research Topics

frontiers

ISSN 1664-8714
ISBN 978-2-83250-120-7
DOI 10.3389/978-2-83250-120-7

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

MULTI-OMICS PROFILING OF UNIQUE NICHES TO REVEAL THE MICROBIAL AND METABOLITE COMPOSITION

Topic Editors:
Roshan Kumar, Magadh University, India
Vasvi Chaudhry, University of Tübingen, Germany
Om Prakash, National Centre for Cell Science, India

Citation: Kumar, R., Chaudhry, V., Prakash, O., eds. (2022). Multi-omics Profiling of Unique Niches to Reveal the Microbial and Metabolite Composition. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-83250-120-7

Table of Contents

05 ***Editorial: Multi-omics Profiling of Unique Niches to Reveal the Microbial and Metabolite Composition***
Roshan Kumar, Vasvi Chaudhry and Om Prakash

08 ***Positive Synergistic Effects of Quercetin and Rice Bran on Human Gut Microbiota Reduces* Enterobacteriaceae *Family Abundance and Elevates Propionate in a Bioreactor Model***
Sudeep Ghimire, Supapit Wongkuna, Ranjini Sankaranarayanan, Elizabeth P. Ryan, G. Jayarama Bhat and Joy Scaria

20 ***Rational Proteomic Analysis of a New Domesticated* Klebsiella pneumoniae *x546 Producing 1,3-Propanediol***
Xin Wang, Lin Zhang, Hong Chen, Pan Wang, Ying Yin, Jiaqi Jin, Jianwei Xu and Jianping Wen

36 ***Multi-Omics Reveal the Efficient Phosphate-Solubilizing Mechanism of Bacteria on Rocky Soil***
Yanqiang Ding, Zhuolin Yi, Yang Fang, Sulan He, Yuming Li, Kaize He, Hai Zhao and Yanling Jin

49 ***Multi-Omics Analysis of Lipid Metabolism for a Marine Probiotic* Meyerozyma guilliermondii *GXDK6 Under High NaCl Stress***
Huijie Sun, Xinghua Cai, Bing Yan, Huashan Bai, Duotao Meng, Xueyan Mo, Sheng He, Guijiao Su and Chengjian Jiang

61 ***β-Glucuronidase Pattern Predicted From Gut Metagenomes Indicates Potentially Diversified Pharmacomicrobiomics***
Francesco Candeliere, Stefano Raimondi, Raffaella Ranieri, Eliana Musmeci, Alfonso Zambon, Alberto Amaretti and Maddalena Rossi

73 ***Inflammatory Bowel Disease and COVID-19: How Microbiomics and Metabolomics Depict Two Sides of the Same Coin***
Gian Mario Cortes, Maria Antonietta Marcialis, Flaminia Bardanzellu, Angelica Corrias, Vassilios Fanos and Michele Mussap

97 ***Digital Therapeutics Care Utilizing Genetic and Gut Microbiome Signals for the Management of Functional Gastrointestinal Disorders: Results From a Preliminary Retrospective Study***
Shreyas V. Kumbhare, Patricia A. Francis-Lyon, Dashyanng Kachru, Tejaswini Uday, Carmel Irudayanathan, Karthik M. Muthukumar, Roshni R. Ricchetti, Simitha Singh-Rambiritch, Juan Ugalde, Parambir S. Dulai, Daniel E. Almonacid and Ranjan Sinha

111 ***Microbial Ecology of Sulfur Biogeochemical Cycling at a Mesothermal Hot Spring Atop Northern Himalayas, India***
Shekhar Nagar, Chandni Talwar, Mikael Motelica-Heino, Hans-Hermann Richnow, Mallikarjun Shakarad, Rup Lal and Ram Krishan Negi

127 ***Insight Into the Molecular Mechanisms Underpinning the Mycoremediation of Multiple Metals by Proteomic Technique***
Priyadarshini Dey, Anushree Malik, Dileep Kumar Singh, Sven-Bastiaan Haange, Martin von Bergen and Nico Jehmlich

141 ***Life Within a Contaminated Niche: Comparative Genomic Analyses of an Integrative Conjugative Element ICE*nah*CSV86 and Two Genomic Islands From* Pseudomonas bharatica *CSV86^T Suggest Probable Role in Colonization and Adaptation***

Balaram Mohapatra, Harshit Malhotra and Prashant S. Phale

160 ***Metagenomics: An Approach for Unraveling the Community Structure and Functional Potential of Activated Sludge of a Common Effluent Treatment Plant***

Gunjan Vasudeva, Harpreet Singh, Sakshi Paliwal and Anil Kumar Pinnaka

frontiers | Frontiers in Microbiology

TYPE Editorial
PUBLISHED 19 August 2022
DOI 10.3389/fmicb.2022.997191

OPEN ACCESS

EDITED AND REVIEWED BY
Ludmila Chistoserdova,
University of Washington,
United States

*CORRESPONDENCE
Roshan Kumar
roshanzhc@gmail.com

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

RECEIVED 18 July 2022
ACCEPTED 04 August 2022
PUBLISHED 19 August 2022

CITATION
Kumar R, Chaudhry V and Prakash O (2022) Editorial: Multi-omics profiling of unique niches to reveal the microbial and metabolite composition. *Front. Microbiol.* 13:997191. doi: 10.3389/fmicb.2022.997191

Editorial: Multi-omics profiling of unique niches to reveal the microbial and metabolite composition

Roshan Kumar[1]*, Vasvi Chaudhry[2] and Om Prakash[3]

[1]Post-Graduate Department of Zoology, Magadh University, Bodh Gaya, India, [2]Department of Microbial Interactions, IMIT/ZMBP, University of Tübingen, Tübingen, Germany, [3]National Centre for Microbial Resource (NCMR), National Centre for Cell Science, Pune, Maharashtra, India

KEYWORDS

multi-omics, microbial composition and distribution, niches, proteomics, metagenomics, transcriptomics, metabolomics

Editorial on the Research Topic
Multi-omics profiling of unique niches to reveal the microbial and metabolite composition

It is an exciting time in the field of Microbiology. With the rise of the era of OMICS, we have witnessed an explosion of modern multi-omics technologies. These include (meta)genomics, transcriptomics, proteomics, meta-metabolomics, and culturomics in different aspects of basic and applied research. The use of OMICS has greatly enhanced our understanding of the structure, structure-related functions, and interactions among microbes with biotic and abiotic components of ecosystems and revealed their role in offering different ecosystem services (Kumar et al., 2022). Considering the extensive use and versatility of OMICS techniques in different disciplines of life science, we created this Research Topic "*Multi-omics profiling of unique niches to reveal the microbial and metabolite composition*" to understand the role of microbes in their respective niches in the presence and absence of certain variables. In this special issue, we selected and published 11 articles contributed by 76 authors. The articles were selected based on their significance in understanding the current state of research on the topic. We take the opportunity here to reflect the significant contributions of these articles.

The first article in this issue demonstrates the impact of dietary supplements, namely quercetin and rice bran, on enriching beneficial gut microbes in a mini-bioreactor array system, i.e., an *in vitro* model system without the confounding host factors (Ghimire et al.). According to the study's findings, rice bran induces substantial changes in gut-microbial diversity when compared with quercetin. In contrast to this, in quercetin supplementation, the *Acidaminococcus intestine* was significantly enriched along with increased iso-butyrate production. Another study in this issue utilizes phosphate-solubilizing bacteria (PSB) isolated from the sweet potato rhizosphere. Multi-omics analysis of these strains demonstrated the upregulation of genes involved in the gluconic acid synthesis and the tricarboxylic acid cycle (Ding et al.). The study concludes that increased gluconic and malic acid, due to up-regulation of genes, is responsible for enhanced phosphate-solubilizing potential.

Further, one study with *Klebsiella pneumoniae* x546, analyzes the production of 1, 3-propanediol (1, 3-PDO) and reports improved production with betaine and shorter fermentation time under controlled pH conditions (Wang et al.). This study used the proteomics approach to identify proteins expressed during fermentation, such as homoserine kinase (ThrB), and in 1, 3-PDO biosynthesis, such as lactate dehydrogenase (LDH), DhaD, DhaK, and BudC.

In another study, the authors reveal the relationship between antibiotic sensitivity and lipid metabolism using a marine bacterium *Meyerozyma guilliermondii* GXDK6 (Sun et al.). They demonstrated that under 10% NaCl stress, expression of AYR1 and NADPH-dependent 1-acyl-dihydroxyacetone phosphate reductase was inhibited, which triggered the intracellular accumulation of glycerol and weakened the budding and proliferation process of cells. They also reported that NaCl stress enhanced the organism's sensitivity to fluconazole by inhibiting the expression of drug target proteins lanosterol 14-alpha demethylase.

Another article examines whether a digital therapeutic intervention tailored to genomic SNPs and gut-microbiome signals could reduce symptoms of functional gastrointestinal disorders (FGIDs) using the datasets from individuals who have successfully lost weight (Kumbhare et al.). A logistic regression model was trained to distinguish FGID status among subjects enrolled in digital therapeutics care programs based on demographic, genetic, and baseline microbiome data.

In their paper, Candeliere et al. assembles, bins, and mines the intestinal metagenomes of 60 healthy individuals. This study evaluated the profiles of β-glucuronidases (GUS), an important enzyme that plays a pivotal role in metabolizing drugs in the gastrointestinal (GI) tract (Pollet et al., 2017) among healthy individuals from five geographically diverse cohorts (Candeliere et al.). Only the Ethiopia cohort (ETH) had significantly fewer GUS-encoding bacteria when compared to the other cohorts. The bacterial species contributing to different GUS categories belonged to genus *Bacteroides*, *Faecalibacterium*, and *Eubacterium*.

An additional metagenomic analysis published in this issue found abundant sulfur-metabolizing genetic repertoires and microbes with established S-cycling capabilities, including *Pseudomonas*, *Thioalkalivibrio*, *Desulfovibrio*, and *Desulfobulbaceae* (Nagar et al.). As evident from the diversity of sulfur-oxidizing bacteria (SOB) and sulfate-reducing bacteria (SRB) with conserved (r)dsrAB, it is likely to be a necessary adaptation for microbial fitness at this site. To get a detailed insight into the metal fungus interactions and to develop the technology of Mycoremediation of heavy metal contaminated sites, another paper in the issue examines the protein expression in *Aspergillus fumigatus* PD-18 in response to heavy metals stress (5 mg/L of Cr, Cd2+, Cu2+, Ni2+, Pb2+, and Zn2+) using the proteomics approach (Dey et al.). Signaling and cellular processing were the most prevalent functional classes under these conditions. Using protein-protein interaction network analysis, authors report that cytochrome-c oxidase and 60S ribosomal protein had crucial roles in detoxifying multi-metals.

A review article explores and compares the effect of metabolic and microbiota alterations on disease progression in patients with COVID-19 and inflammatory bowel disease (Cortes et al.). The study emphasizes that both the conditions are associated with decreased Firmicutes populations, with a specific reduction in *Faecalibacterium prausnitzii* cells, which produces different kinds of short-chain fatty acid (SCFA) under anaerobic conditions. The study also examines how gut dysbiosis, inflammation, oxidative stress, and energy demand affect key metabolites such as tryptophan, phenylalanine, histidine, glutamine, succinate, and citrate levels in the gut.

Using a comparative genomic approach, Mohapatra et al. investigates the presence of three genomic islands, ICEnahCSV86, PBGI-1, and PBGI-2 (GIs, > 50 Kb size), in an aromatic-compound degrading soil bacterium, *Pseudomonas bharatica* CSV86T. Additionally, they elucidate the role of MGEs in the adaptation, niche colonization, and species competitive behavior of strain CSV86T in contaminated environments.

In one metagenomic study by Vasudeva et al., the authors analyze the samples collected from activated sludge of an industrial effluent treatment plant. It highlights the abundance of genes involved in xenobiotic degradation pathways and antimicrobial resistance. In both samples, glycosyl hydrolases and glycosal transferases were the most abundant carbohydrate-active enzyme classes compared to other polysaccharide enzyme classes.

In conclusion, the special issue featured several fascinating articles illustrating the importance of genomics, metagenomics, transcriptomics, proteomics, and metabolomics to understand different niches' microbes and microbial processes. The editors believe that this issue has successfully achieved reporting recent investigations on microbial processes relevant to a variety of microbial processes in a variety of environmental niches.

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Acknowledgments

We would like to thank all the authors who contributed their work to this topic. We acknowledge reviewers for taking the time and necessary effort to review the manuscripts. The editors would like to acknowledge their funding support. RK acknowledges the SERB International Research Experience award (SIR/2022/000474) for the year 2022-2023. OP acknowledges the grant from NCMR (BT/Coord.II/01/03/2016).

VC acknowledges Alexander von Humboldt-Stiftung, Cluster of Excellence (CMFI), and University of Tübingen for support.

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

References

Kumar, R., Nagar, S., Haider, S., Sood, U., Ponnusamy, K., Dhingra, G. G., et al. (2022). Monkey pox virus (MPXV): phylogenomics, host-pathogen interactome, and mutational cascade. *bioRxiv*. doi: 10.1101/2022.07.25.501367

Pollet, R. M., D'agostino, E. H., Walton, W. G., Xu, Y., Little, M. S., Biernat, K. A., et al. (2017). An atlas of β-glucuronidases in the human intestinal microbiome. *Structure* 25**,** 967–977.e965. doi: 10.1016/j.str.2017.05.003

frontiers in Microbiology

ORIGINAL RESEARCH
published: 30 September 2021
doi: 10.3389/fmicb.2021.751225

Positive Synergistic Effects of Quercetin and Rice Bran on Human Gut Microbiota Reduces *Enterobacteriaceae* Family Abundance and Elevates Propionate in a Bioreactor Model

*Sudeep Ghimire[1,2], Supapit Wongkuna[1,2], Ranjini Sankaranarayanan[3], Elizabeth P. Ryan[4], G. Jayarama Bhat[3] and Joy Scaria[1,2]**

[1] Department of Veterinary and Biomedical Sciences, South Dakota State University, Brookings, SD, United States, [2] South Dakota Center for Biologics Research and Commercialization, Brookings, SD, United States, [3] Department of Pharmaceutical Sciences, South Dakota State University, Brookings, SD, United States, [4] Department of Environmental and Radiological Health Sciences, Colorado State University, Fort Collins, CO, United States

OPEN ACCESS

Edited by:
Vasvi Chaudhry, University of Tübingen, Germany

Reviewed by:
Atanu Banerjee, Amity University Gurgaon, India
Utkarsh Sood, The Energy and Resources Institute (TERI), India

***Correspondence:**
Joy Scaria
joy.scaria@sdstate.edu

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

Received: *31 July 2021*
Accepted: *01 September 2021*
Published: *30 September 2021*

Citation:
Ghimire S, Wongkuna S, Sankaranarayanan R, Ryan EP, Bhat GJ and Scaria J (2021) Positive Synergistic Effects of Quercetin and Rice Bran on Human Gut Microbiota Reduces Enterobacteriaceae Family Abundance and Elevates Propionate in a Bioreactor Model. Front. Microbiol. 12:751225. doi: 10.3389/fmicb.2021.751225

Dietary fiber and flavonoids have substantial influence on the human gut microbiota composition that significantly impact health. Recent studies with dietary supplements such as quercetin and rice bran have shown beneficial impacts on the host alongside a positive influence of the gut microbiota. The specific bacterial species impacted by quercetin or rice bran in the diet is not well understood. In this study, we used a minibioreactor array system as a model to determine the effect of quercetin and rice bran individually, as well as in combination, on gut microbiota without the confounding host factors. We found that rice bran exerts higher shift in gut microbiome composition when compared to quercetin. At the species level, *Acidaminococcus intestini* was the only significantly enriched taxa when quercetin was supplemented, while 15 species were enriched in rice bran supplementation and 13 were enriched when quercetin and rice bran were supplemented in combination. When comparing the short chain fatty acid production, quercetin supplementation increased isobutyrate production while propionate dominated the quercetin and rice bran combined group. Higher levels of propionate were highly correlated to the lower abundance of the potentially pathogenic *Enterobacteriaceae* family. These findings suggest that the combination of quercetin and rice bran serve to enrich beneficial bacteria and reduce potential opportunistic pathogens. *In vivo* studies are necessary to determine how this synergy of quercetin and rice bran on microbiota impact host health.

Keywords: gut microbiome, bioreactor, metagenomics, flavonoid, quercetin, rice bran, prebiotic

INTRODUCTION

Gut microbiome influences health and disease (Gentile and Weir, 2018; Maji et al., 2018; Pulikkan et al., 2018) and is highly affected by diet (Turnbaugh et al., 2009; David et al., 2014; Dhakan et al., 2019). Some dietary components are known to enhance host health by promoting the growth of beneficial bacteria in the gut, termed as "prebiotics" (Gibson and Roberfroid, 1995; Slavin, 2013). Quercetin and rice bran have been used as prebiotics separately in mice and pig models, showing significant health benefits to the host (Henderson et al., 2012; Goodyear et al., 2015; Lin et al., 2019). Rice bran, a byproduct of the rice milling process, is available, affordable, sustainable, and a globally produced source of prebiotics. It is known to supplement nutrients (Zarei et al., 2017), increase beneficial bacteria growth, enhance gut mucosal immunity (Henderson et al., 2012), and prevent diseases (Kumar et al., 2012; Goodyear et al., 2015; Lei et al., 2016; Islam et al., 2017). Similarly, quercetin (3,3′,4′,5,7-pentahydroxyflavone) represents an important subgroup of flavonoids found in fruits and leafy vegetables (D'Archivio et al., 2007). Quercetin (QC) has received substantial attention in the past few years from the scientific community for exerting anti-inflammatory effects (Comalada et al., 2005; Stewart et al., 2008) and its potential health-promoting properties in the treatment or prevention of cardiovascular diseases (Egert et al., 2009), lung and colorectal cancers (Linsalata et al., 2010; Darband et al., 2018), and colitis (Hong and Piao, 2018; Lin et al., 2019). Furthermore, the mammalian body can endure high levels of quercetin without any significant adverse health effects (Harwood et al., 2007).

Despite multiple studies corroborating the beneficial effect of quercetin and rice bran on the host, their effect on microbiota varies significantly from study to study and has lower taxonomic resolution (Sheflin et al., 2015; Tamura et al., 2017; Hong and Piao, 2018; Zambrana et al., 2019). Inter-individual variation in *in vivo* studies are often due to multiple host-related factors which makes it difficult to interpret microbiome results (Zoetendal et al., 1998). In contrast, studying the effect of dietary ingredients on the microbiome in the absence of confounding host factors can help to better understand the complex microbial interactions. Bioreactors have been used as model systems to study how dietary ingredients shape microbiomes (Chung et al., 2019). The use of bioreactors allows precise control of the environmental conditions that affect microbiome composition which provide increased reproducibility and reveal microbial interactions in a more defined way. Minibioreactor array is an *in vitro* anaerobic model system that simulates hindgut conditions for growth of complex, stable microbiota without interference of host factors (Auchtung et al., 2015). Furthermore, this system helps to identify microbial biotransformations and allows for measuring metabolites produced (National Academies of Science, Engineering, and Medicine, 2017). We used minibioreactor array systems to gain deeper understanding of how the microbiota responds to quercetin and/or rice bran without host interference, and hypothesized that the combination of quercetin and rice bran will have a synergistic positive effect on the gut microbiota and microbiota metabolism. We show that combined supplementation of quercetin and rice bran in the minibioreactors model for human gut microbiome significantly reduced members of *Enterobacteriaceae* family and resulted in higher propionate production. This study provides novel insights to species-level shifts in the human gut microbiome in the presence of quercetin and/or rice bran supplementation.

MATERIALS AND METHODS

Donor Samples, Mini-Bioreactor Array Preparation and Sample Collection

We obtained fresh fecal samples from six healthy donors with no prior history of antibiotic consumption in the past year. The pooled fecal sample from six individuals was used as the inoculum as our prior study showed that pooled sample represented the individual microbial composition (Ghimire et al., 2020).

The modified BHI medium (Ghimire et al., 2020) was used as a control medium. Heat-stabilized rice bran (RBT 300) was purchased from Rice Bran Technologies (Sacramento, CA, United States), and the extract was prepared as described previously (Kumar et al., 2012) and then added to the control medium (final concentration: 2 mg/mL): designated as RB. Quercetin was added to the media at a final concentration of 75 mg/L and is referred to as QC. The final experimental group consisted of modified BHI medium with the additions of quercetin (75 mg/L) and rice bran extract (2 mg/mL) and is identified as QC + RB. Mini-bioreactors (MBRAs) were sterilized, assembled and the experiment was performed as described previously (Auchtung et al., 2016) with minor modifications. The total experimental volume in each mini-bioreactor was set at 15 ml. To maintain anaerobic conditions, the bioreactors were placed inside an anaerobic chamber (Coy Laboratories) containing with 85% nitrogen, 10% carbon dioxide, and 5% hydrogen. The temperature was maintained at 37°C throughout the experiment. Initially, pH of each of the medium for inflow into the mini-bioreactors was adjusted to 6.8 and not altered throughout the experiment. Two 24 channel Watson-Marlow pumps were used pumping media in and removing excess fermented medium from the bioreactor. For this purpose, the input and output on Watson Marlow pumps were set at 1 and 2 rpm, respectively. The rotating magnetic stirrer was set at 130 rpm. The media (Control, QC, RB, and QC + RB) (**Figure 1A**) were allowed to flow continuously for 24 h each in triplicate. Three hundred microliters of the inoculum was introduced into all wells with a retention time of 16 h. The continuous flow model was operated up to 21 days post-inoculation (**Figure 1B**). Five hundred microliters of the media was collected for sequencing at day 0 (inoculum), days 4, 7, 14, and 21 and directly frozen to −80°C. Also, samples for short chain fatty acid (SCFAs) determination were collected at days 4, 7, 14, and 21 post-inoculation with dietary treatments and controls (**Figure 1B**). Each media condition was run in triplicate and samples were collected at day 4, 7, 14, 21 resulting in 48 samples from four conditions (control, QC, RB, and QC + RB). Duplicate samples from the inoculum was also used

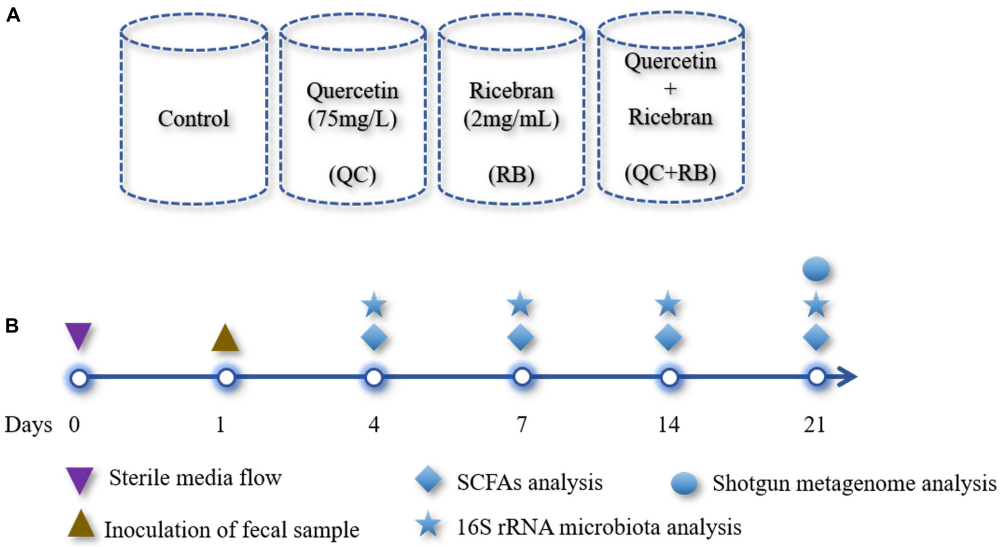


FIGURE 1 | Overview of the study design: **(A)** A schematic diagram shown the design of four different media conditions used in this study. Single substrates (quercetin or rice bran) or the mixture of quercetin and rice bran were used at the final concentration of 75 and 2 mg/mL as shown above in the base medium BHI (see section "Materials and Methods"). Final pH of the media was adjusted to 6.8 ± 0.2 for all conditions. Each condition was run in triplicate and were inoculated with the same fecal inoculum (see section "Materials and Methods"). **(B)** Outline of the bioreactor experiment showing time points for fecal inoculation and sample collection for SCFAs and microbial community analysis.

for sequencing. Therefore, a total of 50 samples were used for DNA isolation and sequencing.

Microbial DNA Extraction and Sequencing

DNA isolation was performed on 50 samples including duplicate inoculum samples. The DNA was extracted from 500 μl of the sample using a Powersoil DNA isolation kit (MoBio Laboratories Inc., CA, United States) following the manufacturer's instructions. After extraction, the quality of DNA was measured using NanoDrop™ one (Thermo Fisher Scientific, DE, United States) and quantified using Qubit Fluorometer 3.0 (Invitrogen, CA, United States). The DNA samples were stored at −20°C until further use. To analyze the variation of the microbial composition over time, all samples were amplicon sequenced using an Illumina MiSeq platform with paired-end V3 chemistry. The library was prepared using an Illumina Nextera XT library preparation kit (Illumina Inc., CA, United States) targeting V3-V4 regions of the 16S rRNA. The libraries were bead normalized and multiplexed before loading into the sequencer. We also performed shotgun metagenome sequencing on twelve samples obtained from day 21 to identify species level taxonomical differences between the experimental groups. We used the Nextera XT kit (Illumina, San Diego, CA, United States) for the preparation of the shotgun metagenome sequencing library. The library was then sequenced using paired end 300 base sequencing chemistry using a MiSeq platform.

Data Analysis

The time-series changes in the microbial communities were analyzed using 16S rRNA community analysis in Quantitative Insights into Microbial Ecology framework (QIIME, Version 2.0) (Bolyen et al., 2019). Briefly, the demultiplexed reads obtained were quality filtered using q2- demux plugin and denoised applying DADA2 (Callahan et al., 2016). All amplicon sequence variants were aligned with *mafft* (Katoh et al., 2002) to construct a phylogeny with fasttree2 (Price et al., 2010). The outputs rooted-tree.qza, table.qza, taxonomy.qza were then imported into R (R Core Team, 2017) for analysis using *phyloseq* (McMurdie and Holmes, 2013). Shannon diversity and bray curtis dissimilarity indices were calculated as alpha and beta diversity metrices. Kruskal–Wallis test at $p = 0.05$ was performed to compare the species richness between the groups. The reads were normalized by rarefying to 30,000 reads using rarefy_even_depth function from phyloseq package in R and the taxonomy was assigned to amplicon sequence variants using the q2-feature-classifier (Bokulich et al., 2018) using Greengenes as the reference (McDonald et al., 2012). Rarefying the reads to 30,000 were enough to estimate total diversity and taxonomy (**Supplementary Figure 1A**). Initially, a total of 947 amplicon sequence variants were identified from 50 samples. The average number of non-chimeric reads per sample obtained in Qiime2 pipeline was 92,512 ± 27,631 (mean ± SD). The amplicon sequence variants obtained were filtered to select those which are present in at least 20% of samples with a count of 10 each or amplicon sequence variants with > 0.001% of total median count reducing the total number to 581.

For shotgun metagenomes, raw fastq sequences were quality controlled using Fastqc[1] and host reads were removed using metaWRAP pipeline (Uritskiy et al., 2018). Filtered reads were

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

analyzed for taxonomy using Kaiju (Menzel et al., 2016) against the proGenomes database (Mende et al., 2017) using default parameters. The percentage abundance of each taxon was plotted using Explicit v2.10.5 (Robertson et al., 2013). The samples from day 21 yielded a total of 39,332,649 reads of which 70. 86% were classified to 161 different species (species with less than 100 reads were removed), whereas 185,188 (5.55%) and 9,278,647 (23.59%) reads remained as unclassified bacteria or unassigned to non-viral species, respectively. In addition, the raw counts obtained for each taxon were used for calculation of alpha and beta diversity using the *phyloseq* package in R. For identification of differentially abundant taxa in QC, RB and QC + RB compared to control medium, the raw counts of abundance of each taxon was then Log10(x + 1) transformed and fed to DESEq2 (Love et al., 2014) package in R. Bacterial taxa which significantly altered from the control were further filtered out with the criteria of at least log2foldchange (Log2FC) of ≥ 2 and p*adj* value > 0.05. The enriched taxa were selectively analyzed for their co-relation to SCFAs phenotypic data using Spearman correlation method in R. In addition, we performed *de novo* assembly of the quality filtered reads using metaSPAdes (SPAdes 3.13.0) (Nurk et al., 2017). The assembled *fasta* sequences were annotated using Prokka (Seemann, 2014) using default parameters. The obtained amino acid *fasta* sequences were searched for the KEGG modules using GHOSTKOALA (Kanehisa et al., 2016) with default parameters. The KEGG modules thus obtained were analyzed for complete, 1 block missing, 2 blocks missing and incomplete pathways and visualized as heatmap using Morpheus web server [2].

Estimation of Short Chain Fatty Acids

For the estimation of SCFAs, 800 μl of samples were collected from each mini-bioreactor, mixed with 160 μl of 25% m-phosphoric acid and frozen at −80°C until further analysis. Later, the frozen samples were thawed and centrifuged (>15,000 × *g*) for 20 min. Five hundred microliters of supernatant was collected in the tubes before loading into the gas chromatography- mass spectrometry (GC-MS) (Agilent Technologies, United States) for analysis (Ghimire et al., 2020). The SCFA concentrations were compared between the groups using the Kruskal–Wallis test followed by the Dunn test with Benjamini-Hochberg correction in R and visualized using GraphPad Prism 6.0.

RESULTS

Rice Bran Produced Larger Shift in Microbiota Composition When Compared to Quercetin

We used a minibioreactor array system to analyze the impact of QC and RB on the gut microbiota using the study design shown in **Figure 1**. The microbial community richness indicated by the Shannon diversity index showed significant differences in the early days (days 4 and 7) due to addition of rice bran in the media whether alone or in combination with quercetin (**Figure 2A**). No significant changes in the richness was observed among the groups on day 14 and day 21, suggesting the stabilization of the microbial communities by day 14. Also, on day 21, shotgun metagenome analysis showed no differences in the richness between the four conditions (**Supplementary Figure 1B**). In contrast, marked differences between the communities were evident as early as day 4 by beta diversity analysis (**Supplementary Figure 2A**). Rice bran (RB) supplementation showed a greater shift in the bacterial community as shown on days 7, 14, and 21 (**Supplementary Figures 2B–D**). Based on Bray-Curtis dissimilarity metrics, RB and QC + RB treatments were similar to each other and clustered separately from control and QC. This reveals that quercetin causes less shift in microbiome composition when compared to rice bran (**Supplementary Figure 2**). Shotgun metagenome sequencing at day 21 also indicated significant differences in the community profile, primarily influenced by the addition of rice bran extract in the medium (**Figure 2B**).

When examined taxonomically, the baseline inoculum composition was dominated by *Prevotellaceae* (29.9 ± 0.87%), followed by *Ruminococcaceae* (18.82 ± 0.81%) and *Lachnospiraceae* (12.66 ± 0.4%). Also, *Enterobacteriaceae* was found to be very low in inoculum (0.33 ± 0.097%). However, the abundance of *Prevotellaceae* was reduced to ∼0.0% after 21 days in all treatment groups suggesting that not all taxa in the inoculum were supported in the bioreactor model. Over time, there was a major shift in the composition of the microbiota between the groups starting as early as day 4 (**Figure 2C**). The control and quercetin medium were dominated by *Enterobacteriaceae*, followed by *Lachnospiraceae* at day 4. However, by day 21, *Lachnospiraceae* dominated the control and QC conditions followed by *Bacteroidaceae*. The abundance of *Enterobacteriaceae* was reduced from 34.16 ± 1.92% and 33.26 ± 1.11% on day 4 to 17.62 ± 4.2% and 18.53 ± 3.88% in control and QC medium, respectively, by day 21 (**Figure 2C**). Even though ∼50% reduction of abundance of *Enterobacteriaceae* was observed for both control and QC medium at day 21 compared to day 4, the population of *Enterobacteriaceae* tend to stabilize. However, *Enterobacteriaceae* were highly reduced in both RB and QC + RB conditions. In RB medium, *Enterobacteriaceae* was 10.212 ± 2.14% at day 4 which was further reduced to 6.33 ± 2.59% by day 21. At day 21, *Veillonellaceae* and *Lachnospiraceae* dominated the RB medium. Similarly, the dominant family were *Lachnospiraceae, Veillonellaceae* and *Bacteroidaceae* in the quercetin and rice bran combination (QC + RB) on day 21. Strikingly, on day 21, the abundance of *Enterobacteriaceae* was reduced by ∼18.22% to 4.20 ± 3.16% compared to day 4 on QC + RB condition. At day 21, when mean abundance of *Enterobacteriaceae* was compared between control, QC, RB and QC + RB, significant reduction was observed in RB and QC + RB (Kruskal–Wallis, $p = 0.0064$).

On day 21, we performed shotgun metagenome analysis to identify the species level differences among the four groups (**Supplementary Figure 3** and **Supplementary Table 1**). Based on DeSEQ2 analysis, 53 species were significantly altered in QC + RB medium (**Figure 3**). The combination (QC + RB)

[2]https://software.broadinstitute.org/morpheus/

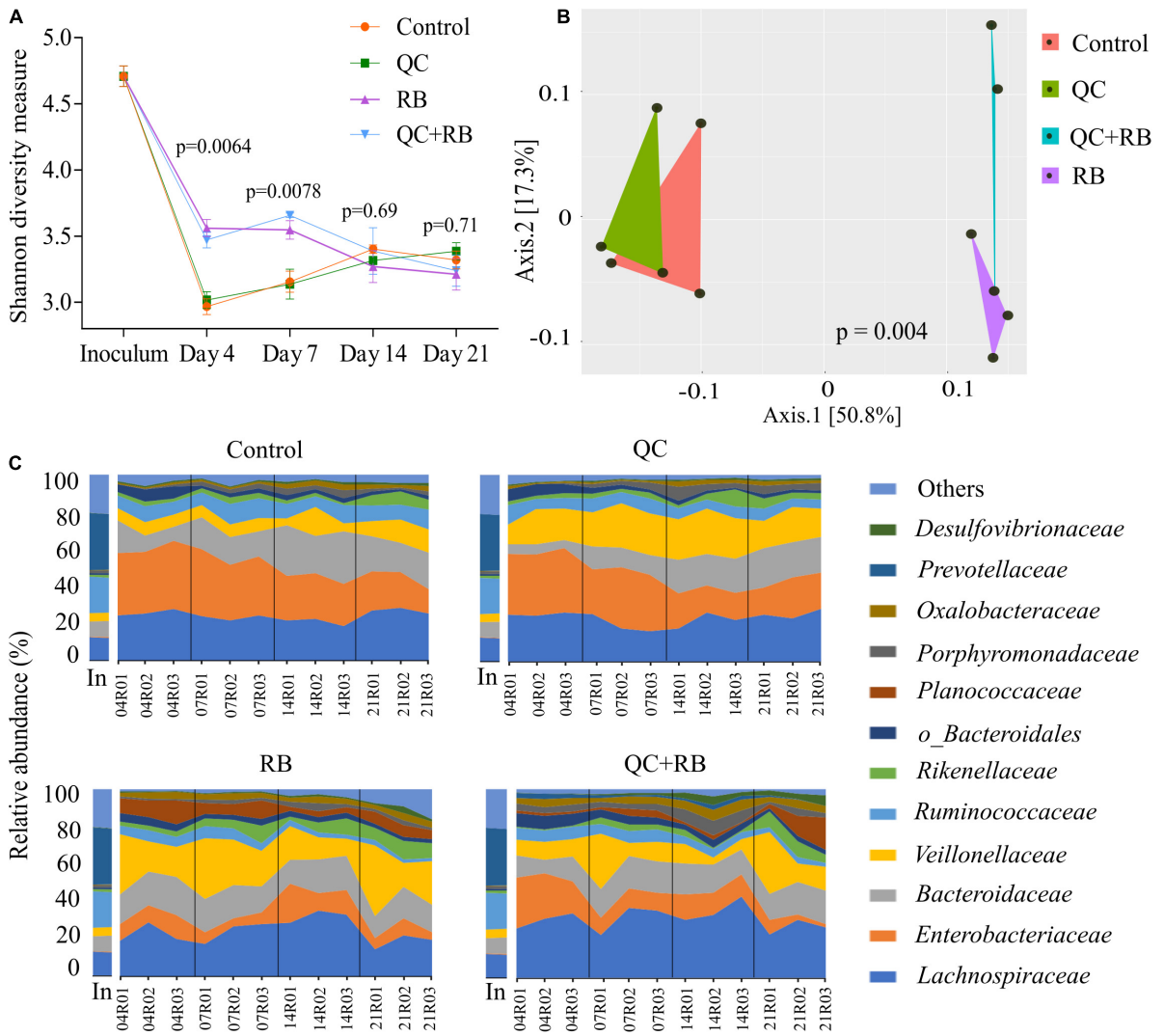


FIGURE 2 | Gut microbiota compositional changes following quercetin and rice bran supplementation **(A)** Alpha diversity of four groups (control, QC, RB, and QC + RB) at day 4, 7, 14, and 21 from amplicon sequencing data. **(B)** Beta diversity among the control, QC, RB and QC + RB groups were obtained using shotgun metagenome sequence analysis on day 21. "MDS" ordination followed by Bray-Curtis dissimilarity calculation was used to visualize the differences between the groups (adonis, $p = 0.004$). **(C)** Temporal family-level composition of the most abundant 12 bacterial taxa at day 4, 7, 14, and 21 and inoculum determined using 16S rRNA community profiling for Control, QC, RB, and QC + RB conditions.

enriched 13 additional species and reduced 13 others when the cut off value of Log2FC $\geq$ 2 and p*adj* value $>$ 0.05 was applied (**Supplementary Table 1**). Specifically, abundance of *Escherichia* sp. TW09308, *E. coli*, *E. albertii*, *Citrobacter koseri*, and *C. youngae* were reduced. The *Citrobacter* species, including *C. rodentium,* were also reduced in RB medium; however, *Escherichia* species were not significantly reduced in this medium (**Figure 3**).

Further, 13 species were altered significantly in the quercetin group (**Figure 3A**), and 70 species in the rice bran group compared to control (**Figure 3B**). With a cut off Log2FC $\geq$ 2 and p*adj* value $>$ 0.05, *Acidaminococcus intestine* was the only enriched taxa in media supplemented with quercetin while *Coprococcus catus, Dorea longicatena, Anaerostipes caccae, Eubacterium limosum,* and *Enterococcus faecalis* were highly reduced. Interestingly, the population of *Flavonifractor plautii,* a flavonoid metabolizing bacterium (Braune and Blaut, 2016; Gupta et al., 2019) was significantly reduced in quercetin supplementation. Rice bran supplementation alone enriched 15 species, including six different species from *Veillonella* (**Supplementary Table 1**).

To gain functional insights into the prebiotic supplementation, we performed comparative pathway analysis between controls and treatments using KEGG modules search for the metagenomic samples at day 21 (**Supplementary Figure 4**). Results from this analysis showed that cysteine biosynthesis (M00338), lysine biosynthesis (M00031) and ornithine biosynthesis (M00763) pathways were absent in QC

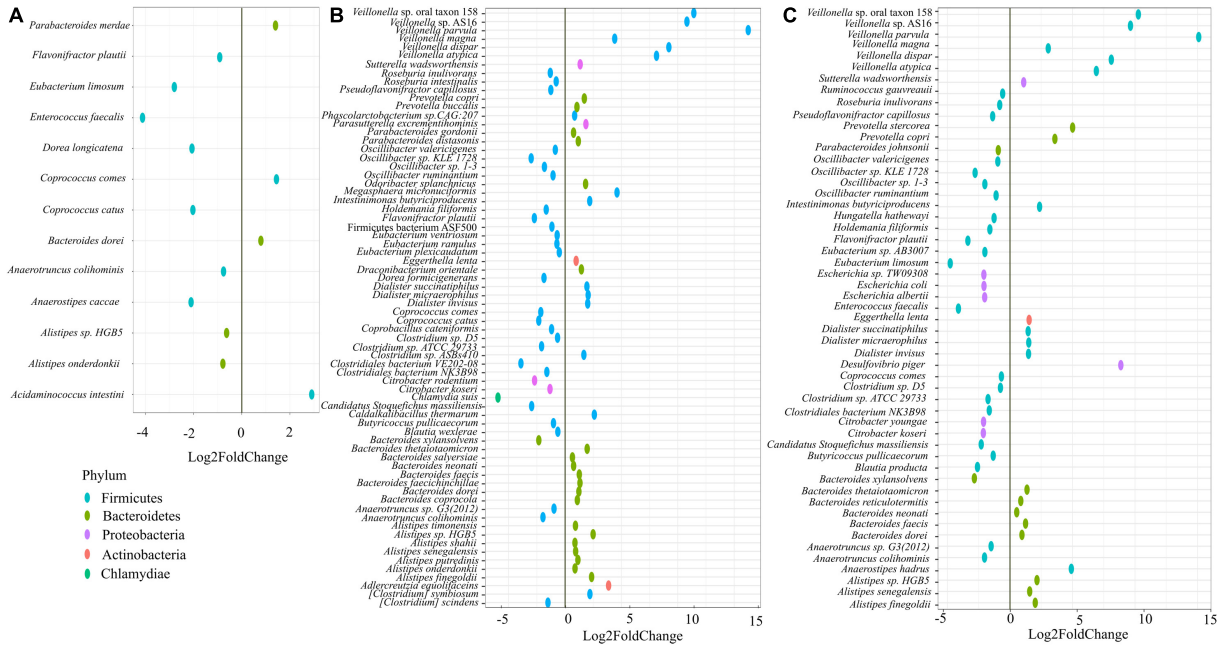


FIGURE 3 | Impact of quercetin and rice bran supplementation on microbiota bacterial species. Bacterial species that are significantly altered in **(A)** quercetin (QC), **(B)** rice bran (RB), and **(C)** QC + RB when compared to the control medium. The differential significance of the species was calculated by using DESEq2 (**Supplementary Table 1**) from shotgun metagenome sequence analysis at day 21. The left column represents bacterial species-level taxa while the colors indicate their corresponding phyla ("Blue" = *Firmicutes*, "Brown" = *Bacteroidetes*, "Pink" = *Proteobacteria*, "Orange" = *Actinobacteria*, "Green" = *Chlamydiae*).

when compared to control. Interestingly, lysine biosynthesis and ornithine biosynthesis pathways were retained in RB but lost in QC + RB suggesting that addition of QC leads to elimination of lysine, and ornithine biosynthesis pathways completely. Also, reductive pentose phosphate cycle (M00165 and M00166), cysteine biosynthesis (M00609), C10-C20 isoprenoid biosynthesis (M00365) and catechol meta-cleavage (M00569) were more complete in RB and QC + RB (**Supplementary Figure 4**). M00569 pathway is responsible for the convertion of catechol moiety found in quercetin to either acetyl-CoA or 4-methylcatechol which is finally converted to propanoyl-CoA suggesting that QC + RB medium enhances amino acids biosynthesis along with production of propionate.

Quercetin and Rice Bran Combination Yield Higher Propionate Levels and Reduces Members of *Enterobacteriaceae* Family

Along with the changes in the taxonomy of the microbial community, diet alters the metabolic profile of a community (Sonnenburg and Sonnenburg, 2014; Sonnenburg and Backhed, 2016). To understand how these dietary substrates have altered the fermentation potential of the microbiota, we measured short chain fatty acids (SCFAs) from each minibioreactor on days 4, 7, 14, and 21. We observed that the amount of each SCFA produced was stable by day 14 and at the endpoint, there were marked differences between the groups (**Figures 4A–E**). At day 21, communities formed in the control and RB weighted strongly toward acetate and butyrate production, respectively (**Supplementary Figure 5**). Statistically at day 21, acetate production in QC + RB group (48.01 ± 3.3 mM) was significantly different from the RB group (33.33 ± 2.44 mM) (**Figure 4F**). The butyrate production in the RB medium was higher but not significantly different compared to the other groups (**Figure 4H**). Similarly, the medium supplemented with quercetin and rice bran (QC + RB) weighted toward the production of propionate and butyrate (**Supplementary Figure 5**). Rice bran alone and combined quercetin and rice bran supplementation significantly raised propionate levels by at least threefold compared to the QC group alone (**Figure 4G**). In contrast, when QC was supplemented in the medium, it shifted the production of minor SCFAs (**Figures 4I,J** and **Supplementary Figure 5**). However, only isobutyrate production was significantly higher in QC compared to QC + RB group (**Figure 4I**).

Short chain fatty acids production by microbiota has been associated with pathogen inhibition to benefit the host (Hung et al., 2013). Physiological levels of SCFAs are reported to reduce *Enterobacteriaceae* members by pH mediated action (Sorbara et al., 2019). Specifically, propionate production by a propionate producing consortium has been shown to reduce antibiotic induced dysbiosis (El Hage et al., 2019). In this study, as higher production of propionate weighted toward QC + RB medium at day 21, we estimated the correlation of abundances of significantly altering members of *Enterobacteriaceae* to propionate levels. Significant high negative correlations were observed between the abundance of the members of *Enterobacteriaceae* family and propionate with

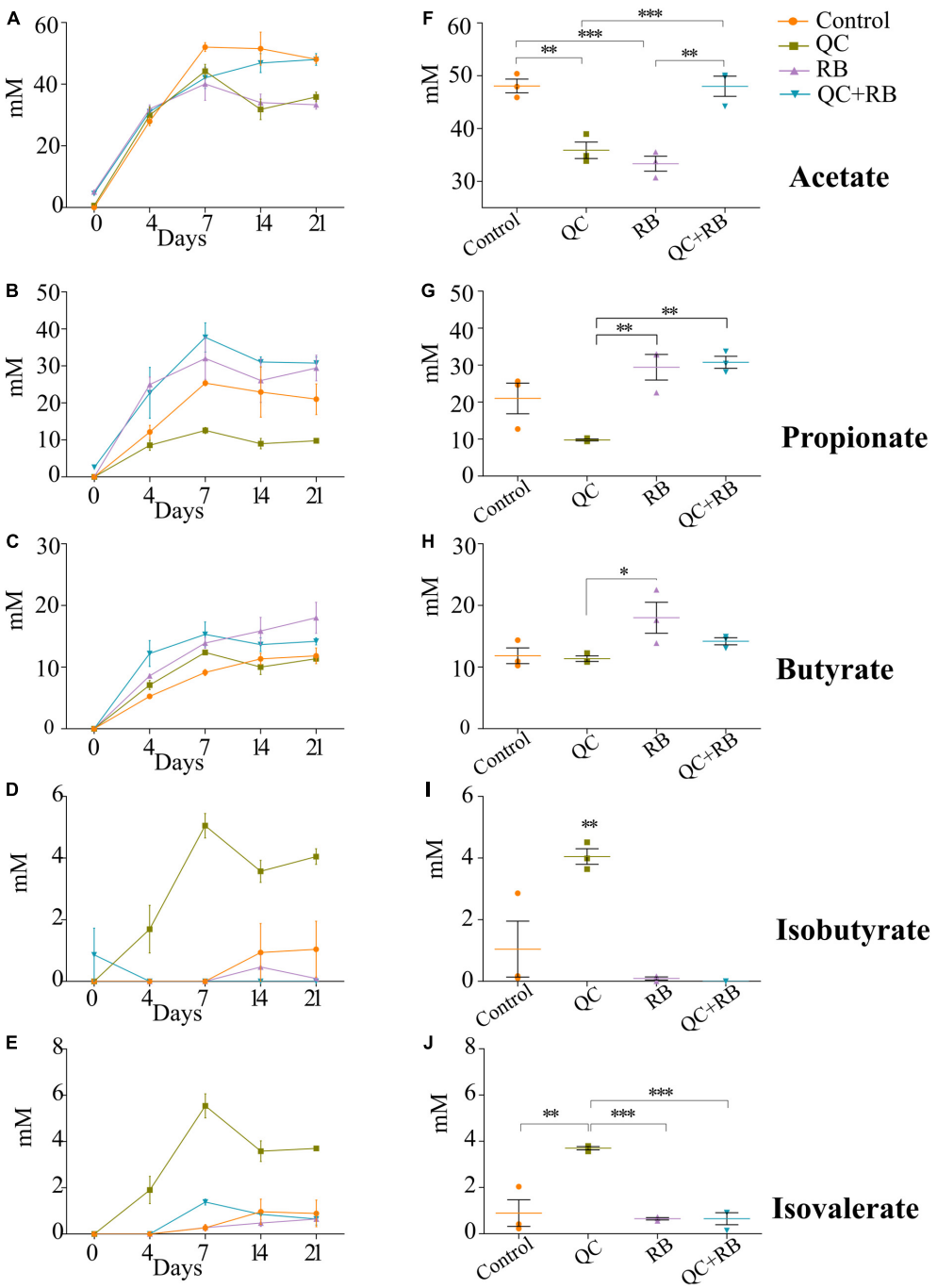


FIGURE 4 | Effect of quercetin (QC) and rice bran (RB) on short Chain Fatty Acids (SCFAs) production in QC, RB, and QC + RB compared to control medium in minibioreactors. **(A–E)** Represent the periodic variation of acetate, propionate, butyrate, isobutyrate and isovalerate at day 0, 4, 7, 14, and 21, respectively. **(F–J)** Represent the comparison of concentrations of acetate, propionate, butyrate, isobutyrate and isovalerate, respectively, from control, QC, RB, and QC + RB medium at day 21 (endpoint). Kruskal–Wallis test was performed between the groups and *post hoc* (Dunn test) analysis was performed to identify the different significant groups. "*, **, and ***" represents significance at $p < 0.05$, $p < 0.01$ and $p < 0.001$ respectively. Error bars represent standard error of the mean of data obtained from three different bioreactors.

media change (**Figure 5**). *Citrobacter rodentium* ($\rho = -0.734$, $p = 0.009$), *C. koseri* ($\rho = -0.769$, $p = 0.0052$), *C. youngae* ($\rho = -0.776$, $p = 0.0046$), *Escherichia albertii* ($\rho = -0.762$, $p = 0.0058$), *E. coli* ($\rho = -0.762$, $p = 0.00587$) and *Escherichia* sp. TW09308 ($\rho = -0.755$, $p = 0.0065$) (**Figures 5A–F**) were greatly reduced in QC + RB medium, whereas propionate levels were at least threefold and 1.5 fold higher compared to QC and control medium.

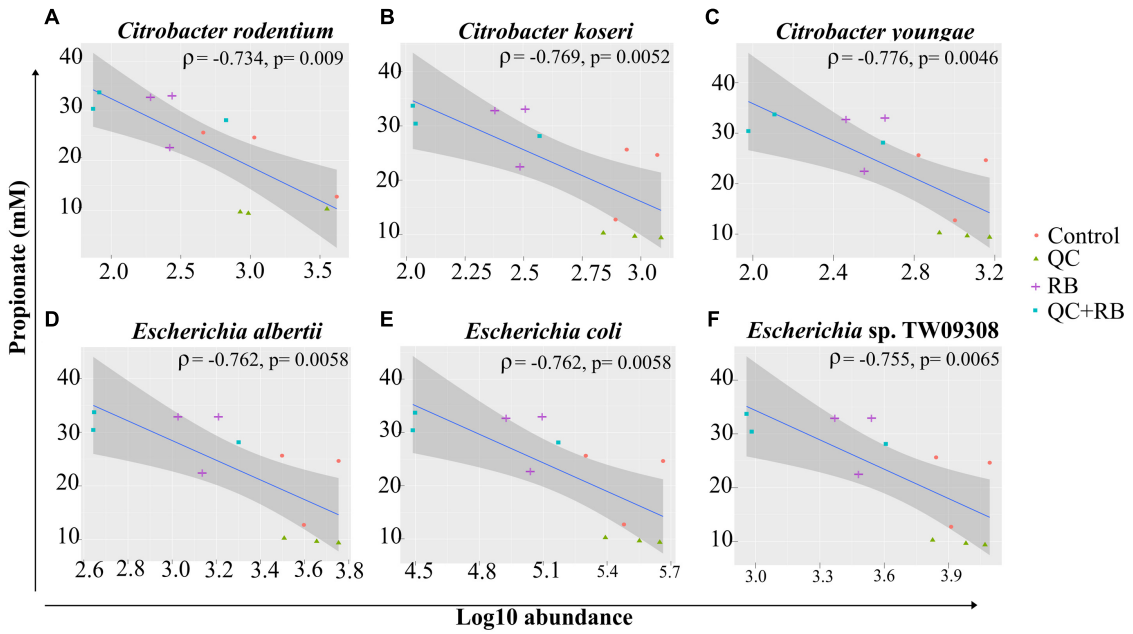


FIGURE 5 | Spearman correlation of log10 abundance of the members of *Enterobacteriaceae* family **(A)** *Citrobacter rodentium*, **(B)** *Citrobacter koseri*, **(C)** *Citrobacter youngae*, **(D)** *Escherichia albertii*, **(E)** *Escherichia coli*, and **(F)** *Escherichia* sp. TW09308 at day 21 determined by shotgun metagenomics sequence analysis to levels of propionate in the medium. A negative correlation is expressed by negative values of correlation coefficient "rho" (ρ) with corresponding *p*-values.

DISCUSSION

In vivo studies have shown changes in microbial communities due to quercetin or rice bran and describe the improvement of colonization resistance and reduction of colon cancer (Kumar et al., 2012; Zhang et al., 2015; Lin et al., 2019). However, these studies are marked by high variations of the microbiota composition likely due to host factors. Additionally, very little is known about the combined effect of quercetin and rice bran on gut microbiota. Thus, we focused on understanding the impact of quercetin and rice bran separately and in combination using a minibioreactor array model (Auchtung et al., 2016). Also, to gain insights into taxonomical composition, we performed 16S rRNA sequencing over time and shotgun metagenome sequencing at the endpoint.

The alpha diversity analysis indicated that the richness of the communities in bioreactors stabilized by day 14 (**Figure 2A**). However, the taxonomical differences were evident as early as day four following inoculation. The changes in the relative abundances of taxa after day 14 were lower in control and quercetin supplemented medium whereas rice bran and combined quercetin with rice bran conditions had a more homogenized microbial composition after day 4 (**Figure 2C**). This indicates that the time required for the stabilization of communities may vary with the substrate used in minibioreactors in contrast to the previously determined timeframe of 1 week following fecal inoculation (Auchtung et al., 2015). Such changes in the time required for stabilization and diversity between the groups can be attributed ecologically to the stochastic rearrangement of the microbial species and their interactions at early stages because of selection pressure by substrate (Vellend, 2010). We also found that members of dominant family such as *Prevotellaceae* in donor microbiota did not retain high abundance in the minibioreactor system. This is probably because the modified BHI medium we used do not contain enough complex carbohydrates to sustain high abundance of species in *Prevotellaceae* family.

Rice bran is a nutrient-dense food with a unique profile and ratio of bioactive phytochemicals such as gamma oryzanol, tocotrienols, ferulic acid, vitamin B, beta-sitosterol and many others (Zarei et al., 2017). It has been reported to be effective in preventing *Salmonella typhimurium* (Kumar et al., 2012; Goodyear et al., 2015; Rubinelli et al., 2017), rotavirus (Yang et al., 2015; Nealon et al., 2017) and norovirus (Lei et al., 2016) infections. However, very few *in vivo* studies have highlighted its effect on gut microbiota and the reports have striking differences. In a clinical trial, Zambrana et al. (2019) showed significant enrichment of *Veillonella*, *Megasphaera* and *Dialister* species at the genus level from gut samples of children from either Nicaragua or Mali at 12 months' time (Zambrana et al., 2019). Another study by Sheflin et al. (2015) showed a significant increase in the abundance of *Methanobrevibacter smithii, Paraprevotella clara, Ruminococcus flavefaciens, Dialister succinatiphilus, Bifidobacterium* sp., *Clostridium glycolicum, Barnesiella intestinihominis, Anaerostipes caccae* and *Ruminococcus bromii* OTUs after heat stabilized rice bran was fed to people (3 g/day) (Sheflin et al., 2015). The differences in the enriched taxa could be because of the unique inoculum used in this study. However, both of the above studies had lower resolution and reported the enrichment

of different taxa which could be attributed to a different variety of rice bran being used and individualized host factors (Stewart et al., 2005; Ridaura et al., 2013; Goodrich et al., 2014). The compounding host factors along with variation in age and geography (Yatsunenko et al., 2012), diet pattern (Wu et al., 2011), lifestyle (Benedict et al., 2016; Cook et al., 2016; Karl et al., 2017), etc., play a crucial role in determining the gut microbiota composition, thus masking the actual effect of the substrate alone on the microbiome. In contrast, this study supplies species-level resolution eliminating host interference and shows *Veillonella Prevotella, Dialister, Bacteroides* and *Alistipes* species are significantly enriched while *Oscillibacter, Eubacterium* and *Citrobacter* species are significantly reduced (**Figure 3**).

Similar to rice bran, previous studies analyzing microbial composition after quercetin supplementation yielded variable results among different hosts. *Enterobacteriaceae* and *Fusobacteriaceae* were reported to be positively related to quercetin supplementation whereas *Sutterellaceae* and *Oscillospiraceae* were found to be negatively correlated (Tamura et al., 2017). However, Lin et al. (2019) reported an increase in abundances of *Bifidobacterium, Bacteroides, Lactobacillus* and *Clostridium*, with a reduction of *Fusobacterium* and *Enterococcus* in mice fed with quercetin (Lin et al., 2019). With higher resolution and removal of host factors, our results contrast with both studies and show enrichment of *Acidaminococcus intestini* and decreased abundances of *E. limosum, E. faecalis, A. caccae, D. longicatena* and *C. catus*. The domination of *Enterobacteriaceae* in medium supplemented with quercetin (**Figure 2C**) might have resulted in lower enrichment of the bacterial taxa as *Enterobacteriaceae* has been reported to affect quercetin metabolism by directly or indirectly inhibiting quercetin degrading bacteria (Tamura et al., 2017).

When the combined effect of quercetin and rice bran were analyzed, we find that majority of the microbial shift is due to the supplementation of rice bran (**Figure 2** and **Supplementary Figures 2**, **3**). Most of the taxa enriched or decreased in the combination were very similar to those affected by rice bran supplementation alone. Compared to quercetin and rice bran supplementation separately, the combination was observed to significantly reduce members of the *Enterobacteriaceae* family (*Escherichia* and *Citrobacter* sp.) (**Supplementary Table 1** and **Figure 5**). *Enterobacteriaceae* consists of class of pathogens that are low in abundance but have potential to grow and dominate during dysbiotic conditions (Seksik et al., 2003; Walker et al., 2011; Taur et al., 2012). The reduction of *Enterobacteriaceae* by combined quercetin and rice bran suggests a possibly beneficial effect on the host.

The reduction of *Enterobacteriaceae* members was highly correlated with greater propionate levels in quercetin and rice bran combined medium (**Figures 4G, 5**). Propionate, along with other SCFAs produced by the gut microbiota, has been previously implicated in regulating the intestinal morphology and functions (Scheppach, 1994). Interestingly, unlike butyrate that is used as an energy source for the colonocytes, the health benefits of propionate are not restricted to the colon. Propionate has been shown to decrease liver lipogenesis, and hepatic and plasma cholesterol levels in rats. It has also been implicated in reducing obesity, by stimulating satiety in mice models. Importantly, propionate has also been demonstrated as a potential anti-inflammatory and anti-cancer agent. The reported anti-inflammatory abilities of propionate are suggested to occur through the inhibition of Nuclear Factor-Kappa B and suppression of IL-6 mRNA and other immune-related gene expression, while its anti-cancer effects are thought to occur through inhibition of Histone Deacetylases (HDACs) and regulation of the AP-1 pathway (Tedelind et al., 2007; Hosseini et al., 2011; Nepelska et al., 2012). Therefore, our observations of increased propionate levels, along with previous reports on the health benefits of propionate, demonstrates the importance of both flavonoids and fiber in the diet. This study, for the first time, reports the combined effect of quercetin and rice bran on the gut microbial composition in the absence of interfering host factors. Here, we report that the gut microbial composition was altered favorably by quercetin and rice bran to result in a significant reduction of opportunistic pathogens that could potentially provide additional health benefits to the host.

DATA AVAILABILITY STATEMENT

The raw sequences are deposited in NCBI under BioProject PRJNA606575.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Institutional Review Board, South Dakota State University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

JS and ER conceived and designed the manuscript. SG, SW, and RS performed the experiments. SG analyzed the data and wrote the first draft with input from all authors. JS, ER, and GB acquired the resources and supervised the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported in part by grants from the South Dakota Governor's Office of Economic Development (SD-GOED) and the by the USDA National Institute of Food and Agriculture projects SD00H702-20 and SD00R646-18 awarded to JS.

ACKNOWLEDGMENTS

Computations supporting this project were performed on high-performance computing systems managed by the Research

Computing Group, part of the Division of Technology and Security at South Dakota State University.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.751225/full#supplementary-material

Supplementary Figure 1 | Exploratory analysis of predicted taxa and alpha diversity of microbiota in various treatments **(A)** Rarefaction curves measuring the bacterial diversity in fecal communities. The curves are based on V3-V4 16S rRNA gene sequences obtained from a total of 48 samples (12 each from control, QC, RB, and QC + RB) from day 4, 7, 14, and 21 and duplicate inoculums **(B)** Alpha diversity measure (Shannon diversity) between Control, supplemented with QC, supplemented with RB and supplemented with QC + RB determined using shotgun metagenome sequence analysis at day 21. Kruskal–Wallis test was performed for the Shannon diversity indices obtained for the groups ($p = 0.27$).

Supplementary Figure 2 | Beta diversity assessment of the microbiota communities formed at days 4, 7, 14, and 21 among four groups of media condition (Control, QC, RB, and QC + RB) determined using 16S rRNA sequence analysis. The communities were ordinated using the "MDS" method followed by Bray-Curtis distance calculation to visualize the difference between the groups. Statistically, PERMANOVA using adonis was calculated on the beta diversity with $p = 0.004$. The upper panel represents the beta diversity in early stages (day 4 and day 7) and the lower panel represents in later stages (day 14 and day 21).

Supplementary Figure 3 | Impact of quercetin (QC) and rice bran (RB) supplementation on bacterial taxonomy. Species-wise variation of the most abundant (top 25) of the bacterial taxa determined by shotgun metagenome sequencing at day 21 (endpoint) for control, QC, RB, and QC + RB conditions in triplicate.

Supplementary Figure 4 | Heatmap depicting the functional pathways in the metagenomes at day 21 obtained from shotgun metagenomic sequencing. KEGG modules obtained from GHOSTKOALA for each metagenomic sequence was plotted as either complete, 1 block missing, 2 blocks missing or incomplete.

Supplementary Figure 5 | Biplot showing ordination of short-chain fatty acid production and bacterial communities formed at day 21 in control, QC, RB, and QC + RB conditions. Abundances of the bacterial communities at day 21 determined by shotgun metagenomics were used with SCFAs profiles of day 21 to generate the plot.

Supplementary Table 1 | List of all differentially abundant taxa in medium with quercetin (QC), medium with rice bran (RB) and medium with both quercetin and rice bran (QC + RB) compared to control. DESEq2 was used in shotgun metagenome sequences on day 21 for identification of differentially present taxa in the quercetin and rice bran supplemented media conditions when compared to control.

REFERENCES

Auchtung, J. M., Robinson, C. D., and Britton, R. A. (2015). Cultivation of stable, reproducible microbial communities from different fecal donors using minibioreactor arrays (MBRAs). *Microbiome* 3:42.

Auchtung, J. M., Robinson, C. D., Farrell, K., and Britton, R. A. (2016). MiniBioReactor Arrays (MBRAs) as a Tool for Studying C. difficile Physiology in the Presence of a Complex Community. *Methods Mol. Biol.* 1476, 235–258. doi: 10.1007/978-1-4939-6361-4_18

Benedict, C., Vogel, H., Jonas, W., Woting, A., Blaut, M., Schurmann, A., et al. (2016). Gut microbiota and glucometabolic alterations in response to recurrent partial sleep deprivation in normal-weight young individuals. *Mol. Metab.* 5, 1175–1186. doi: 10.1016/j.molmet.2016.10.003

Bokulich, N. A., Kaehler, B. D., Rideout, J. R., Dillon, M., Bolyen, E., Knight, R., et al. (2018). Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. *Microbiome* 6:90.

Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., et al. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat. Biotechnol.* 37, 852–857.

Braune, A., and Blaut, M. (2016). Bacterial species involved in the conversion of dietary flavonoids in the human gut. *Gut Microbes* 7, 216–234. doi: 10.1080/19490976.2016.1158395

Callahan, B. J., Mcmurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J., and Holmes, S. P. (2016). DADA2: high-resolution sample inference from Illumina amplicon data. *Nat. Methods* 13, 581–583. doi: 10.1038/nmeth.3869

Chung, W. S. F., Walker, A. W., Vermeiren, J., Sheridan, P. O., Bosscher, D., Garcia-Campayo, V., et al. (2019). Impact of carbohydrate substrate complexity on the diversity of the human colonic microbiota. *FEMS Microbiol. Ecol.* 95:fiy201.

Comalada, M., Camuesco, D., Sierra, S., Ballester, I., Xaus, J., Galvez, J., et al. (2005). In vivo quercitrin anti-inflammatory effect involves release of quercetin, which inhibits inflammation through down-regulation of the NF-kappaB pathway. *Eur. J. Immunol.* 35, 584–592. doi: 10.1002/eji.200425778

Cook, M. D., Allen, J. M., Pence, B. D., Wallig, M. A., Gaskins, H. R., White, B. A., et al. (2016). Exercise and gut immune function: evidence of alterations in colon immune cell homeostasis and microbiome characteristics with exercise training. *Immunol. Cell Biol.* 94, 158–163. doi: 10.1038/icb.2015.108

Darband, S. G., Kaviani, M., Yousefi, B., Sadighparvar, S., Pakdel, F. G., Attari, J. A., et al. (2018). Quercetin: a functional dietary flavonoid with potential chemo-preventive properties in colorectal cancer. *J. Cell Physiol.* 233, 6544–6560. doi: 10.1002/jcp.26595

D'Archivio, M., Filesi, C., Di Benedetto, R., Gargiulo, R., Giovannini, C., and Masella, R. (2007). Polyphenols, dietary sources and bioavailability. *Ann. Ist. Super. Sanita* 43, 348–361.

David, L. A., Maurice, C. F., Carmody, R. N., Gootenberg, D. B., Button, J. E., Wolfe, B. E., et al. (2014). Diet rapidly and reproducibly alters the human gut microbiome. *Nature* 505, 559–563. doi: 10.1038/nature12820

Dhakan, D. B., Maji, A., Sharma, A. K., Saxena, R., Pulikkan, J., Grace, T., et al. (2019). The unique composition of Indian gut microbiome, gene catalogue, and associated fecal metabolome deciphered using multi-omics approaches. *Gigascience* 8:giz004.

Egert, S., Bosy-Westphal, A., Seiberl, J., Kurbitz, C., Settler, U., Plachta-Danielzik, S., et al. (2009). Quercetin reduces systolic blood pressure and plasma oxidised low-density lipoprotein concentrations in overweight subjects with a high-cardiovascular disease risk phenotype: a double-blinded, placebo-controlled cross-over study. *Br. J. Nutr.* 102, 1065–1074. doi: 10.1017/s0007114509359127

El Hage, R., Hernandez-Sanabria, E., Calatayud Arroyo, M., Props, R., and Van De Wiele, T. (2019). Propionate-Producing Consortium Restores Antibiotic-Induced Dysbiosis in a Dynamic in vitro Model of the Human Intestinal Microbial Ecosystem. *Front. Microbiol.* 10:1206. doi: 10.3389/fmicb.2019.01206

Gentile, C. L., and Weir, T. L. (2018). The gut microbiota at the intersection of diet and human health. *Science* 362, 776–780. doi: 10.1126/science.aau5812

Ghimire, S., Roy, C., Wongkuna, S., Antony, L., Maji, A., Keena, M. C., et al. (2020). Identification of Clostridioides difficile-Inhibiting Gut Commensals Using Culturomics, Phenotyping, and Combinatorial Community Assembly. *mSystems* 5, e00620–19.

Gibson, G. R., and Roberfroid, M. B. (1995). Dietary modulation of the human colonic microbiota: introducing the concept of prebiotics. *J. Nutr.* 125, 1401–1412. doi: 10.1093/jn/125.6.1401

Goodrich, J. K., Waters, J. L., Poole, A. C., Sutter, J. L., Koren, O., Blekhman, R., et al. (2014). Human genetics shape the gut microbiome. *Cell* 159, 789–799.

Goodyear, A., Kumar, A., Ehrhart, E. J., Swanson, K. S., Grusak, M. A., Leach, J. E., et al. (2015). Dietary rice bran supplementation prevents *Salmonella* colonization differentially across varieties and by priming intestinal immunity. *J. Funct. Foods* 18, 653–664. doi: 10.1016/j.jff.2015.08.027

Gupta, A., Dhakan, D. B., Maji, A., Saxena, R., P.K, V. P., Mahajan, S., et al. (2019). Association of Flavonifractor plautii, a Flavonoid-Degrading Bacterium, with the Gut Microbiome of Colorectal Cancer Patients in India. *mSystems* 4, e00438–19.

Harwood, M., Danielewska-Nikiel, B., Borzelleca, J. F., Flamm, G. W., Williams, G. M., and Lines, T. C. (2007). A critical review of the data related to the

safety of quercetin and lack of evidence of in vivo toxicity, including lack of genotoxic/carcinogenic properties. *Food Chem Toxicol.* 45, 2179–2205. doi: 10.1016/j.fct.2007.05.015

Henderson, A. J., Kumar, A., Barnett, B., Dow, S. W., and Ryan, E. P. (2012). Consumption of rice bran increases mucosal immunoglobulin A concentrations and numbers of intestinal Lactobacillus spp. *J. Med. Food* 15, 469–475. doi: 10.1089/jmf.2011.0213

Hong, Z., and Piao, M. (2018). Effect of Quercetin Monoglycosides on Oxidative Stress and Gut Microbiota Diversity in Mice with Dextran Sodium Sulphate-Induced Colitis. *Biomed. Res. Int.* 2018:8343052.

Hosseini, E., Grootaert, C., Verstraete, W., and Van De Wiele, T. (2011). Propionate as a health-promoting microbial metabolite in the human gut. *Nutr. Rev.* 69, 245–258. doi: 10.1111/j.1753-4887.2011.00388.x

Hung, C. C., Garner, C. D., Slauch, J. M., Dwyer, Z. W., Lawhon, S. D., Frye, J. G., et al. (2013). The intestinal fatty acid propionate inhibits *Salmonella* invasion through the post-translational control of HilD. *Mol. Microbiol.* 87, 1045–1060. doi: 10.1111/mmi.12149

Islam, J., Koseki, T., Watanabe, K., Ardiansyah, Budijanto, S., Oikawa, A., et al. (2017). Dietary Supplementation of Fermented Rice Bran Effectively Alleviates Dextran Sodium Sulfate-Induced Colitis in Mice. *Nutrients* 9:747. doi: 10.3390/nu9070747

Kanehisa, M., Sato, Y., and Morishima, K. (2016). BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. *J. Mol. Biol.* 428, 726–731. doi: 10.1016/j.jmb.2015.11.006

Karl, J. P., Margolis, L. M., Madslien, E. H., Murphy, N. E., Castellani, J. W., Gundersen, Y., et al. (2017). Changes in intestinal microbiota composition and metabolism coincide with increased intestinal permeability in young adults under prolonged physiological stress. *Am. J. Physiol. Gastrointest. Liver Physiol.* 312, G559–G571.

Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 30, 3059–3066. doi: 10.1093/nar/gkf436

Kumar, A., Henderson, A., Forster, G. M., Goodyear, A. W., Weir, T. L., Leach, J. E., et al. (2012). Dietary rice bran promotes resistance to *Salmonella enterica* serovar Typhimurium colonization in mice. *BMC Microbiol.* 12:71. doi: 10.1186/1471-2180-12-71

Lei, S., Ramesh, A., Twitchell, E., Wen, K., Bui, T., Weiss, M., et al. (2016). High Protective Efficacy of Probiotics and Rice Bran against Human Norovirus Infection and Diarrhea in Gnotobiotic Pigs. *Front. Microbiol.* 7:1699. doi: 10.3389/fmicb.2016.01699

Lin, R., Piao, M., and Song, Y. (2019). Dietary Quercetin Increases Colonic Microbial Diversity and Attenuates Colitis Severity in Citrobacter rodentium-Infected Mice. *Front. Microbiol.* 10:1092. doi: 10.3389/fmicb.2019.01092

Linsalata, M., Orlando, A., Messa, C., Refolo, M. G., and Russo, F. (2010). Quercetin inhibits human DLD-1 colon cancer cell growth and polyamine biosynthesis. *Anticancer Res.* 30, 3501–3507.

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15:550.

Maji, A., Misra, R., Dhakan, D. B., Gupta, V., Mahato, N. K., Saxena, R., et al. (2018). Gut microbiome contributes to impairment of immunity in pulmonary tuberculosis patients by alteration of butyrate and propionate producers. *Environ. Microbiol.* 20, 402–419. doi: 10.1111/1462-2920.14015

McDonald, D., Price, M. N., Goodrich, J., Nawrocki, E. P., Desantis, T. Z., Probst, A., et al. (2012). An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *ISME J.* 6, 610–618. doi: 10.1038/ismej.2011.139

McMurdie, P. J., and Holmes, S. (2013). phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* 8:e61217. doi: 10.1371/journal.pone.0061217

Mende, D. R., Letunic, I., Huerta-Cepas, J., Li, S. S., Forslund, K., Sunagawa, S., et al. (2017). proGenomes: a resource for consistent functional and taxonomic annotations of prokaryotic genomes. *Nucleic Acids Res.* 45, D529–D534.

Menzel, P., Ng, K. L., and Krogh, A. (2016). Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nat. Commun.* 7:11257.

National Academies of Science, Engineering, and Medicine (2017). *Environmental Chemicals, the Human Microbiome, and Health Risk: A Research Strategy*. Washington: National Academies Press.

Nealon, N. J., Yuan, L., Yang, X., and Ryan, E. P. (2017). Rice Bran and Probiotics Alter the Porcine Large Intestine and Serum Metabolomes for Protection against Human Rotavirus Diarrhea. *Front. Microbiol.* 8:653. doi: 10.3389/fmicb.2017.00653

Nepelska, M., Cultrone, A., Beguet-Crespel, F., Le Roux, K., Dore, J., Arulampalam, V., et al. (2012). Butyrate produced by commensal bacteria potentiates phorbol esters induced AP-1 response in human intestinal epithelial cells. *PLoS One* 7:e52869. doi: 10.1371/journal.pone.0052869

Nurk, S., Meleshko, D., Korobeynikov, A., and Pevzner, P. A. (2017). metaSPAdes: a new versatile metagenomic assembler. *Genome Res.* 27, 824–834. doi: 10.1101/gr.213959.116

Price, M. N., Dehal, P. S., and Arkin, A. P. (2010). FastTree 2–approximately maximum-likelihood trees for large alignments. *PLoS One* 5:e9490. doi: 10.1371/journal.pone.0009490

Pulikkan, J., Maji, A., Dhakan, D. B., Saxena, R., Mohan, B., Anto, M. M., et al. (2018). Gut Microbial Dysbiosis in Indian Children with Autism Spectrum Disorders. *Microb. Ecol.* 76, 1102–1114. doi: 10.1007/s00248-018-1176-2

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. Austria: R Foundation for Statistical Computing.

Ridaura, V. K., Faith, J. J., Rey, F. E., Cheng, J., Duncan, A. E., Kau, A. L., et al. (2013). Gut microbiota from twins discordant for obesity modulate metabolism in mice. *Science* 341:1241214.

Robertson, C. E., Harris, J. K., Wagner, B. D., Granger, D., Browne, K., Tatem, B., et al. (2013). Explicet: graphical user interface software for metadata-driven management, analysis and visualization of microbiome data. *Bioinformatics* 29, 3100–3101. doi: 10.1093/bioinformatics/btt526

Rubinelli, P. M., Kim, S. A., Park, S. H., Roto, S. M., Nealon, N. J., Ryan, E. P., et al. (2017). Differential effects of rice bran cultivars to limit *Salmonella* Typhimurium in chicken cecal in vitro incubations and impact on the cecal microbiome and metabolome. *PLoS One* 12:e0185002. doi: 10.1371/journal.pone.0185002

Scheppach, W. (1994). Effects of short chain fatty acids on gut morphology and function. *Gut* 35, S35–S38.

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069. doi: 10.1093/bioinformatics/btu153

Seksik, P., Rigottier-Gois, L., Gramet, G., Sutren, M., Pochart, P., Marteau, P., et al. (2003). Alterations of the dominant faecal bacterial groups in patients with Crohn's disease of the colon. *Gut* 52, 237–242. doi: 10.1136/gut.52.2.237

Sheflin, A. M., Borresen, E. C., Wdowik, M. J., Rao, S., Brown, R. J., Heuberger, A. L., et al. (2015). Pilot dietary intervention with heat-stabilized rice bran modulates stool microbiota and metabolites in healthy adults. *Nutrients* 7, 1282–1300. doi: 10.3390/nu7021282

Slavin, J. (2013). Fiber and prebiotics: mechanisms and health benefits. *Nutrients* 5, 1417–1435. doi: 10.3390/nu5041417

Sonnenburg, E. D., and Sonnenburg, J. L. (2014). Starving our microbial self: the deleterious consequences of a diet deficient in microbiota-accessible carbohydrates. *Cell Metab.* 20, 779–786. doi: 10.1016/j.cmet.2014.07.003

Sonnenburg, J. L., and Backhed, F. (2016). Diet-microbiota interactions as moderators of human metabolism. *Nature* 535, 56–64. doi: 10.1038/nature18846

Sorbara, M. T., Dubin, K., Littmann, E. R., Moody, T. U., Fontana, E., Seok, R., et al. (2019). Inhibiting antibiotic-resistant *Enterobacteriaceae* by microbiota-mediated intracellular acidification. *J. Exp. Med.* 216, 84–98. doi: 10.1084/jem.20181639

Stewart, J. A., Chadwick, V. S., and Murray, A. (2005). Investigations into the influence of host genetics on the predominant eubacteria in the faecal microflora of children. *J. Med. Microbiol.* 54, 1239–1242. doi: 10.1099/jmm.0.46189-0

Stewart, L. K., Soileau, J. L., Ribnicky, D., Wang, Z. Q., Raskin, I., Poulev, A., et al. (2008). Quercetin transiently increases energy expenditure but persistently decreases circulating markers of inflammation in C57BL/6J mice fed a high-fat diet. *Metabolism* 57, S39–S46.

Tamura, M., Hoshi, C., Kobori, M., Takahashi, S., Tomita, J., Nishimura, M., et al. (2017). Quercetin metabolism by fecal microbiota from healthy elderly human subjects. *PLoS One* 12:e0188271. doi: 10.1371/journal.pone.0188271

Taur, Y., Xavier, J. B., Lipuma, L., Ubeda, C., Goldberg, J., Gobourne, A., et al. (2012). Intestinal domination and the risk of bacteremia in patients undergoing

allogeneic hematopoietic stem cell transplantation. *Clin. Infect. Dis.* 55, 905–914. doi: 10.1093/cid/cis580

Tedelind, S., Westberg, F., Kjerrulf, M., and Vidal, A. (2007). Anti-inflammatory properties of the short-chain fatty acids acetate and propionate: a study with relevance to inflammatory bowel disease. *World J. Gastroenterol.* 13, 2826–2832.

Turnbaugh, P. J., Ridaura, V. K., Faith, J. J., Rey, F. E., Knight, R., and Gordon, J. I. (2009). The effect of diet on the human gut microbiome: a metagenomic analysis in humanized gnotobiotic mice. *Sci. Transl. Med.* 1:6ra14. doi: 10.1126/scitranslmed.3000322

Uritskiy, G. V., Diruggiero, J., and Taylor, J. (2018). MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis. *Microbiome* 6:158.

Vellend, M. (2010). Conceptual synthesis in community ecology. *Q. Rev. Biol.* 85, 183–206. doi: 10.1086/652373

Walker, A. W., Sanderson, J. D., Churcher, C., Parkes, G. C., Hudspith, B. N., Rayment, N., et al. (2011). High-throughput clone library analysis of the mucosa-associated microbiota reveals dysbiosis and differences between inflamed and non-inflamed regions of the intestine in inflammatory bowel disease. *BMC Microbiol.* 11:7. doi: 10.1186/1471-2180-11-7

Wu, G. D., Chen, J., Hoffmann, C., Bittinger, K., Chen, Y. Y., Keilbaugh, S. A., et al. (2011). Linking long-term dietary patterns with gut microbial enterotypes. *Science* 334, 105–108. doi: 10.1126/science.1208344

Yang, X., Twitchell, E., Li, G., Wen, K., Weiss, M., Kocher, J., et al. (2015). High protective efficacy of rice bran against human rotavirus diarrhea via enhancing probiotic growth, gut barrier function, and innate immunity. *Sci. Rep.* 5:15004.

Yatsunenko, T., Rey, F. E., Manary, M. J., Trehan, I., Dominguez-Bello, M. G., Contreras, M., et al. (2012). Human gut microbiome viewed across age and geography. *Nature* 486, 222–227. doi: 10.1038/nature11053

Zambrana, L. E., Mckeen, S., Ibrahim, H., Zarei, I., Borresen, E. C., Doumbia, L., et al. (2019). Rice bran supplementation modulates growth, microbiota and metabolome in weaning infants: a clinical trial in Nicaragua and Mali. *Sci. Rep.* 9:13919.

Zarei, I., Brown, D. G., Nealon, N. J., and Ryan, E. P. (2017). Rice Bran Metabolome Contains Amino Acids, Vitamins & Cofactors, and Phytochemicals with Medicinal and Nutritional Properties. *Rice* 10:24.

Zhang, X. A., Zhang, S., Yin, Q., and Zhang, J. (2015). Quercetin induces human colon cancer cells apoptosis by inhibiting the nuclear factor-kappa B Pathway. *Pharmacogn. Mag.* 11, 404–409. doi: 10.4103/0973-1296.153096

Zoetendal, E. G., Akkermans, A. D., and De Vos, W. M. (1998). Temperature gradient gel electrophoresis analysis of 16S rRNA from human fecal samples reveals stable and host-specific communities of active bacteria. *Appl. Environ. Microbiol.* 64, 3854–3859. doi: 10.1128/aem.64.10.3854-3859.1998

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

ORIGINAL RESEARCH
published: 26 November 2021
doi: 10.3389/fmicb.2021.770109

Rational Proteomic Analysis of a New Domesticated *Klebsiella pneumoniae* x546 Producing 1,3-Propanediol

*Xin Wang[1,2,3,4], Lin Zhang[5], Hong Chen[1,2], Pan Wang[1,2], Ying Yin[1,2], Jiaqi Jin[1,2], Jianwei Xu[3,4] and Jianping Wen[1,2]**

[1] Key Laboratory of Systems Bioengineering (Ministry of Education), Tianjin University, Tianjin, China, [2] SynBio Research Platform, Collaborative Innovation Center of Chemical Science and Engineering (Tianjin), School of Chemical Engineering and Technology, Tianjin University, Tianjin, China, [3] Department of Chemistry, National University of Singapore, Singapore, Singapore, [4] Institute of Materials Research and Engineering, Singapore, Singapore, [5] Dalian Petrochemical Research Institute of Sinopec, Dalian, China

OPEN ACCESS

Edited by:
Roshan Kumar,
Magadh University, India

Reviewed by:
Xixian Xie,
Tianjin University of Science & Technology, China
Quan Luo,
Qingdao Institute of Bioenergy and Bioprocess Technology, Chinese Academy of Sciences (CAS), China

***Correspondence:**
Jianping Wen
jpwen@tju.edu.cn

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

Received: 07 September 2021
Accepted: 26 October 2021
Published: 26 November 2021

Citation:
Wang X, Zhang L, Chen H, Wang P, Yin Y, Jin J, Xu J and Wen J (2021) Rational Proteomic Analysis of a New Domesticated Klebsiella pneumoniae x546 Producing 1,3-Propanediol. Front. Microbiol. 12:770109. doi: 10.3389/fmicb.2021.770109

In order to improve the capability of *Klebsiella pneumoniae* to produce an important chemical raw material, 1,3-propanediol (1,3-PDO), a new type of *K. pneumoniae* x546 was obtained by glycerol acclimation and subsequently was used to produce 1,3-PDO. Under the control of pH value using Na^+ pH neutralizer, the 1,3-PDO yield of *K. pneumoniae* x546 in a 7.5-L fermenter was 69.35 g/L, which was 1.5-fold higher than the original strain (45.91 g/L). After the addition of betaine, the yield of 1,3-PDO reached up to 74.44 g/L at 24 h, which was 40% shorter than the original fermentation time of 40 h. To study the potential mechanism of the production improvement of 1,3-PDO, the Tandem Mass Tags (TMT) technology was applied to investigate the production of 1,3-PDO in *K. pneumoniae*. Compared with the control group, 170 up-regulated proteins and 291 down-regulated proteins were identified. Through Gene Ontology and Kyoto Encyclopedia of Genes and Genomes pathway analysis, it was found that some proteins [such as homoserine kinase (ThrB), phosphoribosylglycinamide formyltransferase (PurT), phosphoribosylaminoimidazolesuccinocarboxamide synthase (PurC), etc.] were involved in the fermentation process, whereas some other proteins (such as ProX, ProW, ProV, etc.) played a significant role after the addition of betaine. Moreover, combined with the metabolic network of *K. pneumoniae* during 1,3-PDO, the proteins in the biosynthesis of 1,3-PDO [such as DhaD, DhaK, lactate dehydrogenase (LDH), BudC, etc.] were analyzed. The process of 1,3-PDO production in *K. pneumoniae* was explained from the perspective of proteome for the first time, which provided a theoretical basis for genetic engineering modification to improve the yield of 1,3-PDO. Because of the use of Na^+ pH neutralizer in the fermentation, the subsequent environmental pollution treatment cost was greatly reduced, showing high potential for industry application in the future.

Keywords: *Klebsiella pneumoniae*, 1,3-propanediol production, betaine, Na^+ pH neutralizer, proteomics

INTRODUCTION

The rise of the biodiesel industry leads to the overproduction of glycerol as a by-product, which now threatens the economic feasibility of the industry (Pan et al., 2019; Kim et al., 2020). This situation has prompted scientists to explore the utilization of glycerol as a carbon source to produce 1,3-propanediol (1,3-PDO), which is a precursor of some important commercial polymers such as polyester and polyurethane (Zhou et al., 2019b; Bao et al., 2020; Chen et al., 2020). 1,3-PDO can be produced by chemical synthesis or biosynthesis using *Klebsiella pneumoniae*. Because of its relatively high yield and low environmental pollution, *K. pneumoniae* is preferable to be used in 1,3-PDO production (Li et al., 2019; Mitrea and Vodnar, 2019; Zabed et al., 2019). Researchers have adopted several strategies such as the domestication method, genetic modification, medium optimization, and other methods to significantly increase the output of 1,3-PDO (Zhang et al., 2018; Lee et al., 2019; Ma et al., 2019; Zhou et al., 2019a). However, the industrial-scale production of 1,3-PDO using bacteria is still limited by low efficiency, which seriously hinders the competitiveness of the process (Dexter Tam et al., 2019; Guo et al., 2019; Park et al., 2019).

The proteomic analysis of protein expression patterns under experimental conditions can provide sufficient information on the function and the regulation of metabolic networks, which is important in the reasonable and purposeful exploration of genome and proteome datasets for the pathway analysis of actual biological processes in post–genome research. The Tandem Mass Tags (TMT) technology was one of the most powerful analytical methods with the highest flux, the smallest systematic error, and the most powerful function (Sogame et al., 2014). It could provide more accurate digital signals, higher detection fluxes, and wider detection ranges. A more detailed understanding of the metabolic pathway of *K. pneumoniae* and other species could help to provide a better way to promote the transformation of glycerol into 1,3-PDO in this system. Therefore, it is necessary to apply the TMT technology to the study of 1,3-PDO production by *K. pneumoniae*.

Ca^{2+} salt as a divalent cation can reduce the drastic changes in the activities of various intracellular dehydrogenases in the oxidation pathway, adjust and maintain the intracellular redox pressure, shift the metabolic flow to 1,3-PDO synthesis, and reduce the types of by-products caused by metabolic disorders. Therefore, Ca^{2+} salts have been commonly used as a pH neutralizer in industry (Nakano et al., 2012; Zhang et al., 2016; Tee et al., 2017). However, various Ca^{2+} salt precipitates were formed when using Ca^{2+} neutralizer at the bottom of the fermentation tank, which not only increases the cost of the subsequent product purification, but also causes significant environmental pollution. Considering the environmental pressure caused by the utilization of Ca^{2+} neutralizer, it is critical that Ca^{2+} pH neutralizer be replaced with a new neutralizer without compromising with production efficiency, so as to make the 1,3-PDO bioproduction more environment-friendly. Na^{+} pH neutralizer can reduce the solid pollutants produced after fermentation reaction. As Na_2CO_3 can be synthesized by a chemical method from the electrolysis of high salt wastewater containing Na^{+} and re-extracted for reuse (Shin et al., 2011; Simon et al., 2014), the use of Na_2CO_3 as a pH adjuster in fermentation would promote a new industrial recycling. However, the use of Na^{+} in fermentation leads to an increase in osmotic pressure, thereby restricting the yield of 1,3-PDO (Glaasker et al., 1998; Guerzoni et al., 2001), whereas betaine can slow down the effect of salt stress (Hussain et al., 2020). It can maintain the balance of osmotic pressure inside and outside, thereby maintaining the normal physiological function of the cell (Louesdon et al., 2014). Moreover, the betaine may have an effect on fermentation under Na^{+} conduction.

In this study, Na_2CO_3 was used as the pH neutralizer in fermentation, and betaine was added to alleviate the high osmotic pressure caused by excessive Na^{+}, which would significantly enhance the yield of 1,3-PDO. The 1,3-PDO production further increased to 74.44 g/L and shortened the fermentation time from 40 to 24 h. TMT was used to study the mechanism effects of the introduction of the Na^{+} neutralizer and betaine on the yield of 1,3-PDO during the fermentation. This is the first comprehensive investigation of TMT analysis for the production of 1,3-PDO by *K. pneumoniae* x546, and the results will provide new insights on enhancing the production of 1,3-PDO (genes, proteins, and metabolites), as well as the subsequent industrial strain transformation and process optimization.

MATERIALS AND METHODS

Strains, Media, and Cultivations

Klebsiella pneumoniae American Type Culture Collection (ATCC) 15380 was purchased from the ATCC. Following the previously published adaptive laboratory evolution (Willke and Vorlop, 2008; Gungormusler et al., 2011; Raghunandan et al., 2014), *K. pneumoniae* x546 (domesticated strain with 120–20 g/L glycerol: the strain was first domesticated with a concentration of 120 g/L glycerol and then returned to a concentration of 20 g/L glycerol for domestication) could be obtained (the details could be seen in **Supplementary File 1**). The seed and solid medium (pH 7.0) contained 40 g/L (60, 80, 100, 120, 140, 100–20, 120–20, and 140–20 g/L) glycerol, 4.08 g/L NH_4Cl, 0.57 g/L KCl, 0.95 g/L $NaH_2PO_4{\cdot}2H_2O$, 0.28 g/L Na_2SO_4, 0.25 g/L $MgCl_2{\cdot}6H_2O$, 0.38 g/L citric acid, 0.95 g/L yeast extract, 0.15 g/L Vc, and 4 mL of nutrient solution. Nutrient solution contained 0.035 g/L Na_2MoO_4, 0.029 g/L $ZnCl_2$, 0.29 g/L $CoCl_2{\cdot}6H_2O$, 0.148 g/L $MgSO_4{\cdot}7H_2O$, 0.033 g/L $NiCl_2{\cdot}6H_2O$, and 1.0 mL HCl. The *K. pneumoniae* was domesticated with 40 g/L (60, 80, 100, 120, 140, 100–20, 120–20, and 140–20 g/L), and glycerol was labeled G40, G60 G80, G120, G140, G100–20, G120–20, and G140–20 (Liang et al., 2018).

The production medium was a little different from the seed medium, which contained 40 g/L glycerol, 6.17 g/L NH_4Cl, 0.86 g/L KCl, 1.40 g/L $NaH_2PO_4{\cdot}2H_2O$, 0.32 g/L Na_2SO_4, 0.3 g/L $MgCl_2{\cdot}6H_2O$, 1.06 g/L citric acid, 1.15 g/L yeast extract, 0.25 g/L betaine, 0.11 g/L Vc, 0.23 g/L $C_5H_{11}NO_2$, and 5 mL of nutrient solution. Nutrient solution contained 5.4 g/L $FeCl_3{\cdot}6H_2O$, 0.004 g/L Na_2MoO_4, 0.04 g/L $ZnCl_2$, 0.17 g/L $MnCl_2{\cdot}4H_2O$,

0.47 g/L $CoCl_2 \cdot 6H_2O$, 0.06 g/L H_3BO_4, 0.68 g/L $CuSO_4 \cdot 5H_2O$, and 1.0 mL HCl. The betaine was added only in the production medium for 7.5 L fermenter.

The seed was cultivated in 250-mL flask containing a 100-mL seed medium at 150 revolutions/min (rpm) for 8.5 h at 37°C. The production of 1,3-PDO was carried out in a 250-mL flask with 100 mL working volume at 150 rpm for 48 h at 37°C and in a 7.5-L BioFlo 110 fermenter (New Brunswick Scientific, Edison, NJ, United States) at 400 rpm for 40 h at 37°C after adding 770 g glycerol (with a final 5.4 L working volume). Three biological replicates were used for each fermentation experiment. The pH of the seed medium and fermentation medium was adjusted to 7.0 with 3.125 M Na_2CO_3 solution, respectively.

Determination of 1,3-Propanediol and Glycerol Concentrations

The concentration of 1,3-PDO and glycerol was measured by an HPX-87H column (300 mm × 7.8 mm) (Bio-Rad, Palo Alto, CA, United States) with a differential refractive index detector (SFD GmbH, Schambeck, Germany); 5 mM H_2SO_4 was used as a mobile phase with a flow rate of 0.5 mL/min at a working temperature of 65°C.

Total Protein Extraction

The samples were ground into a powder in liquid nitrogen. Then the powder was suspended in lysis buffer (1% sodium deoxycholate, 8 M urea). The mixture was allowed to settle at 4°C for 30 min during which the sample was vortexed every 5 min and treated by ultrasound at 40 kHz and 40 W for 2 min. After centrifugation at 16,000 rpm at 4°C for 30 min, the concentration of protein supernatant was determined by bicinchoninic acid (BCA) method by BCA Protein Assay Kit (Pierce, Thermo, United States). Protein quantification was performed according to the kit protocol (Chen et al., 2021).

Protein Digestion and Tandem Mass Tags Labeling

Protein digestion was performed according to a standardized procedure, and the resulting peptide mixture was labeled using 10-plex TMT reagent (Thermo Fisher, Scientific). In brief, an aliquot of protein (100 μg) from each sample was mixed with 100 μL of the lysate. Then 10 mM TCEP was added, and the mixture was stored at 37°C for 60 min, followed by the addition of 40 mM iodoacetamide and the storage of the sample in the dark at room temperature for 40 min.

Sixfold volumes of cold acetone were added to precipitate the protein at −20°C for 4 h. After centrifuging at 10,000 rpm for 20 min at 4°C, the pellet was resuspended with 100 μL of 50 mM triethylammonium bicarbonate buffer. Trypsin was added at a trypsin-to-protein mass ratio of 1:50 and incubated at 37°C overnight. One unit of TMT reagent was thawed and reconstituted in 50 μL acetonitrile. After tagging for 2 h at room temperature, hydroxylamine was added to react with mixture for 15 min at room temperature (Dai et al., 2020).

In this work, the strains grown under control (fermentation of the strain domesticated at a glycerol concentration of 40 g/L without betaine) and optimal conditions (fermentation of the strain domesticated at a glycerol concentration of 120–20 g/L with betaine) at 10 h (each sample with two biological replicates) were collected by centrifugation (12,000 rpm, 10 min at 4°C) and frozen in liquid nitrogen, respectively. The samples were labeled as A1, A2, B1, and B2. Finally, all samples were pooled, desalted, and vacuum-dried for subsequent use. To verify the accuracy of proteomic data, quantitative real-time polymerase chain reaction (qRT-PCR) was also done (the details are shown in **Supplementary File 1**).

Liquid Chromatography–Tandem Mass Spectrometry Analysis

Labeled peptides were analyzed by online nano flow liquid chromatography tandem mass spectrometry (MS/MS) using the 9RKFSG2_NCS-3500R system (Thermo Fisher Scientific) connected to the Q_Exactive HF-X system (Thermo Fisher Scientific) via a nanoelectrospray ion source. Briefly, a C18-reversed phase column (75 μm × 25 cm, Thermo Fisher Scientific) was equilibrated with solvent A (A: 2% acetonitrile and 0.1% formic acid) and solvent B (B: 80% acetonitrile and 0.1% formic acid). The peptides were eluted using the following gradient: 0–2 min, 0–3% B; 2–92 min, 5–25% B; 92–102 min, 25–45% B; 102–105 min, 45–100% B; 105–120 min, 100–0% B at a flow rate of 300 μL/min. The Q_Exactive HF-X was operated in the data-dependent acquisition mode to automatically switch between full scan MS and MS/MS acquisition. The survey of full scan MS spectra (m/z 350–1,500) was acquired in the Orbitrap with 70,000 resolutions. The top 20 most intense precursor ions were selected into the collision cell for fragmentation by higher-energy collision dissociation. The MS/MS resolution was set at 35,000 (at m/z 100), with the maximum fill time of 50 ms and a dynamic exclusion of 30 s (Wang et al., 2020a,b).

Protein Identification

The RAW data files were analyzed by Proteome Discoverer 2.2 (Thermo Fisher Scientific) against the *K. pneumoniae* database[1]. The MS/MS search criteria were as follows: a mass tolerance of 20 ppm for MS and 0.02 Da for MS/MS tolerance, trypsin as the enzyme with two-missed cleavages allowed, carbamido methylation of cysteine and the TMT of the N-terminus and lysine side chains of peptides as fixed modification, and methionine oxidation as dynamic modifications, respectively. The false discovery rate for peptide identification was set at ≤0.01. A minimum of one unique peptide identification was used to support protein identification (Wang et al., 2020c).

Statistical Analyses

The thresholds of fold change (FC) (>1.2 or <0.83) and $p < 0.05$ were used to identify differentially expressed proteins (DEPs). Annotation of all identified proteins was performed by Gene Ontology (GO)[2] and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway[3] analyses. DEPs were further used for GO and

[1] www.uniprot.org/taxonomy/?query=ATCC15380&sort=score

[2] http://geneontology.org/

[3] http://www.genome.jp/kegg/

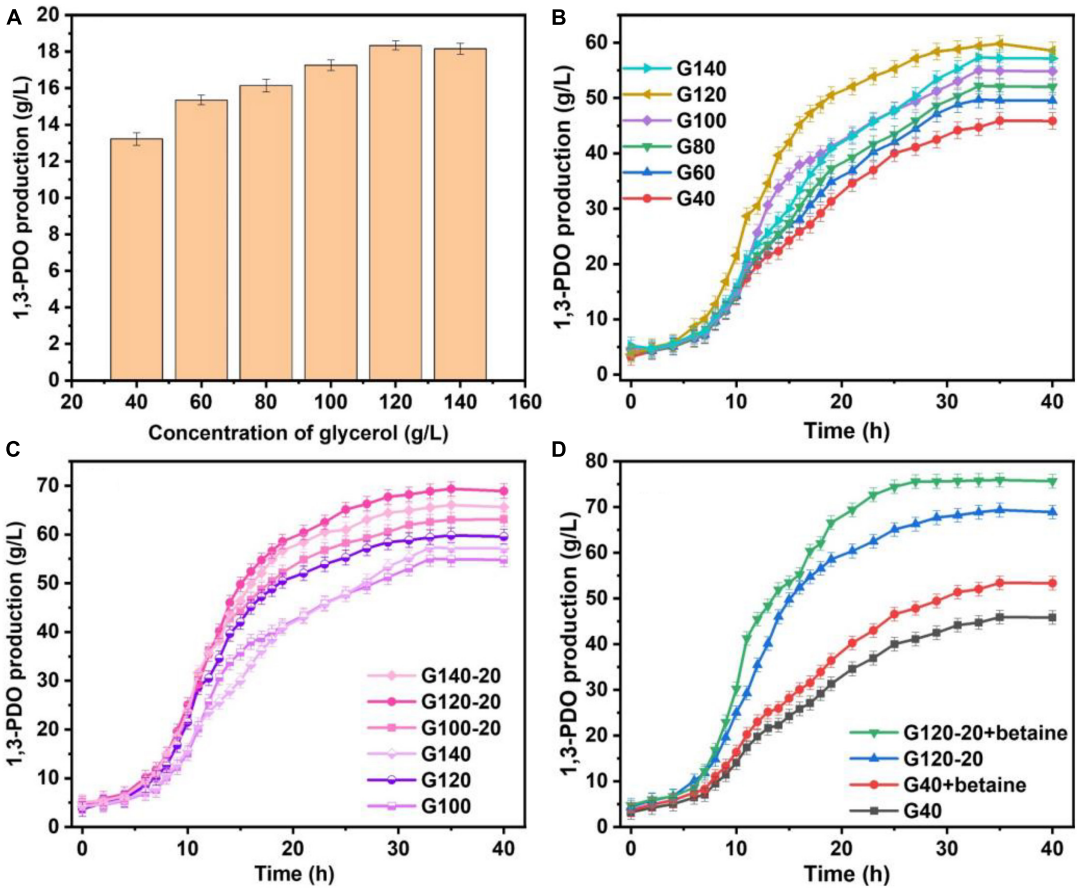


FIGURE 1 | (A) Compa re the production of 1,3-PDO in the 250-mL shaker with glycerine domestication concentration at 40, 60, 80, 100, 120, and 140 g/L. **(B)** The comparison of the yield of 1,3-PDO in the 7.5-L fermentation tank with glycerine domestication concentration at 40, 60, 80, 100, 120, and 140 g/L. **(C)** The production 1,3-PDO in the 7.5-L fermentation tank with glycerine domestication concentration at 100, 100–20, 120, 120–20, 140, and 140–20 g/L. **(D)** The capacity of 1,3-PDO in the 7.5-L fermentation tank with glycerine domestication concentration at 40 g/L, 40 g/L +betaine, 120–20 and 120–20 g/L +betaine.

KEGG enrichment analysis. Protein–protein interaction analysis was performed using the String v10.5.

RESULTS AND DISCUSSION

Comparison of the 1,3-Propanediol Production

The yields of 1,3-PDO produced in 250-mL shaker by the original *K. pneumoniae* and the domesticated *K. pneumoniae* with and without betaine are summarized in **Figure 1A**. The 1,3-PDO production increased from G40 (13.22 g/L) to G120 (18.34 g/L), but decreased at G140. To further investigate the changes in yield during fermentation, the production of 1,3-PDO in the 7.5-L fermentation tank with glycerol domestication concentrations of 40, 60, 80, 100, 120, and 140 g/L were studied, and the results are shown in **Figure 1B**. The 1,3-PDO yields for glycerol domestication concentration from 40 to 120 g/L were 45.91, 49.71, 52.21, 55.02, 59.82, and 57.34 g/L, respectively. When the glycerol acclimation concentration increased to 140 g/L, the 1,3-PDO production decreased. With the increase in the domestication concentration of glycerol for the *K. pneumoniae*, the yield of 1,3-PDO decreased, which was consistent with the previous results (Colin et al., 2000; Yiqiang et al., 2007; Metsoviti et al., 2012; Raghunandan et al., 2014). When the concentration of glycerol increased, metabolism was inhibited, and 1,3-PDO production decreased. Therefore, G100, G120, and G140 were redomesticated under the concentration of 20 g/L glycerol. From **Figure 1C**, the yields of 1,3-PDO increased to 63.11, 69.35, and 66.03 g/L for G100–20, G120–20, and G140–20, respectively. After comparing G120–20 with G40, the yield of 1,3-PDO was improved by 51.06%. During the whole fermentation process, Na_2CO_3 was used as the pH neutralizer, which had an effect on the osmotic pressure of the fermentation liquid and the production yield of 1,3-PDO, so betaine as a fermentation medium was added (Fan et al., 2018). In **Figure 1D**, the 1,3-PDO yields of the G40 (+betaine) and G120–20 (+betaine) reached the 53.42 g/L and 75.92 g/L, respectively, which were higher than those without betaine, consistent with a previous report (Jantama et al., 2010). Thus, betaine could alleviate the osmotic pressure problem and ensure the activity of bacteria during fermentation. Compared to G120–20 (+betaine), which reached a yield of

74.44 g/L at 24 h, the 1,3-PDO production yield of G40 was very low (45.91 g/L) and the fermentation time was very long (40 h), further confirming that betaine could alleviate the increase in the osmotic pressure and counter-suppress the Na^+ effect during fermentation. To explore the enhanced mechanism of 1,3-PDO production after the addition of Na_2CO_3 and betaine, G40 was used as a control group, and TMT was used to compare the differences in protein expression between the control and optimal G120–20 (+betaine).

Protein Identification and Quantitation

After the G40 and G120–20 (+betaine) fermented broths were labeled by TMT, the primary and secondary mass spectra were analyzed statistically. With the help of the Protein Discoverer search library, a total of 3,284 proteins were identified from the four samples. **Supplementary Figure 3A** shows the number distribution of peptides contained in the identified proteins. For example, there were 521 proteins matched with one peptide. The length distribution of the identified peptides is shown in **Supplementary Figure 3B**. For instance, there were 1,860 proteins with a peptide length of eight amino acids. Most of the peptides had 5 to 20 amino acids after enzymatic hydrolysis, accounting for 83.87% of the total, which indicated that the enzymatic hydrolysis was sufficient, and the identification results were reliable. As shown in **Supplementary Figure 3C**, the molecular weight distribution of the identified proteins was determined, especially for these proteins with low molecular weights of less than 20 kDa. The molecular weights of most proteins were from 1 to 60 kDa, and 65 types of macromolecular proteins with molecular weights of more than 100 kDa were identified. **Supplementary Figure 3D** shows the coverage distribution of the identified proteins. The number of amino acids in the peptide was higher than the total number of amino acids in the protein. The identification results were more persuasive with the expansion of coverage distribution. The coverages of polypeptides with more than 10 and 20% of the identified proteins were 78.72 and 62.68%, respectively.

Usually, proteins with differences between G40 and G120–20 were determined based on the FC and the *p*-value. In this work, $p < 0.05$ indicates the difference among the groups. As shown in **Figure 2**, there were 3,284 proteins, including 170 up-regulated proteins (FC > 1.2) and 291 down-regulated proteins (FC < 0.83). These detected proteins were analyzed by GO term and KEGG pathway analysis to identify the biological functions of the differential proteins and the target proteins.

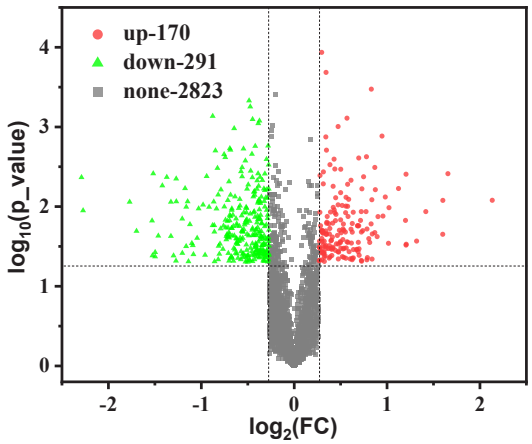


FIGURE 2 | Scatter plot of volcano differences. A volcano plot between the G40 and the G120–20 (+betaine) whose proteins are shown. The horizontal axis is the value of $\log_2$(FC). The vertical axis is the $-\log_{10}$ (*p*-value). The red dots represent up-regulated proteins (FC > 1.2), and the green dots represent down-regulated proteins (FC < 0.83).

The Analysis for Gene Ontology Term

With the GO database, genes and gene products can be classified and annotated as follows: cellular component (CC), molecular function (MF), and Biological Process (BP). It is a bioinformatics analysis tool (Zhong et al., 2019). **Figures 3A,B** show the level 2 of GO classification for 3,284 proteins and 461 differential proteins, respectively. For the 3,284 proteins: 2,190, 2,029, and 1,670 proteins were detected in metabolic process, the cellular process, and single organization process of BP, respectively. There were 1,174 and 1,149 proteins detected in cell and cell part of CC, respectively. There were 2,121 and 1,639 proteins detected in catalytic activity and binding of MF, respectively. As can be seen from **Figure 3B**, in BP: 115 up-regulated and 189 down-regulated differential proteins were detected in metabolic process; 104 up-regulated and 172 down-regulated differential proteins were detected in the cellular process; 85 up-regulated and 154 down-regulated proteins were detected in the single organization process. In CC, 52 up-regulated and 93 down-regulated differential proteins were detected in cell; 51 up-regulated proteins and 89 down-regulated proteins were detected in the cell part. In MF, 108 up-regulated and 192 down-regulated differential proteins were detected mainly in catalytic activity; 86 up-regulated and 139 down-regulated differential proteins were detected in binding. Compared with **Figures 3A,B**, the main functional area was the same, which indicated that these functions played an important role in the production of 1,3-PDO by *K. pneumoniae*. However, because of the excessive number of proteins, the enrichment of differentially abundant proteins requires further analysis.

Gene Ontology functional enrichment analysis can clarify the biological process, cell components, and molecular functions (Zhong et al., 2019). The enrichment of up-regulated and down-regulated proteins for BP is shown in **Figure 3C** and **Table 1**. Up-regulated proteins were related to the ribonucleoside monophosphate metabolic process, nucleoside monophosphate metabolic process, purine nucleoside monophosphate metabolic process, and so on. In MF (**Figure 3E** and **Table 1**), up-regulated proteins were related to phosphotransferase activity (alcohol group as acceptor). In BP (**Figure 3D**), down-regulated proteins were associated with sulfur compound metabolic process, sulfur compound biosynthetic process, sulfur amino acid biosynthetic process, sulfur amino acid metabolic process, methionine biosynthetic process, and methionine metabolic process. Finally, there were 41 up-regulated proteins and 19 down-regulated proteins after further analysis. Therefore, further analysis was

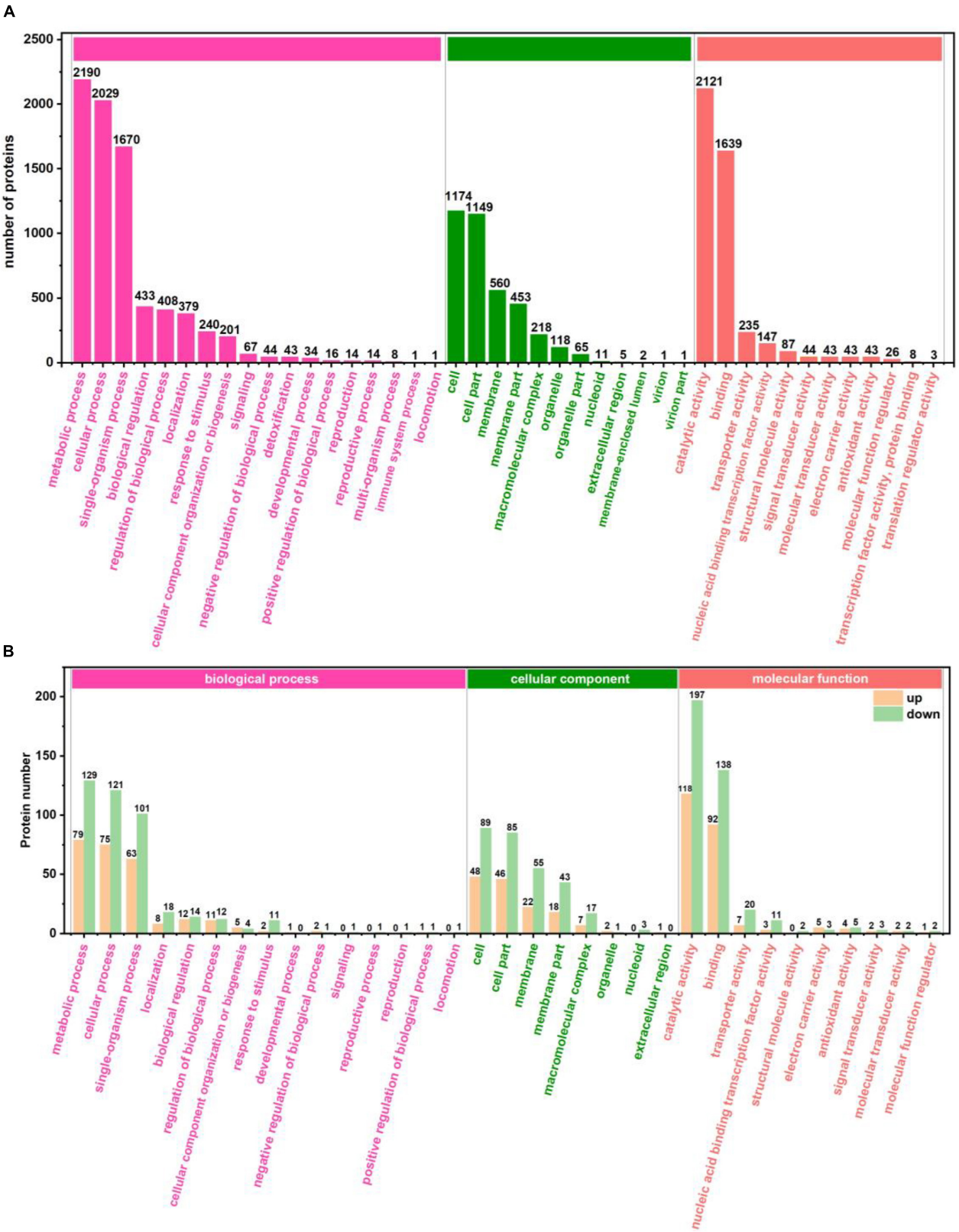


FIGURE 3 | (Continued)

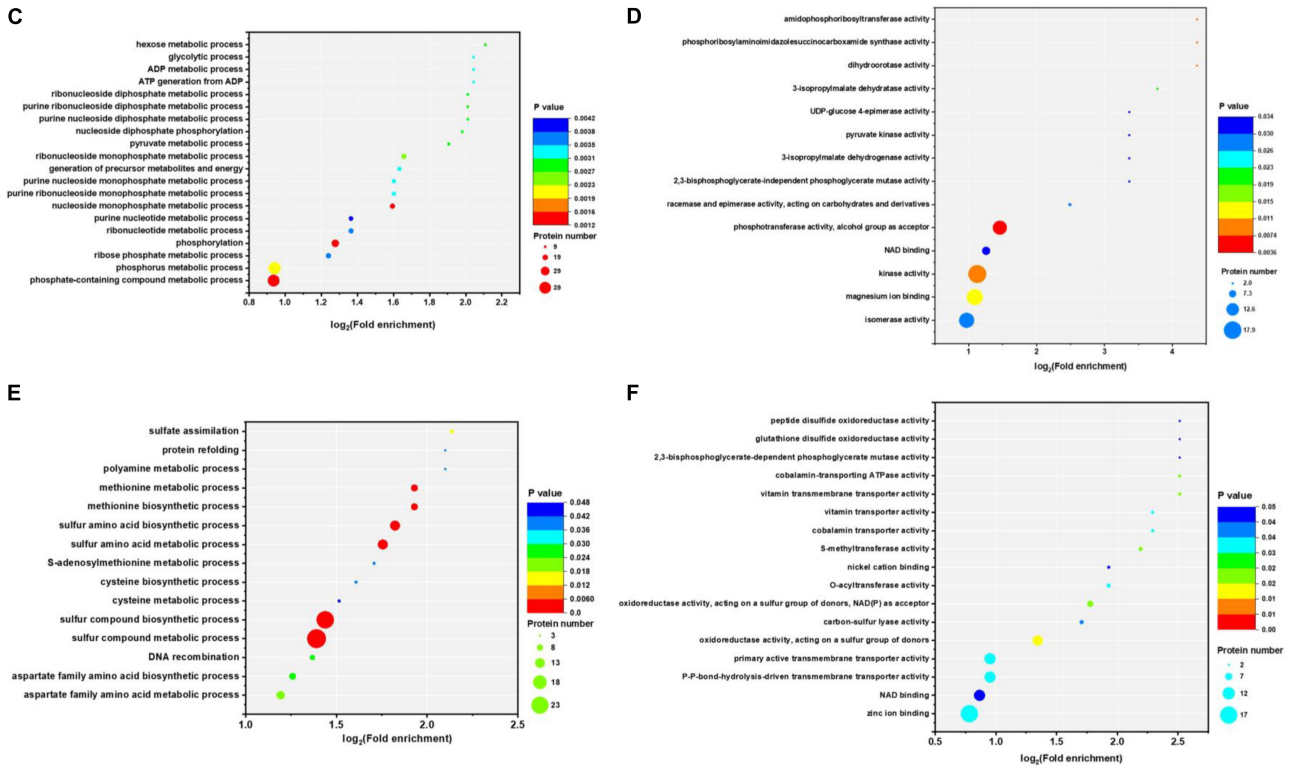


FIGURE 3 | Gene Ontology (GO) classification and enrichment analysis of G40 and G120–20 (+betaine). **(A)** GO classification of 3,284 proteins. The ordinate on the left is $-\log_{10}$ (percent of proteins), and the label on the column is the number of proteins. **(B)** GO classification of 170-up and 291-down proteins. **(C)** The enrichment analysis of the BP (170-up proteins). **(D)** The enrichment analysis of the BP (291-down proteins). **(E)** The enrichment analysis of the MF (170-up proteins). **(F)** The enrichment analysis of the MF (291-down proteins). [The *p*-value (corrected) < 0.005].

needed through the chord diagram of GO term enrichment. As shown in **Figure 4B**, the most up-regulated proteins (1.23 ∼ 3.03-fold) were identified to be NAD-dependent glyceraldehyde-3-phosphate dehydrogenase (GAPDH), 4-hydroxythreonine-4-phosphate dehydrogenase (PdxA), phosphoenolpyruvate-dihydroxyacetone phosphotransferase (DhaL), ThrB, PurT, phosphoglycerate kinase (PGK), 2,3-bisphosphoglycerate-independent phosphoglycerate mutase (GpmI), PurC, and so on. PdxA, whose function is similar to the isocitrate dehydrogenase and isopropylmalate dehydrogenase, can contribute to the phosphotransacetylase (Pta) activity (Sivaraman et al., 2003; Xu et al., 2005). It has been reported that Pta plays a role in the reduction pathway of 1,3-PDO produced by *K. pneumoniae*. It is well known that amino acid metabolism plays an important role in the life of *K. pneumoniae*, which can balance the intracellular pH, generate energy, reduce power, and resist environmental pressures. With the increase in the ThrB, the more threonine is produced. The *K. pneumoniae* can use threonine as a nitrogen source (Reitzer, 2005), and the cavity near ADP is very suitable for homoserine binding, so it can improve the catalytic activity by stabilizing the transition state (Fan et al., 2009; Zhang et al., 2019) and contribute to the 1,3-PDO production. The ligation of amino and carboxylate groups of small molecule metabolites is catalyzed by the ATP-grasp superfamily, which is widespread across primary metabolic processes (Zhang et al., 2008). PurT is a member of the ATP-grasp superfamily, and PurC also has several structural elements in common. With the up-regulation of PurC, the expression levels of diverse proteins involved in purine and pyrimidine synthesis, carbon and energy metabolisms, iron uptake, proteolysis, protein secretion, and signal transduction can be improved. Purine can save energy from the beginning and the consumption of some amino acids (Yuan et al., 2013). As a key enzyme of glycolysis, the up-regulation of GpmI accelerates the catalysis of the interconversion between 3-phosphoglycerate and 2-phosphoglycerate, whereas enolase (Eno) catalyzes the conversion of 2-phosphoglycerate into phosphoenolpyruvate (Yin et al., 2020). GpmI also plays an important role in the carbohydrate transport and metabolism. In addition, PGK not only is a glycolytic enzyme that plays an important role in the growth of biofilm, but also contributes to the formation of surface proteins. In biofilm formation, bacterial cells are embedded in the extracellular matrix, which can protect bacteria from a variety of environmental damages (Wang et al., 2016). Therefore, the tolerance of the strain could be effectively improved during the process of glycerol acclimation, so that the related proteins in the glycolysis pathway were up-regulated, and finally, the production yield of 1,3-PDO was increased. Concurrently, the multiple of down-regulated proteins, such as methylenetetrahydrofolate reductase (MetF), phosphoadenosine phosphosulfate reductase

TABLE 1 | The enrichment analysis of the Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway.

Function	Number of proteins	Log_2(fold enrichment)
GO term		
Up		
BP		
Nucleoside monophosphate metabolic process	17	1.59
Phosphate-containing compound metabolic process	39	0.94
Phosphorylation	24	1.28
Phosphorus metabolic process	40	0.94
Ribonucleoside monophosphate metabolic process	17	1.66
Pyruvate metabolic process	10	1.91
Hexose metabolic process	9	2.11
Nucleoside diphosphate phosphorylation	9	1.98
Purine nucleoside diphosphate metabolic process	9	2.01
Purine ribonucleoside diphosphate metabolic process	9	2.01
Ribonucleoside diphosphate metabolic process	9	2.01
generation of precursor metabolites and energy	14	1.63
Purine ribonucleoside monophosphate metabolic process	14	1.60
Purine nucleoside monophosphate metabolic process	14	1.60
ATP generation from ADP	9	2.04
ADP metabolic process	9	2.04
Glycolytic process	9	2.04
Ribose phosphate metabolic process	18	1.24
Ribonucleotide metabolic process	17	1.36
Purine nucleotide metabolic process	15	1.36
Phosphotransferase activity, alcohol group as acceptor	14	1.46
MF		
Phosphotransferase activity, alcohol group as acceptor	14	1.46
Down		
BP		
Sulfur compound metabolic process	25	1.39
Sulfur compound biosynthetic process	23	1.44
sulfur amino acid biosynthetic process	13	1.82
Sulfur amino acid metabolic process	13	1.76
Methionine biosynthetic process	9	1.93
Methionine metabolic process	9	1.93
MF		
None		
KEGG pathway		
Pathway	**Number of proteins**	**Log_2(fold enrichment)**
Up		
Methane metabolism	11	1.86
Glycolysis/gluconeogenesis	16	1.88
Fructose and mannose metabolism	7	2.10
Down		
Cysteine and methionine metabolism	16	1.46
Sulfur metabolism	12	1.64
Selenocompound metabolism	7	1.90

(CysH), 5-methyltetrahydropteroyltriglutamate-homocysteine *S*-methyltransferase (MetE), methionine adenosyltransferase (MetK) and so on, were down-regulated from 0.40 to 0.79. Commonly, a previous study has shown that high concentrations of homoserine are toxic to cells (Kingsbury and McCusker, 2010). In order to avoid the excessive accumulation of homoserine, decreased MetE, MetF, and MetK levels can inhibit met regulator. To prevent threonine biosynthesis, the up-regulation of ThrB can catalyze the over conversion of homoserine to *o*-phosphate-L-homoserine, which also can inhibit growth. This is a reversible transformation. When the concentration of homoserine decreased, *o*-phosphate-L-homoserine could be converted to homoserine to provide the precursor of methionine (Li et al., 2017). A dynamic balance is beneficial to the growth of cells and the synthesis of 1,3-PDO. For CysH, its down-regulation decrease cysteine (Longo et al., 2016), which reduces the effect of 1,3-PDO production.

The Analysis for Kyoto Encyclopedia of Genes and Genomes Pathway

In organisms, different gene products perform different biological functions through an orderly coordination. Therefore, the pathway information in the KEGG database helped us to understand the biological function of genes at the system level in *K. pneumoniae* (Jia et al., 2021). **Figure 5A** shows the KEGG pathway for differential proteins. The proteins were classified and annotated as follows: metabolism, genetic information processing, environmental information processing, cellular process, organismal systems, and human diseases. In this study, carbohydrate metabolism, amino acid metabolism, and energy metabolism were the most DEPs annotated in metabolism, with 65, 53, and 50 proteins, respectively. Combining with **Figures 5B,C**, there were 38 up-regulated proteins and 27 down-regulated proteins in carbohydrate metabolism. Simultaneously, in genetic information processing, the most DEPs were annotated in (translation), (folding, sorting, and degradation), and (replication and repair); in environmental information processing, the most DEPs were annotated in Immune system and environmental adaptation; in cellular processes, the most DEPs were annotated in cellular community–prokaryotes and cell motility. Despite the superfluous proteins, it required to investigate the KEGG pathway enrichment analysis. As shown in **Figure 5D** and **Table 1**, up-regulated proteins were linked to the methane metabolism, glycolysis/gluconeogenesis, fructose, and mannose metabolism. By contrast, the down-regulated proteins were involved in cysteine and methionine metabolism, sulfur metabolism, and selenocompound metabolism. Ultimately, there were 17 up-regulated proteins and 21 down-regulated proteins after further analysis. Therefore, further analysis was needed through the chord diagram of KEGG pathway enrichment. As can be seen from **Figure 4B**, the up-regulated proteins (glycolysis/gluconeogenesis, methane metabolism, fructose and mannose metabolism, carbon fixation in photosynthetic organisms, taurine and hypotaurine metabolism, pyruvate metabolism, propanoate metabolism, glyoxylate, and dicarboxylate metabolism, inositol phosphate metabolism) are

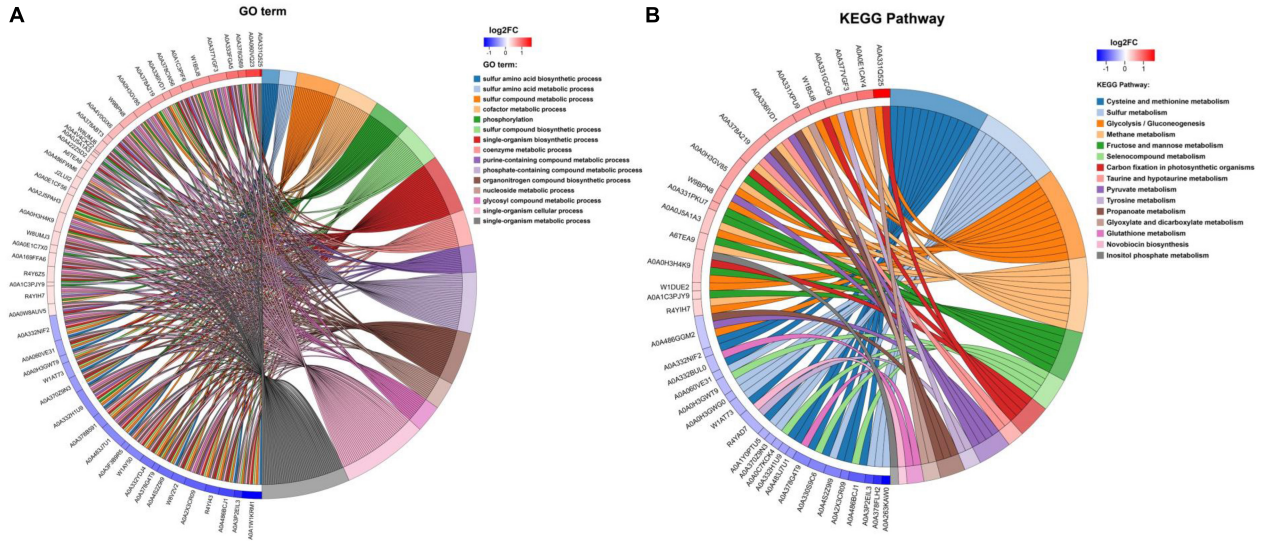


FIGURE 4 | (A) Chord diagram of GO term enrichment. **(B)** Chord diagram of KEGG pathway enrichment. KEGG enrichment chord diagram shows the corresponding relationship among the target proteins, the annotation, and enrichment; the left is protein, and $\log_2$FC is displayed from top to bottom. When $\log_2$FC > 0, the larger the $\log_2$FC is, the greater the differential expression of up-regulated protein is; when $\log_2$FC < 0, the smaller the $\log_2$FC is, the greater the differential expression of down-regulated protein is. The closer the $\log_2$FC is to 0, the smaller the differential expression fold is. The right is KEGG pathway and the score of *z* score. When *z* score > 0, there are more up-regulated proteins than down-regulated proteins involved in this pathway, and this pathway is more likely to be activated. On the contrary, when *z* score < 0, the up-regulated proteins involved in this pathway are less than the down-regulated proteins, and this pathway is more likely to be inhibited. (The same analysis is applied to GO term).

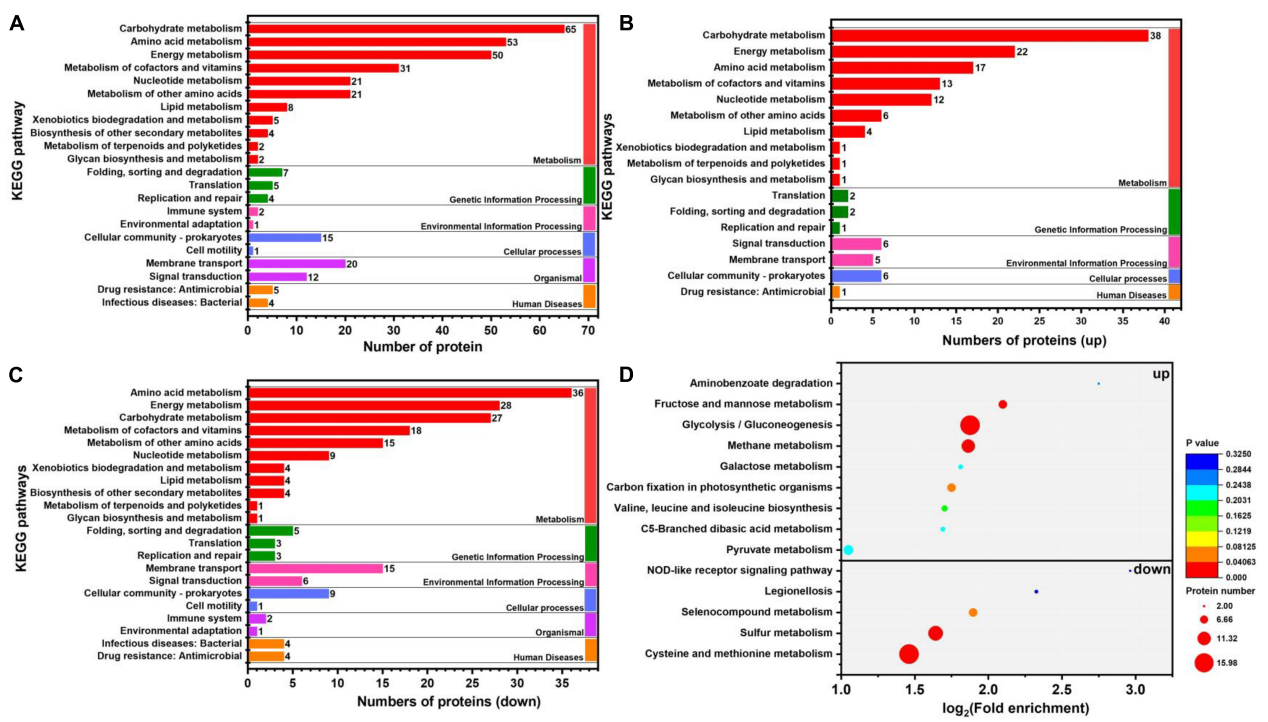


FIGURE 5 | Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway and enrichment analysis of G40 and G120–20 (+betaine). **(A)** KEGG pathway classification of differential protein. **(B)** The enrichment analysis of the KEGG pathway (up proteins). **(C)** The enrichment analysis of the KEGG pathway (down proteins). **(D)** The enrichment analysis of the KEGG pathway (up and down proteins). [The *p*-value (corrected) < 0.05].

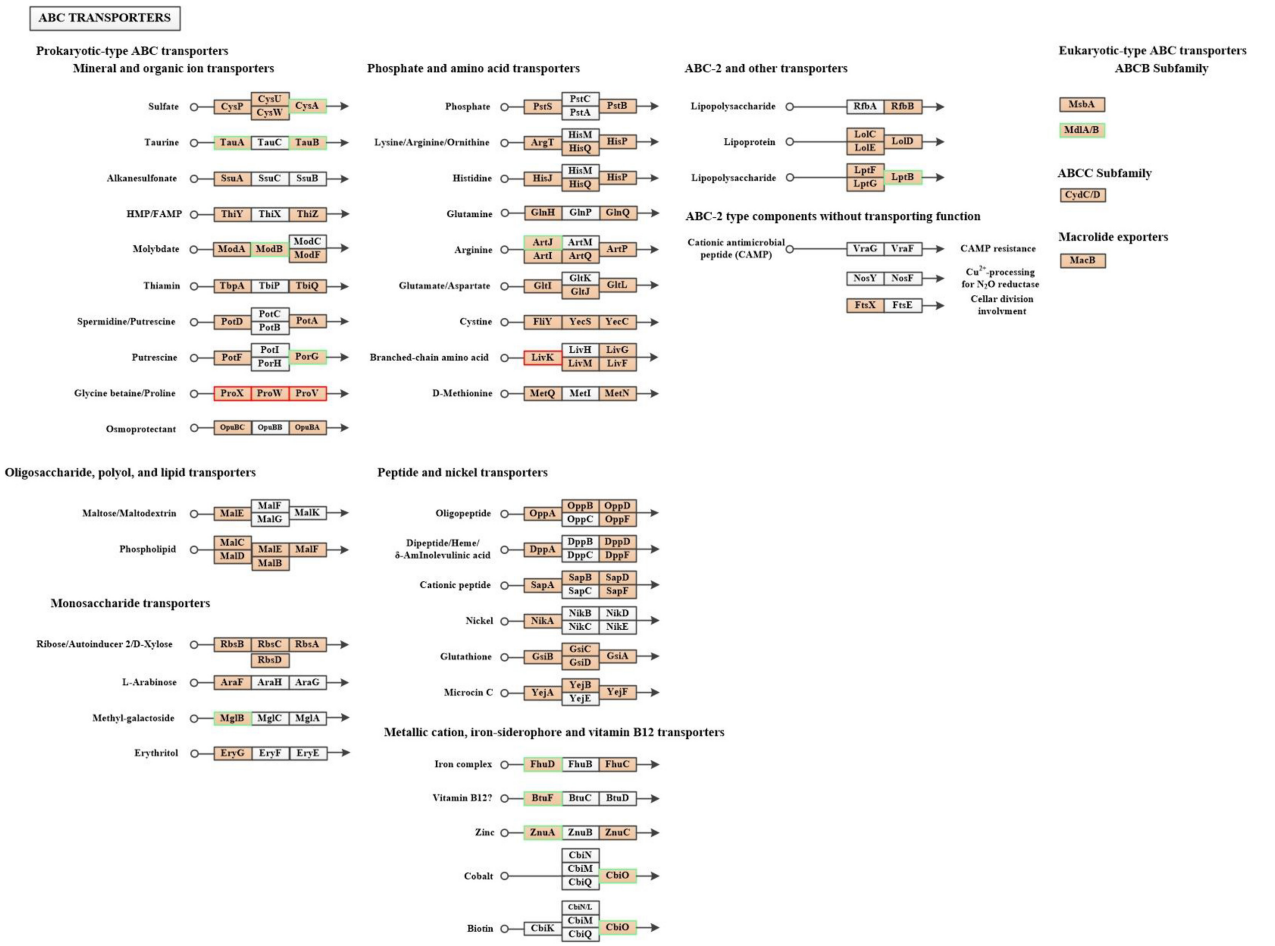


FIGURE 6 | The ABC transporters of the G40 is compared with the G120–20 (+betaine) in KEGG pathway map. Proteins in orange block belong to the 3,284 proteins and in white block belonged to the undetected sample. Up-regulated proteins are marked by red, whereas down-regulated proteins are marked by green in 461 differential proteins. (The abbreviations for full name of all proteins are given in **Supplementary Table 3**).

more than down-regulated proteins, indicating that these KEGG pathways were activated. It suggested that the domesticated strain had an improvement in the production of 1,3-PDO compared to the predomesticated strain. As the degree of acclimation enhanced, the production of 1,3-PDO also increased, corresponding to the results in **Figure 3**. Therefore, the changes in the proteins of the metabolic pathway that produce 1,3-PDO needed to be further analyzed. In addition, the up-regulated and down-regulated proteins were similar to the analysis identified by GO terms, except that taurine import ATP-binding protein (TauA) and taurine-binding periplasmic protein (TauB) decreased, with the FC ranging from 0.35 to 0.43. As these two proteins were related to ABC transporters, further analysis of ABC transporters was given in **Figure 6**, illustrating that 17 differential proteins had changed. The CysA, TauA, TauB, ModB, PorG, MglB, ArtJ, FhuD, BtuF, ZnuA, CbiO, and LptB were down-regulated, whereas the ProX, ProW, ProV, and Livk were up-regulated. ProV and ProW were membrane-associated proteins, and ProV had a considerable sequence identity with ATP-binding proteins from other periplasmic systems. ProX encoded the periplasmic glycine betaine-binding protein (Stirling et al., 1989; May et al., 2010). The biggest FC among them was ProX, which reached 4.39, revealing that betaine played an important role in fermentation.

When Na_2CO_3 was used to adjust the pH in the fermentation process, salt stress appeared with the continuous increase in Na_2CO_3. The osmotic pressure induced by the Na^+ salt increased gradually, which resulted in the outflow of water, the loss of cell swelling pressure, and the change of solute concentration and cell volume (Fan et al., 2018). Compared with $Ca(OH)_2$ used in commercialization as a pH-neutralizing agent (Wang et al., 2017; Liang et al., 2018), the output of 1,3-PDO was greatly affected. After the use of betaine, it was found that the yields were significantly improved as shown in **Figure 1D**. Betaine can be used not only as a stress protector or a stabilizer of intracellular enzymes to resist stress conditions, but also as a methyl donor for methylation. It can be accumulated at high concentrations (through transport or biosynthesis) in the cell to balance the osmotic pressure inside and outside the cell. On the other hand, it can increase the cell growth rate and

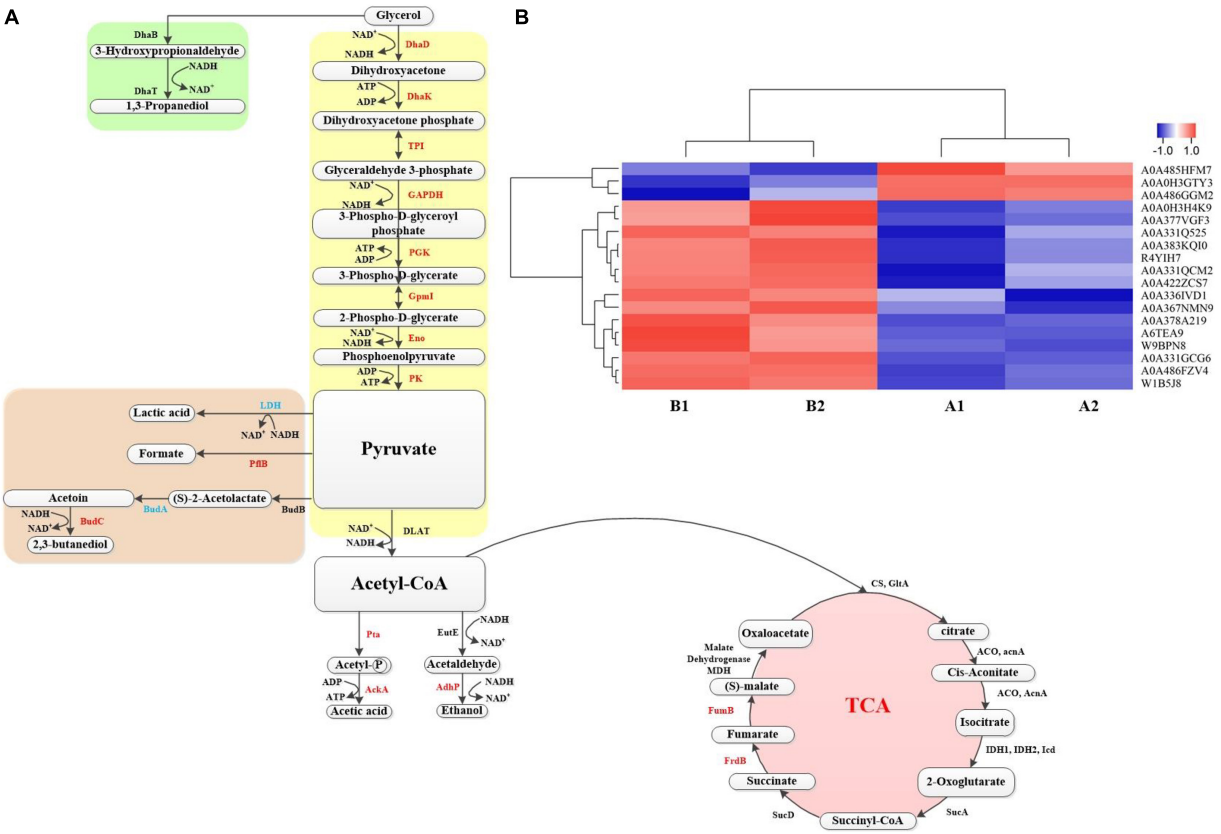


FIGURE 7 | (A) The metabolic pathway for production of 1,3-propanediol by *K. pneumoniae*. The red represents the up-proteins, and the blue represents the down proteins, whereas the black represents the no change. **(B)** Heat map of proteins on the metabolic pathway. Each column in the figure represents a sample, and each row represents a protein. The color in the figure represents the relative expression amount of the protein in the group of samples; red represents the high expression amount of the protein in the sample, and blue represents the low expression amount. For the specific change trend of the expression amount, see the number label under the color bar at the top left.

improve the fermentation performance of the strain under high osmotic stress (high concentration of carbohydrate substrate or product) (Fan et al., 2018). In the 1,3-PDO fermentation process of *K. pneumoniae*, the stress of the Na^+ salt as a neutralizer Na_2CO_3 and the yield of 1,3-PDO were improved by betaine, which offers an alternative way for the industrial pH neutralizer to avoid producing solid pollutants and thus greatly alleviating subsequent sewage treatment and environmental pollution.

Analysis of Several Important Proteins in Metabolic Pathway

In the above, some proteins that play a role in the production of 1,3-PDO by *K. pneumoniae* with the GO and KEGG pathway analysis and the important role of betaine in the fermentation process were analyzed However, protein changes of G120–20 compared with G40 in the metabolic pathway of *K. pneumoniae* producing 1,3-PDO still need further investigation. Combined with KEGG pathway analysis, the metabolic pathway of glycerol in *K. pneumoniae* is shown in **Figure 7A**: bacterial formation pathway, reduction pathway, and oxidation pathway. In the bacterial formation pathway, ATP was consumed, whereas ADP was produced in metabolism. In the reduction pathway: first, glycerol was converted to 3-hydroxyglyceraldehyde by DhaB and then converted to 1,3-PDO by DhaT. In G120–20 g/L (+betaine). Usually, 1,3-PDO oxidoreductase (DhaT) catalyzes the conversion of 3-hydroxypropionaldehyde (3-HPA) to 1,3-PDO, which is a key enzyme in the preparation of 1,3-PDO from glycerol. But DhaT is seriously inactivated by 3-HPA due to the reaction of 3-HPA with the sulfhydryl group of cysteine residue (Li et al., 2016). Although the DhaB and DhaT were not differential proteins in this article, they still had an effect on the production of 1,3-PDO by *K. pneumoniae*.

The oxidation pathway was similar to the glycolysis pathway, in that it generated ATP and reduced $NADH_2$, which was required for bacterial growth. $NADH_2$ was consumed during the 3-HPA-mediated production of 1,3-PDO, which was produced in the oxidation pathway. The oxidation pathway was mainly divided into three phases.

First, glycerol is converted into dihydroxyacetone through DhaD (NAD^+ is required as a coenzyme to produce $NADH_2$). Dihydroxyacetone is phosphorylated into glycolysis under the action of Dhak. As shown in **Figure 7A** and **Table 2**, DhaD and DhaK were up-regulated, and their FCs were 1.38 and 1.30, respectively. According to the report, DhaD encodes glycerol dehydrogenase in *K. pneumoniae* (Tang et al., 1982), and glycerol dehydrogenase and 1,3-propylene glycol

TABLE 2 | The protein of the metabolic pathway for *K. pneumoniae* producing 1,3-PDO.

Accession	Protein name	Description	FC	Log_2FC	*P*	Sum up	Sum down
		Reduction pathway					
W1DMB2	DhaB	Glycerol dehydratase reactivation factor large subunit	1.49	0.58	0.2364	0	0
Q7WRJ3	DhaT	1,3-Ppropanediol oxidoreductase	1.17	0.23	0.2874	0	0
		Oxidation pathway					
A0A367NMN9	DhaD	Glycerol dehydrogenase	1.38	0.47	0.02303	1	0
A6TEA9	DhaK	Dihydroxyacetone kinase	1.30	0.38	0.03098	1	0
A0A0H3H4K9	TPI, tpiA	Triosephosphate isomerase	1.27	0.34	0.03373	1	0
A0A331Q525	GAPDH, GapA	NAD-dependent glyceraldehyde-3-phosphate dehydrogenase	3.03	1.60	0.02229	1	0
W1B5J8	PGK	Phosphoglycerate kinase	1.68	0.75	0.00423	1	0
A0A377VGF3	GpmI	2,3-bisphosphoglycerate-independent phosphoglycerate mutase	1.72	0.78	0.04665	1	0
R4YIH7	Eno	Enolase	1.23	0.30	0.01977	1	0
W9BPN8	PK, Pyk	Pyruvate kinase	1.48	0.57	0.02773	1	0
R4Y5U7	DLAT, AceF, PdhC	Acetyltransferase component of pyruvate dehydrogenase complex	0.35	−1.525	0.04124	0	0
Lactate pathway							
A0A485HFM7	LldD	L-lactate dehydrogenase	0.41	−1.27	0.04810	0	1
A0A486GGM2	LDH	L-lactate dehydrogenase	0.82	−0.28	0.04928	0	1
2,3-butanediol pathway							
A0A378G331	BudB	Acetolactate synthase	–	–	–	0	0
A0A0H3GTY3	BudA	Acetolactate decarboxylase	0.41	−1.27	0.00448	0	1
A0A422ZCS7	BudC	Butanediol dehydrogenase	1.83	0.87	0.02125	1	0
Formic acid pathway							
A0A383KQI0	PflB, PflD	Pyruvate formate-lyase	1.65	0.72	0.02091	1	0
Acetic acid pathway							
A0A336IVD1	Pta	Phosphate acetyltransferase	1.56	0.64	0.04410	1	0
A0A378A219	AckA	Acetate kinase	1.54	0.63	0.01629	1	0
Ethanol pathway							
A0A377VJU1	EutE	Acetaldehyde dehydrogenase	–	–	–	0	0
A0A331GCG6	AdhP	Alcohol dehydrogenase	1.72	0.78	0.00237	1	0
TCA cycle pathway							
A0A170J878	CS, GltA	Citrate synthase	–	–	–	0	0
A0A377ZR87	ACO, AcnA	Aconitate hydratase	–	–	–	0	0
A0A483EZU0	IDH1, IDH2, Icd	Isocitrate dehydrogenase [NADP]	–	–	–	0	0
A0A377 × 309	OGDH, SucA	2-Oxoglutarate dehydrogenase E1 component	–	–	–	0	0
A0A486KFH1	SucD	Succinate—CoA ligase [ADP-forming] subunit alpha	–	–	–	0	0
A0A486FZV4	FrdB	Succinate dehydrogenase iron-sulfur subunit	1.42	0.50	0.00340	1	0
A0A331QCM2	FumB	Fumarate hydratase class I	2.49	1.32	0.02712	1	0
A0A2S6E360	MDH	Malate dehydrogenase (Fragment)	–	–	–	0	0

oxidoreductase are the key enzymes for the conversion of glycerol to 1,3-PDO (Zhao et al., 2009). Therefore, the up-regulation of DhaD was conducive to the production of 1,3-PDO. It has been reported (Raynaud et al., 2011) that the DhaK, DhaL, and DhaM were belong to the PEP-dependent dihydroxyacetone kinases. It can be seen from **Supplementary Table 3**, DhaK [A6TEA9 contains dihydroxyacetone-binding sites (Gutknecht et al., 2001)], DhaL [A0A378G569 contains ADP-binding sites (Gutknecht et al., 2001)] and DhaM [A0A486FQT7, a phosphohistidine protein that can transfer phosphoryl groups from a phosphoryl carrier protein of the phosphotransferase system (HPr or enzyme I) to the DhaL-ADP complex (Gutknecht et al., 2001; Bachler et al., 2005)] were also up-regulated, indicating that dihydroxyacetone kinases were up-regulated. Glycerol dehydratase, 1,3-PDO oxidoreductase, glycerol dehydrogenase, and dihydroxyacetone kinase are

encoded by an operon named dha, and their expression was consistent (Forage and Lin, 1982). It can be seen from **Figures 7A,B**; both were up-regulated at the same time, which also explains that the domesticated strain favors to produce 1,3-PDO.

Second, the dihydroxyacetone phosphate was further oxidized to pyruvate. In this process, triosephosphate isomerase (TPI), GAPDH, PGK, GpmI, Eno, and pyruvate kinase (PK) were up-regulated, and their corresponding FCs were 1.27, 3.03, 1.68, 1.72, 1.23, and 1.48, respectively. TPI plays a vital role in metabolism and is the key to efficient energy production (Zheng et al., 2006). The dihydroxyacetone phosphate was transformed into glyceraldehyde 3-phosphate through TPI. Because of the up-regulation of TPI, the accumulation of dihydroxyacetone phosphate can be reduced, as its toxicity would affect cell growth and survival (Kang et al., 2014). GAPDH catalyzes the conversion of glyceraldehyde 3-phosphate to glycerol 1,3-diphosphate and reduces NAD^+ to NADH. Moreover, the up-regulation of GAPDH can shorten the fermentation time and inhibit the accumulation of some harmful by-products (such as lactic acid) (Yang et al., 2013). By comparing the protein changes in the two cases, most of the proteins in the glycolysis pathway were up-regulated, which helped to provide ATP and NADH for bacteria, promote the growth of the bacteria, and finally increase the yield of 1,3-PDO.

In addition, the by-products of pyruvate metabolism were lactic acid, formic acid, and 2,3-butanediol. The strain domestication may increase in yields of both the main product 1,3-PDO and by-products, so BudC and PflB were up-regulated. AckA and AdhP were also up-regulated when acetyl CoA produced by-products, such as ethanol and acetic acid. LDH and LldD encode L-lactate dehydrogenase (Aguilera et al., 2008; Fu et al., 2016). The down-regulation of lactate dehydrogenase reduces both the consumption of NADH and the formation of the by-product lactic acid and finally improves the output of 1,3-PDO (Xu et al., 2009).

The third step was that pyruvate could be further transferred to produce acetyl-CoA, and acetyl-CoA could enter the tricarboxylic acid (TCA) cycle to produce other small molecular substances. In this process, some energy was consumed, but there was a regeneration process of force reduction. In the whole TCA cycle, FumB and FrdB provide energy for bacteria. Therefore, the up-regulation of FumB and FrdB might accelerate cell growth and increase the yield of 1,3-PDO (Tseng et al., 2001; Xue et al., 2010; Huang et al., 2013; Li et al., 2019).

For the domesticated bacteria, most of the proteins in the metabolic process were up-regulated, resulting in a significant increase in the yield of 1,3-PDO. On the other hand, the 1,3-PDO yield could be further improved by the following genetic modifications, for example, the overexpression of some up-regulated genes (such as *dhaD*, *dhaK*, *tpi*, *gapA*, etc.), blocking the by-product pathway (such as knocking out *budC* to reduce the competition of by-products) and weakening *ldh* to reduce the consumption of NADH.

CONCLUSION

The fermentation capabilities of 1,3-PDO by glycerol domesticated strains at the concentrations of 40, 60, 80, 100, 120, and 140 g/L were compared. It was found that the strain domesticated with 120 g/L glycerol had the highest capability to produce 1,3-PDO, reaching 59.41 g/L. To further improve the yield of 1,3-PDO, the strain domesticated with 120–20 g/L glycerol concentration (*K. pneumoniae* x546) was obtained, and the yield reached 69.35 g/L. In addition, in order to overcome the osmotic pressure problem caused by excessive Na^+ in the fermentation system, betaine was added to the fermentation medium, making the yield further increase to 74.44 g/L and shortening the fermentation time from 40 to 24 h. Based on TMT, it was found that regulating genes, such as *dhaD*, *dhak*, *budC*, *ldh*, and so on, were able to enhance the yield of 1,3-PDO. Moreover, the introduction of Na_2CO_3 and betaine in the fermentation process will render the formation of 1,3-PDO more environment-friendly and facilitate industrial adoption of this technology in the future.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are publicly available. This data can be found here: PRIDE database Project Name: *Klebsiella pneumoniae* x546, ATCC15380, TMT Project accession: PXD028396.

AUTHOR CONTRIBUTIONS

XW carried out the experimental work, analyzed the data, and wrote the manuscript. HC and PW performed the data analysis and participated in the manuscript editing and revise. JJ and LZ helped to partial experiment and figure processing. JX helped to edit the manuscript and involved in discussion in the manuscript preparation. JW was responsible for the experiment design and supervision. All authors read and approved the final manuscript.

FUNDING

This work was financially supported by the National Key Research and Development Program (No. 2018YFA0902200), the China Petrochemical Corporation (No. 2018GKF-0283), the China Scholarship Council (No. 202006250069), and the Frontiers Science Center for Synthetic Biology (Ministry of Education), Tianjin University.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.770109/full#supplementary-material

REFERENCES

Aguilera, L., Campos, E., Gimenez, R., Badia, J., Aguilar, J., and Baldoma, L. (2008). Dual role of LldR in regulation of the lldPRD operon, involved in L-lactate metabolism in *Escherichia coli*. *J. Bacteriol.* 190, 2997–3005. doi: 10.1128/JB.02013-07

Bachler, C., Flukiger-Bruhwiler, K., Schneider, P., Bahler, P., and Erni, B. (2005). From ATP as substrate to ADP as coenzyme. *J. Biol. Chem.* 280, 18321–18325.

Bao, W. J., Wei, R. Q., Liu, X. X., Dong, S. F., Chen, T. Y., Fu, S. L., et al. (2020). Regulation of pyruvate formate lyase-deficient *Klebsiella pneumoniae* for efficient 1,3-propanediol bioproduction. *Curr. Microbiol.* 77, 55–61. doi: 10.1007/s00284-019-01795-5

Chen, W. C., Chuang, C. J., Chang, J. S., Wang, L. F., Soo, P. C., Wu, H. S., et al. (2020). Exploring dual-substrate cultivation strategy of 1,3-propanediol production using *Klebsiella pneumoniae*. *Appl. Biochem. Biotechnol.* 191, 346–359. doi: 10.1007/s12010-019-03208-6

Chen, Z., Zhong, W., Chen, S., Zhou, Y., Ji, P., Gong, Y., et al. (2021). TMT-based quantitative proteomics analyses of sterile/fertile anthers from a genic male-sterile line and its maintainer in cotton (*Gossypium hirsutum* L.). *J. Proteomics* 232:104026. doi: 10.1016/j.jprot.2020.104026

Colin, T., Bories, A., and Moulin, G. (2000). Inhibition of *Clostridium butyricum* by 1,3-propanediol and diols during glycerol fermentation. *Appl. Microbiol. Biotechnol.* 54, 201–205. doi: 10.1007/s002530000365

Dai, J., Yu, X., Han, Y., Chai, L., Liao, Y., Zhong, P., et al. (2020). TMT-labeling proteomics of papillary thyroid carcinoma reveal invasive biomarkers. *J. Cancer* 11, 6122–6132. doi: 10.7150/jca.47290

Dexter Tam, T. L., Ng, C. K., Lim, S. L., Yildirim, E., Ko, J., Leong, W. L., et al. (2019). Proquinoidal-conjugated polymer as an effective strategy for the enhancement of electrical conductivity and thermoelectric properties. *Chem. Mater.* 31, 8543–8550. doi: 10.1021/acs.chemmater.9b03684

Fan, C., Fromm, H. J., and Bobik, T. A. (2009). Kinetic and functional analysis of L-threonine kinase, the PduX enzyme of *Salmonella enterica*. *J. Biol. Chem.* 284, 20240–20248. doi: 10.1074/jbc.M109.027425

Fan, X., Zhang, T., Jie, L. I., Han, H., Gao, L., Zhang, S., et al. (2018). Betaine metabolism in microorganism and its application. *Bull. Fermentation Sci. Technol.* 47:3. doi: 10.16774/j.cnki.issn.1674-2214.2018.03.006

Forage, R. G., and Lin, E. C. (1982). DHA system mediating aerobic and anaerobic dissimilation of glycerol in *Klebsiella pneumoniae* NCIB 418. *J. Bacteriol.* 151, 591–599. doi: 10.1128/jb.151.2.591-599.1982

Fu, J., Huo, G., Feng, L., Mao, Y., Wang, Z., Ma, H., et al. (2016). Metabolic engineering of *Bacillus subtilis* for chiral pure meso-2,3-butanediol production. *Biotechnol. Biofuels* 9:90. doi: 10.1186/s13068-016-0502-5

Glaasker, E., Heuberger, E., Konings, W. N., and Poolman, B. (1998). Mechanism of osmotic activation of the quaternary ammonium compound transporter (QacT) of *Lactobacillus plantarum*. *J. Bacteriol.* 180, 5540–5546. doi: 10.1128/JB.180.21.5540-5546.1998

Guerzoni, M. E., Lanciotti, R., and Cocconcelli, P. S. (2001). Alteration in cellular fatty acid composition as a response to salt, acid, oxidative and thermal stresses in *Lactobacillus helveticus*. *Microbiology* 147, 2255–2264. doi: 10.1099/00221287-147-8-2255

Gungormusler, M., Gonen, C., and Azbar, N. (2011). 1,3-propanediol production potential by a locally isolated strain of *Klebsiella Pneumoniae* in comparison to *Clostridium beijerinckii* NRRL B593 from waste glycerol. *J. Polym. Environ.* 19, 812–817.

Guo, J., Cao, Y. J., Liu, H., Zhang, R. B., Xian, M., and Liu, H. Z. (2019). Improving the production of isoprene and 1,3-propanediol by metabolically engineered *Escherichia coli* through recycling redox cofactor between the dual pathways. *Appl. Microbiol. Biotechnol.* 103, 2597–2608. doi: 10.1007/s00253-018-09578-x

Gutknecht, R., Beutler, R., Garcia-Alles, L. F., Baumann, U., and Erni, B. (2001). The dihydroxyacetone kinase of *Escherichia coli* utilizes a phosphoprotein instead of ATP as phosphoryl donor. *EMBO J.* 20, 2480–2486. doi: 10.1093/emboj/20.10.2480

Huang, C.-J., Wang, Z.-C., Huang, H.-Y., Huang, H.-D., and Peng, H.-L. (2013). YjcC, a c-di-GMP phosphodiesterase protein, regulates the oxidative stress response and virulence of *Klebsiella pneumoniae* CG43. *PLoS One* 8:e66740. doi: 10.1371/journal.pone.0066740

Hussain, S., Chachar, Q., Keerio, M. I., and Shirazi, M. U.-U. (2020). Physiological and biochemical response of wheat genotypes under temperature stress. *Pak. J. Bot.* 52, 365–374.

Jantama, K., Haupt, M. J., Svoronos, S. A., Zhang, X., and Ingram, L. O. (2010). Combining metabolic engineering and metabolic evolution to develop nonrecombinant strains of *Escherichia coli* C that produce succinate and malate. *Biotechnol. Bioeng.* 99, 1140–1153. doi: 10.1002/bit.21694

Jia, C., Lu, X., Gao, J., Wang, R., Sun, Q., and Huang, J. (2021). TMT-labeled quantitative proteomic analysis to identify proteins associated with the stability of peanut milk. *J. Sci. Food Agric.* 101, 6424–6433. doi: 10.1002/jsfa.11313

Kang, T. S., Korber, D. R., and Tanaka, T. (2014). Metabolic engineering of a glycerol-oxidative pathway in *Lactobacillus panis* PM1 for utilization of bioethanol thin stillage: potential to produce platform chemicals from glycerol. *Appl. Environ. Microbiol.* 80, 7631–7639. doi: 10.1128/AEM.01454-14

Kim, C., Lee, J. H., Baek, J., Kong, D. S., Na, J. G., Lee, J., et al. (2020). Small current but highly productive synthesis of 1,3-propanediol from glycerol by an electrode-driven metabolic shift in *Klebsiella pneumoniae* L17. *Chemsuschem* 13, 564–573. doi: 10.1002/cssc.201902928

Kingsbury, J. M., and McCusker, J. H. (2010). Homoserine toxicity in *Saccharomyces cerevisiae* and *Candida albicans* homoserine kinase (thr1 Delta) mutants. *Eukaryot. Cell* 9, 717–728.

Lee, J. H., Jung, H. M., Jung, M. Y., and Oh, M. K. (2019). Effects of gltA and arcA mutations on biomass and 1,3-propanediol production in *Klebsiella pneumoniae*. *Biotechnol. Bioprocess Eng.* 24, 95–102. doi: 10.1007/s12257-018-0246-0

Li, H., Wang, B. S., Li, Y. R., Zhang, L., Ding, Z. Y., Gu, Z. H., et al. (2017). Metabolic engineering of *Escherichia coli* W3110 for the production of l-methionine. *J. Ind. Microbiol. Biotechnol.* 44, 75–88. doi: 10.1007/s10295-016-1870-3

Li, X., Chen, L., Wang, X., and Tian, P. (2019). Physiological investigations of the influences of byproduct pathways on 3-hydroxypropionic acid production in *Klebsiella pneumoniae*. *J. Basic Microbiol.* 59, 1195–1207. doi: 10.1002/jobm.201800640

Li, Z., Ro, S. M., Sekar, B. S., Seol, E., Lama, S., Lee, S. G., et al. (2016). Improvement of 1,3-propanediol oxidoreductase (DhaT) stability against 3-hydroxypropionaldehyde by substitution of cysteine residues. *Biotechnol. Bioprocess Eng.* 21, 695–703. doi: 10.1007/s12257-016-0560-3

Liang, S., Gao, D., Liu, H., Wang, C., and Wen, J. (2018). Metabolomic and proteomic analysis of d-lactate-producing *Lactobacillus delbrueckii* under various fermentation conditions. *J. Ind. Microbiol. Biotechnol.* 45, 681–696. doi: 10.1007/s10295-018-2048-y

Longo, F., Motta, S., Mauri, P., Landini, P., and Rossi, E. (2016). Interplay of the modified nucleotide phosphoadenosine 5'-phosphosulfate (PAPS) with global regulatory proteins in *Escherichia coli*: modulation of cyclic AMP (cAMP)-dependent gene expression and interaction with the HupA regulatory protein. *Chem. Biol. Interact.* 259, 39–47. doi: 10.1016/j.cbi.2016.04.016

Louesdon, S., Charlot-Rouge, S., Juillard, V., Tourdot-Marechal, R., and Beal, C. (2014). Osmotic stress affects the stability of freeze-dried *Lactobacillus buchneri* R1102 as a result of intracellular betaine accumulation and membrane characteristics. *J. Appl. Microbiol.* 117, 196–207. doi: 10.1111/jam.12501

Ma, J. S., Jiang, H., Hector, S. B., Xiao, Z. H., Li, J. L., Liu, R. K., et al. (2019). Adaptability of *Klebsiella pneumoniae* 2e, a newly isolated 1,3-propanediol-producing strain, to crude glycerol as revealed by genomic profiling. *Appl. Environ. Microbiol.* 85:15. doi: 10.1128/AEM.00254-19

May, G., Faatz, E., Lucht, J. M., Haardt, M., and Bremer, E. (2010). Characterization of the osmoregulated *Escherichia coli* proU promoter and identification of ProV as a membrane-associated protein. *Mol. Microbiol.* 3, 1521–1531. doi: 10.1111/j.1365-2958.1989.tb00138.x

Metsoviti, M., Paraskevaidi, K., Koutinas, A., Zeng, A.-P., and Papanikolaou, S. (2012). Production of 1,3-propanediol, 2,3-butanediol and ethanol by a newly

isolated *Klebsiella oxytoca* strain growing on biodiesel-derived glycerol based media. *Process Biochem.* 47, 1872–1882. doi: 10.1016/j.procbio.2012.06.011

Mitrea, L., and Vodnar, D. C. (2019). *Klebsiella pneumoniae*-a useful pathogenic strain for biotechnological purposes: diols biosynthesis under controlled and uncontrolled pH levels. *Pathogens* 8:293. doi: 10.3390/pathogens8040293

Nakano, S., Ugwu, C. U., and Tokiwa, Y. (2012). Efficient production of D-(-)-lactic acid from broken rice by *Lactobacillus delbrueckii* using Ca(OH)(2) as a neutralizing agent. *Bioresour. Technol.* 104, 791–794. doi: 10.1016/j.biortech.2011.10.017

Pan, D. T., Wang, X. D., Shi, H. Y., Yuan, D. C., and Xiu, Z. L. (2019). Ensemble optimization of microbial conversion of glycerol into 1, 3-propanediol by *Klebsiella pneumoniae*. *J. Biotechnol.* 301, 68–78. doi: 10.1016/j.jbiotec.2019.06.001

Park, Y. S., Kang, J., Chung, W. H., Lim, M. Y., Seo, M. J., Nam, Y. D., et al. (2019). Complete genome sequence of acetate-producing *Klebsiella pneumoniae* L5-2 isolated from infant feces. *3 Biotech* 9:84. doi: 10.1007/s13205-019-1578-y

Raghunandan, K., McHunu, S., Kumar, A., Kumar, K. S., Govender, A., Permaul, K., et al. (2014). Biodegradation of glycerol using bacterial isolates from soil under aerobic conditions. *J. Environ. Sci. Health A Tox. Hazard. Subst. Environ. Eng.* 49, 85–92. doi: 10.1080/10934529.2013.824733

Raynaud, C., Lee, J., Sarcabal, P., Croux, C., Meynial-Salles, I., and Soucaille, P. (2011). Molecular characterization of the glycerol-oxidative pathway of *Clostridium butyricum* VPI 1718. *J. Bacteriol.* 193, 3127–3134. doi: 10.1128/JB.00112-11

Reitzer, L. (2005) Catabolism of amino acids and related compounds. *EcoSal Plus* 1:2. doi: 10.1128/ecosalplus.3.4.7

Shin, W. S., June, C. S., Lee, Y., Park, Y., Choi, J., and Jeon, W. (2011). Synthesis of low concentration of NaOH solution using Na^+ ion in the concentrated water from membrane separation process. *Korean Chem. Eng. Res.* 49, 810–815. doi: 10.9713/kcer.2011.49.6.810

Simon, A., Fujioka, T., Price, W. E., and Nghiem, L. D. (2014). Sodium hydroxide production from sodium carbonate and bicarbonate solutions using membrane electrolysis: a feasibility study. *Sep. Purif. Technol.* 127, 70–76. doi: 10.1016/j.seppur.2014.02.020

Sivaraman, J., Li, Y. G., Banks, J., Cane, D. E., Matte, A., and Cygler, M. (2003). Crystal structure of *Escherichia coli* PdxA, an enzyme involved in the pyridoxal phosphate biosynthesis pathway. *J. Biol. Chem.* 278, 43682–43690. doi: 10.1074/jbc.M306344200

Sogame, Y., Kojima, K., Takeshita, T., Kinoshita, E., and Matsuoka, T. (2014). Identification of cAMP-dependent phosphorylated proteins involved in the formation of environment-resistant resting cysts by the terrestrial ciliate *Colpoda cucullus*. *ISJ Invertebrate Surviv. J.* 11, 213–218.

Stirling, D. A., Hulton, C. S., Waddell, L., Park, S. F., Stewart, G. S., Booth, I. R., et al. (1989). Molecular characterization of the proU loci of *Salmonella typhimurium* and *Escherichia coli* encoding osmoregulated glycine betaine transport systems. *Mol. Microbiol.* 3, 1025–1038. doi: 10.1111/j.1365-2958.1989.tb00253.x

Tang, J., Forage, R. G., and Lin, E. (1982). Immunochemical properties of NAD+-linked glycerol dehydrogenases from *Escherichia coli* and *Klebsiella pneumoniae*. *J. Bacteriol.* 152, 1169–1174. doi: 10.1128/jb.152.3.1169-1174.1982

Tee, Z. K., Jahim, J. M., Tan, J. P., and Kim, B. H. (2017). Preeminent productivity of 1,3-propanediol by *Clostridium butyricum* JKT37 and the role of using calcium carbonate as pH neutraliser in glycerol fermentation. *Bioresour. Technol.* 233, 296–304. doi: 10.1016/j.biortech.2017.02.110

Tseng, C. P., Yu, C. C., Lin, H. H., Chang, C. Y., and Kuo, J. T. (2001). Oxygen- and growth rate-dependent regulation of *Escherichia coli* fumarase (FumA, FumB, and FumC) activity. *J. Bacteriol.* 183, 461–467. doi: 10.1128/JB.183.2.461-467.2001

Wang, M., Wang, G., Zhang, T., Fan, L., and Tan, T. (2017). Multi-modular engineering of 1,3-propanediol biosynthesis system in *Klebsiella pneumoniae* from co-substrate. *Appl. Microbiol. Biotechnol.* 101, 647–657. doi: 10.1007/s00253-016-7919-4

Wang, S., Yang, Y., Zhao, Y., Zhao, H., Bai, J., Chen, J., et al. (2016). Sub-MIC tylosin inhibits *Streptococcus suis* biofilm formation and results in differential protein expression. *Front. Microbiol.* 7:384. doi: 10.3389/fmicb.2016.00384

Wang, C., Chen, L., Cai, Z. C., Chen, C., Liu, Z., Liu, X., et al. (2020a). Comparative proteomic analysis reveals the molecular mechanisms underlying the accumulation difference of bioactive constituents in *Glycyrrhiza uralensis* Fisch under salt stress. *J. Agric. Food Chem.* 68, 1480–1493. doi: 10.1021/acs.jafc.9b04887

Wang, Z., Wang, L., Zhou, J., Zou, J., and Fan, L. (2020b). New insights into the immune regulation and tissue repair of *Litopenaeus vannamei* during temperature fluctuation using TMT-based proteomics. *Fish Shellfish Immunol.* 106, 975–981. doi: 10.1016/j.fsi.2020.09.014

Wang, Z., Li, M. X., Xu, C. Z., Zhang, Y., Deng, Q., Sun, R., et al. (2020c). Comprehensive study of altered proteomic landscape in proximal renal tubular epithelial cells in response to calcium oxalate monohydrate crystals. *BMC Urol.* 20:136. doi: 10.1186/s12894-020-00709-z

Willke, T., and Vorlop, K. (2008). Biotransformation of glycerol into 1,3-propanediol. *Eur. J. Lipid Sci. Technol.* 110, 831–840. doi: 10.1002/ejlt.200800057

Xu, Q. S., Jancarik, J., Lou, Y., Kuznetsova, K., Yakunin, A. F., Yokota, H., et al. (2005). Crystal structures of a phosphotransacetylase from *Bacillus subtilis* and its complex with acetyl phosphate. *J. Struct. Funct. Genomics* 6, 269–279. doi: 10.1007/s10969-005-9001-9

Xu, Y.-Z., Guo, N.-N., Zheng, Z.-M., Ou, X.-J., Liu, H.-J., and Liu, D.-H. (2009). Metabolism in 1,3-propanediol fed-batch fermentation by a D-lactate deficient mutant of *Klebsiella pneumoniae*. *Biotechnol. Bioeng.* 104, 965–972. doi: 10.1002/bit.22455

Xue, X., Li, W., Li, Z., Xia, Y., and Ye, Q. (2010). Enhanced 1,3-propanediol production by supply of organic acids and repeated fed-batch culture. *J. Ind. Microbiol. Biotechnol.* 37, 681–687. doi: 10.1007/s10295-010-0711-z

Yang, T., Rao, Z., Zhang, X., Xu, M., Xu, Z., and Yang, S.-T. (2013). Improved production of 2,3-butanediol in *Bacillus amyloliquefaciens* by over-expression of glyceraldehyde-3-phosphate dehydrogenase and 2,3-butanediol dehydrogenase. *PLoS One* 8:e76149. doi: 10.1371/journal.pone.0076149

Yin, L., Luo, X., Zhang, Y., Zheng, W., Yin, F., and Fu, Y. (2020). Comparative proteomic analysis of *Rhizopus oryzae* hyphae displaying filamentous and pellet morphology. *3 Biotech* 10:469. doi: 10.1007/s13205-020-02458-0

Yiqiang, P., Lin, L., Baishan, F., and Wenzhen, Z. (2007). The selection of 1,3-propanediol produced bacterium enduring high glycerol concentration mutate and its immobidi ferment study. *Acta Laser Biol. Sin.* 16, 754–758.

Yuan, Z. H., Wang, L., Sun, S. T., Wu, Y., and Qian, W. (2013). Genetic and proteomic analyses of a *Xanthomonas campestris* pv. campestris purC mutant deficient in purine biosynthesis and virulence. *J. Genet. Genomics* 40, 473–487. doi: 10.1016/j.jgg.2013.05.003

Zabed, H. M., Zhang, Y. F., Guo, Q., Yun, J. H., Yang, M. M., Zhang, G. Y., et al. (2019). Co-biosynthesis of 3-hydroxypropionic acid and 1,3-propanediol by a newly isolated *Lactobacillus reuteri* strain during whole cell biotransformation of glycerol. *J. Clean. Prod.* 226, 432–442. doi: 10.1016/j.jclepro.2019.04.071

Zhang, L., Fan, Y., Li, X., Liao, S., Wang, P., and Qiao, K. (2016). Effects of two alkaline pH regulators on the fermentation of 1,3-propanediol. *Chem. Ind. Eng. Prog.* 35, 2542–2546.

Zhang, L. J., Bao, W. J., Wei, R. Q., Fu, S. L., and Gong, H. (2018). Inactivating NADH:quinone oxidoreductases affects the growth and metabolism of *Klebsiella pneumoniae*. *Biotechnol. Appl. Biochem.* 65, 857–864. doi: 10.1002/bab.1684

Zhang, Y., Ma, C., Dischert, W., Soucaille, P., and Zeng, A.-P. (2019). Engineering of phosphoserine aminotransferase increases the conversion of l-homoserine to 4-hydroxy-2-ketobutyrate in a glycerol-independent pathway of 1,3-propanediol production from glucose. *Biotechnol. J.* 14:e1900003. doi: 10.1002/biot.201900003

Zhang, Y., Morar, M., and Ealick, S. E. (2008). Structural biology of the purine biosynthetic pathway. *Cell. Mol. Life Sci.* 65, 3699–3724. doi: 10.1007/s00018-008-8295-8

Zhao, L., Ma, X., Zheng, Y., Zhang, J., Wei, G., and Wei, D. (2009). Over-expression of glycerol dehydrogenase and 1,3-propanediol oxidoreductase in *Klebsiella pneumoniae* and their effects on conversion of glycerol into 1,3-propanediol in

resting cell system. *J. Chem. Technol. Biotechnol.* 84, 626–632. doi: 10.1002/jctb.2092

Zheng, P., Sun, J., van den Heuvel, J., and Zeng, A.-P. (2006). Discovery and investigation of a new, second triose phosphate isomerase in *Klebsiella pneumoniae*. *J. Biotechnol.* 125, 462–473. doi: 10.1016/j.jbiotec.2006.03.034

Zhong, Q. P., Wang, B., Wang, J., Liu, Y. F., Fang, X., and Liao, Z. L. (2019). Global proteomic analysis of the resuscitation state of *Vibrio parahaemolyticus* compared with the normal and viable but non-culturable state. *Front. Microbiol.* 10:1045. doi: 10.3389/fmicb.2019.01045

Zhou, S., Lama, S., Sankaranarayanan, M., and Park, S. (2019b). Metabolic engineering of *Pseudomonas denitrificans* for the 1,3-propanediol production from glycerol. *Bioresour. Technol.* 292:121933. doi: 10.1016/j.biortech.2019.121933

Zhou, S., Huang, Y. H., Mao, X. L., Li, L. L., Guo, C. Y., Gao, Y. L., et al. (2019a). Impact of acetolactate synthase inactivation on 1,3-propanediol fermentation by *Klebsiella pneumoniae*. *PLoS One* 14:e0200978. doi: 10.1371/journal.pone.0200978

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

ORIGINAL RESEARCH
published: 09 December 2021
doi: 10.3389/fmicb.2021.761972

Multi-Omics Reveal the Efficient Phosphate-Solubilizing Mechanism of Bacteria on Rocky Soil

Yanqiang Ding[1†], Zhuolin Yi[1†], Yang Fang[1], Sulan He[2], Yuming Li[2], Kaize He[1], Hai Zhao[1] and Yanling Jin[1]*

[1]*CAS Key Laboratory of Environmental and Applied Microbiology, Environmental Microbiology Key Laboratory of Sichuan Province, Chengdu Institute of Biology, Chinese Academy of Sciences, Chengdu, China,* [2]*Sweetpotato Institute, Nanchong Academy of Agricultural Sciences, Nanchong, China*

OPEN ACCESS

Edited by:
Om Prakash, National Centre for Cell Science, India

Reviewed by:
Maria Carolina Quecine, University of São Paulo, Brazil
Ewoud Ewing, Karolinska Institutet (KI), Sweden
Harold Alexander Vargas Hoyos, University of Antioquia, Colombia
Hillol Chakdar, National Bureau of Agriculturally Important Microorganisms (ICAR), India

***Correspondence:**
Yanling Jin
jinyl@cib.ac.cn

[†]These authors have contributed equally to this work

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

Received: *20 August 2021*
Accepted: *22 November 2021*
Published: *09 December 2021*

Citation:
Ding Y, Yi Z, Fang Y, He S, Li Y, He K, Zhao H and Jin Y (2021) Multi-Omics Reveal the Efficient Phosphate-Solubilizing Mechanism of Bacteria on Rocky Soil. Front. Microbiol. 12:761972. doi: 10.3389/fmicb.2021.761972

Phosphate-solubilizing bacteria (PSB) can alleviate available phosphorus (AP)-deficiency without causing environmental pollution like chemical phosphate fertilizers. However, the research and application of PSB on the barren rocky soil is very rare. We screened six PSB from sweetpotato rhizosphere rocky soil. Among them, *Ochrobactrum haematophilum* FP12 showed the highest P-solubilizing ability of 1,085.00 mg/L at 7 days, which was higher than that of the most reported PSB. The assembled genome of PSB FP12 was 4.92 Mb with P-solubilizing and plant growth-promoting genes. In an AP-deficient environment, according to transcriptome and metabolomics analysis, PSB FP12 upregulated genes involved in gluconic acid synthesis and the tricarboxylic acid cycle, and increased the concentration of gluconic acid and malic acid, which would result in the enhanced P-solubilizing ability. Moreover, a series of experiments in the laboratory and field confirmed the efficient role of the screened PSB on significantly increasing AP in the barren rocky soil and promoting sweetpotato yield. So, in this study, we screened highly efficient PSB, especially suitable for the barren rocky soil, and explored the P-solubilizing mechanism. The research will reduce the demand for chemical phosphate fertilizers and promote the environment-friendly agricultural development.

Keywords: barren rocky soil, application potential, multi-omics, phosphate-solubilizing bacteria, phosphate-solubilizing mechanism

INTRODUCTION

Available phosphorus (AP) is deficient in 74% of the arable soil of China and in more than 40% of the arable soil globally, because about 99% of the total phosphorus (TP) is dissolved (Balemi and Negisho, 2012; Yu et al., 2019). This leads to a huge demand for chemical phosphate fertilizers. According to the report of the Food and Agriculture Organization of the United Nations, the world's demand for phosphate fertilizer has reached 45.86 million tons in 2020.[1] However, the utilization rate of chemical phosphate fertilizers is as low as 10–25% (Roberts and Johnston, 2015). Moreover, excessive chemical phosphate fertilizer would result in soil nutrition imbalance, heavy metal accumulation, and water eutrophication, meanwhile the production of chemical phosphate fertilizer often causes some other environmental pollution (Huang et al., 2017;

[1]http://www.fao.org/

Park et al., 2021). Therefore, developing methods to reduce the demand for chemical phosphate fertilizers and improve the bioavailability of soil TP is of significance to both agricultural production and environmental protection.

Microorganisms play an important role in the biogeochemical cycle. Among them, phosphate-solubilizing bacteria (PSB) can degrade insoluble inorganic and organic phosphorus into AP, which would then be easily absorbed and utilizated by plants (Granada et al., 2018; Wei et al., 2018; Parastesh et al., 2019; Bi et al., 2020). The previous researches screened out some PSB, which showed good P-solubilizing ability (115–716 mg/L) under laboratory culture conditions (Yu et al., 2019; Kaur and Kaur, 2020). However, it was reported that, in about 70% of the experiments, PSB did not play a direct role in supplementing soil AP for plants (Gyaneshwar et al., 2002). Considering possible effects of the different organic matter, pH, and other characteristics of the soil on the colonization and function of PSB, it is very necessary to develop suitable PSB for different types of soil.

Sedimentary rock covers 75–80% of the Earth's crust, forming parent materials for a large majority of soils (Wilkinson et al., 2009). In the Sichuan Basin, China, more than 70% of the hills (approximately 5.0×10^{10} m^2) are covered by barren rocky soil, a kind of weathered shale. Rocky soil has high potential for agricultural utilization (Manning and Theodoro, 2018). However, the AP in the rocky soil is very deficient, moreover, in which chemical phosphate fertilizer is more easily washed away and cause serious environmental pollution (Huang et al., 2017). There have been some PSB fertilizers for arable soil, but there are few studies on PSB suitable for the barren rocky soil.

The unclear mechanism of PSB is one of the important factors blocking the application of PSB (Elhaissoufi et al., 2021). Normally, direct oxidation of glucose to gluconic acid is thought to be the main P-solubilizing way in Gram-negative bacteria (Sashidhar and Podile, 2010; Luduena et al., 2018). Whereas, other studies found that pyruvate, lactic acid, succinic acid, and citric acid were the main organic acids secreted by PSB when solubilizing P (Li et al., 2017; Zeng et al., 2017). Furthermore, there is controversy regarding the effect of P availability on the P-solubilizing ability of PSB (Zeng et al., 2017). In addition, there is few systematic study on how PSB respond to P-deficiency and synthesize and secrete organic acids (Zeng et al., 2017). Fortunately, genomics, transcriptome, metabolomics, and other omics technologies have recently provided powerful methods to shed light on the P-solubilizing mechanism and application research.

The main objectives of this study were: (i) to screen highly efficient PSB suitable for the barren rocky soil, (ii) to explore the P-solubilizing mechanism at the level of genome, transcriptome, and metabolomics, and (iii) to verify the application effect of the PSB on the barren rocky soil.

MATERIALS AND METHODS

Screening for PSB

Sweetpotato rhizosphere rocky soil samples were collected at the Yingxi experimental base of the Nanchong Academy of Agricultural Sciences, Sichuan, China (30°52'N, 106°02'E) using the methods described in a previous study (Ding et al., 2020). The soil type is purplish soils according to classification and codes for Chinese soil (GB/T 17296-2009). Soil samples (2.0 g) and sterile water (30 ml) were mixed to obtain a bacterial suspension. Bacterial strains were cultured and purified on Luria-Bertani medium according to the dilution coating method and streak plate method (Sanders, 2012).

The purified strains were inoculated on modified PKOC2 medium with sterilized toothpicks (Li et al., 2019), and the plates were inverted and incubated in a 28°C incubator. The strains with P-solubilizing circles were selected for re-screening. Holes with a diameter of 6.00 mm were punched in the screening medium. Next, 100 µl of the cultured bacterial solution with the same OD_{600} value were inoculated into the holes. After 7 days of incubation in a 28°C incubator, the diameter of the P-solubilizing circle was measured.

About 2 mm of cultured bacterial solution with the same OD_{600} value were inoculated into a 100 ml triangle flask containing 30 ml of Monkina inorganic P culture medium (Yh, 2014), with three replicates per strain. After incubation at 28°C and 150 rpm in a shaking incubator for 7 days, the pH of the medium was determined using a pH meter (PHS-3C, Fangzhou, China), and the AP of the medium was determined using the molybdenum blue method (Murphy and Riley, 1962).

Identification of PSB

For six strains of PSB (BP10, BP11, BP23, FP2, FP12, and FP16), 2 ml of the bacterial solution were used for 16S rRNA gene sequencing. The genomic DNA were extracted using the SDS method (Natarajan et al., 2016). The amplified 16S rRNA genes were obtained with specific primers (27F: 5'-AGTTTGATCMTGGCTCAG-3'; 1492R: 5'-GGTTACCTT GTTACGACTT-3'), and sequenced using the Sanger sequencing platform (Chen et al., 2015). The 16S rRNA gene sequences were aligned to the NCBI[2] database to search for the species with the highest similarity, wherein the selection "16S ribosomal RNA (Bacteria and Archaea)" was made. The EZBioCloud platform was further used for strain identification (Yoon et al., 2017a).

Phosphate-Solubilizing Profile of PSB

To explore the P-solubilizing mechanism, PSB FP12, the efficient PSB selected from this study, was cultured in four kinds of media with different P sources. Four media were used as PKOC2 medium [the insoluble P medium (IPM)], PKOC2 medium with the replace of $Ca_3(PO_4)_2$ with 3.02 g $NaH_2PO_4 \bullet 2H_2O$ [the available P medium (APM)], PKOC2 medium with the addition of 3.02 g $NaH_2PO_4 \bullet 2H_2O$ [the insoluble and available P medium (IAPM)], and the PKOC2 medium without $Ca_3(PO_4)_2$ [no P medium (NPM)]. Each group had three replicates. After incubation at 28°C and 150 rpm in a shaking incubator for 7 days, the pH and AP of the medium were similarly determined as above.

[2]https://blast.ncbi.nlm.nih.gov/

Organic Acids Targeted Metabolomics

At 1 and 2 days, as the pH and dissolved P changed largely, the solution from groups IPM and APM were sampled and filtered through a 0.2 μm filter membrane for organic acid-targeted metabolomics analysis. Each group had three replicates. The following organic acids were measured, i.e., gluconic acid, 2-keto-gluconic acid, malic acid, acetic acid, α-ketoglutaric acid, lactic acid, formic acid, citric acid, succinic acid, pyruvate, oxalic acid, and tartaric acid.

The organic acid-targeted metabolomics was performed using an Agilent liquid chromatograph (Aglient1200, United States). The liquid chromatograph column was the LAEQ-462572 Athena C18-WP 4.6 * 250 mm, with a column temperature of 30°C (van Hees et al., 1999). The detection wavelength was 210 nm (Zitouni et al., 2020). Chromatographic standard samples were used to draw standard curves to ensure that $R^2 > 0.99$.

Genome Sequencing of PSB FP12

The genomic DNA of PSB FP12 was extracted using the SDS method (Natarajan et al., 2016). Then the high-quality genomic DNA was sequenced using the Illumina NovaSeq sequencing platform (Illumnia, Inc., San Diego, CA, PE150). The raw sequencing data was filtered using Readfq (Version 10)[3] with default parameters. The filtered clean data was analysed for PSB FP12 genome *de novo* assembly using SOAPdenovo (Version 2.04) with default parameters (Luo et al., 2012). Then GapCloser (Version 1.12)[4] was used to fill gaps in the initial assembly genome. And the short contigs below 500 bp were filtered out.

GeneMarkS (Version 4.17) was used to predict the coding genes of the genome (Besemer et al., 2001). Diamond was used to align the predicted protein sequence with GO, KEGG, COG, NR, Pfam, TCDB, and Swiss-Prot (e-value ≤ 1e-5; Buchfink et al., 2015). For the alignment of each protein sequence, the match with the highest score (default identity ≥ 40%, coverage ≥ 40%) was used to select the functional annotation. A map of the circular genome was drawn with BRIG (Version 0.95; Alikhan et al., 2011). The Average Nucleotide Identity (ANI) of the genome of PSB FP12 and related species was calculated using the ANI Calculator (Yoon et al., 2017b).

Transcriptome of PSB FP12

In order to study the transcriptional expression changes of PSB FP12 in a P-deficient environment, we analyzed the transcriptome of PSB FP12 in the insoluble P medium (IPM group) and the available P medium (APM group). At 1 and 2 days, the solution from groups IPM and APM were sampled, and each group was sampled in triplicate. The RNA of PSB FP12 was extracted using the TRIZOL method (Rio et al., 2010). Qualified libraries were sequenced on the Illumina Novaseq sequencing platform using the paired-end sequencing method (PE150). FastQC_v0.11.3 was used for quality control of the raw sequencing data. The filtered sequence was aligned with the rRNA database using Bowtie2 to remove the rRNA sequence (Langmead and Salzberg, 2012).

Qualified sequencing data were aligned to the PSB FP12 genome using Hisat2 (Kim et al., 2015), and the count value and the Reads Per Kilobase per Million mapped reads (RPKM) values of the genes were calculated using HTseq-count and AWK script (Anders et al., 2015). EdgeR was used to calculate the differentially expressed genes [$p < 0.05$ and $|\mathrm{Log_2}$ (fold change)$| > 1$; Robinson et al., 2009]. Gene Ontology (GO) classification and enrichment analysis of differentially expressed genes was performed on the WEGO 2.0 platform (Ye et al., 2018). KEGG enrichment analysis of differentially expressed genes was performed based on the KEGG database.[5]

Treating the Barren Rocky Soil With PSB

To verify the application effects of those selected PSB on the barren rocky soil, a series of experiments were performed. The barren rocky soil was collected at the Yingxi, Nanchong, Sichuan, China (30°52′N, 106°02′E). The soil type is purplish soils according to Classification and codes for Chinese soil (GB/T 17296-2009; **Supplementary Figure S1**). The barren rocky soil properties were showed in **Supplementary Table S1**. About 1 ml of single PSB (BP10, BP11, BP23, FP2, FP12, or FP16; approximately 10^9 CFU/ml) or the mixed PSB (MB group) were added to 10 g of sterilized barren rocky soil. Around 1 ml sterile water without PSB was added in the control (CK) group. After 10 days of incubation in a 28°C incubator, the soil pH and AP were measured using the methods described in a previous study (Ding et al., 2020).

Addition of Sweetpotato Root Exudates

One of the main factors limiting the colonization of PSB in the barren rocky soil is the lack of organic matter. Plant root exudates are important sources of soil organic matter, and plants can affect the colonization of rhizosphere microorganisms through root exudates (Sasse et al., 2018). Therefore, the effect of plant root exudates was also considered when developing PSB suitable for the barren rocky soil.

Sweetpotato root exudates were collected according to methods described previously (Ding et al., 2020). BP10, BP11, BP23, FP2, FP12, and FP16 PSB (approximately 10^9 CFU/ml) were mixed in equal proportions and then centrifuged. The supernatant was discarded, and the remaining pellet was resuspended in sterile water. Around 10 ml of the resuspended PSB was added to 100 g of sterilized barren rocky soil, then 10 ml of sweetpotato root exudates was added as a treatment group (BR group), and 10 ml of sterile water was added as a control group (BW group). Meanwhile, 20 ml of sterile water was added to 100 g of sterilized soil, without PSB, as another control (BCK group). After 10 days of incubation in a 28°C incubator, the soil pH and AP were measured using the methods described in a previous study (Ding et al., 2020).

[3]http://github.com/cjfields/readfq

[4]https://sourceforge.net/projects/soapdenovo2/files/GapCloser/

[5]https://www.kegg.jp/

Pot Experiments

A greenhouse pot experiment was performed in a greenhouse (30°63′N, 104°07′E) at Chengdu Institute of Biology, Chinese Academy of Sciences. A strain of sweetpotato [*Ipomoea batatas* (L.) Lam.] cultivar Nanshu 88 was planted in a pot (high, 26.5 cm; diameter 24.0 cm) containing 6.5 kg of barren rocky soil. In the BP group, 75 ml of mixed PSB (approximately 10^9 CFU/ml) was applied to each pot; in the CK group, 75 ml of sterile water without PSB was applied. Each group had eight pots. The greenhouse temperature was 25°C.

The field pot experiment was performed at the Yingxi experimental base (30°52′N, 106°02′E), and designed same as in the greenhouse pot experiment. During the experiment, the local monthly precipitation was 64–180 mm, and the mean temperature was 18.5–28.4°C.

At 100 days, the sweetpotato and soil in the greenhouse and field pot experiments were weighed and sampled. The P and potassium (K) content of sweetpotato roots and vines were measured using an elemental content analyzer with inductively coupled plasma optical emission spectrometry (ICP-OES; Optima 8300, PerkinElmer, United States). The nitrogen (N) content of sweetpotato roots and vines, as well as the pH, total carbon (TC), total nitrogen (TN), and AP of soil were measured using the methods described in a previous study (Ding et al., 2020).

Statistics Analysis

Significant differences between samples were evaluated using Tukey's honest significant difference test on R (version 3.5.0). Spearman's correlation coefficients between AP and pH were also calculated using R (version 3.5.0).

RESULTS

Screening and Identification of PSB

After preliminary screening, 26 bacterial strains that grew well on the screening media and had obvious P-solubilizing circles were selected for re-screening. These 26 PSB strains were inoculated into the round holes of the screening media to observe the P-solubilizing circle (**Supplementary Figure S2A**). At 7 days, the P-solubilizing circle of BP11 was the largest one, with a diameter of 25.75 mm, and the ratio of the P-solubilizing circle to the diameter of the round hole (D/d) was 4.29. In addition, the P-solubilizing circle of PSB FP12 was also observed with a large diameter of 23.75 mm and D/d of 3.96 (**Supplementary Table S2**).

These 26 PSB strains were inoculated and screened in the liquid media, and the AP content was measured at 7 days (**Supplementary Figure S2B**). Most strains showed significantly higher values of AP than the control group. The AP of FP12 reached 744.00 mg/L at 7 days, which was significantly higher than that of other strains ($p < 0.001$; **Supplementary Figure S2B**; **Supplementary Table S3**) and was further optimized to 1,085.00 mg/L (**Figure 1**).

Combined with the re-screening results of the P-solubilizing ability on plate and shake flask, we finally selected BP10, BP11, BP23, FP2, FP12, and FP16 for further analysis. Through 16S rRNA sequencing, PSB BP10, BP11, BP23, FP2, FP12, and FP16 were identified as *Stenotrophomonas maltophilia*, *Achromobacter xylosoxidans*, *Achromobacter xylosoxidans*, *Stenotrophomonas maltophilia*, *Ochrobactrum haematophilum*, and *Cellulosimicrobium cellulans*, respectively (**Supplementary Table S4**).

The Relationship Between pH and Solubilized AP

To study whether the P-solubilizing mechanism is related to acid production, the relationship between solubilized AP and pH was analyzed. The correlation analysis showed that there was a significant negative correlation between the solubilized AP and pH; Spearman ρ was −0.92 ($p < 2.2e{-}16$). The trend analysis showed that the solubilized AP and pH correlated well in line with the logarithmic function: $pH = -0.565 \ln(AP) + 7.7973$, $R^2 = 0.8801$.

As shown in **Figure 1A**, PSB FP12 decreased the pH of the NPM and IPM groups rapidly, which was significantly lower than that of the APM group. This result indicated that the AP-deficient environment could enhance the acidification of PSB FP12, so as to decrease the pH of the medium. We also found that the pH of the IAPM and APM group media were significantly higher than those of the IPM and NPM group media (**Figure 1A**), indicating that AP might weaken the acidification of PSB FP12.

Along with the decrease of pH, PSB FP12 rapidly increased AP in the IPM group. The AP of the IPM group significantly increased to 216.50 mg/L at 1 days, 838.33 mg/L at 2 days, and 1,085.00 mg/L at 7 days, respectively. The AP of the IPM group was always significantly higher than that of the IPCK group (**Figure 1B**).

Organic Acid-Targeted Metabolomics of PSB FP12

The organic acid concentrations of the IPM and APM groups at 1 and 2 days were measured. Gluconic acid, malic acid, acetic acid, α-ketoglutaric acid, lactic acid, formic acid, citric acid, and succinic acid were detected, while pyruvate, oxalic acid, tartaric acid, and 2-keto-gluconic acid were not detected. Among them, gluconic acid showed the highest content, followed by malic acid, acetic acid, and α-ketoglutaric acid (**Figure 2**).

The concentrations of organic acids in the IPM group were generally higher than these in the APM group. The concentrations of gluconic acid in the IPM group were 5,302.04 and 7,263.62 μg/ml at 1 and 2 days, respectively, which were significantly higher than these (2,992.47 μg/ml at 1 days, 5,642.76 μg/ml at 2 days) in the APM group. The concentrations of malic acid in the IPM group were 2,671.39 and 3,556.95 μg/ml at 1 and 2 days, respectively, which were significantly higher than these (1,791.03 μg/ml at 1 days, 1,778.81 μg/ml at 2 days) in the APM group. The concentrations of acetic acid and lactic acid in the IPM group were also significantly higher than these in the APM group (**Figure 2**).

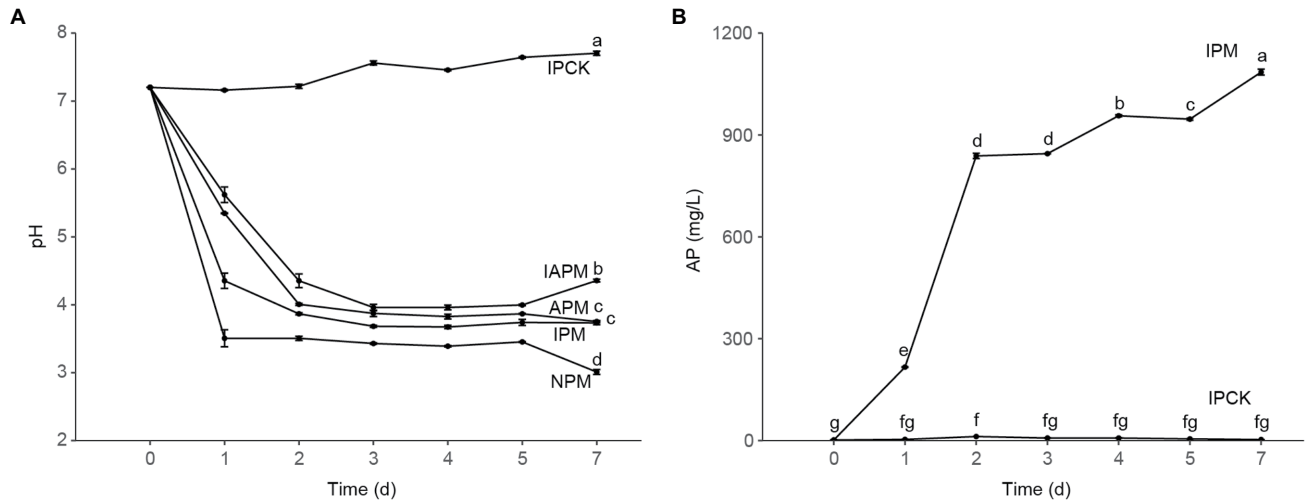


FIGURE 1 | The pH and solubilized available phosphorus (AP). **(A)** The change of pH in mediums with different phosphorus sources. Phosphate-solubilizing bacteria (PSB) FP12 was cultured in four kinds of media, i.e., insoluble phosphorus medium (IPM); available phosphorus medium (APM); insoluble and available phosphorus medium (IAPM); medium without phosphorus (NPM). IPCK is insoluble phosphorus medium without PSB. **(B)** The AP dissolved by phosphate-solubilizing bacteria FP12 in IPM group. Different letters denote significant difference from a Tukey's HSD test ($p < 0.05$).

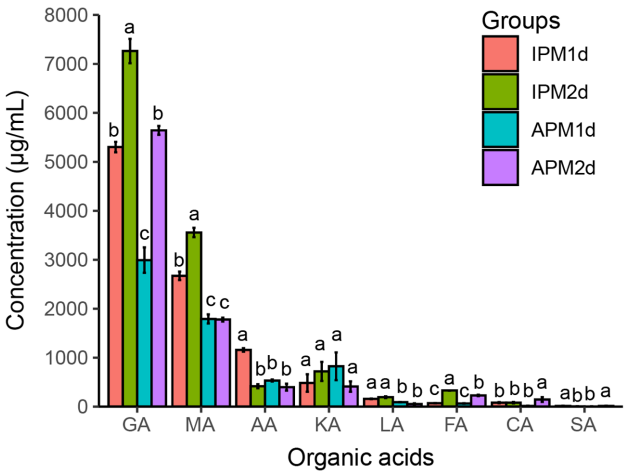


FIGURE 2 | The concentration of organic acids in IPM and APM groups. IPM1d and IPM2d represent 1 day and 2 days after inoculation of phosphate-solubilizing bacteria FP12 in the insoluble phosphorus medium, respectively. APM 1d and APM 2d represent 1 day and 2 days after inoculation of phosphate-solubilizing bacteria FP12 in the available phosphorus medium, respectively. GA, gluconic acid; MA, malic acid; AA, acetic acid; KA, α-ketoglutaric acid; LA, lactic acid; FA, formic acid; CA, citric acid; and SA, succinic acid. Different letters denote significant difference from a Tukey's HSD test ($p < 0.05$).

Genome of PSB FP12

The final assembled genome size was 4.92 Mb with a scaffold N50 of 0.32 Mb, and a GC content of 57.05%. Total 4,714 genes were predicted in the PSB FP12 genome, with a total length of 4.27 Mb, accounting for 86.92% of the genome (**Figure 3**). While compared with the published genomes, PSB FP12 genome showed the highest ANI of 97.84% with the *O. haematophilum* strain FI11154 genome (**Supplementary Table S5**), which indicated that they are same species.

In the PSB FP12 genome, genes for the main enzymes and their coenzymes involved in gluconic acid and 2-keto-gluconic acid synthesis were found to be: glucose dehydrogenase (GDH) gene *gdhB* (Quinoprotein glucose dehydrogenase, FP12_GM000535), pyrroloquinoline quinone (PQQ) synthesis protein series genes pqqA-E (FP12_GM003319, FP12_GM003320, FP12_GM003321, FP12_GM003322, and FP12_GM003323), and gluconate dehydrogenase gene *kduD* (FP12_GM002270; **Figure 3**).

In addition, the indole acetic acid (IAA) synthetic gene [aldehyde dehydrogenase (*ALDH*)] and siderophores biosynthesis genes were also identified in the PSB FP12 genome.

Transcriptome of PSB FP12

The GO enrichment results showed that the differentially expressed genes between the IPM and APM groups may be related to the response of the PSB to the AP-deficient environment, the regulation of their own metabolism, and the synthesis and secretion of organic acids (**Supplementary Figures S3, S4**). KEGG enrichment results showed that the gene expression of the organic acid synthesis pathways, such as the gluconic acid synthesis pathway, the TCA cycle, and the pyruvate metabolic pathway, were upregulated in the AP-deficient environment (**Figure 4**; **Supplementary Table S6**).

In our study, the expression of the GDH gene *gdhB* was significantly higher than that in the APM group at 1 days ($p < 0.0001$), and the expression levels of the PQQ synthesis protein genes, i.e., *pqqB*, *pqqC*, *pqqD*, and *pqqE* in the IP group were higher than those in the APM group (**Figure 4**).

Transcriptome analysis revealed that the expression of genes related to the TCA cycle in the IPM group was generally

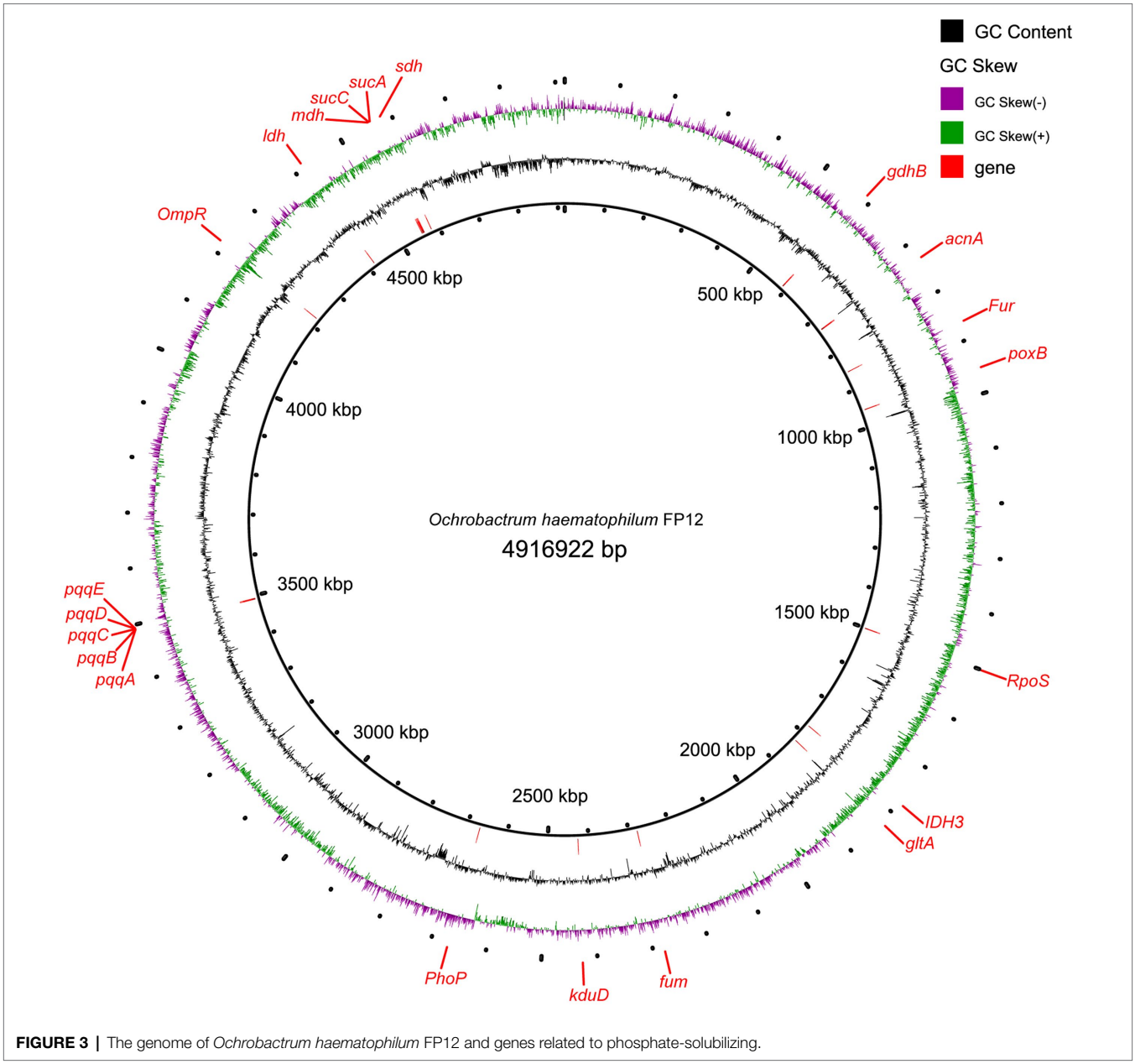


FIGURE 3 | The genome of *Ochrobactrum haematophilum* FP12 and genes related to phosphate-solubilizing.

higher than those in the APM group, such as citrate synthase (CS), aconitic hydratase (ACO), isocitrate dehydrogenase (IDH), α-ketoglutarate dehydrogenase (OGDH), succinyl-CoA synthetase (SUC), succinate dehydrogenase (SDH), and fumarate hydratase (FH) genes (**Figure 4**). However, the expression level of the malate dehydrogenase (MDH) gene, *mdh*, in the IPM group was 671.88 RPKM at 1 days, which was lower than that (1088.42 RPKM) in the APM group, and this gene also showed lower expression levels in the IPM group (217.49 RPKM) than in the APM group (309.99 RPKM) at 2 days (**Figure 4**).

The expression of pyruvate dehydrogenase (POX) gene in the IPM group was significantly higher than that in the APM group ($p < 0.01$; **Figure 4**). The expression of the L-lactate dehydrogenase (LDH) gene in the IPM group was significantly higher than that in the APM group ($p < 0.01$; **Figure 4**).

Bacteria induce an acid tolerance response by producing acid shock proteins and changing cell membrane fluidity. We found that the expression levels of these regulatory proteins in the IPM group were higher than those in the APM group (**Figure 4**). The expression levels of RpoS in IPM1d and IPM2d were 2682.65 RPKM and 5549.95 RPKM, respectively, which were significantly higher than those in APM group (1371.87 RPKM and 3997.41 RPKM; $p < 0.01$). The expression level of OmpR in IPM1d was 584.40 RPKM, which was higher than APM1d (556.62 RPKM); and the expression level of OmpR in IPM2d was 414.22 RPKM, which was significantly higher than that of APM1d (223.08 RPKM; $p < 0.05$; **Figure 4**).

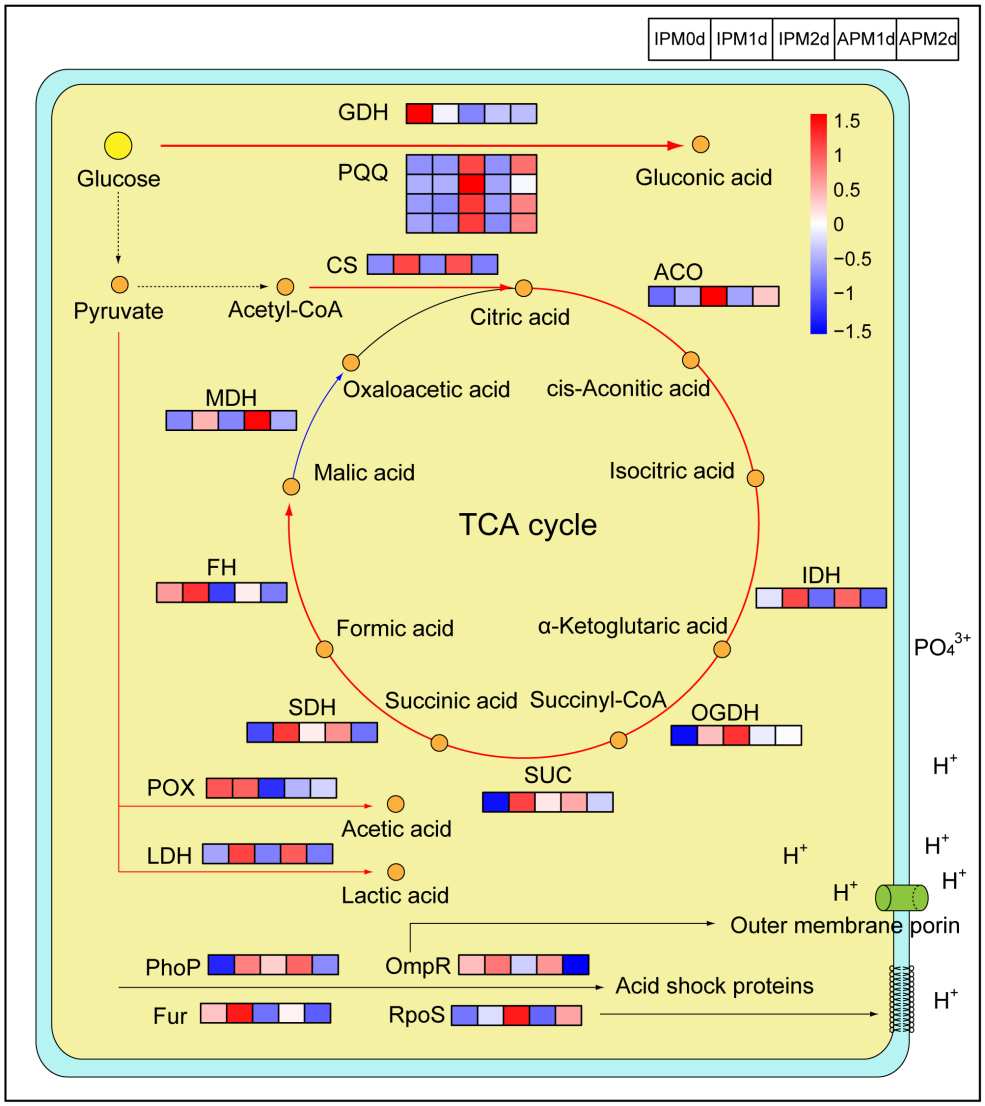


FIGURE 4 | The expression level of genes related to organic acids metabolism and acid tolerance. IPM0d, IPM1d and IPM2d represent 0 day, 1 day, and 2 days after inoculation of phosphate-solubilizing bacteria FP12 in the insoluble phosphorus medium, respectively. APM 1d and APM 2d represent 1 day and 2 days after inoculation of phosphate-solubilizing bacteria FP12 in the available phosphorus medium, respectively. Red represents high expression, and blue represents low expression. GDH, glucose dehydrogenase; PQQ, pyrroloquinoline quinone; CS, citrate synthase; ACO, aconitic hydratase; IDH, isocitrate dehydrogenase; OGDH, α-ketoglutarate dehydrogenase; SUC, succinyl-CoA synthetase; SDH, succinate dehydrogenase; FH, fumarate hydratase; MDH, malate dehydrogenase; POX, pyruvate dehydrogenase; and LDH, L-lactate dehydrogenase.

Treating Barren Rocky Soil With PSB

The six strains of PSB and mixed PSB significantly increased the AP in the barren rocky soil and their AP contents were significantly higher than those in the control group (**Figure 5A**). Among them, PSB FP12 showed the highest P-solubilizing efficiency, which significantly increased the soil AP from 1.07 to 3.08 mg/kg. The mixed PSB also significantly increased the soil AP to 2.76 mg/kg (**Figure 5A**). Meanwhile, we found that PSB treatment significantly reduced soil pH. The pH of BP10, BP11, BP23, FP2, FP12, FP16, and the mixed PSB groups significantly reduced from 8.54 to 8.17, 8.10, 8.19, 8.06, 8.17, 8.16, and 8.13, respectively.

Effects of Sweetpotato Root Exudates on P Solubilization

The addition of sweetpotato root exudates significantly improved the P-solubilizing effect of the PSB (**Figure 5B**). The AP in the BR group increased from 1.07 to 2.45 mg/kg, which was significantly higher than that (2.10 mg/kg) in the BW group. The AP in the BR and BW groups was significantly higher than that in the BCK group (1.26 mg/kg; **Figure 5B**). Meanwhile, we found that the pH of barren rocky soil in the BR group decreased significantly from 8.54 to 8.35, which was significantly lower than that in the BW group (8.56).

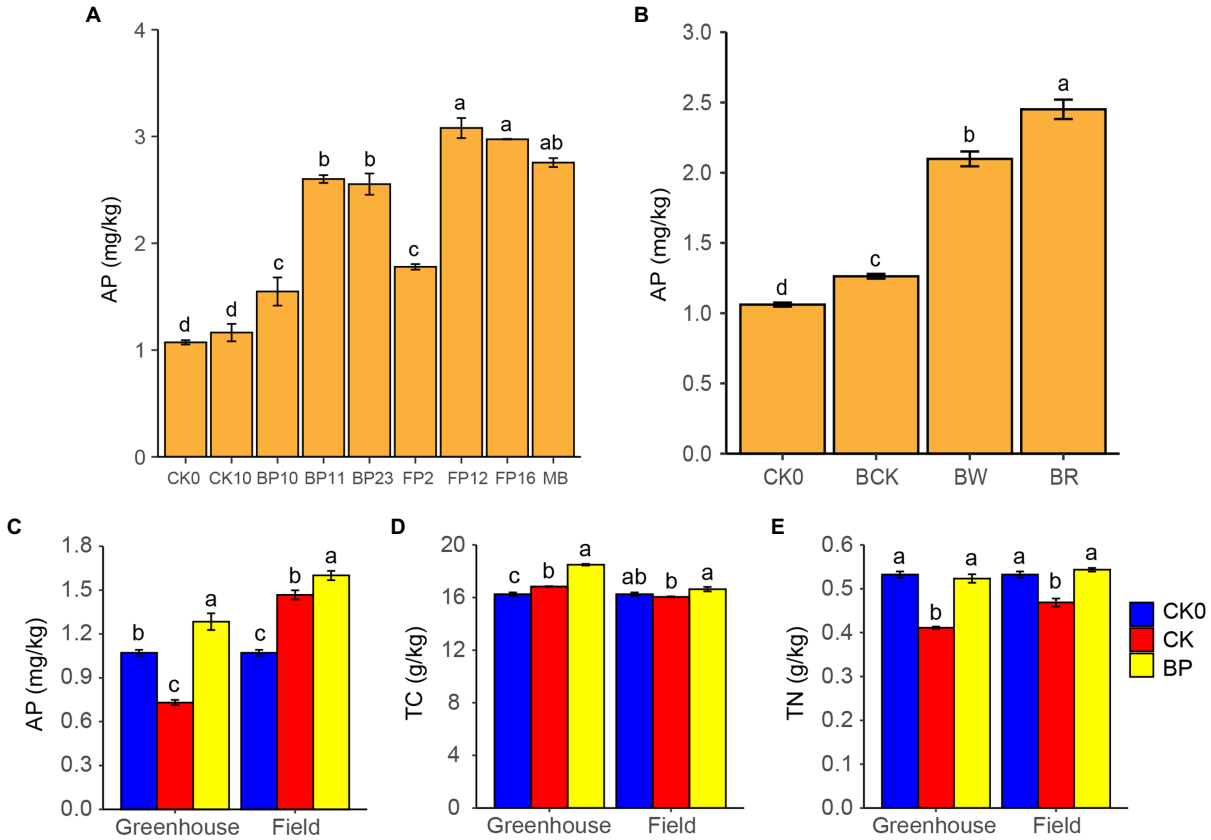


FIGURE 5 | Phosphate-solubilizing bacteria application potential. **(A)** AP in the barren rocky soil after treatment by phosphate-solubilizing bacteria, CK0, control group 0 day; CK10, control group 10 days; MB, mixed phosphate-solubilizing bacteria. **(B)** Effect of sweetpotato root exudates on phosphate-solubilizing, CK0, control group 0 day; BCK, control group without mixed phosphate-solubilizing bacteria 10 days; BW, mixed phosphate-solubilizing bacteria with sterile water; BR, mixed phosphate-solubilizing bacteria with sweetpotato root exudates. **(C-E)** Soil properties in greenhouse and field pot experiments, AP, available phosphorus; TC, total carbon; TN, total nitrogen; CK0 group, control group 0 day; CK group, control group without phosphate-solubilizing bacteria 100 days; and BP group, mixed phosphate-solubilizing bacteria was applied 100 days. Different letters denote significant difference from a Tukey's HSD test ($p < 0.05$).

Greenhouse and Field Pot Experiments

The application of PSB significantly increased the AP in the barren rocky soil. In the greenhouse and field pot experiment, AP in the barren rocky soil increased significantly from 1.07 to 1.28 mg/kg and 1.63 mg/kg in the BP group, respectively, which was both significantly higher than those (0.73 and 1.47 mg/kg) in the CK group ($p < 0.05$; **Figure 5C**). In addition, if the P absorbed by sweetpotato is taken into account (4.52 and 7.10 mg P per sweetpotato in the greenhouse and field pot experiment, respectively), the PSB actually releases more P.

The application of PSB also affected the barren rocky soil properties such as pH, TC, and TN. Applying PSB decreased the pH of the barren rocky soil, which is beneficial to reduce alkalization of barren rocky soil. Meanwhile, applying PSB increased the TC and TN of barren rocky soil (**Figures 5D,E**).

Applying PSB increased the tuberous roots and vines yields, as well as N, P, and K content in sweetpotato. In the greenhouse and field pot experiment, the yields of tuberous roots of sweetpotato were 26.08 and 36.43 g in the BP group, respectively, which were significantly higher than those (16.21 and 29.38 g) in the CK group ($p < 0.05$). Meanwhile, applying PSB increased the ratio of tuberous roots to total biomass of sweetpotato. Moreover, applying PSB even significantly increased the N, P, and K content in the sweetpotato roots and vines (**Table 1**; **Supplementary Figure S5**).

DISCUSSION

Our previous study found that sweetpotato grew well in AP-deficient rocky soil, and some PSB may be present in the sweetpotato rhizosphere (Ding et al., 2020). In this study, we screened six highly efficient PSB from sweetpotato rhizosphere rocky soil. Among them, *O. haematophilum* FP12 showed the highest P-solubilizing ability of 1,085.00 mg/L at 7 days. To our knowledge, it showed higher P-solubilizing ability than most of previously reported PSB (Chen et al., 2006; Sarikhani et al., 2019; Kaur and Kaur, 2020).

Multi-Omics Reveal the Highly Efficient P-Solubilizing Mechanism of PSB FP12

Organic acid secretion is an important P-solubilizing way in PSB (Zeng et al., 2016). We found that the concentrations of

TABLE 1 | Physiological data of sweetpotato in greenhouse and field pot experiments.

		Tuberous roots (g)	Fibrous roots (g)	Vines (g)	Roots (g/kg)			Vines (g/kg)		
					N	P	K	N	P	K
Greenhouse pot	CK	16.21 ± 5.95	2.78 ± 0.98	14.26 ± 4.02	4.65 ± 0.02	0.31 ± 0.01	15.30 ± 0.01	8.84 ± 0.01	0.57 ± 0.00	17.64 ± 0.02
	BP	26.08 ± 4.45*	3.93 ± 1.94	17.89 ± 3.37	7.41 ± 0.00*	0.38 ± 0.00*	16.05 ± 0.03*	12.84 ± 0.01*	0.61 ± 0.00*	18.74 ± 0.01*
Field pot	CK	29.38 ± 6.23	11.37 ± 2.68	26.88 ± 6.51	4.97 ± 0.01	0.54 ± 0.00	16.57 ± 0.01	7.83 ± 0.01	0.61 ± 0.00	17.40 ± 0.01
	BP	36.43 ± 4.76*	10.58 ± 5.02	31.25 ± 6.94	7.54 ± 0.01*	0.57 ± 0.00*	16.91 ± 0.01*	11.36 ± 0.02*	0.63 ± 0.00	17.91 ± 0.03*

CK group, control group without phosphate-solubilizing bacteria 100 days; BP group, mixed phosphate-solubilizing bacteria was applied 100 days.
*$p < 0.05$.

organic acids in the IPM group were generally higher than those in the APM group, especially gluconic acid and malic acid, whose concentrations in the IPM group were 1.29–2.00 times higher than those in the APM group (**Figure 2**). These results may explain why the pH of the IPM group was lower than that of the APM group (**Figure 1A**). Thus, AP-deficiency could promote the secretion of organic acid by PSB, while AP inhibited it. This result is consistent with the previous finding (Zeng et al., 2017).

Direct oxidation of glucose to gluconic acid is thought to be the main P-solubilizing way in Gram-negative bacteria (Sashidhar and Podile, 2010; Luduena et al., 2018). We found that gluconic acid was the most prevalent organic acid in the media (**Figure 2**), which plays a major role in the rapid P-solubilization by PSB FP12. Whereas, a study in the PSB *Burkholderia multivorans* WS-FJ9 found that pyruvate was the main organic acid secreted when solubilizing P (Zeng et al., 2017). Another study in the PSB *Enterobacter cloacae* RW8 found that lactic acid, succinic acid, and citric acid were the main organic acids secreted (Li et al., 2017). These results showed that different PSB would secrete the varied types and concentrations of organic acids to solubilize P.

The genome contains the genetic information that allows bacteria to function. Genes related to various organic acid synthesis, acid shock regulatory proteins (RpoS, OmpR, PhoP, and Fur), acid/alkaline phosphatase, IAA synthesis, and siderophores biosynthesis were identified in the PSB FP12 genome (**Figure 3**). The acid/alkaline phosphatase suggested that PSB FP12 may degrade organic P. And the IAA synthesis and siderophores biosynthesis genes suggested that PSB FP12 may promote plant growth through producing IAA and siderophores, in addition to solubilizing P (Luduena et al., 2018).

A previous study has illustrated the important role of gluconic acid synthesis in P-solubilizing (Farhat et al., 2013). In our study, the expression of the GDH genes in the IPM group were higher than that in the APM group (**Figure 4**), which may result to the higher gluconic acid concentration in the IPM group (**Figure 2**). Our results differ from those in *B. multivorans* WS-FJ9, whose expression of gluconic acid synthesis genes are not affected by P availability (Zeng et al., 2017). This difference may be due to the different P-solubilizing mechanisms among PSB strains. For example, the PSB *B. multivorans* WS-FJ9 mainly secreted pyruvate to solubilize P, not gluconic acid (Zeng et al., 2017).

The TCA cycle produces a variety of organic acids, such as citric acid, α-ketoglutaric acid, succinic acid, and malic acid (Akram, 2014; Vuoristo et al., 2016). Transcriptome analysis revealed that the expression of genes related to the TCA cycle in the IPM group was generally higher than that in the APM group, excepting the MDH gene (**Figure 4**). MDH catalyzes malic acid to oxaloacetate (Sutherland and McAlisterhenn, 1985). Thus, low expression of MDH gene may be the reason why malic acid concentration in the IPM group was higher than that in the APM group (**Figure 2**). Pyruvate can be converted to organic acids such as acetic acid and lactic acid through the pyruvate metabolic pathway (Melo et al., 2018). Transcriptome analysis revealed that the expression of related genes (POX, LDH) was higher than that in the APM group (**Figure 4**), which might induce the higher acetic acid and lactic acid concentration in the IPM group (**Figure 2**).

Bacteria induce an acid tolerance response by producing acid shock proteins and changing cell membrane fluidity. Regulatory proteins such as RpoS, OmpR, PhoP, and Fur can affect the production of acid shock proteins (Hall and Foster, 1996). In addition, RpoS reduces cell membrane fluidity (Spector and Kenyon, 2012), and OmpR directly affects the outer membrane porin OmpC and OmpF to respond to external acids (Csonka and Hanson, 1991). We found that the expression levels of these regulatory proteins in the IPM group were significantly higher than those in the APM group ($p < 0.05$; **Figure 4**). These results indicated that PSB FP12 enhanced its acid tolerance response in the IPM group, which helped it survive in an acidic environment.

Therefore, in an AP-deficient environment, PSB FP12 upregulated the expression of genes in the organic acid synthesis pathway and secreted more organic acids such as gluconic acid, malic acid, and acetic acid (**Figures 2**, **4**). The expression patterns of related genes and the secretion of organic acids in PSB FP12 were different from some of those previously reported PSB, indicating that different PSB have various P-solubilizing mechanisms.

The Screened PSB Significantly Increased the AP in the Barren Rocky Soil From the Laboratory to the Field

The application of PSB is affected by many factors, such as soil properties, rhizosphere colonization, and the local climate (Gyaneshwar et al., 2002). A previous study found that P-solubilizing

microorganisms did not play a direct role in supplementing soil AP for plants in about 70% of the experiments (Gyaneshwar et al., 2002). Therefore, we further studied the application effect of PSB through experiments such as treating the barren rocky soil with PSB and adding sweetpotato root exudates, as well as greenhouse potting and field potting (**Supplementary Table S7**). The barren rocky soil experiment indicated that all of these screened PSB had P-solubilizing effects on the barren rocky soil. Considering that the composite strains have stronger adaptability to complex environments, mixed PSB was used for further experiments (Scheuerl et al., 2020).

We found that adding sweetpotato root exudates improved the P-solubilizing effect of the PSB (**Figure 5B**). Plant root exudates are important sources of nutrients such as carbon and N for rhizosphere microorganisms, and plants can also affect the colonization of rhizosphere microorganisms through root exudates (Sasse et al., 2018). The sweetpotato root exudates might increase the P-solubilizing effect through promoting the colonization of PSB.

The results of the greenhouse and field pot experiments were consistent. The application of PSB increased the AP, TC, and TN in the barren rocky soil (**Figures 5C–E**). And the P-solubilizing effect of PSB was higher than that of most of previously reported PSB (Zhu et al., 2018). In addition, based on the preliminary 16S rRNA sequencing results of our follow-up field experiment, the relative abundance of the PSB could be more than 1%, which was significantly higher than that of the control (unpublished data). Moreover, P is an essential macro element required for plants, which is especially important for early stage root development, stem strength, and yield (Satyaprakash et al., 2017). In this study, the application of PSB increased the yield of sweetpotato tuberous roots, biomass and the ratio of tuberous roots to total biomass, and the content of N, P, and K in the roots and vines (**Table 1**). These results indicated that in the greenhouse and field environments, the PSB could improve the barren rocky soil fertility and promote sweetpotato growth. In adition, some PSB also showed a good growth-promoting effect on a variety of crops, such as *O. haematophilum* could promote the growth of corn and tobacco (Pereira and Castro, 2014; Gao et al., 2016). It would be an interesting and valuable work to study the effect of these screened PSB, especially the FP12, on other crops in future.

Chemical phosphate fertilizers applied to the barren rocky soil is easy to be washed away and cause serious environmental pollution (Huang et al., 2017). In addition, our preliminary research found that commercialized PSB did not work well in the barren rocky soil. Here, the screened PSB can effectively solubilize P and continuously alleviate the AP-deficiency in the barren rocky soil, showing great potential in reducing the application of chemical phosphate fertilizer and the related environmental pollution. Therefore, to pave the way toward their actual application in the rocky soil, more efficient PSB would be screened and the possible interaction among these screened PSB would be studied in future. More importantly, the biosafety potential of those screened PSB, especially the *O. haematophilum* FP12, would also be carefully and thoroughly analyzed before promoting their large-scale applications in the rocky soil.

CONCLUSION

We screened six strains of PSB from sweetpotato rhizosphere rocky soil. Among them, the PSB *O. haematophilum* FP12 had the best P-solubilizing ability, reaching 1,085.00 mg/L at 7 days. The assembled genome of PSB FP12 was 4.92 Mb in length, containing 4,714 genes, including P-solubilizing and plant growth-promoting genes. In AP-deficient environment, PSB FP12 upregulated the expression of organic acid synthesis genes and produced more organic acids, leading to P-solubilization. Moreover, these PSB improved the barren rocky soil fertility and promoted sweetpotato yield in the laboratory and field experiments. This study provides efficient and adaptable PSB strains for barren rocky soil, which will reduce the environmental pollution caused by excessive chemical phosphate fertilizer and promote the environment-friendly agricultural development.

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 at: https://www.ncbi.nlm.nih.gov/, PRJNA662158, PRJNA662170, and MW454804-MW454809.

AUTHOR CONTRIBUTIONS

YD: conceptualization, methodology, data curation, software, writing – original draft, and writing – review and editing. ZY: methodology, data curation, formal analysis, and writing – review and editing. YF: investigation and data curation. SH and YL: investigation. KH: resources and writing – review and editing. HZ: writing – review and editing and funding acquisition. YJ: writing – review and editing, supervision, project administration, and funding acquisition. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by the China Agriculture Research System of MOF and MARA (CARS-10-B24), the National Key R&D Program of China (2018YFF0213505), the National Natural Science Foundation of China (31770395), the Science & Technology Program of Sichuan Province (2017HH0077), and Innovation Academy for Seed Design, CAS.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.761972/full#supplementary-material

REFERENCES

Akram, M. (2014). Citric acid cycle and role of its intermediates in metabolism. *Cell Biochem. Biophys.* 68, 475–478. doi: 10.1007/s12013-013-9750-1

Alikhan, N.-F., Petty, N. K., Zakour, N. L. B., and Beatson, S. A. (2011). BLAST ring image generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics* 12:402. doi: 10.1186/1471-2164-12-402

Anders, S., Pyl, P. T., and Huber, W. (2015). HTSeq—a python framework to work with high-throughput sequencing data. *Bioinformatics* 31, 166–169. doi: 10.1093/bioinformatics/btu638

Balemi, T., and Negisho, K. (2012). Management of soil phosphorus and plant adaptation mechanisms to phosphorus stress for sustainable crop production: a review. *J. Soil Sci. Plant Nutr.* 12, 547–561. doi: 10.4067/s0718-95162012005000015

Besemer, J., Lomsadze, A., and Borodovsky, M. (2001). GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Res.* 29, 2607–2618. doi: 10.1093/nar/29.12.2607

Bi, Q.-F., Li, K.-J., Zheng, B.-X., Liu, X.-P., Li, H.-Z., Jin, B.-J., et al. (2020). Partial replacement of inorganic phosphorus (P) by organic manure reshapes phosphate mobilizing bacterial community and promotes P bioavailability in a paddy soil. *Sci. Total Environ.* 703:134977. doi: 10.1016/j.scitotenv.2019.134977

Buchfink, B., Xie, C., and Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* 12, 59–60. doi: 10.1038/nmeth.3176

Chen, Y. L., Lee, C. C., Lin, Y. L., Yin, K. M., Ho, C. L., and Liu, T. (2015). Obtaining long 16S rDNA sequences using multiple primers and its application on dioxin-containing samples. *BMC Bioinformatics* 16:S13. doi: 10.1186/1471-2105-16-s18-s13

Chen, Y. P., Rekha, P. D., Arun, A. B., Shen, F. T., Lai, W. A., and Young, C. C. (2006). Phosphate solubilizing bacteria from subtropical soil and their tricalcium phosphate solubilizing abilities. *Appl. Soil Ecol.* 34, 33–41. doi: 10.1016/j.apsoil.2005.12.002

Csonka, L. N., and Hanson, A. D. (1991). Prokaryotic osmoregulation: genetics and physiology. *Annu. Rev. Microbiol.* 45, 569–606. doi: 10.1146/annurev.mi.45.100191.003033

Ding, Y., Jin, Y., He, K., Yi, Z., Tan, L., Liu, L., et al. (2020). Low nitrogen fertilization alter rhizosphere microorganism community and improve sweetpotato yield in a nitrogen-deficient rocky soil. *Front. Microbiol.* 11:678. doi: 10.3389/fmicb.2020.00678

Elhaissoufi, W., Ghoulam, C., Barakat, A., Zeroual, Y., and Bargaz, A. (2021). Phosphate bacterial solubilization: a key rhizosphere driving force enabling higher P use efficiency and crop productivity. *J. Adv. Res.* doi: 10.1016/j.jare.2021.08.014

Farhat, M. B., Fourati, A., and Chouayekh, H. (2013). Coexpression of the pyrroloquinoline quinone and glucose dehydrogenase genes from *Serratia marcescens* CTM 50650 conferred high mineral phosphate-solubilizing ability to *Escherichia coli*. *Appl. Biochem. Biotechnol.* 170, 1738–1750. doi: 10.1007/s12010-013-0305-0

Gao, L., Kong, F., Feng, C., Wang, J., Gao, J., Shen, G., et al. (2016). Isolation, characterization, and growth promotion of phosphate-solubilizing bacteria associated with *Nicotiana tabacum* (Tobacco). *Pol. J. Environ. Stud.* 25, 993–1003. doi: 10.15244/pjoes/61820

Granada, C. E., Passaglia, L. M. P., de Souza, E. M., and Sperotto, R. A. (2018). Is phosphate solubilization the forgotten child of plant growth-promoting rhizobacteria? *Front. Microbiol.* 9:2054. doi: 10.3389/fmicb.2018.02054

Gyaneshwar, P., Kumar, G. N., Parekh, L. J., and Poole, P. S. (2002). Role of soil microorganisms in improving P nutrition of plants. *Plant Soil* 245, 83–93. doi: 10.1023/A:1020663916259

Hall, H. K., and Foster, J. W. (1996). The role of fur in the acid tolerance response of *Salmonella typhimurium* is physiologically and genetically separable from its role in iron acquisition. *J. Bacteriol.* 178, 5683–5691. doi: 10.1128/jb.178.19.5683-5691.1996

Huang, J., Xu, C.-C., Ridoutt, B. G., Wang, X.-C., and Ren, P.-A. (2017). Nitrogen and phosphorus losses and eutrophication potential associated with fertilizer application to cropland in China. *J. Clean. Prod.* 159, 171–179. doi: 10.1016/j.jclepro.2017.05.008

Kaur, R., and Kaur, S. (2020). Variation in the phosphate solubilizing bacteria from virgin and the agricultural soils of Punjab. *Curr. Microbiol.* 77, 2118–2127. doi: 10.1007/s00284-020-02080-6

Kim, D., Langmead, B., and Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* 12, 357–360. doi: 10.1038/nmeth.3317

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923

Li, X., Wang, X., Chen, X., Cai, L., Zeng, Q., Shu, J., et al. (2017). Transcriptome profiling analysis of the phosphate-solubilizing mechanism of the white clover rhizosphere strain RW8. *Acta Pratacul. Sin.* 26, 168–179. doi: 10.11686/cyxb2016399

Li, H., Yao, T., Zhang, R., Zhang, J., Li, Z., Rong, L., et al. (2019). Isolation and screening of phosphate-solubilizing bacteria from the rhizosphere of *Trifolium pratense* and culture medium optimization. *Acta Pratacul. Sin.* 28, 170–179. doi: 10.11686/cyxb2018064

Luduena, L. M., Anzuay, M. S., Angelini, J. G., McIntosh, M., Becker, A., Rupp, O., et al. (2018). Strain *Serratia* sp. S119: a potential biofertilizer for peanut and maize and a model bacterium to study phosphate solubilization mechanisms. *Appl. Soil Ecol.* 126, 107–112. doi: 10.1016/j.apsoil.2017.12.024

Luo, R. B., Liu, B. H., Xie, Y. L., Li, Z. Y., Huang, W. H., Yuan, J. Y., et al. (2012). SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience* 1:18. doi: 10.1186/2047-217x-1-18

Manning, D. A. C., and Theodoro, S. H. (2018). Enabling food security through use of local rocks and minerals. *Extract. Indus. Soc.* 7, 480–487. doi: 10.1016/j.exis.2018.11.002

Melo, N., Mulder, K. C., Nicola, A. M., Carvalho, L. S., Menino, G. S., Mulinari, E., et al. (2018). Effect of pyruvate decarboxylase knockout on product distribution using *Pichia pastoris* (*Komagataella phaffii*) engineered for lactic acid production. *Bioengineering* 5:17. doi: 10.3390/bioengineering5010017

Murphy, J., and Riley, J. P. (1962). A modified single solution method for the determination of phosphate in natural waters. *Anal. Chim. Acta* 27, 31–36. doi: 10.1016/S0003-2670(00)88444-5

Natarajan, V. P., Zhang, X., Morono, Y., Inagaki, F., and Wang, F. (2016). A modified SDS-based DNA extraction method for high quality environmental DNA from seafloor environments. *Front. Microbiol.* 7:986. doi: 10.3389/fmicb.2016.00986

Parastesh, F., Alikhani, H. A., and Etesami, H. (2019). Vermicompost enriched with phosphate–solubilizing bacteria provides plant with enough phosphorus in a sequential cropping under calcareous soil conditions. *J. Clean. Prod.* 221, 27–37. doi: 10.1016/j.jclepro.2019.02.234

Park, H. J., Kim, S. U., Jung, K. Y., Lee, S., Choi, Y. D., Owens, V. N., et al. (2021). Cadmium phytoavailability from 1976 through 2016: changes in soil amended with phosphate fertilizer and compost. *Sci. Total Environ.* 762:143132. doi: 10.1016/j.scitotenv.2020.143132

Pereira, S. I., and Castro, P. M. (2014). Diversity and characterization of culturable bacterial endophytes from *Zea mays* and their potential as plant growth-promoting agents in metal-degraded soils. *Environ. Sci. Pollut. Res. Int.* 21, 14110–14123. doi: 10.1007/s11356-014-3309-6

Rio, D. C., Ares, M., Hannon, G. J., and Nilsen, T. W. (2010). Purification of RNA using TRIzol (TRI reagent). *Cold Spring Harb. Protoc.* 2010:pdb.prot5439. doi: 10.1101/pdb.prot5439

Roberts, T. L., and Johnston, A. E. (2015). Phosphorus use efficiency and management in agriculture. *Resour. Conserv. Recycl.* 105, 275–281. doi: 10.1016/j.resconrec.2015.09.013

Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2009). edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140. doi: 10.1093/bioinformatics/btp616

Sanders, E. R. (2012). Aseptic laboratory techniques: plating methods. *J. Vis. Exp.* 63:e3064. doi: 10.3791/3064

Sarikhani, M. R., Khoshru, B., and Greiner, R. (2019). Isolation and identification of temperature tolerant phosphate solubilizing bacteria as a potential microbial fertilizer. *World J. Microbiol. Biotechnol.* 35:126. doi: 10.1007/s11274-019-2702-1

Sashidhar, B., and Podile, A. R. (2010). Mineral phosphate solubilization by rhizosphere bacteria and scope for manipulation of the direct oxidation pathway involving glucose dehydrogenase. *J. Appl. Microbiol.* 109, 1–12. doi: 10.1111/j.1365-2672.2009.04654.x

Sasse, J., Martinoia, E., and Northen, T. (2018). Feed your friends: do plant exudates shape the root microbiome? *Trends Plant Sci.* 23, 25–41. doi: 10.1016/j.tplants.2017.09.003

Satyaprakash, M., Nikitha, T., Reddi, E., Sadhana, B., and Vani, S. S. (2017). Phosphorous and phosphate solubilising bacteria and their role in plant nutrition. *Int. J. Curr. Microbiol. App. Sci.* 6, 2133–2144. doi: 10.20546/ijcmas.2017.604.251

Scheuerl, T., Hopkins, M., Nowell, R. W., Rivett, D. W., Barraclough, T. G., and Bell, T. (2020). Bacterial adaptation is constrained in complex communities. *Nat. Commun.* 11:754. doi: 10.1038/s41467-020-14570-z

Spector, M. P., and Kenyon, W. J. (2012). Resistance and survival strategies of *Salmonella enterica* to environmental stresses. *Food Res. Int.* 45, 455–481. doi: 10.1016/j.foodres.2011.06.056

Sutherland, P., and McAlisterhenn, L. (1985). Isolation and expression of the *Escherichia coli* gene encoding malate dehydrogenase. *J. Bacteriol.* 163, 1074–1079. doi: 10.1128/jb.163.3.1074-1079.1985

van Hees, P. A. W., Dahlén, J., Lundström, U. S., Borén, H., and Allard, B. (1999). Determination of low molecular weight organic acids in soil solution by HPLC. *Talanta* 48, 173–179. doi: 10.1016/S0039-9140(98)00236-7

Vuoristo, K. S., Mars, A. E., Sanders, J. P. M., Eggink, G., and Weusthuis, R. A. (2016). Metabolic engineering of TCA cycle for production of chemicals. *Trends Biotechnol.* 34, 191–197. doi: 10.1016/j.tibtech.2015.11.002

Wei, Y., Zhao, Y., Lu, Q., Cao, Z., and Wei, Z. (2018). Organophosphorus-degrading bacterial community during composting from different sources and their roles in phosphorus transformation. *Bioresour. Technol.* 264, 277–284. doi: 10.1016/j.biortech.2018.05.088

Wilkinson, B. H., McElroy, B. J., Kesler, S. E., Peters, S. E., and Rothman, E. D. (2009). Global geologic maps are tectonic speedometers-rates of rock cycling from area-age frequencies. *Geol. Soc. Am. Bull.* 121, 760–779. doi: 10.1130/B26457.1

Ye, J., Zhang, Y., Cui, H. H., Liu, J. W., Wu, Y. Q., Cheng, Y., et al. (2018). WEGO 2.0: a web tool for analyzing and plotting GO annotations, 2018 update. *Nucleic Acids Res.* 46, W71–W75. doi: 10.1093/nar/gky400

Yh, Y. (2014). Isolation and identification of phosphate-solubilizing microorganism from Soil. Henan Agricultural University.

Yoon, S.-H., Ha, S.-M., Kwon, S., Lim, J., Kim, Y., Seo, H., et al. (2017a). Introducing EzBioCloud: a taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. *Int. J. Syst. Evol. Microbiol.* 67, 1613–1617. doi: 10.1099/ijsem.0.001755

Yoon, S.-H., Ha, S.-M., Lim, J., Kwon, S., and Chun, J. (2017b). A large-scale evaluation of algorithms to calculate average nucleotide identity. *Antonie Van Leeuwenhoek* 110, 1281–1286. doi: 10.1007/s10482-017-0844-4

Yu, L. Y., Huang, H. B., Wang, X. H., Li, S., Feng, N. X., Zhao, H. M., et al. (2019). Novel phosphate-solubilising bacteria isolated from sewage sludge and the mechanism of phosphate solubilisation. *Sci. Total Environ.* 658, 474–484. doi: 10.1016/j.scitotenv.2018.12.166

Zeng, Q., Wu, X., Wang, J., and Ding, X. (2017). Phosphate solubilization and gene expression of phosphate-solubilizing bacterium *Burkholderia multivorans* WS-FJ9 under different levels of soluble phosphate. *J. Microbiol. Biotechnol.* 27, 844–855. doi: 10.4014/jmb.1611.11057

Zeng, Q., Wu, X., and Wen, X. (2016). Effects of soluble phosphate on phosphate-solubilizing characteristics and expression of GCD gene in *Pseudomonas frederiksbergensis* JW-SD2. *Curr. Microbiol.* 72, 198–206. doi: 10.1007/s00284-015-0938-z

Zhu, J., Li, M., and Whelan, M. (2018). Phosphorus activators contribute to legacy phosphorus availability in agricultural soils: a review. *Sci. Total Environ.* 612, 522–537. doi: 10.1016/j.scitotenv.2017.08.095

Zitouni, H., Hssaini, L., Ouaabou, R., Viuda-Martos, M., Hernández, F., Ercisli, S., et al. (2020). Exploring antioxidant activity, organic acid, and phenolic composition in strawberry tree fruits (*Arbutus unedo* L.) growing in Morocco. *Plants* 9:1677. doi: 10.3390/plants9121677

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.

Glossary

AP	Available phosphorus
P	Phosphorus
PSB	Phosphate-solubilizing bacteria
TP	Total phosphorus
ANI	Average Nucleotide Identity
RPKM	Reads Per Kilobase per Million mapped reads
K	Potassium
N	Nitrogen
TC	Total carbon
TN	Total nitrogen
D/d	The ratio of the P-solubilizing circle to the diameter of the round hole
PQQ	Pyrroloquinoline quinone
IAA	Indole acetic acid
ALDH	Aldehyde dehydrogenase
GDH	Glucose dehydrogenase
CS	Citrate synthase
ACO	Aconitic hydratase
IDH	Isocitrate dehydrogenase
OGDH	α-ketoglutarate dehydrogenase
SUC	Succinyl-CoA synthetase
SDH	Succinate dehydrogenase
FH	Fumarate hydratase
MDH	Malate dehydrogenase
POX	Pyruvate dehydrogenase
LDH	L-lactate dehydrogenase

frontiers in Genetics

ORIGINAL RESEARCH
published: 13 January 2022
doi: 10.3389/fgene.2021.798535

Multi-Omics Analysis of Lipid Metabolism for a Marine Probiotic *Meyerozyma guilliermondii* GXDK6 Under High NaCl Stress

Huijie Sun[1], *Xinghua Cai*[1], *Bing Yan*[2], *Huashan Bai*[1], *Duotao Meng*[1], *Xueyan Mo*[1], *Sheng He*[3], *Guijiao Su*[1] *and Chengjian Jiang*[1,2]*

[1]State Key Laboratory for Conservation and Utilization of Subtropical Agro-bioresources, Guangxi Research Center for Microbial and Enzyme Engineering Technology, College of Life Science and Technology, Guangxi University, Nanning, China, [2]Guangxi Key Lab of Mangrove Conservation and Utilization, Guangxi Mangrove Research Center, Guangxi Academy of Sciences, Beihai, China, [3]Guangxi Birth Defects Prevention and Control Institute, Maternal and Child Health Hospital of Guangxi Zhuang Autonomous Region, Nanning, China.

Investigating microbial lipid regulation contributes to understanding the lipid-dependent signal transduction process of cells and helps to improve the sensitivity of microorganisms to environmental factors by interfering with lipid metabolism, thus beneficial for constructing advanced cell factories of novel molecular drugs. Integrated omics technology was used to systematically reveal the lipid metabolism mechanism of a marine *Meyerozyma guilliermondii* GXDK6 under high NaCl stress and test the sensitivity of GXDK6 to antibiotics when its lipid metabolism transformed. The omics data showed that when GXDK6 perceived 10% NaCl stress, the expression of *AYR1* and NADPH-dependent 1-acyldihydroxyacetone phosphate reductase was inhibited, which weaken the budding and proliferation of cell membranes. This finding was further validated by decreased 64.39% of OD_{600} under 10% NaCl stress when compared with salt-free stress. In addition, salt stress promoted a large intracellular accumulation of glycerol, which was also verified by exogenous addition of glycerol. Moreover, NaCl stress remarkably inhibited the expression of drug target proteins (such as lanosterol 14-alpha demethylase), thereby increasing sensitivity to fluconazole. This study provided new insights into the molecular mechanism involved in the regulation of lipid metabolism in *Meyerozyma guilliermondii* strain and contributed to developing new methods to improve the effectiveness of killing fungi with lower antibiotics.

Keywords: ***Meyerozyma guilliermondii*, lipid metabolism, integrated omics technology, signal transduction, antibiotics**

OPEN ACCESS

Edited by:
Om Prakash,
National Centre for Cell Science, India

Reviewed by:
Xuehuan Feng,
University of Nebraska-Lincoln, United States
Pan Liao,
Purdue University, United States

***Correspondence:**
Chengjian Jiang
jiangcj0520@vip.163.com

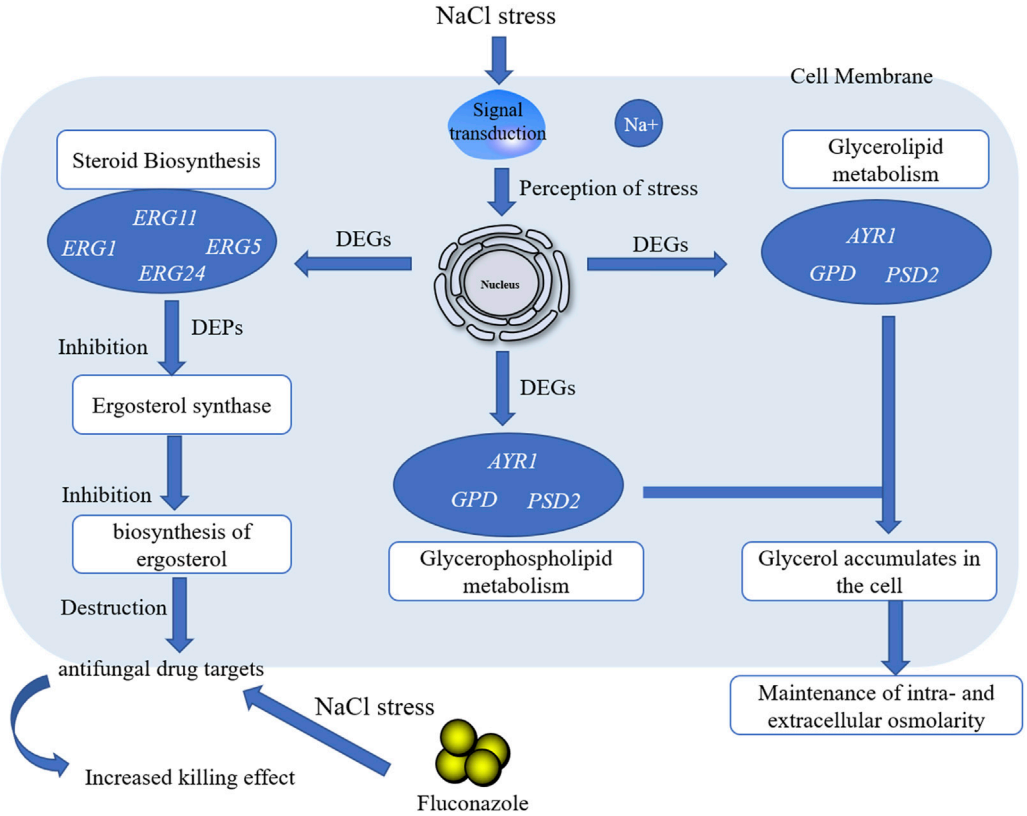
Specialty section:
This article was submitted to Evolutionary and Genomic Microbiology, a section of the journal Frontiers in Genetics

Received: *20 October 2021*
Accepted: *24 December 2021*
Published: *13 January 2022*

Citation:
Sun H, Cai X, Yan B, Bai H, Meng D, Mo X, He S, Su G and Jiang C (2022) Multi-Omics Analysis of Lipid Metabolism for a Marine Probiotic Meyerozyma guilliermondii *GXDK6 Under High NaCl Stress. Front. Genet. 12:798535. doi: 10.3389/fgene.2021.798535*

INTRODUCTION

Microbial lipid metabolism plays a key role in cell signaling, osmotic pressure balance, and the construction of cell membrane structures (Wang S. et al., 2021; Klug and Daum, 2014). Interfering with the lipid metabolism of microorganisms could significantly change their sensitivity to the environment and transform their life activities, thereby contributing to the targeted development of new microbial cell factories (Ferreira et al., 2018). The lipid metabolism in the Kyoto Encyclopedia of Genes and Genomes (KEGG) database is currently known to have 17 branch pathways. Among



GRAPHICAL ABSTRACT |

them, steroid biosynthesis and sphingolipid metabolism are relevant to the microorganisms' cytoskeleton, which regulates their membrane fluidity and permeability (Kraft, 2017; Xu and Li, 2020). Glycerolipid metabolism is closed related to the construction of cell membrane structure (phospholipids and other lipids, varying by organelle type) and energy storage (primarily in the form of triacylglycerol) (Zhang and Reue, 2017). Glycerophospholipid metabolism is mainly linked to cell growth and proliferation (Nebauer et al., 2004). In addition, fatty acid metabolism (such as fatty acid biosynthesis, fatty acid degradation, fatty acid elongation, and biosynthesis of unsaturated fatty acids) are closely related to microbial cell signal transduction, and they are widely considered to be responsible for participating in and responding to extracellular environmental signal factors (Ross et al., 2003; Fan et al., 2016). These pathways of lipid metabolism are known to be indispensable for the basic life activities of microorganisms by supporting microbes to make precise physiological responses in different environments. Therefore, the lipid metabolism of microorganisms is important to forming the cytoskeleton and responding to environmental signal factors, thus enabling microorganisms to exhibit unique life activities in different environments.

The microbial lipid metabolism mechanism has gained increasing interest. Guo *et al.* reported that when *Saccharomyces cerevisiae* was stressed by strong acid, its lipid products were remodeled, which was helpful for improving cell viability under acid stress (Guo et al., 2018). Randez-Gil *et al.* demonstrated that lipid metabolism was involved in regulating the composition of cell membranes in *S. cerevisiae* under low-temperature conditions (Randez-Gil et al., 2018); Zhou *et al.* engineered *S. cerevisiae* to produce L-alkenes by manipulating fatty acid metabolism, enzyme selection, and the electron transfer system and expressing a transporter (Zhou et al., 2018). Therefore, the regulation of lipid metabolism is a useful method for microorganisms to enhance their survival ability under different adverse environments. However, few reports on the mechanism of lipid metabolism focused on how the lipids work and/or how this mechanism contributed to cells (Casanovas et al., 2015; Wang et al., 2018; Lu et al., 2021). A global regulatory network of lipid metabolism has not been revealed yet, resulting in random transformation, which is not conducive to accurately constructing advanced cell factories (Galafassi et al., 2015; Qi et al., 2017; Tian et al., 2019; Li C. et al., 2021). Therefore, mapping the lipid regulation network of microorganisms is an important prerequisite for constructing advanced cell factories by interfering with lipid metabolism.

In the previous study, a marine *Meyerozyma guilliermondii* GXDK6 that contained abundant lipid metabolism genes was obtained from subtropical marine mangrove microorganisms,

and it showed excellent salt-tolerant survivability (14% NaCl and 18% KCl) (Mo et al., 2021). On the basis of its typical physicochemical characteristics, GXDK6 was hypothesized to have a unique regulatory network of lipid metabolism. GXDK6 can precisely regulate the expression of genes and/or proteins related to lipid metabolism to support its survival under salt stress conditions, which lead to greater sensitivity to certain environmental factors. The lipid regulation mechanism of GXDK6 under high-salt stress was systematically investigated by using an integrative omics technology (whole-genomic, transcriptomic, and proteomic) to test these hypotheses. The systematic analysis of the regulation network of lipid metabolism in *M. guilliermondii* strain provide a feasible reference for the construction of lipid cell factories.

MATERIALS AND METHODS

Experimental Strain

A yeast *M. guilliermondii* GXDK6 was obtained from the subtropical marine mangrove sediments, and was released by China General Microbiological Culture Collection Center (CGMCC) with the preservation number CGMCC No. 16007.

Effect of NaCl Stress on the Survival of *M. guilliermondii* Strain

GXDK6 was cultured at different NaCl concentrations (0, 5, and 10%) with 37°C and 200 rpm for 0–48 h, respectively. Its growth curve was evaluated by turbidimetric method. The cells were collected by centrifugation at 12,000 rpm for 10 min and then washed three times with sterile saline (0, 5, and 10% NaCl). The GXDK6 was further prepared into cell suspension and observed using scanning electron microscopy (SEM, Thermoelectric Company, American) to explore the influence of NaCl stress on the morphology.

Whole-Genome Analysis for Lipid Metabolism Genes in *M. guilliermondii* Strain

GXDK6 was incubated by using an enrichment technique and GBM liquid medium (0.2% yeast extract, 0.2% beef extract, 0.5% polypeptone, 0.6% sucrose, pH 7.0) with 200 rpm and 30°C for 8 h. The cells were then collected by centrifugation at 8,000 rpm and 4°C for 10 min and washed repeatedly with 0.1 mol/L PBS buffer. The purity of DNA in GXDK6 (extracted by CTAB method) was verified by PCR (Polymerase Chain Reaction) and agarose gel electrophoresis, and then the ITS gene was amplified by PCR (Mo et al., 2021). The ITS sequencing data of GXDK6 were deposited to the National Microbiology Data Center database (http://nmdc.cn) under the Accession Number of NMDCN000022O. The whole-genome sequencing and analysis of GXDK6 were performed by the BGI Genomics Co., Ltd (Shenzhen, China). An online software FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc) was used for the quality control of the second-generation sequencing downtime data. To determine the percentage of single-copy genes in the total single-copy genes, an online software BUSCO (http://busco.ezlab.org, v3.0.2) was conducted to complete the sequence comparison of the genome sequences. The genes related to lipid metabolism in GXDK6 were blasted and analyzed by whole genome method. The pathway enrichment analysis of genes regulating lipid metabolism was annotated according to the KEGG database.

Transcriptome Analysis of *M. guilliermondii* Strain

GXDK6 was incubated for 16 h at 0, 5, and 10% NaCl, respectively. The cells were then collected by centrifugation at 12,000 rpm for 10 min. Total RNA was extracted using Trizol method (Villa-Rodríguez et al., 2018) (Trizol reagent kit, Invitrogen, Carlsbad, CA, United States). RNA quality was assessed on an Agilent 2,100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, United States) and checked using rnase free agarose gel electrophoresis. After total RNA was extracted, eukaryotic mRNA was enriched by Oligo (dT) beads, while prokaryotic mRNA was enriched by removing rRNA by Ribo-ZeroTM Magnetic Kit (Epicentre, Madison, WI, United States). Then the enriched mRNA was fragmented into short fragments using fragmentation buffer and reverse transcripted into cDNA with random primers. Second-strand cDNA were synthesized by DNA polymerase I, rnase H, dNTP and buffer. Then the cDNA fragments were purified with QiaQuick PCR extraction kit (Qiagen, Venlo, Netherlands), end repaired, poly(A) added, and ligated to Illumina sequencing adapters. The ligation products were size selected by agarose gel electrophoresis, PCR amplified, and sequenced using Illumina HiSeq2500 by Gene Denovo Biotechnology Co. (Guangzhou, China). RNAs differential expression analysis was performed by DESeq2 (Love et al., 2014) software between two different groups (and by edgeR between two samples) (Robinson et al., 2010). The genes/transcripts with the parameter of false discovery rate (FDR) below 0.05 and absolute fold change≥2 were considered differentially expressed genes/transcripts. All samples for transcriptome sequencing were set up in three parallel.

Proteomic Analysis of *M. guilliermondii* Strain

The incubation and cell collection of GXDK6 were consistent with the above method. The proteins in GXDK6 were extracted by SDT Lysis methods (Zhu et al., 2014) and separated on 12.5% SDS-PAGE gel. Protein bands were visualized by Coomassie Blue R-250 staining. The detergent, DTT and other low-molecular-weight components were removed using UA buffer by repeated ultrafiltration (Sartorius, 30 kD). Then 100 μL iodoacetamide (100 mM IAA in UA buffer) was added to block reduced cysteine residues. After washing, the protein suspensions were digested with 4 μg trypsin (Promega) in 40 μL 0.1 M TEAB buffer overnight at 37°C, and the resulting peptides were collected as a filtrate. The peptide content was estimated by UV light spectral density at 280 nm using an extinctions coefficient of 1.1

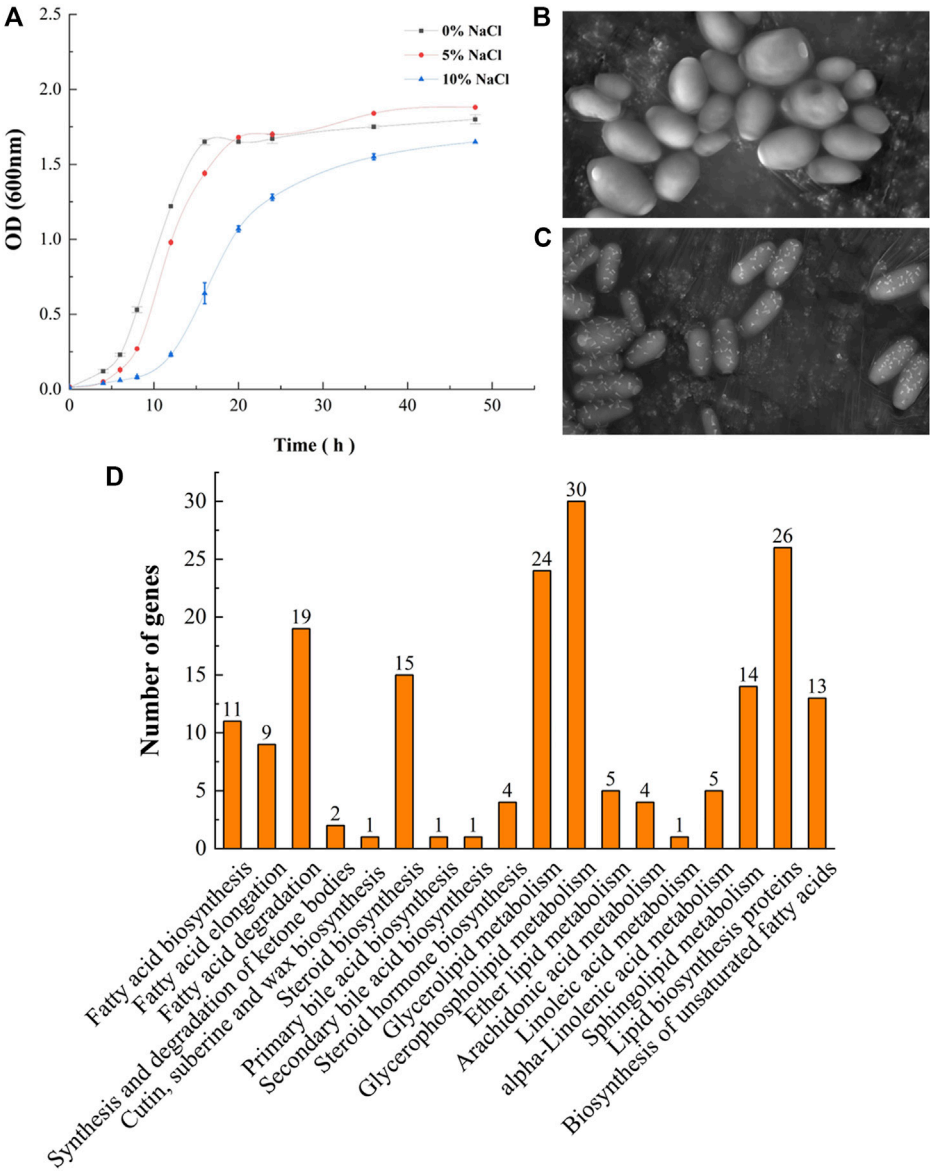


FIGURE 1 | Growth of GXDK6 under NaCl stress. **(A)** Growth curve of GXDK6 under different NaCl stresses; **(B)** morphology of GXDK6 under 0% NaCl; **(C)** morphology of GXDK6 under 10% NaCl; **(D)** Gene enrichment analysis of lipid metabolism.

of 0.1% (g/L) solution. Sequencing of the extracted proteins was conducted by Gene Denovo Biotechnology Co. (Guangzhou, China) by using a tandem mass tag (TMT)-based quantitative proteomics (Myers et al., 2019). TMT labeled peptides were fractionated by RP chromatography using the Agilent 1,260 infinity II HPLC. LC-MS/MS analysis was performed on a Q Exactive plus mass spectrometer (Thermo Fisher Scientific) that was coupled to Easy nLC (Thermo Fisher Scientific) for 60/90 min. MS/MS raw files were processed using MASCOT engine (Matrix Science, London, United Kingdom; version 2.6) embedded into Proteome Discoverer 2.2, and searched against the NCBInr/UniProt database. Proteins with Fold change>1.2 and p value (Student's t test) < 0.05 were considered as differentially expressed proteins.

Fluconazole Resistance Test of *M. guilliermondii* Strain

GXDK6 was inoculated under 0, 5 and 10% NaCl stress containing fluconazole (64 μg/ml). The culture conditions under 0, 5 and 10% NaCl stress without fluconazole were set as the controls. Turbidity method and spread plate method was used to investigate the effect of salt stimulation on the drug resistance ability of GXDK6.

RT-qPCR Verification of Transcriptome Data

According to transcriptome and whole-genome data, *GUT1* (encoding glycerol kinase), *ADH7* (encoding NADP-dependent alcohol dehydrogenase 7), *GPP1* (encoding glycerol-1-phosphate phosphohydrolase 1) were selected and designed (Primer Premier 5.0 software) for qPCR verification in this work (**Supplementary Table**

TABLE 1 | KEGG pathway enrichment analysis of lipid metabolism in GXDK6.

Pathway	Number of genes
Fatty acid biosynthesis	11
Fatty acid elongation	9
Fatty acid degradation	19
Synthesis and degradation of ketone bodies	2
Cutin, suberine and wax biosynthesis	1
Steroid biosynthesis	15
Primary bile acid biosynthesis	1
Secondary bile acid biosynthesis	1
Steroid hormone biosynthesis	4
Glycerolipid metabolism	24
Glycerophospholipid metabolism	30
Ether lipid metabolism	5
Arachidonic acid metabolism	4
Linoleic acid metabolism	1
alpha-Linolenic acid metabolism	5
Sphingolipid metabolism	14
Lipid biosynthesis proteins	26
Biosynthesis of unsaturated fatty acids	13

S18) [the internal reference gene refers to Díaz et al. (2013)]. The reaction volume of RT-qPCR was 20 μL. The mixture was heated to 95°C for 2 min, followed by 40 cycles of denaturation at 95°C for 15 s, and annealing/extension at 60°C (depending on the primers) for 30 s. The cycling temperature was then increased by 0.3°C every 10 s from 63 to 95°C to obtain the melting curve. Three sets of parallel trials were set up for each gene. Real-time PCR was carried out using a high-throughput real-time fluorescence quantitative PCR instrument ROCHE 480 (ROCHE, United States). Meanwhile, fluorescence detection was performed on LightCycler® 480 Software release 1.5.1.62 SP3, and the number of gene amplification cycles was counted in the Software. Relative quantitation was calculated using the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001).

Effect of Adding Exogenous Glycerol on the Growth of *M. guilliermondii* Strain

Exogenous addition of glycerol was used to verify the transcriptome and proteome results. The salt tolerance of GXDK6 was investigated by adding 120 mg/L, 1,200 mg/L and 12,000 mg/L of exogenous glycerol to salt-containing medium. The liquid medium without glycerol and containing 0, 5 and 10% NaCl was also set up as a control. Subsequently, OD_{600} was detected to evaluate the effect of adding exogenous glycerol on the growth of GXDK6.

Data Analysis

Experimental data were processed with SPSS 25, Origin 2018, Excel 2019, TBtools, and Diamond software. The relevant query databases were Uniprot, SWISS-MODEL, KEGG, and NCBI. Significant difference was at a p-value < 0.05.

RESULTS AND DISCUSSION

The Growth Status was Changed When *M. guilliermondii* Strain Perceived Salt Stress

The lag phase of GXDK6 under 5–10% NaCl stress was prolonged to 4 h, which was approximately 2 h longer than that of the controls, indicating that the DNA replication and transcription of GXDK6 were inhibited by NaCl stress (**Figure 1A**). Besides, the growth of GXDK6 in log phase was remarkably inhibited by NaCl stress. With the incubation time at 16 h as an example, the OD_{600} of GXDK6 from non-NaCl stress condition was 1.567; it decreased to 1.281 (decreased by 18.25%) under 5% NaCl stress and 0.558 under 10% NaCl stress (decreased by 64.39%), suggesting that the metabolism and protein expression of GXDK6 were transformed by NaCl stress. However, the OD_{600} value of GXDK6 in stable phase showed an inconspicuous difference under 0–5% NaCl stress ($p > 0.05$), demonstrating that GXDK6 has a strong salt-tolerant ability and may need to take a while to adapt to salt stress. Moreover, the morphology of GXDK6 under NaCl stress was observed by SEM. The results showed that the cell morphology of GXDK6 was round, smooth, and oval under non-salt stress (**Figure 1B**). However, it showed a contracted, elongated, or rod-shaped form under 10% NaCl stress (**Figure 1C**), suggesting that the morphological changes in GXDK6 may respond to NaCl stress due to its regulation of lipid metabolism and membrane permeability (Wang D. et al., 2021). For testing of this hypothesis, the transcriptome and proteome of GXDK6 were further investigated under NaCl stress.

Genome-Wide Analysis Revealed That Lipid Metabolism-Related Genes Were Significantly Enriched in Pathways Contributing to Salt-Tolerant Survival

117 genes relevant to the lipid metabolism of GXDK6 were annotated, which accounted for 8.77% of the total metabolic regulation genes (1,334 genes, provided in **Supplementary Table S1**; **Supplementary Figure S1**), indicating that GXDK6 had the ability to produce rich lipids. Particularly, mangrove-associated microorganisms resulted in characterization of almost 1,000 new metabolites, among them, ~850 derived from fungi, and ~120 from bacteria (Ancheeva et al., 2018; Blunt et al., 2018). Interestingly, whole-genome analysis found *M. guilliermondii* strain had abundant genes [e.g., *ERG1* encoding squalene monooxygenase, *ERG5* encoding cytochrome P450 61, *ERG6* encoding sterol 24-C-methyltransferase, *ERG9* encoding squalene synthase, *ERG11* encoding lanosterol 14-alpha demethylase, and *ERG24* encoding delta (14)-sterol reductase] regulating steroid biosynthesis pathway, which regulate the biosynthesis of sterols and terpenoids. Therefore, GXDK6 could have a specific regulatory network to biosynthesize important terpenoids. Previous studies have shown that GXDK6 had the ability to biosynthesize nerol (an acyclic monoterpene), which was widely used in food, cosmetics and pharmaceuticals as the valuable fragrance (Mo et al., 2021), aptly validating the above speculation.

In addition, these genes (e.g., *GPD* encoding glycerol-3-phosphate dehydrogenase, *ERG24*, and *GPP1*) were mainly involved in regulating glycerophospholipid metabolism, lipid biosynthesis proteins, glycerolipid metabolism, fatty acid degradation, steroid biosynthesis, sphingolipid metabolism, the biosynthesis of unsaturated fatty acids, fatty acid biosynthesis, and fatty acid elongation (**Figure 1D**; **Table 1**). Among them, glycerophospholipid metabolism, glycerolipid metabolism, fatty

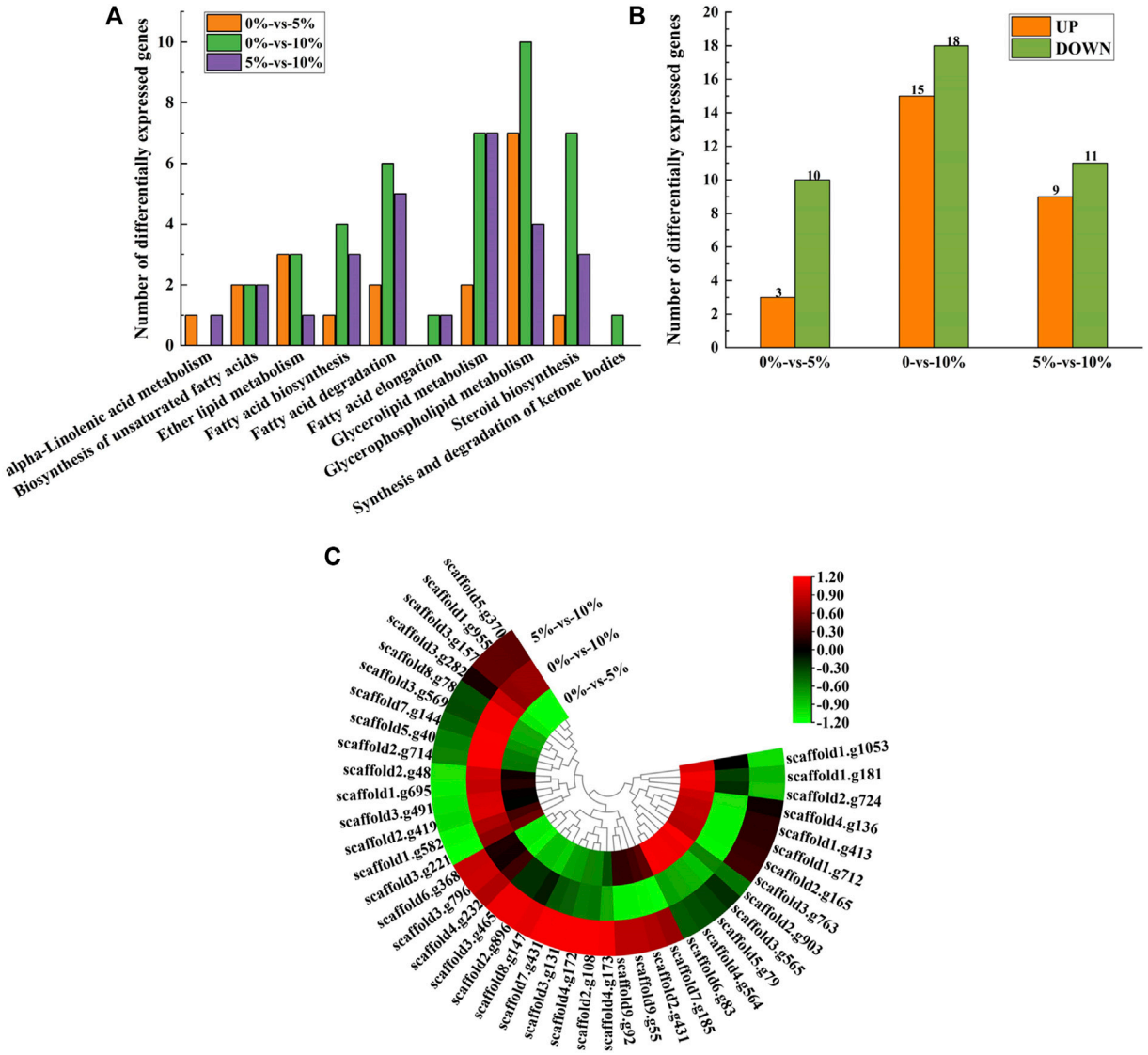


FIGURE 2 | Transcriptome analysis of lipid metabolism. **(A)** Expression of differentially expressed genes at 0, 5, and 10% NaCl; **(B)** Enrichment of pathways under different NaCl stress; **(C)** Heatmap of differentially expressed genes.

acid degradation, the biosynthesis of unsaturated fatty acids, fatty acid biosynthesis, and fatty acid elongation played an important role in the defense and survival of microorganisms (Chen et al., 2015). Vigorous lipid metabolism was speculated to have contributed to salt-tolerant survival. Therefore, the transcriptome and proteome of GXDK6 were further studied.

Transcriptome Analysis Revealed That ERG24 Was Significantly Down-Regulated in *M. guilliermondii* Strain After Perceived Salt Stress

As shown in **Figure 2B**, three genes were upregulated and 10 genes were downregulated under 5% NaCl stress compared with those under salt-free stress, while 15 genes were upregulated and 18 genes were downregulated under 10% NaCl stress. These differentially expressed genes (e.g., *GPD*, *GUT*, and *GPP*) under 10% NaCl stress were significantly enriched in glycerophospholipid metabolism (accounting for 30.30%), glycerolipid metabolism (accounting for 21.21%), steroid biosynthesis (accounting for 21.21%), fatty acid degradation (accounting for 18.18%), fatty acid biosynthesis (accounting for 12.12%), ether lipid metabolism (accounting for 9.09%), the biosynthesis of unsaturated fatty acids (accounting for 6.06%), fatty acid elongation (accounting for 3.03%), and synthesis and degradation of ketone bodies (accounting for 3.03%, **Figure 2A**). Numerous differentially expressed genes (e.g., *GPD*, *GUT*, and *GPP*) that regulate lipid metabolism were significantly enriched in pathways that contributed to salt-tolerant survival (e.g., glycerophospholipid metabolism,

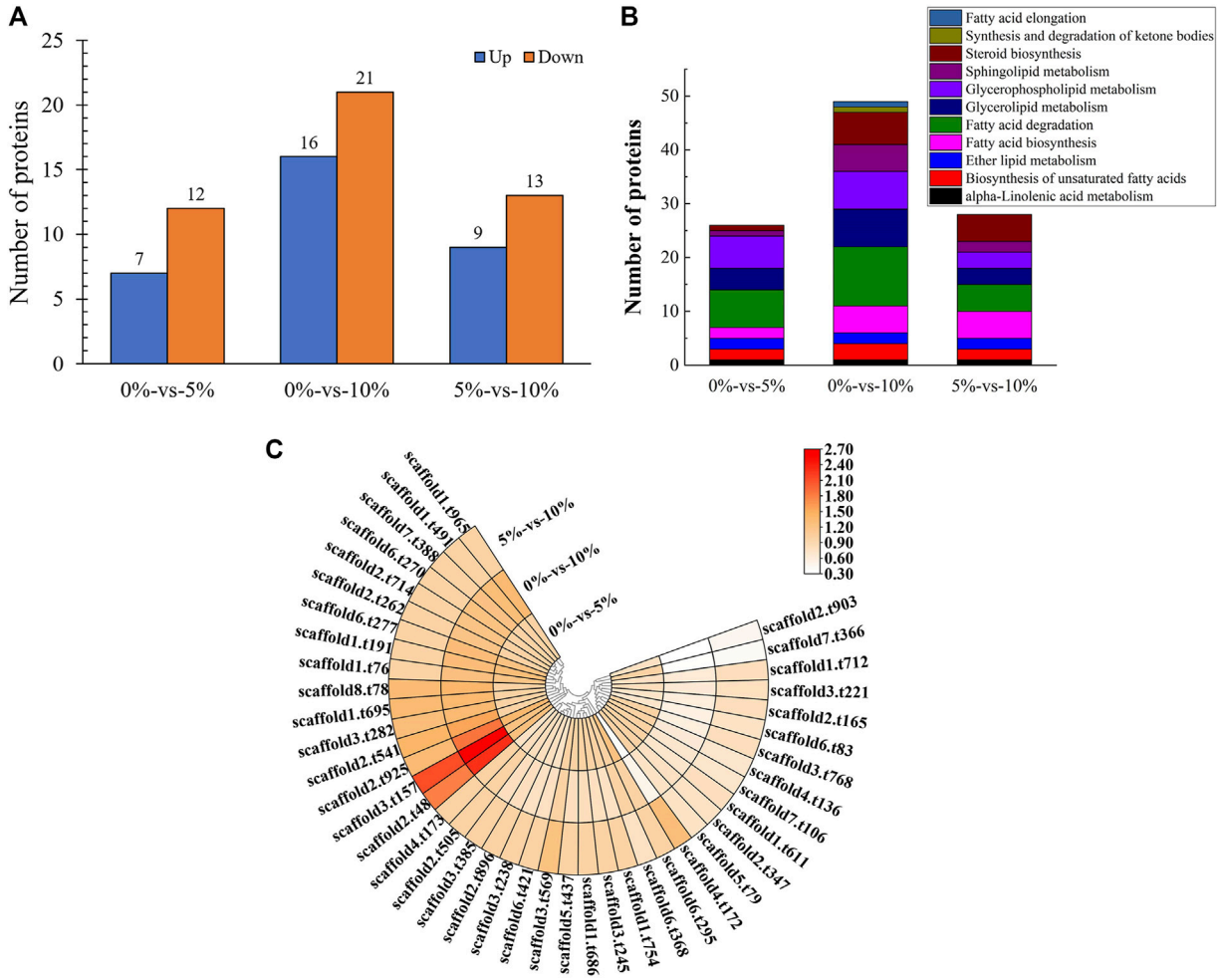


FIGURE 3 | Proteomic analysis of lipid metabolism. **(A)** Number of differentially expressed proteins under different salt stress; **(B)** Pathway enrichment of differentially expressed proteins; **(C)** Differentially expressed proteins under different salt stress.

glycerolipid metabolism, fatty acid degradation) (Xia et al., 2019; Yang et al., 2019) under 10%NaCl stress, which meant *M. guilliermondii* strain had good salt resistance, providing a reference for the application in high salt environments (e.g., soy sauce brewing).

With the condition of 10% NaCl stress as an example, the most significantly upregulated gene was *GPD* (upregulated by 4.00 folds) (**Figure 2C**, provided in **Supplementary Table S4**). *GPD* belonged to the HOG-MAPK pathway, which contributed to the timely reception of signals generated by salt stimulation in *M. guilliermondii* strain and induced the expression of targeted regulatory genes in cells, thereby contributing to salt-tolerant survival (Akhtar et al., 2000). According to the analysis of glycerol metabolism, genes (e.g., *GUT*, *ALD5*, and *GPP*) were also significantly up-regulated under 10% NaCl stress (**Supplementary Figure S2**). It would be reasonable to speculate that *M. guilliermondii* strain regulate glycerol metabolism and accumulate to maintain intra- and extracellular osmotic pressure, thereby promoting the survival of GXDK6 under salt stress. Similar results had also been demonstrated by Duskova et al. (2015), Capusoni et al. (2019), and Lee and Levin (2015). We have also confirmed through experiments that exogenous addition of glycerol enhanced the salt tolerance of GXDK6 under stress (**Supplementary Figure S3**).

However, the most significantly downregulated gene was *ERG24* (downregulated by 3.15 folds, provided in **Supplementary Table S4**). Delta (14)-sterol reductase (regulated by *ERG24*) is a drug target of fenpropimorph, which is involved in step 2 of the subpathway that synthesizes zymosterol from lanosterol. Due to the significant down-regulation of this gene under salt stress, it was speculated that NaCl stress may change the sensitivity of *M. guilliermondii* strain to drugs. This speculation was indirectly verified by Bhuiyan et al. (2007).

In order to verify the reliability of the transcriptome data, *GUT1*, *GPP1* and *ADH7* related to glycerol metabolism were selected for RT-qPCR verification. It can be seen from

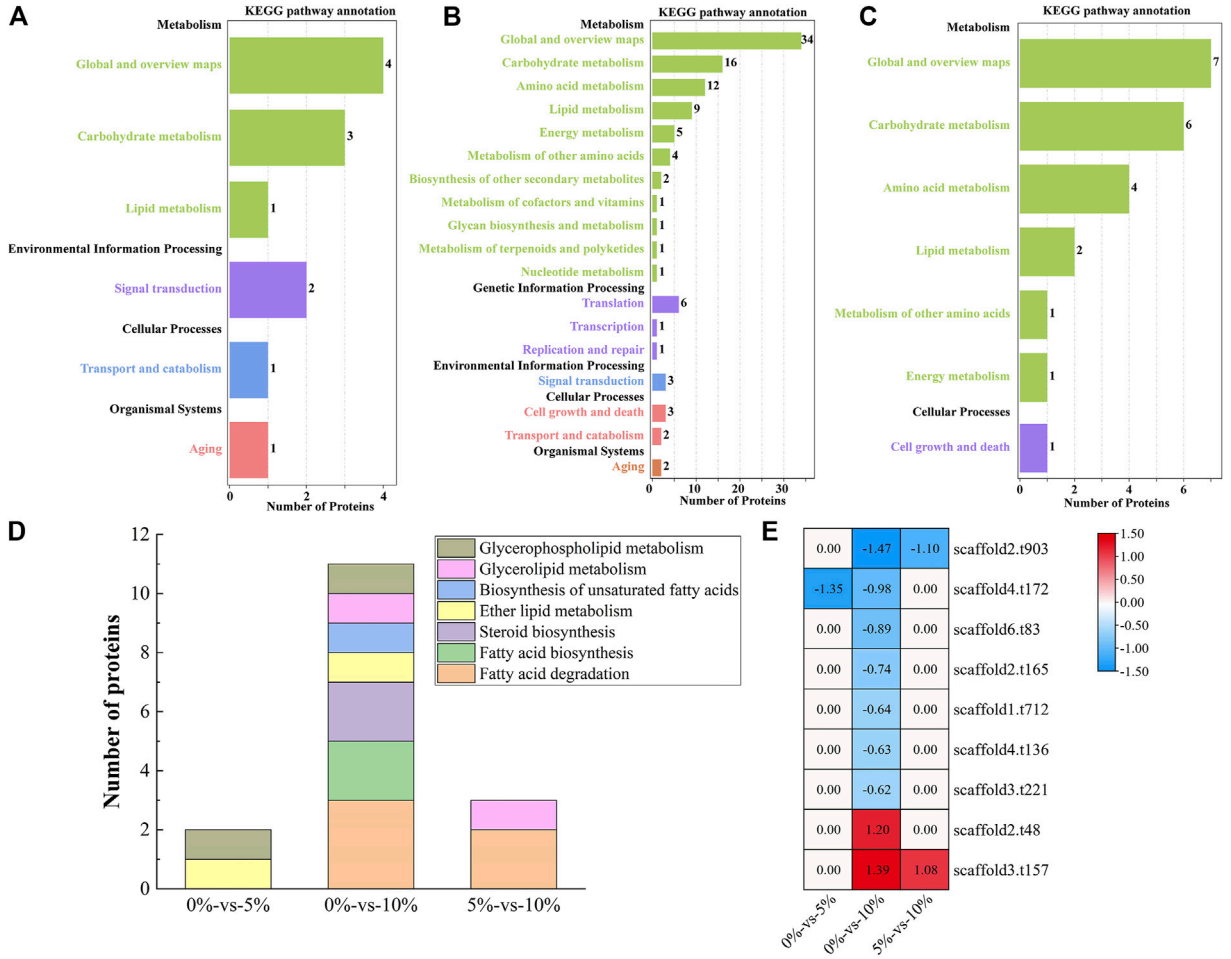


FIGURE 4 | Association analysis of multi-omics data relevant to lipid metabolism. **(A)** Multi-omics association analysis under the condition of 0%-vs-5%; **(B)** Multi-omics association analysis under the condition of 0%-vs-10%; **(C)** Multi-omics association analysis under the condition of 5%-vs-10%; **(D)** KO enrichment of related differentially expressed genes and differentially expressed proteins; **(E)** Related differentially expressed genes and differentially expressed proteins.

Supplementary Figure S4 that the transcription levels of the three genes verified by RT-qPCR are up-regulated, which is consistent with the results of the transcriptome data, indicating the reliability of the transcriptome data. Furthermore, we analyzed the proteomic data of lipid metabolism in GXDK6.

Proteomic Analysis Revealed That NaCl Stress Inhibited Spore Germination of *M. guilliermondii* Strain

16 proteins were significantly upregulated and 21 proteins were significantly downregulated, which were all linked to lipid metabolism in GXDK6 under 10% NaCl stress (**Figure 3A**). These differentially expressed proteins were enriched in fatty acid degradation (accounting for 29.73%), glycerophospholipid metabolism (accounting for 18.92%), glycerolipid metabolism (accounting for 18.92%), steroid biosynthesis (accounting for 16.22%), sphingolipid metabolism (accounting for 13.51%), fatty acid biosynthesis (accounting for 13.51%), the biosynthesis of unsaturated fatty acids (accounting for 8.11%), ether lipid metabolism (accounting for 5.41%), alpha-linolenic acid metabolism (accounting for 2.70%), synthesis and degradation of ketone bodies (accounting for 2.70%), and fatty acid elongation (accounting for 2.70%, **Figure 3B**). Among them, the top three significantly upregulated proteins were alcohol dehydrogenase 3 (upregulated by 1.20 folds, promoting the dehydrogenation of aldehydes to produce alcohols, provided in **Supplementary Table S11**), aldehyde dehydrogenase 5 (upregulated by 1.39 folds, reducing the accumulation of aldehydes, provided in **Supplementary Table S11**), and 3-ketoacyl-thiolase peroxisomal (upregulated by 0.89 folds, a kind of peroxisomal involved in the pathway of fatty acid metabolism). The top three significantly downregulated proteins were NADPH-dependent 1-acyldihydroxyacetone phosphate reductase (downregulated by 0.98 folds; plays a role in cell wall biogenesis, but this effect may be indirect by affecting the activities of cell wall synthesis enzymes), methylsterol

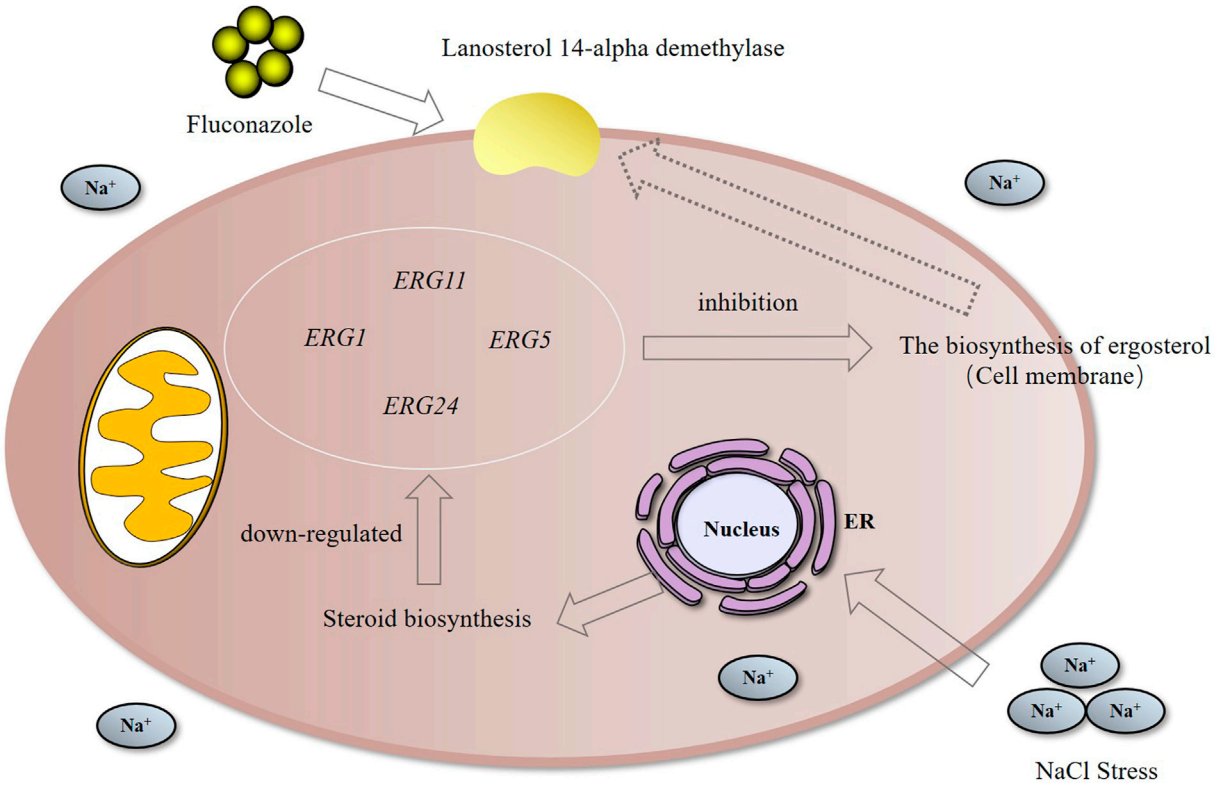


FIGURE 5 | Drug resistance of GXDK6 to fluconazole under NaCl stress.

monooxygenase (downregulated by 1.43 folds; involved in the pathway of ergosterol biosynthesis, which is a part of steroid metabolism), and alcohol dehydrogenase 2 (downregulated by 1.47 folds; this isozyme preferentially catalyzes the conversion of ethanol to acetaldehyde and acts on various primary unbranched aliphatic alcohols) under 10% NaCl (**Figure 3C**, provided in **Supplementary Table S11**). Aldehyde dehydrogenase 5 (regulated by *ALDH5*) and alcohol dehydrogenase (regulated by *ADH2*, and *ADH3*) had functions of protecting cells and maintaining cellular osmotic pressure homeostasis.

Among them, NADPH-dependent 1-acyldihydroxyacetone phosphate reductase (regulatory gene *AYR1*, which regulated the budding and division of cell membranes) (Athenstaedt and Daum, 2000) regulated phosphatidic acid biosynthesis. According to significant down-regulated expression of this protein under high salt stress, it was reasonable to speculate that the budding division of the cell membrane was inhibited, thus inhibiting cell proliferation. This finding further revealed the reason for the decrease in the number of cells under 10% NaCl. The test showed that with the incubation time at 16 h as an example, the OD_{600} of GXDK6 decreased by 64.39% under 10% NaCl stress when compared with salt-free stress. The presence of NaCl during the cultivation process caused a slight decrease in the growth rate. This result was also confirmed by Capusoni et al. (2019) and Petrovic et al. (2002). The analysis of the growth state of *M. guilliermondii* strain under salt stress could provide a reference for its application in high-salt environments.

In addition, the methylsterol monooxygenase regulates the biosynthesis of ergosterol (Ward et al., 2018), which contributes to maintaining the rigidity of the plasma membrane (Schützhold et al., 2016), and serves as the precursor for manufacturing steroid drugs (Liu et al., 2019). However, the methylsterol monooxygenase (subcellularly located in the endoplasmic reticulum membrane) was significantly down-regulated under 10% NaCl stress. It was reasonable to speculate that *M. guilliermondii* strain can regulate the expression of membrane proteins to enhance resistance to stress environments.

Transcriptomic and Proteomic Analysis Revealed That Lanosterol 14-Alpha Demethylase was Down-Regulated in *M. guilliermondii* Strain After Perceived Salt Stress

One differentially expressed gene (*AYR1*) and one differentially expressed protein (NADPH-dependent 1-acyldihydroxyacetone phosphate reductase, which could convert acyl and alkyl dihydroxyacetone-phosphate into glycerolipids and ether lipids) were associated in the pathway of lipid metabolism under 5% NaCl compared with those under salt-free stress (**Figure 4A**, provided in **Supplementary Table S14**), while nine differentially expressed genes and nine differentially

TABLE 2 | Drug resistance of GXDK6 to fluconazole under NaCl stress.

Conditions	Dilution times	Cfu/ml
0%NaCl+0 μg/ml fluconazole	10^{-6}	$2.1*10^{8}$
5%NaCl+0 μg/ml fluconazole	10^{-6}	$1.0*10^{7}$
10%NaCl+0 μg/ml fluconazole	10^{-5}	$9.0*10^{5}$
0%NaCl+64 μg/ml fluconazole	10^{-5}	$1.0*10^{6}$
5%NaCl+64 μg/ml fluconazole	10^{-3}	$1.1*10^{5}$
10%NaCl+64 μg/ml fluconazole	10^{-2}	$2.3*10^{4}$

expressed proteins were associated under 10% NaCl (**Figures 4B,E**, provided in **Supplementary Table S15**). When GXDK6 was under 10% NaCl, two differentially expressed genes and two differentially expressed protein were associated in lipid metabolism compared with those under 5% NaCl stress (**Figure 4C**, provided in **Supplementary Table S16**). With the condition of 10% NaCl stress as an example, the related genes and proteins were mainly enriched in fatty acid degradation (accounting for 33.33%), fatty acid biosynthesis (accounting for 22.22%), steroid biosynthesis (accounting for 22.22%), ether lipid metabolism (accounting for 11.11%), the biosynthesis of unsaturated fatty acids (accounting for 11.11%), glycerolipid metabolism (accounting for 11.11%), and glycerophospholipid metabolism (accounting for 11.11%, **Figure 4D**). These differentially expressed proteins were alcohol dehydrogenase 2 (downregulated by 1.47 folds), NADPH-dependent 1-acyl dihydroxyacetone phosphate reductase (downregulated by 0.98 folds), squalene monooxygenase (downregulated by 0.89 folds), fatty acid synthase alpha subunit (downregulated by 0.74 folds), acetyl-CoA carboxylase (downregulated by 0.64 folds), lanosterol 14-alpha demethylase (downregulated by 0.63 folds), delta (12)-fatty-acid desaturase (downregulated by 0.62 folds), alcohol dehydrogenase 3 (upregulated by 1.20 folds), and aldehyde dehydrogenase (upregulated by 1.39 folds).

Squalene monooxygenase was the key rate-limiting enzyme for the synthesis of sterols and triterpenoids (Han et al., 2020). The expression of this enzyme was downregulated under 10% NaCl stress, indicating that the synthesis of sterol substances was changed under salt stress. Lanosterol 14-alpha demethylase was a key enzyme in ergosterol biosynthesis. When ergosterol synthesis was blocked, the cell membranes were damaged, thus making antifungal drugs, such as azoles, work (Kauffman and Carver, 1997; Sagatova et al., 2015). Studies have found that the protein was significantly differentially expressed under salt stress. It was speculated that the killing effect of antifungal drugs would be affected by NaCl stress. Therefore, we explored the changes of resistance in GXDK6 under salt stress.

NaCl Stress Enhanced the Killing Effect of Fluconazole on *M. guilliermondii* Strain

As shown in **Supplementary Table S4**, the differentially expressed genes were *ERG1* (downregulated by 2.34 folds, encoding squalene monooxygenase), *ERG5* (downregulated by 1.09 folds, encoding squalene monooxygenase), *ERG11* (downregulated by -1.17 folds, encoding lanosterol 14-alpha demethylase), and *ERG24* [downregulated by 3.15 folds, encoding delta (14)-sterol reductase]. The above genes were mainly involved in the biosynthesis of ergosterol (Konecna et al., 2016; Guo et al.,2018). The critical role of ergosterol in maintaining cell membrane integrity and its uniqueness in fungi make it a highly attractive antifungal drug target (Li Y. et al., 2021). Previous studies had shown that *ERG24* was an antifungal target site for azole drugs (Jia et al., 2002). Flowers et al. (2015) found that single mutations in *ERG11* (an important drug target) contributed to azole resistance in *Candida albicans*. In this study, these genes were significantly downregulated under salt stress. Therefore, we speculated that ergosterol, as a target of antifungal drugs, its synthesis pathway was inhibited under salt stress, leading to improve killing effect of fluconazole against *M. guilliermondii* strain (Long et al., 2017). This speculation was also confirmed in the present work, and the results were shown in **Figure 5**. The OD_{600} of GXDK6 under 10% NaCl and contained 64 μg/ml fluconazole was 0.319. However, it decreased to 0.124 under 64 μg/ml fluconazole, which was 61.13% lower than the condition of salt-free but contained 64 μg/ml fluconazole (provided in **Supplementary Figure S5**). Moreover, the number of colonies of GXDK6 under 10% NaCl and without fluconazole were $9.0*10^{5}$ cfu/ml (provided in **Supplementary Figure S6**). However, it decreased to $2.3*10^{4}$ cfu/ml under 64 μg/ml fluconazole, which was 97.70% lower than the condition of salt-free but contained 64 μg/ml fluconazole (number of colonies = $1.0*10^{6}$ cfu/ml, **Table 2**). It was considered that the genes regulating synthesis of ergosterol were down-regulated under salt stress, which lead to down-regulation of the corresponding proteins expression, thereby weakening the synthesis of ergosterol. The results demonstrated that salt stress promoted the fungicidal effect of fluconazole, which provided a new approach to enhance the killing effect of fluconazole on fungi.

CONCLUSION

The regulation mechanism of lipid metabolism in *M. guilliermondii GXDK6* was revealed by integrative omics technology. The key genes and proteins that regulated steroid biosynthesis and glycerolipid metabolism under salt stress were identified. High-concentration salt stress could change the biosynthesis of ergosterol and inhibited cell proliferation. Further results demonstrated that the killing effect of fluconazole on *M. guilliermondii* GXDK6 was significantly enhanced under salt stimulation. Final results showed that salt stress disturbed lipid metabolism, which reduced the expression of antifungal drugs targets, thus making *M. guilliermondii* strain more sensitive to fluconazole. This work revealed the regulatory changes of lipid metabolism in *M. guilliermondii* strain under salt stress, which also provided a new strategy to enhance the fungicidal effect of antifungal drugs.

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

CJ designed research. HS conducted the experiments and wrote the manuscript. XC conducted the experiments and revised the manuscript. HB and DM arranged and analyzed the experimental data. XM and SH provided the technical support. BY provided theoretical direction of multi-omics research. GS corrected the language errors of the manuscript. All the authors read and approved the final manuscript.

FUNDING

This research was supported by the National Natural Science Foundation of China (Grant No. 31760437), Science and Technology Basic Resources Investigation Program of China (Grant No. 2017FY100704), the Funding Project of Chinese Central Government Guiding to the Guangxi Local Science and Technology Development (Grant No. GUIKEZY21195021), Natural Science Fund for Distinguished Young Scholars of Guangxi Zhuang Autonomous Region of China (Grant No. 2019GXNSFFA245011), China-ASEAN International Innovative Center for Health Industry of Traditional Chinese Medicine (AD20297142), and the Innovation Project of Guangxi Graduate Education (Grant No. YCBZ2021012 and YCSW2021064).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.798535/full#supplementary-material

REFERENCES

Akhtar, N., Påhlman, A. K., Larsson, K., Corbett, A. H., and Adler, L. (2000). SGD1 Encodes an Essential Nuclear Protein of *Saccharomyces cerevisiae* that Affects Expression of the *GPD1* Gene for Glycerol 3-phosphate Dehydrogenase. *FEBS Lett.* 483 (2-3), 87–92. doi:10.1016/s0014-5793(00)02087-1

Ancheeva, E., Daletos, G., and Proksch, P. (2018). Lead Compounds from Mangrove-Associated Microorganisms. *Mar. Drugs* 16 (9), 319. doi:10.3390/md16090319

Athenstaedt, K., and Daum, G. (2000). 1-Acyldihydroxyacetone-phosphate Reductase (Ayr1p) of the Yeast *Saccharomyces cerevisiae* Encoded by the Open reading Frame YIL124w Is a Major Component of Lipid Particles. *J. Biol. Chem.* 275 (1), 235–240. doi:10.1074/jbc.275.1.235

Bhuiyan, M. S. A., Eckstein, J., Barbuch, R., and Bard, M. (2007). Synthetically Lethal Interactions Involving Loss of the Yeast *ERG24*: the Sterol C-14 Reductase Gene. *Lipids* 42 (1), 69–76. doi:10.1007/s11745-006-1001-4

Blunt, J. W., Carroll, A. R., Copp, B. R., Davis, R. A., Keyzers, R. A., and Prinsep, M. R. (2018). Marine Natural Products. *Nat. Prod. Rep.* 35 (1), 8–53. doi:10.1039/c7np00052a

Capusoni, C., Arioli, S., Donzella, S., Guidi, B., Serra, I., and Compagno, C. (2019). Hyper-osmotic Stress Elicits Membrane Depolarization and Decreased Permeability in Halotolerant marine *Debaryomyces Hansenii* Strains and in *Saccharomyces cerevisiae*. *Front. Microbiol.* 10, 64. doi:10.3389/fmicb.2019.00064

Casanovas, A., Sprenger, R. R., Tarasov, K., Ruckerbauer, D. E., Hannibal-Bach, H. K., Zanghellini, J., et al. (2015). Quantitative Analysis of Proteome and Lipidome Dynamics Reveals Functional Regulation of Global Lipid Metabolism. *Chem. Biol.* 22 (3), 412–425. doi:10.1016/j.chembiol.2015.02.007

Chen, K., Li, E., Xu, Z., Li, T., Xu, C., Qin, J. G., et al. (2015). Comparative Transcriptome Analysis in the Hepatopancreas Tissue of pacific white Shrimp litopenaeus Vannamei Fed Different Lipid Sources at Low Salinity. *PLoS One* 10 (12), e0144889. doi:10.1371/journal.pone.0144889

Díaz, C., Molina, A. M., Nähring, J., and Fischer, R. (2013). Characterization and Dynamic Behavior of Wild Yeast during Spontaneous Wine Fermentation in Steel Tanks and Amphorae. *Biomed. Res. Int.* 2013, 540465. doi:10.1155/2013/540465

Duskova, M., Borovikova, D., Herynkova, P., Rapoport, A., and Sychrova, H. (2015). The Role of Glycerol Transporters in Yeast Cells in Various Physiological and Stress Conditions. *FEMS Microbiol. Lett.* 362 (3), 1–8. doi:10.1093/femsle/fnu041

Fan, J., Li, X., Issop, L., Culty, M., and Papadopoulos, V. (2016). ACBD2/ECI2-mediated Peroxisome-Mitochondria Interactions in Leydig Cell Steroid Biosynthesis. *Mol. Endocrinol.* 30 (7), 763–782. doi:10.1210/me.2016-1008

Ferreira, R., Teixeira, P. G., Siewers, V., and Nielsen, J. (2018). Redirection of Lipid Flux toward Phospholipids in Yeast Increases Fatty Acid Turnover and Secretion. *Proc. Natl. Acad. Sci. U S A.* 115 (6), 1262–1267. doi:10.1073/pnas.1715282115

Flowers, S. A., Colón, B., Whaley, S. G., Schuler, M. A., and Rogers, P. D. (2015). Contribution of Clinically Derived Mutations in ERG11 to Azole Resistance in *Candida Albicans*. *Antimicrob. Agents Chemother.* 59 (1), 450–460. doi:10.1128/AAC.03470-14

Galafassi, S., Toscano, M., Vigentini, I., Zambelli, P., Simonetti, P., Foschino, R., et al. (2015). Cold Exposure Affects Carbohydrates and Lipid Metabolism, and Induces Hog1p Phosphorylation in *Dekkera Bruxellensis* Strain CBS 2499. *Antonie Van Leeuwenhoek* 107 (5), 1145–1153. doi:10.1007/s10482-015-0406-6

Guo, Z. P., Khoomrung, S., Nielsen, J., and Olsson, L. (2018). Changes in Lipid Metabolism Convey Acid Tolerance in *Saccharomyces cerevisiae*. *Biotechnol. Biofuels* 11, 297. doi:10.1186/s13068-018-1295-5

Han, J. Y., Jo, H. J., and Choi, Y. E. (2020). Overexpression of the Squalene Epoxidase Gene (*PgSE1*) Resulted in Enhanced Production of Ginsenosides and Phytosterols in Transgenic Ginseng. *Plant Biotechnol. Rep.* 14, 673–682. doi:10.1007/s11816-020-00643-4

Jia, N., Arthington-Skaggs, B., Lee, W., Pierson, C. A., Lees, N. D., Eckstein, J., et al. (2002). *Candida Albicans* Sterol C-14 Reductase, Encoded by the *ERG24* Gene, as a Potential Antifungal Target Site. *Antimicrob. Agents Chemother.* 46 (4), 947–957. doi:10.1128/AAC.46.4.947-957.2002

Kauffman, C. A., and Carver, P. L. (1997). Use of Azoles for Systemic Antifungal Therapy. *Adv. Pharmacol.* 39, 143–189. doi:10.1016/s1054-3589(08)60071-x

Klug, L., and Daum, G. (2014). Yeast Lipid Metabolism at a Glance. *FEMS Yeast Res.* 14 (3), 369–388. doi:10.1111/1567-1364.12141

Konecna, A., Toth Hervay, N., Valachovic, M., and Gbelska, Y. (2016). *ERG6* Gene Deletion Modifies *Kluyveromyces Lactis* Susceptibility to Various Growth Inhibitors. *Yeast* 33 (12), 621–632. doi:10.1002/yea.3212

Kraft, M. L. (2017). Sphingolipid Organization in the Plasma Membrane and the Mechanisms that Influence it. *Front Cel Dev Biol* 4, 154. doi:10.3389/fcell.2016.00154

Lee, J., and Levin, D. E. (2015). Rgc2 Regulator of Glycerol Channel Fps1 Functions as a Homo- and Heterodimer with Rgc1. *Eukaryot. Cel* 14 (7), 719–725. doi:10.1128/EC.00073-15

Li, C., Zhao, D., Yan, J., Zhang, N., and Li, B. (2021). Metabolomics Integrated with Transcriptomics: Assessing the central Metabolism of marine Red Yeast *Sporobolomyces Pararoseus* under Salinity Stress. *Arch. Microbiol.* 203 (3), 889–899. doi:10.1007/s00203-020-02082-9

Li, Y., Dai, M., Zhang, Y., and Lu, L. (2021). The Sterol C-14 Reductase Erg24 Is Responsible for Ergosterol Biosynthesis and Ion Homeostasis in *Aspergillus fumigatus*. *Appl. Microbiol. Biotechnol.* 105 (3), 1253–1268. doi:10.1007/s00253-021-11104-5

Liu, J. F., Xia, J. J., Nie, K. L., Wang, F., and Deng, L. (2019). Outline of the Biosynthesis and Regulation of Ergosterol in Yeast. *World J. Microbiol. Biotechnol.* 35 (7), 98. doi:10.1007/s11274-019-2673-2

Livak, K. J., and Schmittgen, T. D. (2001). Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* 25 (4), 402–408. doi:10.1006/meth.2001.1262

Long, N., Xu, X., Zeng, Q., Sang, H., and Lu, L. (2017). Erg4A and Erg4B Are Qequired for Conidiation and Azole Resistance via Regulation of Ergosterol Biosynthesis in *Aspergillus fumigatus*. *Appl. Environ. Microbiol.* 83 (4), e02924–16. doi:10.1128/AEM.02924-16

Love, M. I., Huber, W., and Anders, S. (2014). Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2. *Genome Biol.* 15 (12), 550. doi:10.1186/s13059-014-0550-8

Lu, H., Chen, H., Tang, X., Yang, Q., Zhang, H., Chen, Y. Q., et al. (2021). Metabolomics Analysis Reveals the Role of Oxygen Control in the Nitrogen Limitation Induced Lipid Accumulation in *Mortierella Alpina*. *J. Biotechnol.* 325, 325–333. doi:10.1016/j.jbiotec.2020.10.004

Mo, X., Cai, X., Hui, Q., Sun, H., Yu, R., Bu, R., et al. (2021). Whole Genome Sequencing and Metabolomics Analyses Reveal the Biosynthesis of Nerol in a Multi-Stress-Tolerant *Meyerozyma Guilliermondii* GXDK6. *Microb. Cel Fact* 20 (1), 4. doi:10.1186/s12934-020-01490-2

Myers, S. A., Klaeger, S., Satpathy, S., Viner, R., Choi, J., Rogers, J., et al. (2019). Evaluation of Advanced Precursor Determination for Tandem Mass Tag (TMT)-based Quantitative Proteomics across Instrument Platforms. *J. Proteome Res.* 18 (1), 542–547. doi:10.1021/acs.jproteome.8b00611

Nebauer, R., Birner-Grünberger, R., and Daum, G. (2004). 3 Biogenesis and Cellular Dynamics of Glycerophospholipids in the Yeast *Saccharomyces cerevisiae*. *Top. Curr. Genet.* 6, 125–168. doi:10.1007/978-3-540-40999-1_4

Petrovic, U., Gunde-Cimerman, N., and Plemenitas, A. (2002). Cellular Responses to Environmental Salinity in the Halophilic Black Yeast *Hortaea Werneckii*. *Mol. Microbiol.* 45 (3), 665–672. doi:10.1046/j.1365-2958.2002.03021.x

Qi, Y., Liu, H., Yu, J., Chen, X., and Liu, L. (2017). Med15B Regulates Acid Stress Response and Tolerance in *Candida Glabrata* by Altering Membrane Lipid Composition. *Appl. Environ. Microbiol.* 83 (18), e01128–17. doi:10.1128/AEM.01128-17

Randez-Gil, F., Córcoles-Saez, I., Estruch, F., and Prieto, J. A. (2018). Lipid Metabolism Regulation and its Relationship with Cold Stress Response in Yeast. *Cryobiology* 85, 127. doi:10.1016/j.cryobiol.2018.10.041

Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: a Bioconductor Package for Differential Expression Analysis of Digital Gene Expression Data. *Bioinformatics* 26 (1), 139–140. doi:10.1093/bioinformatics/btp616

Ross, J. A., Maingay, J. P., Fearon, K. C., Sangster, K., and Powell, J. J. (2003). Eicosapentaenoic Acid Perturbs Signalling via the NFkappaB Transcriptional Pathway in Pancreatic Tumour Cells. *Int. J. Oncol.* 23 (6), 1733–1738. doi:10.3892/ijo.23.6.1733

Sagatova, A. A., Keniya, M. V., Wilson, R. K., Monk, B. C., and Tyndall, J. D. (2015). Structural Insights into Binding of the Antifungal Drug Fluconazole to *Saccharomyces cerevisiae* Lanosterol 14α-Demethylase. *Antimicrob. Agents Chemother.* 59 (8), 4982–4989. doi:10.1128/AAC.00925-15

Schützhold, V., Hahn, J., Tummler, K., and Klipp, E. (2016). Computational Modeling of Lipid Metabolism in Yeast. *Front. Mol. Biosci.* 3, 57. doi:10.3389/fmolb.2016.00057

Tian, Y., Xia, Z., Li, M., Zhang, G., Cui, H., Li, B., et al. (2019). The Relationship between Microwave Radiation Injury and Abnormal Lipid Metabolism. *Chem. Phys. Lipids* 225, 104802. doi:10.1016/j.chemphyslip.2019.104802

Villa-Rodríguez, E., Ibarra-Gámez, C., and de Los Santos-Villalobos, S. (2018). Extraction of High-Quality RNA from Bacillus Subtilis with a Lysozyme Pre-treatment Followed by the Trizol Method. *J. Microbiol. Methods* 147, 14–16. doi:10.1016/j.mimet.2018.02.011

Wang, D., Zhang, M., Huang, J., Zhou, R., Jin, Y., Zhao, D., et al. (2021). Heat Preadaptation Improved the Ability of *Zygosaccharomyces Rouxii* to Salt Stress: a Combined Physiological and Transcriptomic Analysis. *Appl. Microbiol. Biotechnol.* 105 (1), 259–270. doi:10.1007/s00253-020-11005-z

Wang, G., Li, D., Miao, Z., Zhang, S., Liang, W., and Liu, L. (2018). Comparative Transcriptome Analysis Reveals Multiple Functions for Mhy1p in Lipid Biosynthesis in the Oleaginous Yeast *Yarrowia Lipolytica*. *Biochim. Biophys. Acta Mol. Cel Biol Lipids* 1863 (1), 81–90. doi:10.1016/j.bbalip.2017.10.003

Wang, S., Tang, K., Lu, Y., Tian, Z., Huang, Z., Wang, M., et al. (2021). Revealing the Role of Glycerophospholipid Metabolism in Asthma through Plasma Lipidomics. *Clin. Chim. Acta* 513, 34–42. doi:10.1016/j.cca.2020.11.026

Ward, D. M., Chen, O. S., Li, L., Kaplan, J., Bhuiyan, S. A., Natarajan, S. K., et al. (2018). Altered Sterol Metabolism in Budding Yeast Affects Mitochondrial Iron-Sulfur (Fe-S) Cluster Synthesis. *J. Biol. Chem.* 293 (27), 10782–10795. doi:10.1074/jbc.RA118.001781

Xia, Z., Zhou, X., Li, J., Li, L., Ma, Y., Wu, Y., et al. (2019). Multiple-omics Techniques Reveal the Role of Glycerophospholipid Metabolic Pathway in the Response of *Saccharomyces cerevisiae* against Hypoxic Stress. *Front. Microbiol.* 10, 1398. doi:10.3389/fmicb.2019.01398

Xu, S., and Li, Y. (2020). Yeast as a Promising Heterologous Host for Steroid Bioproduction. *J. Ind. Microbiol. Biotechnol.* 47 (9-10), 829–843. doi:10.1007/s10295-020-02291-7

Yang, H. L., Liao, Y. Y., Zhang, J., and Wang, X. L. (2019). Comparative Transcriptome Analysis of Salt Tolerance Mechanism of *Meyerozyma Guilliermondii* W2 under NaCl Stress. *3 Biotech.* 9 (7), 286. doi:10.1007/s13205-019-1817-2

Zhang, P., and Reue, K. (2017). Lipin Proteins and Glycerolipid Metabolism: Roles at the ER Membrane and beyond. *Biochim. Biophys. Acta Biomembr* 1859 (9 Pt B), 1583–1595. doi:10.1016/j.bbamem.2017.04.007

Zhou, Y. J., Hu, Y., Zhu, Z., Siewers, V., and Nielsen, J. (2018). Engineering 1-alkene Biosynthesis and Secretion by Dynamic Regulation in Yeast. *ACS Synth. Biol.* 7 (2), 584–590. doi:10.1021/acssynbio.7b00338

Zhu, Y., Xu, H., Chen, H., Xie, J., Shi, M., Shen, B., et al. (2014). Proteomic Analysis of Solid Pseudopapillary Tumor of the Pancreas Reveals Dysfunction of the Endoplasmic Reticulum Protein Processing Pathway. *Mol. Cel Proteomics* 13 (10), 2593–2603. doi:10.1074/mcp.M114.038786

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

ORIGINAL RESEARCH
published: 03 March 2022
doi: 10.3389/fmicb.2022.826994

β-Glucuronidase Pattern Predicted From Gut Metagenomes Indicates Potentially Diversified Pharmacomicrobiomics

*Francesco Candeliere[1], Stefano Raimondi[1], Raffaella Ranieri[1], Eliana Musmeci[1], Alfonso Zambon[2], Alberto Amaretti[1,3] and Maddalena Rossi[1,3]**

[1] Department of Life Sciences, University of Modena and Reggio Emilia, Modena, Italy, [2] Department of Chemistry and Geological Sciences, University of Modena and Reggio Emilia, Modena, Italy, [3] Biogest-Siteia, University of Modena and Reggio Emilia, Modena, Italy

OPEN ACCESS

Edited by:
Roshan Kumar, Magadh University, India

Reviewed by:
Priyadarshini Dey, Ramaiah Institute of Technology, India
Balaram Mohapatra, Indian Institute of Technology Bombay, India

***Correspondence:**
Maddalena Rossi
maddalena.rossi@unimore.it

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

Received: *01 December 2021*
Accepted: *21 January 2022*
Published: *03 March 2022*

Citation:
Candeliere F, Raimondi S, Ranieri R, Musmeci E, Zambon A, Amaretti A and Rossi M (2022) β-Glucuronidase Pattern Predicted From Gut Metagenomes Indicates Potentially Diversified Pharmacomicrobiomics. Front. Microbiol. 13:826994. doi: 10.3389/fmicb.2022.826994

β-glucuronidases (GUS) of intestinal bacteria remove glucuronic acid from glucoronides, reversing phase II metabolism of the liver and affecting the level of active deconjugated metabolites deriving from drugs or xenobiotics. Two hundred seventy-nine non-redundant GUS sequences are known in the gut microbiota, classified in seven structural categories (NL, L1, L2, mL1, mL2, mL1,2, and NC) with different biocatalytic properties. In the present study, the intestinal metagenome of 60 healthy subjects from five geographically different cohorts was assembled, binned, and mined to determine qualitative and quantitative differences in GUS profile, potentially affecting response to drugs and xenobiotics. Each metagenome harbored 4–70 different GUS, altogether accounting for 218. The amount of intestinal bacteria with at least one GUS gene was highly variable, from 0.7 to 82.2%, 25.7% on average. No significant difference among cohorts could be identified, except for the Ethiopia (ETH) cohort where GUS-encoding bacteria were significantly less abundant. The structural categories were differently distributed among the metagenomes, but without any statistical significance related to the cohorts. GUS profiles were generally dominated by the category NL, followed by mL1, L2, and L1. The GUS categories most involved in the hydrolysis of small molecules, including drugs, are L1 and mL1. Bacteria contributing to these categories belonged to *Bacteroides ovatus*, *Bacteroides dorei*, *Bacteroides fragilis*, *Escherichia coli*, *Eubacterium eligens*, *Faecalibacterium prausnitzii*, *Parabacteroides merdae*, and *Ruminococcus gnavus*. Bacteria harboring L1 GUS were generally scarcely abundant (<1.3%), except in three metagenomes, where they reached up to 24.3% for the contribution of *E. coli* and *F. prausnitzii.* Bacteria harboring mL1 GUS were significantly more abundant (mean = 4.6%), with *Bacteroides* representing a major contributor. Albeit mL1 enzymes are less active than L1 ones, *Bacteroides* likely plays a pivotal role in the deglucuronidation, due to its remarkable abundance in the microbiomes. The observed broad interindividual heterogeneity of GUS profiles, particularly of the L1 and mL1

categories, likely represent a major driver of pharmacomicrobiomics variability, affecting drug response and toxicity. Different geographical origins, genetic, nutritional, and lifestyle features of the hosts seemed not to be relevant in the definition of glucuronidase activity, albeit they influenced the richness of the GUS profile.

Keywords: β-glucuronidase, human gut microbiota, metagenome, WGS, whole genome sequencing, drug metabolism, pharmacomicrobiomics

INTRODUCTION

Humans and their colon microbiota evolved together, establishing a close symbiotic interrelationship, fruitful for both. The gut microbiota is implicated in a number of biological processes such as resistance to colonization (Ruan et al., 2020), immune system modulation (Saldana-Morales et al., 2021), synthesis of essential vitamins and nutrients (Oliphant and Allen-Vercoe, 2019), and breakdown of undigested polysaccharides and proteins (El Kaoutari et al., 2013; Huang et al., 2017; Raimondi et al., 2021). Furthermore, it encodes a broad diversity of enzymes capable of processing foreign compounds (e.g., phytochemicals, environmental pollutants, pharmaceuticals, and other xenobiotics) and their endogenous metabolites, adding significant chemical diversity and modifying lifetimes, bioavailability, and biological activity (Rossi et al., 2013; Koppel et al., 2017). In this context, pharmacomicrobiomics is an emerging field focusing on the interplay of microbiome and drug metabolism and response (Doestzada et al., 2018; Hassan et al., 2021).

Hundreds of bacterial enzymes are dedicated to the hydrolysis of carbohydrates and glycoconjugates that are not digested in the upper gut and reach the colon, where they are broken down by the microbiota (Flint et al., 2012). Among these enzymes, β-glucuronidases (GUS) remove glucuronic acid from glucoronides, reversing the phase II metabolism carried out by liver enzymes on endo- and xeno-biotics in order to facilitate their excretion from the body (Ervin and Redinbo, 2020). Glucuronic acid is then utilized by bacteria as a carbon and energy source, being channeled into the Entner–Doudoroff pathway that catabolizes sugar acids into pyruvate (Peekhaus and Conway, 1998). The deglucuronidated compounds can be reabsorbed through the gut epithelium and reach the plasma, in a process called enterohepatic circulation (Roberts et al., 2002; Pellock and Redinbo, 2017). Thus, bacterial GUS affect the pharmacokinetics of compounds such as polyphenols, xenobiotics, and drugs and participate in the regulation of the levels of circulating metabolites, altering the pharmacological properties and the biological activities of xenobiotics and potentially impacting on their beneficial and/or toxic effects on health (Biernat et al., 2018; Wang et al., 2019; Awolade et al., 2020).

GUS were first identified in 1934 in *Escherichia coli* and other Enterobacteriaceae (Masamune, 1934; Oshima, 1934), but later, they have been detected in several bacterial taxa belonging to all the main phyla within the gut microbiota: Bacteroidetes, Firmicutes, Proteobacteria, and Actinobacteria (McBain and Macfarlane, 1998; Russell and Klaenhammer, 2001; Nakamura et al., 2002; Gloux et al., 2011). Nowadays, it is known that intestinal bacteria encode different GUS types with structural differences affecting function, biocatalytic properties, and substrate specificity (Biernat et al., 2019; Parvez et al., 2021). The driving force for such evolution and diversification of bacterial GUS has been the availability of dietary and endogenous glucuronides to the commensal microbiota (Pellock and Redinbo, 2017). In particular, the glucuronides of several endogenous metabolites (such as bilirubin, estrogen and androgen hormones, neurotransmitters, and bile acids) are produced by liver UDP-glucuronosyltransferase and abundantly excreted into the intestinal lumen (Liston et al., 2001; Meech et al., 2012; Jarrar and Lee, 2021). The massive sequencing of the human intestinal metagenomes in the Human Microbiome Project (HMP) (Turnbaugh et al., 2007) and bioinformatic mining tools enabled the identification of a wide repertoire of GUS encoded by human gut bacteria. The so-called GUSome has been proposed, encompassing 279 non-redundant GUS sequences (Pollet et al., 2017), 93.5% of which have been taxonomically assigned to Bacteroidetes (52%), Firmicutes (43%), Verrucomicrobia (1.5%), and Proteobacteria (0.5%) (Pollet et al., 2017).

Bacterial GUS present a conserved folding, with two structural elements (loop 1 and loop 2), adjacent to the active site, that differ in length and amino acid composition and permit classification into seven GUS structural categories: NL, L1, L2, mL1, mL2, mL1,2, and NC (Pollet et al., 2017). The enzymes of diverse categories differ in size, substrate-binding modules, active site features, and subcellular localization. Most of the intestinal GUS belong to the category NL (57.3%), followed by mL1, L2, L1, mL2, NC, and mL1,2 in decreasing order (Pollet et al., 2017). The dimension of the loops is pivotal for substrate recognition and affects the biocatalytic properties of the enzymes. Categories L1, mL1, and L2 are more efficient to catalyze the deglucuronidation of small substrates in comparison to categories mL2, mL1,2, and NL (Wallace et al., 2015; Biernat et al., 2019). Differences in the cellular localization are related to the category: L1 enzymes lack signal peptide and are intracellular, whereas L2, mL2, and mL1,2 GUS are likely extracellular. For the GUS belonging to categories mL1 and NL, the presence of signal peptide is linked to the phylum: absent in Firmicutes and present in Bacteroidetes.

The microbial composition of intestinal microbiota impacts GUS abundance and diversity, with major effects on the metabolism of drugs and xenobiotics likely responsible for different individual responses (Elmassry et al., 2021). This study wanted to determine the qualitative and quantitative differences of GUS-encoding genes among metagenomes of healthy subjects. It aimed to investigate the interindividual variability of GUS-encoding bacteria in the gut, mining 60 intestinal publicly

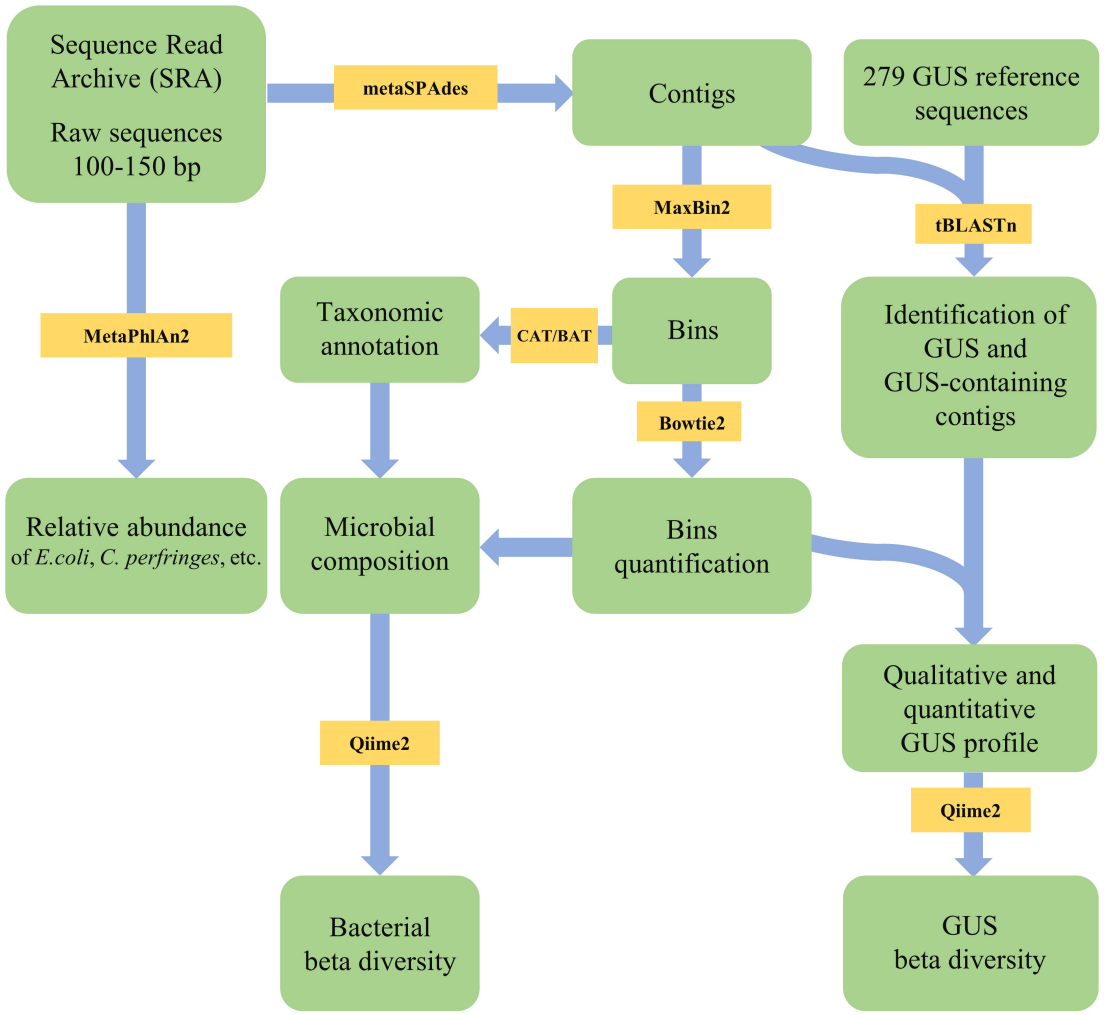


FIGURE 1 | Flowsheet of the pipeline applied in the present study for metagenome analysis and β-glucuronidases (GUS) search.

available metagenomes of healthy subjects. To circumvent the bias arising from diverse genetic, nutritional, and lifestyle features, the metagenomes belonging to five geographically different cohorts were retrieved and processed for GUS profiling. This approach provided preliminary information of interindividual differences of the GUS repertoire, with awareness that transformation of drugs and xenobiotics is subjected to regulation of the expression. The results herein presented could promote intentional manipulation of gut microbiota to enhance drug effectiveness in order to reduce adverse drug interactions or other approaches of personalized therapy to obtain maximum efficacy and minimum toxicity.

MATERIALS AND METHODS

Metagenomes

Sixty publicly available metagenomes of gut microbiota from healthy adults were collected from the NCBI Sequence Read Archive (SRA), with the accession numbers listed in **Supplementary Table 1**. The subjects were ascribed to five cohorts from five different countries: China (CHN), Ethiopia (ETH), Spain (ESP), United States of America (USA), and Sweden (SWE). The selected metagenomes were sequenced through whole-genome shotgun sequencing on Illumina paired-end platforms and produced reads ranging between 100 and 150 bp in length.

Assembly and Binning

The FASTQ files were checked for quality and primer presence with FastQC v0.11.8 (Andrews, 2010), in order to assure that only high-quality reads (length > 50 bp; quality score > 20) were further analyzed. When necessary, the tool Cutadapt v1.16 (minimum length 50; quality cutoff 20) (Martin, 2011) was used for quality filtering. The cohort ESP required primer removal, which was carried out through Trimmomatic (Bolger et al., 2014) with ILLUMINACLIP setting. The reads were assembled in contigs using metaSPAdes v 3.9 (Nurk et al., 2017) with

default parameters. The contigs were binned with MaxBin2 v2.2.7 (Wu et al., 2016) to obtain metagenome-assembled genomes (MAGs). MaxBin2 measures the tetranucleotide frequencies of the contigs and their coverages to classify them into individual bins. It employs single-copy marker gene prediction to determine the completeness of bins (Wu et al., 2014, 2016). According to MaxBin2 default parameters, only contigs at least 1,000 bp long were utilized for binning, and those shorter were discarded from further analysis. MAGs were taxonomically identified with the CAT/BAT tool (von Meijenfeldt et al., 2019). Each bin was mapped against the raw reads using Bowtie2 (Langmead and Salzberg, 2013) to assess the relative abundance. Except for CAT/BAT that was run locally, the steps were conducted on Galaxy platform[1] (Afgan et al., 2018).

BACTERIAL COMPOSITION AND ALPHA AND BETA DIVERSITY

The relative abundance of taxonomically identified MAGs was used to define the abundance profile of bacterial taxa in each metagenome. A BIOM file was produced and imported into Qiime2 (Bolyen et al., 2019) to compute beta diversity according to Bray–Curtis dissimilarity. The beta distance matrix was utilized for principal coordinate analysis (PCoA). Bacterial composition at a deeper taxonomic level was assessed by MetaPhLan2 (Segata et al., 2012; Truong et al., 2015) for the species *Clostridium perfringens*, *Eubacterium eligens*, *Lactobacillus rhamnosus*, *Ruminococcus gnavus*, *Streptococcus agalactiae*, *Bacteroides uniformis*, *Bacteroides ovatus*, *Bacteroides dorei*, *Bacteroides fragilis*, and *Parabacteroides merdae*, known to encode several deeply characterized GUS (Pellock et al., 2018; Biernat et al., 2019; Ervin et al., 2019). Alpha diversity has been calculated using Shannon index, Chao-1 index, and Pielou's evenness with the tool Past v 4.08 (Hammer et al., 2001).

β-Glucuronidase Identification and Profiling

The 279 sequences of GUS identified and classified by Pollet et al. (2017), listed in **Supplementary Material 1**, were blasted to the binned metagenomes using tBLASTn with an *e*-value 10^{-100} (Altschul et al., 1990). The results were filtered at a high identity percentage (≥98.5%). Redundant hits mapping on the same position of the same contig were discarded.

The abundance of each GUS was correlated to the abundance of the bin containing the contig where the GUS sequence was mapped to. In particular, the abundance of each GUS was calculated taking into account the number of reads mapping on the corresponding bin. The Jaccard similarity was computed to estimate the beta diversity based on GUS profiles and subjected to PCoA.

Statistical Analysis

Statistical analysis using ANOVA ($p < 0.05$) followed by Tukey's *post hoc* test was conducted to compare cohorts in terms of

[1]https://usegalaxy.eu

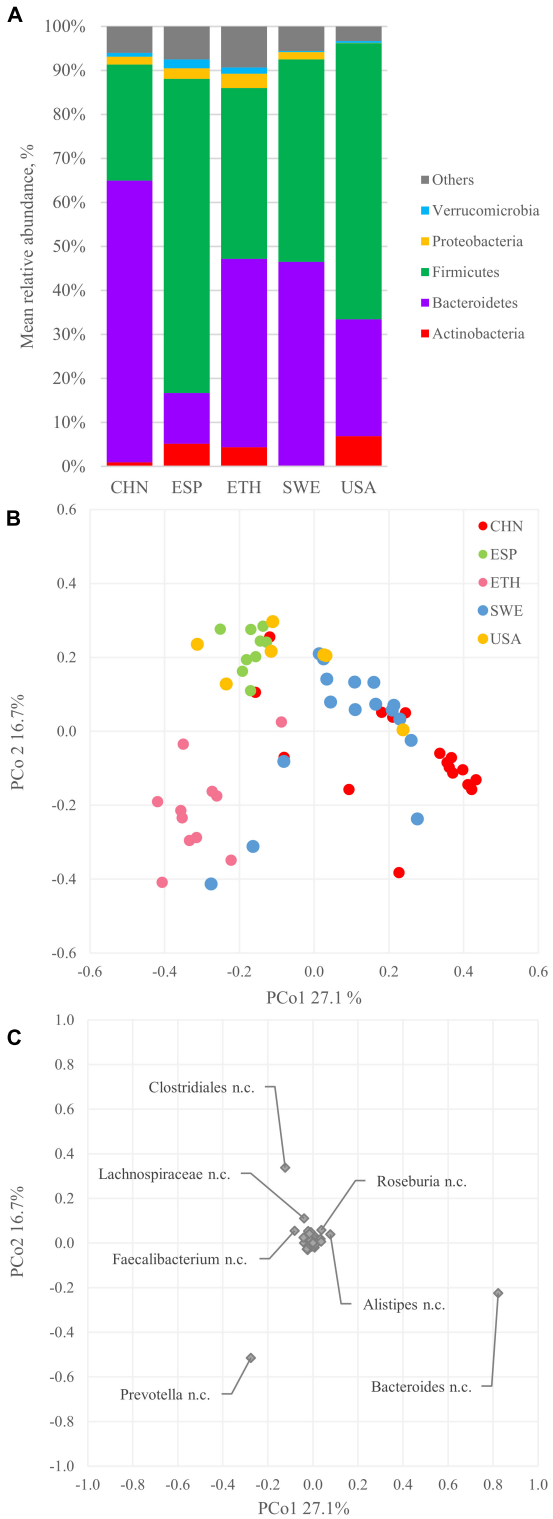


FIGURE 2 | Bacterial composition and beta diversity of gut metagenomes of 60 healthy adults. **(A)** Mean relative abundance of the main phyla identified in the whole dataset and in the five cohorts. **(B)** principal coordinate analysis (PCoA) plot of beta diversity based on Bray–Curtis dissimilarity index of the microbial composition. **(C)** PCoA plot of the species contribution to metagenome differentiation.

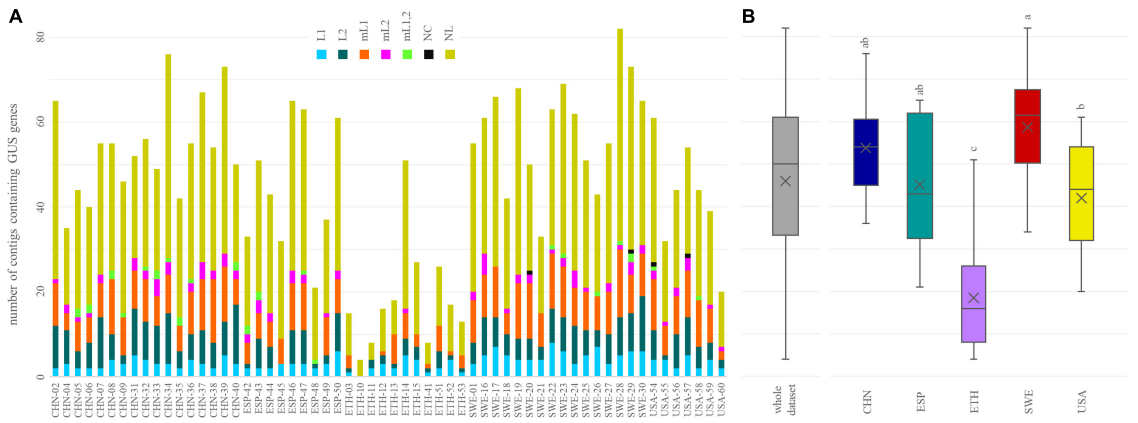


FIGURE 3 | The number of β-glucuronidases (GUS) identified in each metagenome **(A)** and in the whole dataset and in each cohort **(B)**. In **(B)**, cohorts sharing the same letter did not significantly differ ($P \geq 0.05$, ANOVA, Tukey).

GUS profiles, abundance of bacteria harboring GUS genes, and relative abundance of each GUS structural category. Alpha diversity indices of cohorts were compared with the Kruskal–Wallis test followed by Dunn's multiple-comparison test. In beta diversity analysis of microbiome composition and GUS profile, the statistical significance among cohorts was analyzed with PERMANOVA statistical test ($p < 0.05$).

RESULTS

Metagenomic Analysis

Sixty metagenomes of gut microbiota from healthy subjects, sequenced with Illumina paired-end technology, were retrieved and scanned according to the flowsheet reported in **Figure 1** to search the genes encoding the 279 GUS proteins identified by Pollet et al. (2017). The metagenomes encompassed 44 ± 35 million reads (mean ± SD), with lengths ranging between 100 and 150 bp (**Supplementary Figure 1A**). Assembly of metagenomes with MetaSPAdes yielded on average 408,745 ± 131,363 contigs per metagenome (mean ± SD) (**Supplementary Figure 1B**).

Reference-free binning with MaxBin2 recovered from 29 to 179 binned genomes per subject, with a mean value of 72 (**Supplementary Figure 1C**). For most metagenomes (>75%), the reads associated with a bin, mapped by Bowtie2, accounted for more than 80% (mean 85.2%) (**Supplementary Figure 1D**).

Bacterial Composition and Beta Diversity

The bins were quantified with Bowtie2 and assigned a taxonomic designation with CAT/BAT. The dominant phyla were Firmicutes and Bacteroidetes, with the former generally outnumbering the others and the latter dominating only CHN metagenomes (**Figure 2A**). The relative amounts of Actinobacteria and Proteobacteria were quite different among subjects, lying in the range of 0.2–6.9% and 0.1–3.3%, respectively. Verrucomicrobia ranged from 0.2 to 2.0% of the whole bacterial population. Bins ascribed to other phyla or lacking taxonomic attribution (labeled as "others") ranged from 3.3 to 9.3%. At deeper taxonomic level, the quantity of unclassified bins increased; thus, the profiling was less accurate. Among families, Bacteroidaceae were among the most abundant, with a mean of 22.5% in 60 metagenomes, resulting in its prevalence in the CHN cohort. Prevotellaceae were remarkably higher in the ETH cohort compared to the others. A similar distribution was observed in genera distribution, with *Bacteroides* prevailing in the CHN cohort and *Prevotella* in the ETH cohort (**Supplementary Figure 2**). For each cohort, core genera present in at least 85% of subjects were identified. Genera *Alistipes*, *Bacteroides*, *Faecalibacterium*, and *Ruminococcus* were identified in all the cohort (**Supplementary Figure 3**).

Alpha diversity of metagenomes was evaluated with Shannon index, Chao-1 index, and Pielou's evenness (**Supplementary Figure 4**). Shannon index showed a significant difference ($p < 0.05$) between the CHN and ETH cohorts. Chao-1 richness highlighted higher values for the SWE cohort, significantly different from the CHN and USA cohorts. Pielou's evenness presented high values for all cohorts, with CHN showing a wide distribution and being significantly different from ESP and ETH.

The beta diversity was assessed according to the Bray–Curtis dissimilarity index and analyzed with PCoA. The plot in **Figure 2B** displays the two most informative dimensions of the PCoA space, describing 27.1 and 16.7% of the diversity in the dataset. According to PERMANOVA, the grouping in cohorts was significant ($p < 0.05$), even though extensive overlapping of some cohorts was observed (e.g., ESP and USA). The CHN and ETH cohorts were separated along with PCo1, lying mostly at positive and negative values, respectively. Subjects belonging to the ESP, SWE, and USA cohorts mostly lie at positive PCo2 values, unlike the ones belonging to CHN and ETH, mostly located at negative PCo2.

The genus *Bacteroides* mainly contributed to PCo1 positive values that characterized the CHN subjects (**Figure 2C**), according to the prevalence of Bacteroidetes over Firmicutes in this cohort (**Figure 2A**). On the other side, *Prevotella* negatively contributed to the PCo1 autovector. Along with PCo2, the main

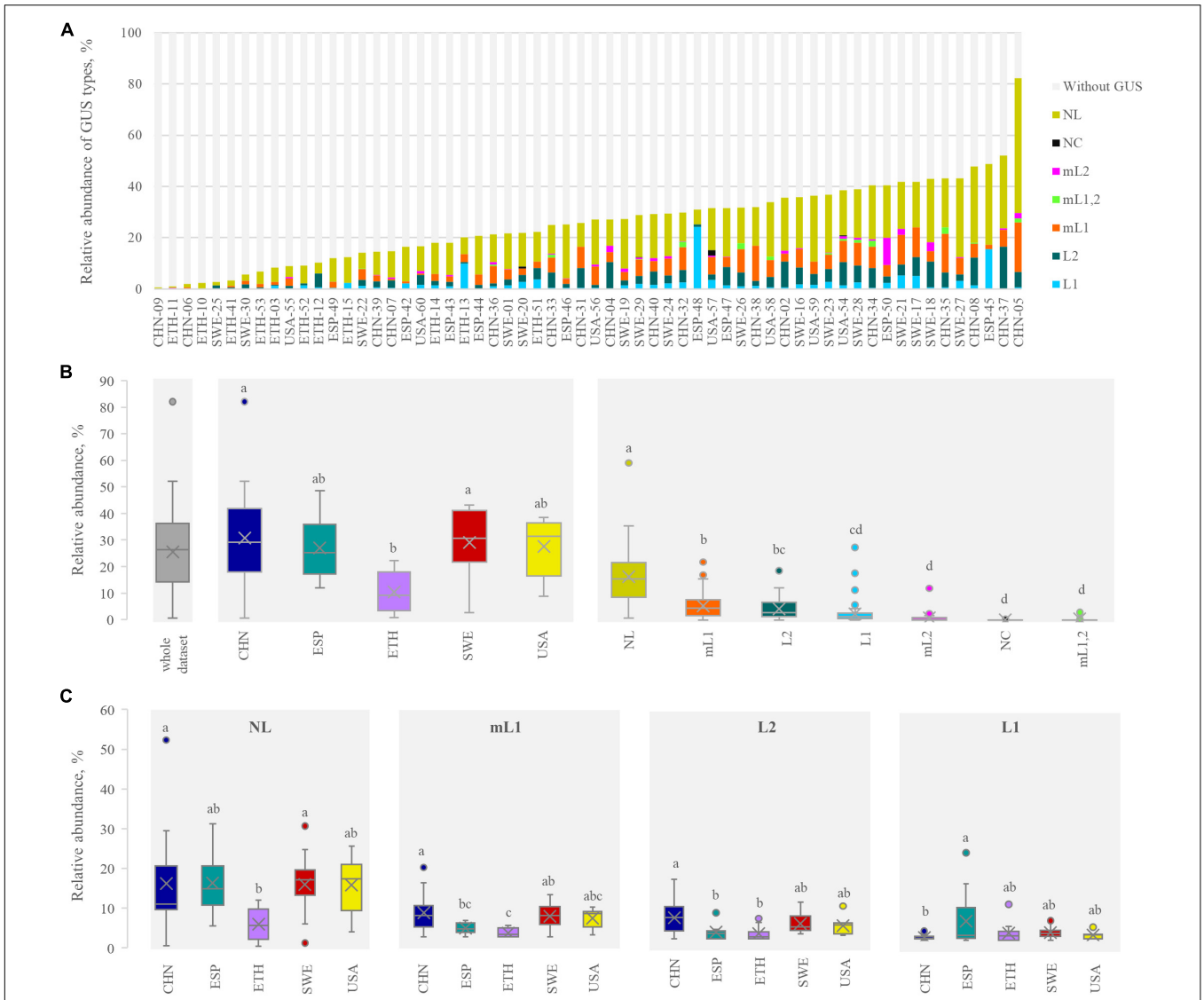


FIGURE 4 | β-glucuronidases (GUS) abundance profile in each metagenome **(A)** and in the whole dataset, in each cohort, and for each structural category **(B)**. The abundance of the four main categories in each cohort is shown in **(C)**. In **(B,C)** cohorts or categories sharing the same letter did not significantly differ ($P \geq 0.05$, ANOVA, Tukey). In determining the GUS abundance profile, for the bins bearing more than one GUS, the abundance was multiplied by the number of GUS therein identified.

positive contribution came from Clostridiales, while a negative one came from *Prevotella* and *Bacteroides*.

β-Glucuronidases Types and Categories

tBLASTn search within the whole sets of contigs pinpointed 218 of the 279 GUS sequences of the inventory of Pollet et al. (2017). Each metagenome encompassed 4 to 82 contigs containing at least a GUS sequence (**Figure 3A**). The number of different GUS types per subject ranged from 4 (ETH-10) to 70 (SWE-28), with a mean of 40. The richness in different GUS was similar among the cohorts ($p < 0.05$), except in ETH subjects, which presented significantly lower values ($p < 0.05$) (**Figure 3B**).

Genes encoding GUS of the categories NL, mL1, L2, and L1 were found in all or the vast majority of the metagenomes (≥57), while mL2, mL1,2, and NC GUS genes occurred less frequently (43, 21, and 5 metagenomes, respectively). In terms of both the overall number of sequences and the number of sequences per sample, NL was the richest category, followed by mL1, L2, and L1 (**Figure 3A**). NL accounted for 129 of the 218 sequences, reaching up to 50 different sequences per sample, while mL1, L2, and L1 respectively accounted for 33, 30, and 13 different sequences and reached up to 16, 14, and 13 sequences per sample. mL2, mL1,2, and NC were represented only by 7, 4, and 2 different sequences, respectively. Despite the different distribution of structural categories among the subjects, the grouping in cohorts was not significant ($p > 0.05$, ANOVA).

The relative abundance of the intestinal bacteria harboring at least a GUS gene was calculated, linking each GUS gene

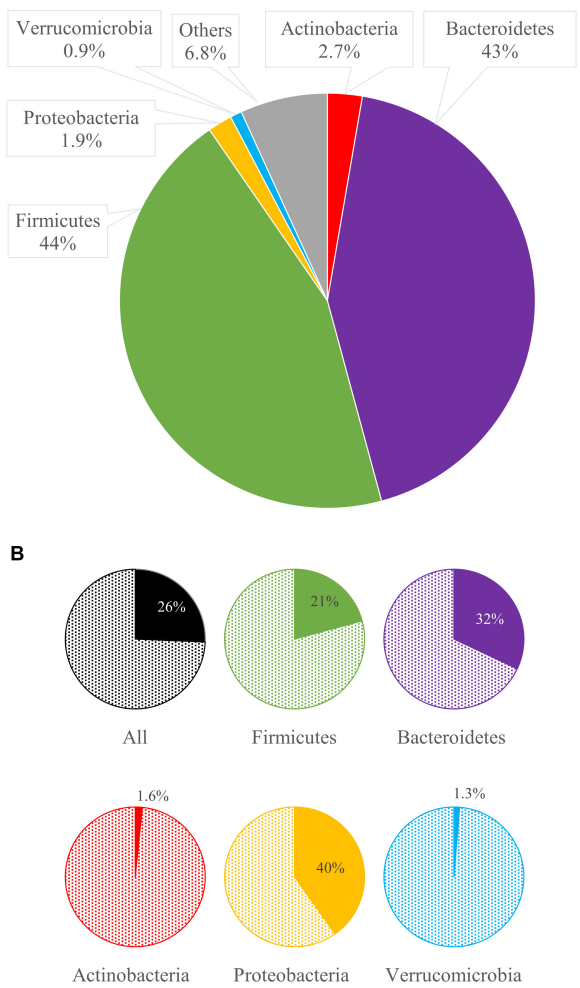


FIGURE 5 | The mean relative abundance of bacterial phyla **(A)** and, for each phylum, the mean abundance of bacteria bearing β-glucuronidases (GUS) genes **(B)**.

with the relative abundance of the corresponding bin, in its turn obtained by the number of reads mapping in the bin. GUS-encoding bacteria ranged from 0.7% (CHN-09) to 82.2% (CHN-05) (**Figure 4A**), with a mean abundance of 25.7%. The bacteria harboring GUS genes were significantly less abundant in the ETH than in the other cohorts ($p < 0.05$) (**Figure 4B**). However, the dataset presented a high variability, even within the same cohort. For instance, the CHN cohort encompassed both subjects where GUS-encoding bacteria presented the lowest and the highest abundance (CHN-9 and CHN-05, respectively).

With regard to the relative abundance of the GUS-encoding bacteria, NL was the most represented, followed by mL1, L2, and L1 (on average, 14.4, 4.6, 3.7, and 2.1%, respectively). Bacteria harboring GUS genes of mL2, NC, and mL1,2 categories were less abundant, accounting on average for less than 1%. Despite the low mean abundance of bacteria harboring L1 genes, this class of GUS presented the highest variability, with encoding bacteria ranging from 0 to 24.3%. The abundance of bacteria encoding each structural category of GUS was similar among the cohorts, with the sole exceptions of NL and mL1, which were less abundant in the ETH cohort ($p < 0.05$, ANOVA, Tukey) (**Figure 4C**).

Contribution of the Taxa to β-Glucuronidases Abundance

Abundances, frequencies, and taxa mostly contributing to GUS profile were explored. Taxonomic assignment of GUS was done according to the GUS types classified by Pollet et al. (2017). The abundance of taxa encoding each GUS type was calculated by summing the relative abundance of each bin harboring at least one GUS gene, normalized among the whole set of bins, encompassing or not the GUS genes.

Among the 60 metagenomes, 25.7% of the bins encoded for at least one GUS gene. The phyla Bacteroidetes and Firmicutes dominated all the microbiomes and encompassed many bacterial species encoding for at least one GUS gene (32.0 and 20.8%, respectively, **Figures 2A**, **5**).

The genus *Bacteroides* was the main contributor of the intestinal GUS pool, encoding 120 of the 218 GUS identified in this study. Among the 20 most relevant GUS sequences (**Table 1**), 13 were from *Bacteroides*, including the three most abundant ones (i.e., no. NL-11, mL1-176, and NL-36), which were also the most frequently occurring in the metagenomes. In particular, *Bacteroides* spp. encoding NL-11 reached up to 8.8% in CHN-37 and occurred in 42 metagenomes (**Table 1**). *Bacteroides vulgatus* encoding mL1-176 was the second most abundant GUS-encoding bin, accounting for 5.3% of total bacteria and being detected in 36 out of 60 metagenomes. *B. dorei* encoding mL1-177 occurred in 13 subjects with a mean abundance of 1.8% and reached more than 10% in CHN-05.

Among Firmicutes, the genus *Faecalibacterium* was a major contributor to the abundance of GUS genes. As a whole, the metagenomes encompassed 44 bins of *F. prausnitzii* and 197 bins of *Faecalibacterium* sp., accounting together, on average, for 6.2% of the microbiome, 2.9% encoding at least one GUS gene. In particular, among the 241 bins of *Faecalibacterium*, 125 harbored at least one GUS sequence, mainly belonging to categories L1, NL, or, more rarely, mL1. Eighty-three bins harbored L1 GUS genes, accounting, as a whole, for 0.55% of the metagenomes. In some cases, two diverse GUS were found in the same bin, generally L1 and NL. GUS sequences L1-223 and NL-67 of *Faecalibacterium prausnitzii* were frequently encountered, both being present in 37 of the 60 metagenomes. *F. prausnitzii* encoding GUS L1-223 and GUS NL-67 presented the highest abundance in SWE-21 (2.7%) and in ESP-42 (2.1%), respectively. Other GUS originating from Firmicutes came from the Clostridiales *Eubacterium*, *Ruminococcus*, *Roseburia*, and *Fusicatenibacter* or from unidentified Firmicutes. *Eubacterium* sp. CAG:180 encoding GUS NL-76 occurred in 14 metagenomes and was remarkably abundant in ESP-45 (14.2%) (**Supplementary Material 2** and **Supplementary Figure 5**).

Actinobacteria and Proteobacteria were minor phyla within the analyzed metagenomes, which scarcely contributed to the pool of GUS, since only a minority of bacteria ascribed to

TABLE 1 | List of the main β-glucuronidases (GUS) sequences identified.

GUS ID	Type	Origin	Mean%	Max% (subject)	Frequency, no. subjects (%)
11	NL	*Bacteroides* spp.	5.6	8.8 (CHN-37)	42 (70)
176	mL1	*Bacteroides vulgatus*	5.3	6.0 (CHN-35)	36 (60)
36	NL	*Bacteroides uniformis*	4.4	8.8 (CHN-37)	30 (50)
220	L1	*Escherichia coli*	3.3	23.3 (ESP-48)	11 (18)
87	NL	*Bacteroides dorei*	2.4	8.8 (CHN-05)	23 (38)
17	NL	*Bacteroides uniformis*	2.4	4.0 (SWE-16)	27 (45)
242	L2	*Bacteroides uniformis*	2.2	5.7 (CHN-08)	22 (37)
177	mL1	*Bacteroides dorei*	1.8	10.6 (CHN-05)	13 (22)
47	NL	*Bacteroides massiliensis*	1.8	4.9 (CHN-35)	16 (27)
67	NL	*Faecalibacterium prausnitzii*	1.7	2.1 (ESP-42)	37 (62)
35	NL	*Bacteroides uniformis*	1.6	2.8 (SWE-21)	15 (25)
223	L1	*Faecalibacterium prausnitzii*	1.6	2.6 (SWE-21)	37 (62)
10	NL	*Bacteroides ovatus*	1.6	2.3 (SWE-26)	25 (42)
185	mL1	*Bacteroides ovatus*	1.6	4.6 (CHN-05)	22 (37)
173	mL1	*Bacteroides massiliensis*	1.5	4.1 (CHN-38)	17 (28)
257	L2	*Bacteroides ovatus*	1.5	2.3 (SWE-26)	26 (43)
53	NL	*Parabacteroides merdae*	1.5	2.5 (CHN-08)	31 (52)
76	NL	*Eubacterium* sp. CAG:180	1.5	14.2 (ESP-45)	14 (23)
126	NL	*Firmicutes*	1.4	2.2 (ESP-47)	36 (60)
134	NL	*Firmicutes*	1.3	2.2 (ESP-47)	34 (57)
180	mL1	*Parabacteroides merdae*	1.3	2.5 (CHN-08)	26 (43)
261	L2	*Bacteroides cellulosilyticus* CAG:158	1.2	4.9 (CHN-08)	16 (27)

The sequences reported represent the 10 GUS with higher mean abundance, abundance in single microbiome, and frequency in the set of GUS-encoding bacteria.

these phyla-encoded GUS (1.6 and 1.3%, respectively). Among Proteobacteria, *E. coli* encoding GUS L1-220 mostly participated in GUSome, being detected in 11 metagenomes, remarkably abundant in ESP-48 (23.3%), ESP-45 (12.8%), and ETH-13 (8.5%).

The 40.2% of Verrucomicrobia harbored GUS genes (**Figure 5**). However, bacteria ascribed to this phylum represented only 0.9% within the set of microbiomes. *Akkermansia* was the main GUS-encoding genus ascribed to this phylum. Among the whole dataset, 13 bins were ascribed to *Akkermansia* sp., accounting for 0.6%, 10 of which harbored GUS genes. In particular, the sequence of GUS mL2-218 was identified in eight bins of *Akkermansia muciniphila*, representing 0.4% of the whole dataset, whereas another two bins of *Akkermansia* encoded the gene mL2-209. A remarkably high amount of *A. muciniphila* encoding GUS mL2-218 (10.6%) was found only in subject ESP-50.

The abundance of specific species known to encode several deeply characterized L1-GUS, such as *C. perfringens*, *E. eligens*, *L. rhamnosus*, *R. gnavus*, *S. agalactiae*, *B. uniformis*, *B. ovatus*, *B. dorei*, *B. fragilis*, and *P. merdae*, was inferred by MetaPhlAn2 metagenomic analysis because CAT/BAT failed to name these bins with a species designation. *L. rhamnosus*, *S. agalactiae*, and *C. perfringens* lie below the limit of detection in most of the metagenomes ($\geq$ 53/60) and, when found, were present at very low abundances ($\leq$ 0.15%). *E. eligens* and *R. gnavus* were more represented (observed in 51 and 39 subjects, respectively) with a mean abundance of 1.0% and 0.3%. *B. uniformis* was the most abundant among the above-mentioned Bacteroidetes, being detected in 55/60 metagenomes at the mean abundance of 2.7%, followed by *B. ovatus* (1.1%; 51/60), *B. dorei* (1.0%; 51/60), and *B. fragilis* (0.5%, 32/60), while *P. merdae* was pinpointed in 47 subjects with a mean abundance of 1.2%.

The beta diversity was computed based on the presence of the 218 GUS sequences. Jaccard metrics were utilized, with a qualitative approach to prevent abundant *Bacteroides* from concealing the differences among metagenomes and cohorts. The plot in **Figure 6A** displays the PCoA space of beta diversity in the two most informative dimensions, describing 11.8 and 6.4% of the diversity. Cohort grouping based on the presence of GUS sequences was statistically significant according to PERMANOVA ($p < 0.05$). The CHN and SWE cohorts were characterized by a negative value of PCo1, and the ETH by a positive one. Most of the GUS that negatively contributed to PCo1 originated from *Bacteroides*, mainly *B. uniformis* and *B. vulgatus*. On the other side, the main positive contributors to PCo1 were GUS sequences from *Prevotella*, *Prevotella copri*, *F. prausnitzii*, *Eubacterium*, and *E. coli* (**Figure 6B**).

DISCUSSION

Pharmacomicrobiomics investigates the interplay of microbiome diversity and drug disposition and response and may provide an important basis in personalized medicine (Doestzada et al., 2018; Hassan et al., 2021). The role of bacterial GUS on drug bioavailability and biological effects (i.e., the reactivation and absorption vs the excretion) encouraged our deep analysis of

interindividual variability of the intestinal GUS, originating from different microbiome compositions. With this aim, metagenome data of 60 healthy adults from different geographic provenance were utilized to predict and compare their GUSome, i.e., the abundance and the diversity of GUS genes and structural categories.

The abundance of intestinal bacteria harboring at least a GUS gene was highly variable among subjects, ranging from a very small minority (e.g., 0.7% in CHN-09) to an overwhelming majority (e.g., 82.2% in CHN-05) of the intestinal microbial community. No significant difference between cohorts could be identified, except for ETH, where GUS-encoding bacteria were significantly less abundant. The GUSome (i.e., the abundance and the diversity of GUS genes within the metagenome) was generally dominated by category NL, followed by mL1, L2, and L1, in agreement with literature information (Pollet et al., 2017). In some subjects, particularly in the ETH cohort, a low level of GUS sequences was in relation to the low abundance of bacteria encoding NL and mL1 GUS.

NL and L2 GUS are involved in processing commonly available large substrates (e.g., mucins, glycosaminoglycans, plant polysaccharides, etc.) (Pollet et al., 2017). On the one hand, their abundance and frequency in the microbiome are presumably related to a role in the energetic metabolism of intestinal bacteria and may be a positive adaptive trait related to the evolution of bacteria in the colonic environment. On the other hand, GUS belonging to categories L1 and mL1 are involved in the hydrolysis of small molecules, including drugs (Pollet et al., 2017; Biernat et al., 2019).

L1 GUS process estrone and estradiol glucuronides, affecting the estrogen profile and promoting the onset of hormonal disorders (Baker et al., 2017; Ervin et al., 2019). L1 GUS also participate in the toxicity of irinotecan since they are key effectors of metabolite SN-38 reactivation and have been targeted by specific inhibitors to protect the intestinal epithelial cells and to reduce chemotherapy-induced diarrhea (Bhatt et al., 2020; Jariwala et al., 2020; Parvez et al., 2021; Wang et al., 2021). Moreover, L1 GUS are responsible for the deconjugation of glucuronidated non-steroidal anti-inflammatory drugs (NSAIDs), which is among the most commonly used medications worldwide (Maseda and Ricciotti, 2020).

In most of the microbiomes analyzed in this study, the bacteria encoding L1 GUS genes presented a relatively low abundance. The ratio of bacteria encoding L1 genes was less than 2% in 34 subjects and accounted for 2–5% in 12, resulting significantly higher in the SPA cohort than in the CHN one. Three microbiomes presented a very high abundance of L1 GUS-encoding bacteria (ESP-48 24.3%, ESP-45 15.6%, and ETH-13 9.9%). The most relevant L1 GUS, in terms of frequency and abundance of the corresponding bins, were L1-220 GUS of *E. coli* and L1-223 of *F. prausnitzii* (**Table 1**). *E. coli* containing the L1-220 GUS gene were present in 11 out of 60 metagenomes and generally occurred in low concentrations. These results are consistent with data reported in a recent study aimed to characterize intestinal *E. coli* from healthy adults (Raimondi et al., 2019), revealing that approx. a fifth of *E. coli* isolates did not present β-glucuronidase activity. In the microbiomes where L1-encoding *E. coli* presented a high abundance (e.g., ESP-48 23.3%, ESP-45 12.8%, and ETH-13 8.5%), it is expected that L1 GUS from this commensal pathobiont could heavily interact with drug glucuronides, affecting efficacy and/or toxicity. *Faecalibacterium* sp. and *F. prausnitzii*, which are commensals associated with beneficial health effects, were much more frequent and abundant, representing on average 6.2% of the microbiomes herein analyzed. Despite their abundance, only a negligible ratio of *Faecalibacterium* encoded L1 GUS, accounting for 0.55% of the set of bacteria. However, the small portion of *Faecalibacterium* encoding L1-223 remained the major player in L1-catalyzed deglucuronidations, being identified in the majority of the subjects. On the other side, *E. coli* encoding L1-220 could participate in deconjugation in few microbiomes where it is exceptionally abundant. However, L1-220-encoding *E. coli* lie below the limit of detection in most cases.

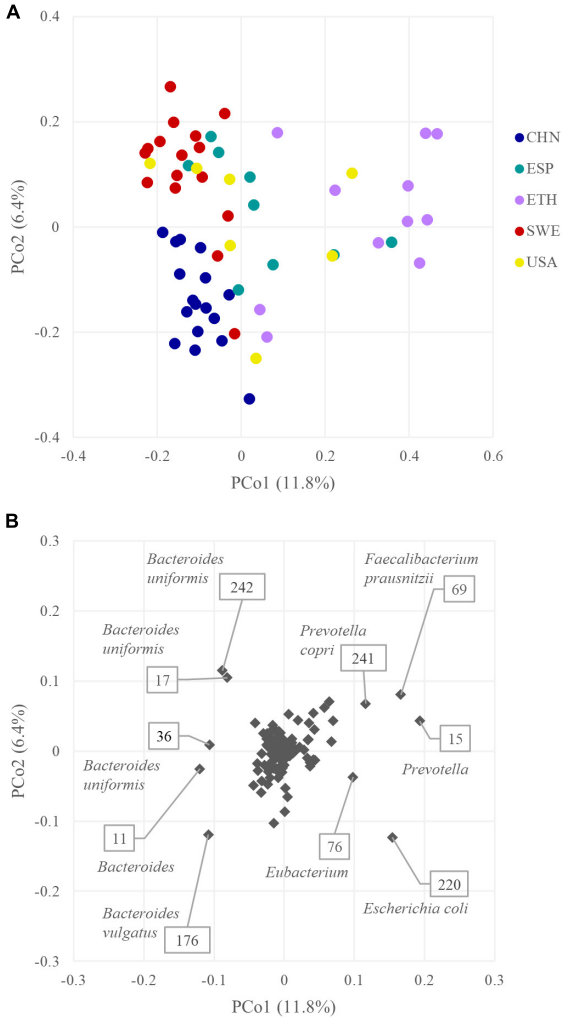


FIGURE 6 | (A) PCoA plot of beta diversity based on Jaccard dissimilarity index of the β-glucuronidases (GUS) profiles of the 60 metagenomes. **(B)** PCoA plot of the GUS contribution to GUSome differentiation.

Literature reports several L1-GUS that have been deeply characterized in some Firmicutes (i.e., *C. perfringens*, *E. eligens*, *L. rhamnosus, R. gnavus*, and *S. agalactiae*) and Bacteroidetes (i.e., *B. ovatus*, *B. dorei*, *B. fragilis*, and *P. merdae*) (Pollet et al., 2017; Biernat et al., 2019; Jariwala et al., 2020). CAT/BAT did not pinpoint nearly any bin of these species, likely because of the high fragmentation of metagenomic assemblies; thus, they were searched using MetaPhlAn2, which allowed higher accuracy of taxonomic profiling. *C. perfringens*, *L. rhamnosus*, and *S. agalactiae* were found in a minority of metagenomes at a negligible concentration. On the other hand, *E. eligens* and *R. gnavus* occurred more frequently and abundantly in the metagenomes and thus are expected to contribute to the pool of L1 GUS.

L1 GUS are involved in the deconjugation of glucuronidated NSAIDs, such as diclofenac. The glucuronide of diclofenac is synthesized in the liver and excreted in the gut lumen (Maseda and Ricciotti, 2020). L1 GUS from *E. eligens* is the most active in hydrolyzing this glucuronide (kcat 138 s^{-1}), followed by the corresponding of *S. agalactiae*, *C. perfingens*, *F. prausnitzii*, *E. coli*, and *R. gnavus* (kcat from 97 to 30.7 s^{-1}), while that from *L. rhamnosus* presents a catalytic efficiency approx. one magnitude lower (Biernat et al., 2019). Taking into account the abundance of these species, *E. eligens*, *F. prausnitzii*, and *E. coli* are expected to mostly contribute to the release of the diclofenac aglycone from the glucuronide. The peculiar richness of some of these species, as detected in our dataset for *E. coli*, likely modifies the clearance of the drug, facilitating reuptake and recirculation. A similar pattern of catalytic activity has also been assessed for SN-38, with L1 GUS reactivating this toxic metabolite of the anticancer drug irinotecan and causing consequent gastrointestinal toxicity (Jariwala et al., 2020; Parvez et al., 2021).

Glucuronides of small drugs can be also hydrolyzed by mL1 GUS (Wallace et al., 2015; Biernat et al., 2019). For instance, GUS mL1-188 of *B. fragilis* was described to possess a remarkably high activity against p-nitrophenol-β-D-glucuronide, diclofenac-glucuronide, and SN38-glucuronide, albeit lower than L1 enzymes (Biernat et al., 2019; Bhatt et al., 2020). In the microbiomes analyzed in this study, the abundance of mL1 GUS-encoding bacteria was significantly higher than L1, with the CHN cohort significantly richer than the ESP and ETH ones. Among the 16 mL1 GUS sequences herein identified, the one encoded by *B. vulgatus* (mL1-176) was the main in terms of abundance and frequency, followed by those encoded by *B. dorei*, *B. ovatus*, and *B. massiliensis* (mL1-177, mL1-185, and mL1-173, respectively). mL1 GUS 188 of *B. fragilis*, the structure and activity of which has been deeply characterized (Wallace et al., 2015; Biernat et al., 2019), was retrieved in 13 out of 60 metagenomes, but its mean abundance was 0.15%. Since the genus *Bacteroides* generally occurs at high levels in the microbiomes and encodes the main mL1 GUS, it is conceivable that this genus plays an important role in the regeneration of parent compounds or active metabolites in the gut, evoking their major contribution to drug reactivation.

Within the observed broad interindividual heterogeneity of GUS profiles, the described differences in terms of L1 and mL1 GUS likely represent the major drivers of the variability in pharmacomicrobiomics, affecting the level of active deconjugated molecules and thus influencing drug response and toxicity. Strains belonging to the species *B. ovatus*, *B. dorei*, *B. fragilis*, *E. coli*, *E. eligens*, *F. prausnitzii*, *P. merdae*, and *R. gnavus* emerged as pivotal in the differentiation of catalytic activity toward small glucuronides in the host and can be claimed as the main players in the reactivation of drug metabolites by the various intestinal microbial ecosystems. The targeted control of gut microbiota, by modulating metabolic diversity, represents a future perspective to govern and reduce the deconjugation activity. This strategy, also based on the administration of selected probiotics, could mitigate extreme and severe differences among patients in terms of adverse drug responses, responsible for numerous disease states.

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

FC, AA, SR, and MR conceived the study. FC, EM, and RR carried out the bioinformatic analysis. AA, SR, FC, and AZ conducted data interpretation and presentation. FC and MR wrote the manuscript with contributions from SR, RR, EM, AZ, and AA. All authors contributed to the article and approved the submitted version.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.826994/full#supplementary-material

REFERENCES

Afgan, E., Baker, D., Batut, B., van den Beek, M., Bouvier, D., Cech, M., et al. (2018). The galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res.* 46, W537–W544. doi: 10.1093/nar/gky379

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. doi: 10.1016/S0022-2836(05)80360-2

Andrews, S. (2010). *FastQC: A Quality Control Tool for High Throughput Sequence Data*. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc (accessed January 20, 2022).

Awolade, P., Cele, N., Kerru, N., Gummidi, L., Oluwakemi, E., and Singh, P. (2020). Therapeutic significance of β–glucuronidase activity and its inhibitors: a review. *Eur. J. Med. Chem.* 187:111921. doi: 10.1016/j.ejmech.2019.111921

Baker, J. M., Al-Nakkash, L., and Herbst-Kralovetz, M. M. (2017). Estrogen–gut microbiome axis: physiological and clinical implications. *Maturitas* 103, 45–53. doi: 10.1016/j.maturitas.2017.06.025

Bhatt, A. P., Pellock, S. J., Biernat, K. A., Walton, W. G., Wallace, B. D., Creekmore, B. C., et al. (2020). Targeted inhibition of gut bacterial β–glucuronidase activity enhances anticancer drug efficacy. *Proc. Natl. Acad. Sci. U.S.A.* 117, 7374–7381. doi: 10.1073/pnas.1918095117

Biernat, K. A., Li, B., and Redinbo, M. R. (2018). Microbial unmasking of plant glycosides. *mBio* 9:e02433-17. doi: 10.1128/mBio.02433-17

Biernat, K. A., Pellock, S. J., Bhatt, A. P., Bivins, M. M., Walton, W. G., Tran, B., et al. (2019). Structure, function, and inhibition of drug reactivating human gut microbial β-glucuronidases. *Sci. Rep.* 9:825. doi: 10.1038/s41598-018-36069-w

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., et al. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat. Biotechnol.* 37, 852–857. doi: 10.1038/s41587-019-0209-9

Doestzada, M., Vila, A. V., Zhernakova, A., Koonen, D. P. Y., Weersma, R. K., Touw, D. J., et al. (2018). Pharmacomicrobiomics: a novel route towards personalized medicine? *Protein Cell* 9, 432–445. doi: 10.1007/s13238-018-0547-2

El Kaoutari, A., Armougom, F., Gordon, J. I., Raoult, D., and Henrissat, B. (2013). The abundance and variety of carbohydrate–active enzymes in the human gut microbiota. *Nat. Rev. Microbiol.* 11, 497–504. doi: 10.1038/nrmicro3050

Elmassry, M. M., Kim, S., and Busby, B. (2021). Predicting drug-metagenome interactions: variation in the microbial β–glucuronidase level in the human gut metagenomes. *PLoS One* 16:e0244876. doi: 10.1371/journal.pone.0244876

Ervin, S. M., and Redinbo, M. R. (2020). The gut microbiota impact cancer etiology through "Phase IV Metabolism" of xenobiotics and endobiotics. *Cancer Prev. Res. (Philadelphia, PA)* 13, 635–642. doi: 10.1158/1940-6207.CAPR-20-0155

Ervin, S. M., Li, H., Lim, L., Roberts, L. R., Liang, X., Mani, S., et al. (2019). Gut microbial β–glucuronidases reactivate estrogens as components of the estrobolome that reactivate estrogens. *J. Biol. Chem.* 294, 18586–18599. doi: 10.1074/jbc.RA119.010950

Flint, H. J., Scott, K. P., Duncan, S. H., Louis, P., and Forano, E. (2012). Microbial degradation of complex carbohydrates in the gut. *Gut Microbes* 3, 289–306. doi: 10.4161/gmic.19897

Gloux, K., Berteau, O., El Oumami, H., Béguet, F., Leclerc, M., and Doré, J. (2011). A metagenomic β–glucuronidase uncovers a core adaptive function of the human intestinal microbiome. *Proc. Natl. Acad. Sci. U.S.A.* 108, 4539–4546. doi: 10.1073/pnas.1000066107

Hammer, Ø, Harper, D. A., and Ryan, P. D. (2001). PAST: paleontological statistics software package for education and data analysis. *Palaeontol. Electron.* 4, 1–9.

Hassan, R., Allali, I., Agamah, F. E., Elsheikh, S. S. M., Thomford, N. E., Dandara, C., et al. (2021). Drug response in association with pharmacogenomics and pharmacomicrobiomics: towards a better personalized medicine. *Brief. Bioinform.* 22:bbaa292. doi: 10.1093/bib/bbaa292

Huang, X., Nie, S., and Xie, M. (2017). Interaction between gut immunity and polysaccharides. *Crit. Rev. Food Sci. Nutr.* 57, 2943–2955. doi: 10.1080/10408398.2015.1079165

Jariwala, P. B., Pellock, S. J., Goldfarb, D., Cloer, E. W., Artola, M., Simpson, J. B., et al. (2020). Discovering the microbial enzymes driving drug toxicity with activity–based protein profiling. *ACS Chem. Biol.* 15, 217–225. doi: 10.1021/acschembio.9b00788

Jarrar, Y., and Lee, S. J. (2021). The Functionality of UDP–glucuronosyltransferase genetic variants and their association with drug responses and human diseases. *J. Pers. Med.* 11:554. doi: 10.3390/jpm11060554

Koppel, N., Maini Rekdal, V., and Balskus, E. P. (2017). Chemical transformation of xenobiotics by the human gut microbiota. *Science* 356:eaag2770. doi: 10.1126/science.aag2770

Langmead, B., and Salzberg, S. (2013). Fast gapped–read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923

Liston, H. L., Markowitz, J. S., and DeVane, C. L. (2001). Drug glucuronidation in clinical psychopharmacology. *J. Clin. Psychopharmacol.* 21, 500–515. doi: 10.1097/00004714-200110000-00008

Martin, M. (2011). Cutadapt removes adapter sequences from high–throughput sequencing reads. *EMBnet J.* 17, 10–12. doi: 10.14806/ej.17.1.200

Masamune, H. (1934). Biochemical studies on carbohydrates. IV. On an enzyme which catalyses the hydrolysis of biosynthetic osides of glucuronic acid. *J. Biochem.* 19, 353–375.

Maseda, D., and Ricciotti, E. (2020). NSAID–gut microbiota interactions. *Front. Pharmacol.* 11:1153. doi: 10.3389/fphar.2020.01153

McBain, A. J., and Macfarlane, G. T. (1998). Ecological and physiological studies on large intestinal bacteria in relation to production of hydrolytic and reductive enzymes involved in formation of genotoxic metabolites. *J. Med. Microbiol.* 47, 407–416. doi: 10.1099/00222615-47-5-407

Meech, R., Miners, J. O., Lewis, B. C., and MacKenzie, P. I. (2012). The glycosidation of xenobiotics and endogenous compounds: versatility and redundancy in the UDP glycosyltransferase superfamily. *Pharmacol. Ther.* 134, 200–218. doi: 10.1016/j.pharmthera.2012.01.009

Nakamura, J., Kubota, Y., Miyaoka, M., Saitoh, T., Mizuno, F., and Benno, Y. (2002). Comparison of four microbial enzymes in *Clostridia* and *Bacteroides* isolated from human feces. *Microbiol. Immunol.* 46, 487–490. doi: 10.1111/j.1348-0421.2002.tb02723.x

Nurk, S., Meleshko, D., Korobeynikov, A., and Pevzner, P. A. (2017). MetaSPAdes: a new versatile metagenomic assembler. *Genome Res.* 27, 824–834. doi: 10.1101/gr.213959.116

Oliphant, K., and Allen-Vercoe, E. (2019). Macronutrient metabolism by the human gut microbiome: major fermentation by–products and their impact on host health. *Microbiome* 7:91. doi: 10.1186/s40168-019-0704-8

Oshima, G. (1934). Biochemical studies on carbohydrates. XII. On β–glucuronidase, 2nd communication. *J. Biochem.* 20, 361–370.

Parvez, M. M., Basit, A., Jariwala, P. B., Gáborik, Z., Kis, E., Heyward, S., et al. (2021). Quantitative investigation of irinotecan metabolism, transport, and gut microbiome activation. *Drug Metab. Dispos.* 49, 683–693. doi: 10.1124/dmd.121.000476

Peekhaus, N., and Conway, T. (1998). What's for dinner?: entner–doudoroff metabolism in *Escherichia coli*. *J. Bacteriol.* 180, 3495–3502. doi: 10.1128/JB.180.14.3495-3502.1998

Pellock, S. J., and Redinbo, M. R. (2017). Glucuronides in the gut: sugar–driven symbioses between microbe and host. *J. Biol. Chem.* 292, 8569–8576. doi: 10.1074/jbc.R116.767434

Pellock, S. J., Walton, W. G., Biernat, K. A., Torres-Rivera, D., Creekmore, B. C., Xu, Y., et al. (2018). Three structurally and functionally distinct β–glucuronidases from the human gut microbe *Bacteroides* uniformis. *J. Biol. Chem.* 293, 18559–18573. doi: 10.1074/jbc.RA118.005414

Pollet, R. M., D'Agostino, E. H., Walton, W. G., Xu, Y., Little, M. S., Biernat, K. A., et al. (2017). An atlas of β–glucuronidases in the human intestinal microbiome. *Structure* 25, 967.e–977.e. doi: 10.1016/j.str.2017.05.003

Raimondi, S., Calvini, R., Candeliere, F., Leonardi, A., Ulrici, A., Rossi, M., et al. (2021). Multivariate analysis in microbiome description: correlation of human gut protein degraders, metabolites, and predicted metabolic functions. *Front. Microbiol.* 12:723479. doi: 10.3389/fmicb.2021.723479

Raimondi, S., Righini, L., Candeliere, F., Musmeci, E., Bonvicini, F., Gentilomi, G., et al. (2019). Antibiotic resistance, virulence factors, phenotyping, and genotyping of *E. coli* isolated from the feces of healthy subjects. *Microorganisms* 7:251. doi: 10.3390/microorganisms7080251

Roberts, M. S., Magnusson, B. M., Burczynski, F. J., and Weiss, M. (2002). Enterohepatic circulation: physiological, pharmacokinetic and clinical implications. *Clin. Pharmacokinet.* 41, 751–790. doi: 10.2165/00003088-200241100-00005

Rossi, M., Amaretti, A., Leonardi, A., Raimondi, S., Simone, M., and Quartieri, A. (2013). Potential impact of probiotic consumption on the bioactivity of dietary phytochemicals. *J. Agric. Food Chem.* 61, 9551–9558. doi: 10.1021/jf402722m

Ruan, W., Engevik, M. A., Spinler, J. K., and Versalovic, J. (2020). Healthy human gastrointestinal microbiome: composition and function after a decade of exploration. *Dig. Dis. Sci.* 65, 695–705. doi: 10.1007/s10620-020-06118-4

Russell, W. M., and Klaenhammer, T. R. (2001). Identification and cloning of gusA, encoding a new β–glucuronidase from *Lactobacillus gasseri* ADH. *Appl. Environ. Microbiol.* 67, 1253–1261. doi: 10.1128/AEM.67.3.1253-1261.2001

Saldana-Morales, F. B., Kim, D. V., Tsai, M. T., and Diehl, G. E. (2021). Healthy intestinal function relies on coordinated enteric nervous system, immune system, and epithelium responses. *Gut Microbes* 13, 1–14. doi: 10.1080/19490976.2021.1916376

Segata, N., Waldron, L., Ballarini, A., Narasimhan, V., Jousson, O., and Huttenhower, C. (2012). Metagenomic microbial community profiling using unique clade–specific marker genes. *Nat. Methods* 9, 811–814. doi: 10.1038/nmeth.2066

Truong, D. T., Franzosa, E. A., Tickle, T. L., Scholz, M., Weingart, G., Pasolli, E., et al. (2015). MetaPhlAn2 for enhanced metagenomic taxonomic profiling. *Nat. Methods* 12, 902–903. doi: 10.1038/nmeth.3589

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. doi: 10.1038/nature06244

von Meijenfeldt, F., Arkhipova, K., Cambuy, D. D., Coutinho, F. H., and Dutilh, B. E. (2019). Robust taxonomic classification of uncharted microbial sequences and bins with CAT and BAT. *Genome Biol.* 20:217. doi: 10.1186/s13059-019-1817-x

Wallace, B. D., Roberts, A. B., Pollet, R. M., Ingle, J. D., Biernat, K. A., Pellock, S. J., et al. (2015). Structure and inhibition of microbiome β–glucuronidases essential to the alleviation of cancer drug toxicity. *Chem. Biol.* 22, 1238–1249. doi: 10.1016/j.chembiol.2015.08.005

Wang, L., Sun, R., Zhang, Q., Luo, Q., Zeng, S., Li, X., et al. (2019). An update on polyphenol disposition *via* coupled metabolic pathways. *Expert Opin. Drug Metab. Toxicol.* 15, 151–165. doi: 10.1080/17425255.2019.1559815

Wang, P., Jia, Y., Wu, R., Chen, Z., and Yan, R. (2021). Human gut bacterial β–glucuronidase inhibition: an emerging approach to manage medication therapy. *Biochem. Pharmacol.* 190:114566. doi: 10.1016/j.bcp.2021.114566

Wu, Y. W., Tang, Y. H., Tringe, S. G., Simmons, B. A., and Singer, S. W. (2014). MaxBin: an automated binning method to recover individual genomes from metagenomes using an expectation–maximization algorithm. *Microbiome* 2:26. doi: 10.1186/2049-2618-2-26

Wu, Y., Simmons, B. A., and Singer, S. W. (2016). MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* 32, 605–607. doi: 10.1093/bioinformatics/btv638

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

REVIEW
published: 21 March 2022
doi: 10.3389/fmicb.2022.856165

Inflammatory Bowel Disease and COVID-19: How Microbiomics and Metabolomics Depict Two Sides of the Same Coin

*Gian Mario Cortes[1], Maria Antonietta Marcialis[1], Flaminia Bardanzellu[1], Angelica Corrias[1], Vassilios Fanos[1] and Michele Mussap[2]**

[1]Neonatal Intensive Care Unit, Department of Surgical Sciences, University of Cagliari, Monserrato, Italy, [2]Laboratory Medicine, Department of Surgical Sciences, School of Medicine, University of Cagliari, Monserrato, Italy

The integrity of the gastrointestinal tract structure and function is seriously compromised by two pathological conditions sharing, at least in part, several pathogenetic mechanisms: inflammatory bowel diseases (IBD) and coronavirus disease 2019 (COVID-19), caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection. IBD and COVID-19 are marked by gut inflammation, intestinal barrier breakdown, resulting in mucosal hyperpermeability, gut bacterial overgrowth, and dysbiosis together with perturbations in microbial and human metabolic pathways originating changes in the blood and fecal metabolome. This review compared the most relevant metabolic and microbial alterations reported from the literature in patients with IBD with those in patients with COVID-19. In both diseases, gut dysbiosis is marked by the prevalence of pro-inflammatory bacterial species and the shortfall of anti-inflammatory species; most studies reported the decrease in *Firmicutes*, with a specific decrease in obligately anaerobic producers short-chain fatty acids (SCFAs), such as *Faecalibacterium prausnitzii*. In addition, *Escherichia coli* overgrowth has been observed in IBD and COVID-19, while *Akkermansia muciniphila* is depleted in IBD and overexpressed in COVID-19. In patients with COVID-19, gut dysbiosis continues after the clearance of the viral RNA from the upper respiratory tract and the resolution of clinical symptoms. Finally, we presented and discussed the impact of gut dysbiosis, inflammation, oxidative stress, and increased energy demand on metabolic pathways involving key metabolites, such as tryptophan, phenylalanine, histidine, glutamine, succinate, citrate, and lipids.

Keywords: inflammatory bowel disease, Crohn's disease, ulcerative colitis, SARS-CoV-2, COVID-19, metabolomics, microbiomics

OPEN ACCESS

Edited by:
Roshan Kumar, Magadh University, India

Reviewed by:
Younes Zaid, Université de Montréal, Canada
Andrew S. Day, University of Otago, New Zealand
Busra Aktas, Mehmet Akif Ersoy University, Turkey

****Correspondence:***
Michele Mussap
mumike153@gmail.com

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

Received: *16 January 2022*
Accepted: *21 February 2022*
Published: *21 March 2022*

Citation:
Cortes GM, Marcialis MA, Bardanzellu F, Corrias A, Fanos V and Mussap M (2022) Inflammatory Bowel Disease and COVID-19: How Microbiomics and Metabolomics Depict Two Sides of the Same Coin. Front. Microbiol. 13:856165. doi: 10.3389/fmicb.2022.856165

INTRODUCTION

Since the onset of the pandemic outbreak caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), it emerged that frailty, elderly, and pre-existing chronic diseases, such as chronic kidney disease, hypertension, cardiovascular disease, and diabetes, are risk factors for the development of severe and/or fatal coronavirus disease 2019 (COVID-19;

Grasselli et al., 2020; Falandry et al., 2021). Theoretically, patients with immune-mediated inflammatory diseases, such as inflammatory bowel disease (IBD), might be at increased risk of developing severe COVID-19. However, current knowledge on the pathophysiology of IBD and COVID-19 points out that patients with IBD are not at increased risk or have adverse outcomes for COVID-19 (Neurath, 2020). Strong evidence supporting this conclusion emerge from clinical studies published elsewhere (Allocca et al., 2020; D'Amico et al., 2020), including the discovery that biological therapies (e.g., monoclonal antibodies) may play a protective role against the cytokine storm observed in the course of the SARS-CoV-2 infection (Allocca and Craviotto, 2021). IBD and COVID-19 may share many alterations in molecular mechanisms, microbial communities, and biochemical pathways; "omics" technologies may considerably contribute to decipher mechanisms inducing these alterations, improving patient care and outcome. Microbiome and metabolome were primarily investigated in IBD, and similar but relatively few studies were conducted in patients with COVID-19; in this review, we examined analogies and differences in gut microbiota and body fluids metabolome between IBD and COVID-19 with the aim to identify microbial and metabolic hallmarks linking IBD, COVID-19 and SARS-CoV-2 infection.

SARS-CoV-2 INFECTION IN PATIENTS WITH IBD

IBD is an umbrella term encompassing a group of disorders, namely, Crohn's disease (CD), ulcerative colitis (UC), and inflammatory bowel disease, type unclassified (IBDU; Satsangi et al., 2006). IBD is marked by chronic relapsing–remitting or continuously active idiopathic inflammation and bowel injuries; both adults and children exhibit an immunological dysregulation. The etiology of IBD is multifactorial, including the contribution of genetic, environmental, host factors and their reciprocal interactions (Flynn and Eisenstein, 2019); recent data indicate a worldwide 0.3% incidence and prevalence of IBD (Ng et al., 2017). The therapeutic treatment of IBD with immunomodulators (Park et al., 2020) and biologics (Neurath, 2019) may activate a transient or persistent immunocompromised state inducing opportunistic infections, especially when multiple drugs are prescribed simultaneously (Bonovas et al., 2016; Shah et al., 2017; Irving et al., 2021). Several research groups investigated whether patients with IBD may be or not more susceptible to developing SARS-CoV-2 infection (Monteleone and Ardizzone, 2020), how they should be managed in the context of the COVID-19 pandemic, and the risks and benefits of the therapeutic treatment with immunomodulators, especially in the pediatric age (Dipasquale et al., 2020; Sultan et al., 2020). An early analysis of data collected from the international registry Surveillance Epidemiology of Coronavirus Under Research Exclusion for Inflammatory Bowel Disease (SECURE-IBD) showed that among 525 pediatric and adult patients with IBD and confirmed COVID-19, 31% were hospitalized, 7% developed severe COVID-19, and 3% died (Brenner et al., 2021). Data from the registry evidenced that among patients with IBD, corticosteroids treatment may be a key risk factor for severe COVID-19 (Brenner et al., 2020), confirming results previously reported elsewhere (Mazza et al., 2020). A recent multicenter study enrolling 1816 patients with IBD treated with biologic therapy over the first 2 months of the pandemic reported an overall COVID-19 incidence of 3.9 per 1,000 patients with a 57% hospitalization rate and 29% case fatality rate (CFR; Ardizzone et al., 2021). In a cohort of 1912 patients with an IBD median duration of 17 years, the crude incidence rate of COVID-19 was at 6.2 cases per 1,000 patients, lower than that found in the general population (6.6 cases per 1,000 individuals); the mortality rate was 0.9 per 1,000 and 1 per 1,000 in patients with IBD and the general population, respectively (Taxonera et al., 2020). Although the CFR for IBD cases with COVID-19 was higher than in the general population (16.7 vs. 13.2%, respectively), the statistical difference was not significant. Finally, a meta-analysis including 9,177 patients with IBD from eight studies reported an incidence of 0.3% for COVID-19; 8.6% required admission to the intensive care unit, and the mortality rate was 6.3% (Aziz et al., 2020). In a cohort of patients with IBD, the rate of positive results for anti-SARS-CoV-2 antibodies (approximately 4.6% for IgG and IgM, and 6% for IgA) was found higher than that in healthcare professionals without inflammatory diseases (approximately 1.6% for IgG and IgM, and 1% for IgA); interestingly, no SARS-CoV-2-infected patients with IBD developed symptomatic COVID-19 (Łodyga et al., 2021). Further studies reported similar results, confirming that SARS-CoV-2 seroprevalence among individuals with IBD is closely comparable to that in subjects without IBD (Norsa et al., 2020; Bertè et al., 2021).

GUT MICROBIOTA IN IBD

Gut microbiota plays a key role in health and disease; it actively impacts multiple host systems and organs. A balanced gut microbial ecosystem with high biodiversity is associated with the beneficial effects of a myriad of symbiotic interactions between intra- and inter-microbial species, genera, families, phyla, and between microbes and host systems and organs, such as the immune system (Zheng et al., 2020), the brain, and the lung (Morais et al., 2021; Sencio et al., 2021). Conversely, perturbations in gut microbial communities, namely, dysbiosis, induce detrimental effects on these networks and are associated with diseases (Durack and Lynch, 2019). Gut dysbiosis can be defined as the loss of the overall microbial biodiversity with the imbalance between beneficial commensal and opportunistic pathogens, resulting in excessive production of pro-inflammatory mediators (Wei et al., 2021). A large body of literature investigated and evaluated extensively gut dysbiosis in individuals with IBD; the most frequently observed alterations are the overgrowth of pro-inflammatory bacterial species (e.g., *Escherichia coli*) associated with the shortfall of anti-inflammatory species (e.g., *Faecalibacterium prausnitzii*). The latter are involved in the generation of short-chain fatty acids (SCFAs), namely,

butyrate, propionate, and acetate (Zuo and Ng, 2018; Khan et al., 2019; Aldars-García et al., 2021; Alshehri et al., 2021). Regrettably, data on gut microbiota composition are partially heterogeneous between studies, and results could be categorized as (a) fully concordant between studies; (b) roughly concordant with some exceptions; (c) discordant between studies. **Table 1** recapitulates the most relevant data on gut dysbiosis in IBD, obtained from a great proportion of available studies from the literature (Favier et al., 1997; Seksik et al., 2003; Macfarlane et al., 2004; Martin et al., 2004; Ott et al., 2004; Gophna et al., 2006; Manichanh et al., 2006; Scanlan et al., 2006; Frank et al., 2007; Andoh et al., 2009; Kang et al., 2010; Rehman et al., 2010; Schwiertz et al., 2010; Willing et al., 2010; Joossens et al., 2011; Mondot et al., 2011; Rausch et al., 2011; Walker et al., 2011; Michail et al., 2012; Morgan et al., 2012; Nemoto et al., 2012; Vigsnæs et al., 2012; Fujimoto et al., 2013; Kabeerdoss et al., 2013, 2015; Kumari et al., 2013; Prideaux et al., 2013; Sha et al., 2013; Tong et al., 2013; Varela et al., 2013; Gevers et al., 2014; Hedin et al., 2014; Machiels et al., 2014; Walters et al., 2014; Wang et al., 2014; Hoarau et al., 2016; Jacobs et al., 2016; Mar et al., 2016; Takahashi et al., 2016; Chen et al., 2017; Halfvarson et al., 2017; Pascal et al., 2017; Santoru et al., 2017; Sokol et al., 2017; Vrakas et al., 2017; Zhang et al., 2017, 2021; de Meij et al., 2018; Laserna-Mendieta et al., 2018; Nishino et al., 2018; Franzosa et al., 2019; Heidarian et al., 2019; Lloyd-Price et al., 2019; Yilmaz et al., 2019; Alam et al., 2020; Ryan et al., 2020; Shahir et al., 2020; Clooney et al., 2021). No specific pattern of dysbiosis in patients with IBD has been definitively established; nevertheless, there is a broad agreement between studies on the imbalance of gut bacterial abundance in IBD. In particular, most studies report the depletion of *Clostridium* genus, *C. leptum* (cluster IV), *C. coccoides* (cluster XIVa) groups, *F. prausnitzii*, *E. rectale*, *R. bromii* species, *Ruminococcaceae*, *Lachnospiraceae* families, and the overgrowth of *Enterococcus* and *Fusobacterium* genera, *E. coli* and *F. nucleatum* species, *Enterobacteriaceae*, *Veillonellaceae* families. Controversial results may derive from many variables affecting gut microbiota composition, including the pre-existence of chronic diseases, the intensive therapeutic treatment in critically ill patients, especially with antibiotics to prevent secondary bacterial infections, and sudden and radical changes to eating habits (**Table 2**). Gut microbiome composition in IBD is strongly influenced by complex interactions between microbial communities and genetically altered host functional pathways (Huang et al., 2014; Knights et al., 2014). In CD, gut dysbiosis is more pronounced than in UC and is marked by a lower microbial diversity, a more altered microbiome composition, and a more unstable microbial community (Pascal et al., 2017). Microbial diversity and abundance significantly differ between feces and gut mucosa, as reported in early studies (Lepage et al., 2005; Gillevet et al., 2010; Morgan et al., 2012) and confirmed in more recently published papers (Lo Presti et al., 2019; Ryan et al., 2020). In IBD, inflammation alters the mucosal barrier inducing bacterial translocation; in patients with CD, bacterial translocation is revealed by the increase in several bacterial families within the submucosa compared to the corresponding superjacent mucosa at the advancing disease margin (Chiodini et al., 2016). Mucosal and fecal microbiome differences may explain, at least in part, some discrepancies between studies; for example, the increase in fecal *F. prausnitzii* (**Table 1**) corresponds to the decreased proportion of this bacterium at the mucosal surface (Walters et al., 2014). Differences in the abundance of various bacterial species and families (e.g., *Lactobacilli*, *C. leptum group*, *E. coli*, and *F. prausnitzii*) were observed between ulcerated (inflamed) and non-ulcerated (non-inflamed) mucosa (Zhang et al., 2007; Li et al., 2012) as well as between patients with high clinical activity indexes and/or sigmoidoscopy scores and patients with low clinical activity indexes and/or sigmoidoscopy scores (Fite et al., 2013). Other studies reported no difference in microbiota composition and enrichment between inflamed and non-inflamed mucosa (Kabeerdoss et al., 2015; Nishino et al., 2018). The abundance of some bacteria, such as the genus of *Faecalibacterium* and the family of *Enterobacteriaceae*, significantly differs between ileal CD and colonic CD (Dicksved et al., 2008; Naftali et al., 2016); for example, *F. prausnitzii* is markedly reduced in CD localized in the ileum compared with colonic localization (Willing et al., 2010). Interestingly, in current smokers with CD, the abundance of *Bacteroides–Prevotella* genera is higher than in non-smokers with CD (Benjamin et al., 2012a). Significant differences in several microbial taxa can be observed between young adults with IBD and IBD adults aged 60 years or older; in particular, *Bifidobacterium* genus decrease with age, and *Bacteroides* genus increase with age, probably reflecting body mass index and diet changes over time (Morgan et al., 2012). Researchers have a unanimous consensus on the *E. coli* overgrowth in IBD. *E. coli* overgrowth has been found in children with severe IBD (Schwiertz et al., 2010; Michail et al., 2012; Gevers et al., 2014; de Meij et al., 2018) and in adults with CD (Mondot et al., 2011). In patients with CD, the high prevalence of *E. coli* strictly adhering to the ileal mucosa has led to the identification of a new group of *E. coli* strains (Nadalian et al., 2021). This pathogenic group, called adherent-invasive *E. coli* (AIEC), has the ability to adhere and colonize enterocytes as well as to internalize into macrophages and replicate within their cytoplasm, inducing the release of tumor necrosis factor-α (TNF-α) and the cytotoxic response of Th17 and $CD8^+$ (Lee et al., 2019). Thus, AIEC is involved in the pathogenesis of IBD, specifically CD (Palmela et al., 2018; Chervy et al., 2020), by promoting inflammatory diseases that originated from the adaptative evolution of the genome (Nash et al., 2010; Ellermann et al., 2015). The recognition of AIEC is unusual in patients with UC; rather, UC is associated with the intestinal enrichment of a heterogeneous, diarrheagenic group of *E. coli* strains, termed diffusely adherent *E. coli* (DAEC); this group was found expressed not only in children and young adults with UC but even in those with CD (Walczuk et al., 2019).

GUT MICROBIOTA IN SARS-CoV-2 INFECTION

Although the lung is considered the main entry route for SARS-CoV-2, the gastrointestinal tract is equally a key target

TABLE 1 | Gut dysbiosis in patients with IBD and COVID-19 compared with healthy subjects (s, stool sample; m, mucosal biopsy; e, endoscopic lavage).

Bacterial taxa	Inflammatory bowel disease (IBD)		Coronavirus Disease 2019 (COVID-19)	
	Enriched	**Underrepresented**	**Enriched**	**Underrepresented**
Firmicutes				
• *Clostridium*	Morgan et al., 2012 (s,m)	Gophna et al., 2006 (m), Michail et al., 2012 (s), Tong et al., 2013 (e), Walters et al., 2014 (s), Gevers et al., 2014 (m), Chen et al., 2017 (s), Nishino et al., 2018 (m)	Zuo et al., 2020 (s), Tao et al., 2020 (s) [*C. hathewayi*]	
• *Clostridium cluster IV (C. leptum)*		Seksik et al., 2003 (s), Scanlan et al., 2006 (s), Manichanh et al., 2006 (s), Andoh et al., 2009 (s), Schwiertz et al., 2010 (s), Mondot et al., 2011 (s), Morgan et al., 2012 (s,m), Kabeerdoss et al., 2013 (s), Sha et al., 2013 (s), Kumari et al., 2013 (s), Fujimoto et al., 2013 (s), Wang et al., 2014 (s,m), Hedin et al., 2014 (s), Kabeerdoss et al., 2015 (m), Vrakas et al., 2017 (m), Laserna-Mendieta et al., 2018 (s), Shahir et al., 2020 (m)		Tang et al., 2020 (s), Yeoh et al., 2021 (s) [during antibiotic therapy]
• *Clostridium cluster XIVa (C. coccoides)*		Seksik et al., 2003 (s), Scanlan et al., 2006 (s), Andoh et al., 2009 (s), Schwiertz et al., 2010 (s), Joossens et al., 2011 (s), Morgan et al., 2012 (s,m), Nemoto et al., 2012 (s), Sha et al., 2013 (s), Kumari et al., 2013 (s), Prideaux et al., 2013 (m), Hedin et al., 2014 (s), Machiels et al., 2014 (s), Kabeerdoss et al., 2015 (m), Vrakas et al., 2017 (m), Clooney et al., 2021 (s)		
• *F. prausnitzii*	Walters et al., 2014 (s)	Frank et al., 2007 (m), Schwiertz et al., 2010 (s), Willing et al., 2010 (s), Rehman et al., 2010 (m), Mondot et al., 2011 (s), Joossens et al., 2011 (s), Rausch et al., 2011 (m), Morgan et al., 2012 (s,m), Fujimoto et al., 2013 (s), Varela et al., 2013 (s), Tong et al., 2013 (e), Kabeerdoss et al., 2013 (s), Kumari et al., 2013 (s), Prideaux et al., 2013 (m), Machiels et al., 2014 (s), Wang et al., 2014 (s,m), Hedin et al., 2014 (s), Gevers et al., 2014 (m), Takahashi et al., 2016 (s), Mar et al., 2016 (s), Hoarau et al., 2016 (s), Jacobs et al., 2016 (s), Sokol et al., 2017 (s), Vrakas et al., 2017 (m), Pascal et al., 2017 (s), Halfvarson et al., 2017 (s), Santoru et al., 2017 (s), Laserna-Mendieta et al., 2018 (s), Franzosa et al., 2019 (s), Lloyd-Price et al., 2019 (s,m), Yilmaz et al., 2019 (m), Heidarian et al., 2019 (s), Ryan et al., 2020 (m), Clooney et al., 2021 (s), Zhang et al., 2021 (s,m)		Tang et al., 2020 (s), Zuo et al., 2020 (s) [during antibiotic therapy], Tao et al., 2020 (s), Yeoh et al., 2021 (s), Gaibani et al., 2021 (s)
• *E. rectale*		Macfarlane et al., 2004 (m), Mondot et al., 2011 (s), Kumari et al., 2013 (s), Gevers et al., 2014 (m), Kabeerdoss et al., 2015 (m), Franzosa et al., 2019 (s), Clooney et al., 2021 (s)		Tang et al., 2020 (s), Zuo et al., 2020 (s) [during antibiotic therapy], Yeoh et al., 2021 (s)
• *Enterococcus*	Macfarlane et al., 2004 (m), Kang et al., 2010 (s), Mondot et al., 2011 (s), Nemoto et al., 2012 (s), Tong et al., 2013 (e), Gevers et al., 2014 (m), Takahashi et al., 2016 (s), Mar et al., 2016 (s), Zhang et al., 2017 (s), Pascal et al., 2017 (s), Franzosa et al., 2019 (s)		Tang et al., 2020 (s), Wu et al., 2021 (s), Gaibani et al., 2021 (s)	
• *Ruminococcaceae*		Morgan et al., 2012 (s,m), Hedin et al., 2014 (s), Mar et al., 2016 (s), Halfvarson et al., 2017 (s), Sokol et al., 2017 (s), Zhang et al., 2021 (s,m), Nishino et al., 2018 (m),		Gu et al., 2020 (s), He et al., 2021 (s), Gaibani et al., 2021 (s)
• *R. bromii*		Frank et al., 2007 (m), Mondot et al., 2011 (s), Prideaux et al., 2013 (m), Hoarau et al., 2016 (s), Sokol et al., 2017 (s), Nishino et al., 2018 (m), Ryan et al., 2020 (m),		Yeoh et al., 2021 (s)

(Continued)

TABLE 1 | Continued

Bacterial taxa	Inflammatory bowel disease (IBD)		Coronavirus Disease 2019 (COVID-19)	
	Enriched	**Underrepresented**	**Enriched**	**Underrepresented**
• *R. gnavus*	Willing et al., 2010 (s), Joossens et al., 2011 (s), Machiels et al., 2014 (s), Hoarau et al., 2016 (s), Sokol et al., 2017 (s), Nishino et al., 2018 (m), Franzosa et al., 2019 (s), Lloyd-Price et al., 2019 (s,m), Yilmaz et al., 2019 (m), Ryan et al., 2020 (m), Clooney et al., 2021 (s),	Frank et al., 2007 (m), Gevers et al., 2014 (m)	Yeoh et al., 2021 (s)	
• *Lachnospiraceae*	Alam et al., 2020 (s)	Frank et al., 2007 (m), Rausch et al., 2011 (m), Kumari et al., 2013 (s), Prideaux et al., 2013 (m), Mar et al., 2016 (s), Chen et al., 2017 (s), Sokol et al., 2017 (s), Nishino et al., 2018 (m), Yilmaz et al., 2019 (m), Ryan et al., 2020 (m),		Zuo et al., 2020 (s) [during antibiotic therapy], Gu et al., 2020 (s), Zuo et al., 2021 (s), Gaibani et al., 2021 (s), He et al., 2021 (s), Wu et al., 2021 (s)
• *R. hominis*		Tong et al., 2013 (e), Machiels et al., 2014 (s), Franzosa et al., 2019 (s), Lloyd-Price et al., 2019 (s,m)		
• *D. forminigenerans*		Franzosa et al., 2019 (s), Zhang et al., 2021 (s,m)		Yeoh et al., 2021 (s) [during antibiotic therapy]
• *Lactobacillus*	Willing et al., 2010 (s), Kang et al., 2010 (s), Fujimoto et al., 2013 (s), Wang et al., 2014 (s,m), Kabeerdoss et al., 2015 (m), Zhang et al., 2021 (s,m)	Ott et al., 2004 (m), Frank et al., 2007 (m), Rausch et al., 2011 (m), Vigsnaes et al., 2012 (s), Sha et al., 2013 (s), Vrakas et al., 2017 (m), Zhang et al., 2017 (s)	Gu et al., 2020 (s), Tao et al., 2020 (s), Yeoh et al., 2021 (s), Wu et al., 2021 (s) Gaibani et al., 2021 (s)	Tang et al., 2020 (s)
• *Veillonellaceae*	Macfarlane et al., 2004 (m), Michail et al., 2012 (s), Gevers et al., 2014 (m), Santoru et al., 2017 (s), Lloyd-Price et al., 2019 (s,m), Alam et al., 2020 (s), Ryan et al., 2020 (m)		Gu et al., 2020 (s), Gaibani et al., 2021 (s)	
Proteobacteria				
• *Enterobacteriaceae*	Seksik et al., 2003 (s), Frank et al., 2007 (m), Andoh et al., 2009 (s), Michail et al., 2012 (s), Nishino et al., 2018 (m), Alam et al., 2020 (s), Ryan et al., 2020 (m), Shahir et al., 2020 (m)		Tang et al., 2020 (s)	
• *E. coli*	Martin et al., 2004 (m), Gophna et al., 2006 (m), Schwiertz et al., 2010 (s), Willing et al., 2010 (s), Rehman et al., 2010 (m), Mondot et al., 2011 (s), Michail et al., 2012 (s), Morgan et al., 2012 (s,m), Tong et al., 2013 (e), Sha et al., 2013 (s), Gevers et al., 2014 (m), Wang et al., 2014 (s,m), Kabeerdoss et al., 2015 (m), Hoarau et al., 2016 (s), Takahashi et al., 2016 (s), Zhang et al., 2017 (s), Pascal et al., 2017 (s), Santoru et al., 2017 (s), Vrakas et al., 2017 (m), Chen et al., 2017 (s), de Meij et al., 2018 (s), Franzosa et al., 2019 (s), Lloyd-Price et al., 2019 (s,m), Zhang et al., 2021 (s,m)			
• *Shigella*	Willing et al., 2010 (s), Kang et al., 2010 (s), Morgan et al., 2012 (s,m)			

(Continued)

TABLE 1 | Continued

Bacterial taxa	Inflammatory bowel disease (IBD)		Coronavirus Disease 2019 (COVID-19)	
	Enriched	**Underrepresented**	**Enriched**	**Underrepresented**
• *P. mirabilis*	Zhang et al., 2021 (s,m)			
• *Sutterella*	Frank et al., 2007 (m), Michail et al., 2012 (s), Pascal et al., 2017 (s)	Rausch et al., 2011 (m)	Yeoh et al., 2021 (s) [during antibiotic therapy]	
Fusobacteria				
• *Fusobacterium*	Michail et al., 2012 (s), Alam et al., 2020 (s), Zhang et al., 2021 (s,m)			
• *F. nucleatum*	Gevers et al., 2014 (m), Pascal et al., 2017 (s), Santoru et al., 2017 (s), Clooney et al., 2021 (s)			
Bacteroidetes				
• *Bacteroides*	Andoh et al., 2009 (s), Walker et al., 2011 (m), Wang et al., 2014 (s,m), Kabeerdoss et al., 2015 (m), Hoarau et al., 2016 (s), Vrakas et al., 2017 (m)	Seksik et al., 2003 (s), Ott et al., 2004 (m), Rehman et al., 2010 (m), Nemoto et al., 2012 (s), Sha et al., 2013 (s), Fujimoto et al., 2013 (s), Gevers et al., 2014 (m), Takahashi et al., 2016 (s), Mar et al., 2016 (s), Sokol et al., 2017 (s), Heidarian et al., 2019 (s)	Zuo et al., 2020 (s), Yeoh et al., 2021 (s) [*B. dorei* during antibiotic therapy], He et al., 2021 (s)	Tao et al., 2020 (s), Zuo et al., 2021 (s), Wu et al., 2020 (s) [during patient's admission in ICU]
• *B. fragilis*	Gophna et al., 2006 (m), Walters et al., 2014 (s), Ryan et al., 2020 (m), Shahir et al., 2020 (m)	Manichanh et al., 2006 (s), Scanlan et al., 2006 (s), Macfarlane et al., 2004 (m), Kang et al., 2010 (s), Sha et al., 2013 (s), Jacobs et al., 2016 (s), de Meij et al., 2018 (s)		
• *B. vulgatus*		Macfarlane et al., 2004 (m), Gevers et al., 2014 (m), Ryan et al., 2020 (m)	Yeoh et al., 2021 (s)	
• *B. ovatus*		Frank et al., 2007 (m), Macfarlane et al., 2004 (m), Santoru et al., 2017 (s), Shahir et al., 2020 (m)	Yeoh et al., 2021 (s)	
• *Prevotellaceae*	Manichanh et al., 2006 (s), Andoh et al., 2009 (s), Walker et al., 2011 (m), Kabeerdoss et al., 2015 (m), Alam et al., 2020 (s)	Seksik et al., 2003 (s), Ott et al., 2004 (m), Kang et al., 2010 (s), Rausch et al., 2011 (m), Sha et al., 2013 (s), Fujimoto et al., 2013 (s), Prideaux et al., 2013 (m), Hedin et al., 2014 (s), Takahashi et al., 2016 (s), Mar et al., 2016 (s), Hoarau et al., 2016 (s), Santoru et al., 2017 (s), Sokol et al., 2017 (s), Nishino et al., 2018 (m), Heidarian et al., 2019 (s)		Gaibani et al., 2021 (s)
• *Alistipes*	Walker et al., 2011 (m), Rausch et al., 2011 (m), Shahir et al., 2020 (m)	Frank et al., 2007 (m), Willing et al., 2010 (s), Mondot et al., 2011 (s), Gevers et al., 2014, Ryan et al., 2020 (m), (m), Chen et al., 2017 (s), Sokol et al., 2017 (s), Halfvarson et al., 2017 (s), Nishino et al., 2018 (m), de Meij et al., 2018 (s), Franzosa et al., 2019 (s), Lloyd-Price et al., 2019 (s,m), Shahir et al., 2020 (m)		Zuo et al., 2021 (s), Yeoh et al., 2021 (s) [*A. putredinis* during antibiotic therapy]
Verrucomicrobia				
• *A. muciniphila*		Vigsnaes et al., 2012 (s), Jacobs et al., 2016 (s), Santoru et al., 2017 (s), de Meij et al., 2018 (s)	Gaibani et al., 2021 (s)	
Actinobacteria				
• *Bifidobacterium*	Willing et al., 2010 (s), Wang et al., 2014 (s,m), Takahashi et al., 2016 (s),	Favier et al., 1997 (s), Seksik et al., 2003 (s), Schwiertz et al., 2010 (s), Kang et al., 2010 (s), Sha et al., 2013 (s), Sokol et al., 2017 (s), Zhang et al., 2017 (s), Vrakas et al., 2017 (m), Yilmaz et al., 2019 (m), Alam et al., 2020 (s)	Tao et al., 2020 (s), Wu et al., 2021 (s)	Tang et al., 2020 (s)
• *B. bifidum*		Macfarlane et al., 2004 (m), Mondot et al., 2011 (s), Gevers et al., 2014 (m)		
• *B. adolescentis*	Jacobs et al., 2016 (s), Mar et al., 2016 (s)	Macfarlane et al., 2004 (m), Joossens et al., 2011 (s), Hedin et al., 2014 (s), Gevers et al., 2014 (m), Machiels et al., 2014 (s)		Yeoh et al., 2021 (s)
• *Collinsella*	Willing et al., 2010 (s)	Joossens et al., 2011 (s), Santoru et al., 2017 (s), Pascal et al., 2017 (s), Nishino et al., 2018 (m),	Zuo et al., 2021 (s), Gaibani et al., 2021 (s)	Yeoh et al., 2021 (s), Wu et al., 2021 (s)

of the virus (Xiao et al., 2020; Qian et al., 2021). Indeed, the brush border of human enterocytes exhibits the highest expression of the SARS-CoV-2 receptor angiotensin-converting enzyme 2 (ACE2; Qi et al., 2020); even the transmembrane serine protease 2 (TMPRSS2), mediating the entry of SARS-CoV-2, is expressed on the luminal surface of enterocytes from ileum and colon as well as on the epithelial and gland cells of the esophagus (Knyazev et al., 2021). Thus, it is not surprising that SARS-CoV-2 infects human gut enterocytes (Lamers et al., 2020), promoting gut mucosal inflammatory infiltration with activated immune cells and cytokines (Lehmann et al., 2021). These findings support (a) the frequently observed gastrointestinal symptoms in patients with COVID-19, including severe abdominal pain, diarrhea, nausea, and vomiting (Devaux et al., 2021); (b) the persistence of viral RNA in patient's stool even after the virus clearance from oropharyngeal swab (Morone et al., 2020); (c) the likelihood of SARS-CoV-2 transmission by the fecal–oral route (Cheung et al., 2020; Guo et al., 2021); (d) gut dysbiosis in asymptomatic infected individuals and patients with COVID-19 (Yamamoto et al., 2021). SARS-CoV-2 gut colonization significantly alters the gut microbial ecosystem, leading to dysbiosis; on the other hand, gut dysbiosis, due to aging, unhealthy lifestyle habits, and pre-existing chronic diseases (e.g., hypertension, type-2 diabetes, autoimmune diseases, and metabolic syndrome), is a key risk factor for developing COVID-19 (Magalhães et al., 2021). In hospitalized patients with COVID-19, the reduction in gut microbiota diversity, the depletion of beneficial bacterial symbionts, and the enrichment of opportunistic pathogens closely correlate with the host immune response and, in turn, with the disease severity and the clinical outcome (Villapol, 2020). The variety of the therapeutic treatment for COVID-19 may impact changes in gut microbiome composition, as highlighted in a recent review (Aktas and Aslim, 2021). Gut dysbiosis continues after the clearance of the viral RNA from the upper respiratory tract and the resolution of clinical symptoms. These associations result from complex interactions between the gut and the lung microbiota, the so-called gut–lung axis (de Oliveira et al., 2021). In particular, a balanced gut microbial ecosystem enhances the pulmonary defense against viral infections, for example, by stimulating the lung's synthesis of type I interferons (Cyprian et al., 2021). Conversely, gut dysbiosis negatively influences the progression of the viral infection, COVID-19 development, and patient outcome (Hussain et al., 2021).

Since SCFAs inhibit the overgrowth of opportunistic pathogens, activate the adaptive immune response by enhancing antiviral immunity, and contribute to maintaining the integrity of the intestinal mucosal barrier, the depletion of SCFAs producer bacteria is closely associated with COVID-19 severity and adverse outcome; therefore, the number of SCFAs producer bacteria could predict the severity of the disease (Tang et al., 2020). Based on the observation that the *Enterococcus*/*Enterobacteriaceae* ratio is altered in approximately 74% of patients with severe/critical COVID-19, being significantly increased in non-survivors compared with survivors, it was proposed that this index may be useful to predict death in critically ill patients (Tang et al., 2020). However, a strong limitation may be the heterogeneity of this ratio. *Enterococcus* is a genus belonging to the *Enterococcaceae* family (*Firmicutes* phylum), while *Enterobacteriaceae* family belongs to *Proteobacteria* phylum. Despite the limited number of studies on the gut microbiome in patients with COVID-19, the pattern of gut dysbiosis associated with the disease has been partially defined (**Table 1**). When compared with non-infected individuals, gut dysbiosis in patients with COVID-19 is marked by the depletion of *C. leptum* (cluster IV) group, *F. prausnitzii*, and *E. rectale* species, *Ruminococcaceae*, and *Lachnospiraceae* families, in conjunction with the overgrowth of *Enterococcus* genus, *Veillonellaceae*, and *Enterobacteriaceae* families (Tang et al., 2020; Tao et al., 2020; Zuo et al., 2020, 2021; Gaibani et al., 2021; He et al., 2021; Wu et al., 2021; Yeoh et al., 2021). The abundance of *Coprobacillus*, *C. ramosum*, *C. hathewayi* (Zuo et al., 2020), and *Enterococcus* (Gaibani et al., 2021) was found positively correlated with COVID-19 severity; conversely, an inverse correlation was observed between the disease severity and the abundance of *C. leptum* (cluster IV) group, *Lactobacillus*, *Bifidobacterium*, *C. butyricum* (Tang et al., 2020), *Bilophila*, *Citrobacter* (Tao et al., 2020), *Bacteroides* (Gaibani et al., 2021), *F. prausnitzii*, *E. rectale* (Tang et al., 2020; Yeoh et al., 2021), *B. bifidum*, and *B. adolescentis* (Yeoh et al., 2021). In addition, the abundance of some microbial taxa, including *Erysipelotrichaceae bacterium 2_2_44A* (Zuo et al., 2020), *P. copri*, *E. dolichum* (Wu et al., 2021), *C. aerofaciens*, *C. tanakaei*, *S. infantis* (Zuo et al., 2021) positively correlates with the fecal SARS-CoV-2 load. Conversely, the abundance of *B. dorei*, *B. thetaiotaomicron*, *B. massiliens*, *B. ovatus* (Zuo et al., 2020), *S. anginosus*, *Dialister*, *Alistipes*, *Ruminococcus*, *C. citroniae*, *Bifidobacterium*, *Haemophilus*, *H. parainfluenzae* (Wu et al., 2021), *P. merdae*, *B. stercoris*, *A. onderdonkii*, and *Lachnospiraceae bacterium 1_1_57FAA* (Zuo et al., 2021) was inversely correlated with the fecal viral load. The increased gut colonization of bacteria usually resident in the oral and respiratory tracts, such as *Actinomyces* (Gu et al., 2020; Zuo et al., 2020; Gaibani et al., 2021; Wu et al., 2021) and *Granulicatella* (Wu et al., 2021), is associated with COVID-19 and its severity, confirming the active interchange between the gut, oral, and respiratory tract microbiota. The close relationship between gut microbiota and the immune-mediate response to SARS-CoV-2 infection and COVID-19 progression and outcome is supported by the correlation between the abundance of some gut microbial taxa and biomarkers of inflammation. For example, in critical patients with COVID-19, a negative correlation was observed between serum C-reactive protein (CRP) levels and the gut abundance of *C. butyricum* and *F. prausnitzii* (Tang et al., 2020). Fecal inflammatory cytokine IL-18 concentration was positively correlated with the abundance of *Peptostreptococcus*, *Fusobacterium*, and *Citrobacter* in patients with COVID-19 (Tao et al., 2020). Similarly, *B. dorei* and *A. muciniphila* abundance was found positively correlated with the serum level of IL-1β, IL-6, and C-X-C motif ligand 8 (CXCL8), whereas *F. prausnitzii*, *E. rectale*, and *B. adolescentis* were negatively correlated with serum level of TNF-α, IL-10, C-C motif ligand 2 (CCL2), and CXCL10 (Yeoh et al., 2021).

TABLE 2 | Main factors affecting the variability of the results between published studies on gut microbiota composition in IBD.

1. Host genetic polymorphisms and gene expression
2. Mucosal immune system interactions (e.g., with Treg/Th17, PRRs, TLRs, and NLRs)
3. Disease phenotype based on clinical activity indices and sigmoidoscopy scores
4. Type of disease (e.g., Crohn's Disease and Ulcerative Colitis)
5. Type of biological sample (e.g., stool, endoscopic biopsies, or resection specimens)
6. Site of the biopsy sampling (e.g., terminal ileum and large bowel)
7. Host demographics (e.g., gender and aging)
8. Environmental stimuli and patient's life style (e.g., smoking)
9. Diet and medications (e.g., antibiotics)
10. Inter-individual variability between patients
11. Methods for the microbiome analysis (e.g., fluorescence *in situ*, terminal restriction fragment length polymorphism, 16S rDNA sequencing, and whole-genome sequence)

METABOLOMICS IN IBD AND COVID-19

Metabolomics is an evolving "omic" discipline allowing the identification and the quantification of endogenous and exogenous products of the cellular metabolism, namely, metabolites, within a biological system in a high-throughput manner (Liu and Locasale, 2017). The set of metabolites recognizable in a biological matrix is called metabolome or metabolic profile; it is a highly personalized readout of the current metabolism and metabolic activity that occurred in the past (Zamboni et al., 2015). Qualitative and quantitative data on metabolites reveal basic information on changes and perturbations of metabolic pathways deriving from interactions between genome, environment, microbiome, nutrients, and the intake of drugs and toxicants in health and disease states (Ashrafian et al., 2020). An updated PubMed literature search, querying the keyword metabolomics, results in approximately 26,200 studies, including *in vitro* experimental studies, studies on animal models, and clinical studies on patients with various diseases (Kang et al., 2021). Approximately 200 studies used the metabolomic approach in patients with IBD (Gallagher et al., 2021) and 25 in patients with COVID-19 (Mussap and Fanos, 2021). The most relevant findings are reported below.

Tryptophan Metabolism

Tryptophan (TRP) is an essential amino acid mainly derived from the diet and involved in serotonin, melatonin, and niacin biosynthesis, as detailed in **Figure 1**. More than 95% of TRP is metabolized along the kynurenine pathway; in the liver, the enzyme tryptophan dioxygenase (TDO) converts tryptophan into kynurenine. In the brain and the immunocompetent cells, the conversion is catalyzed by indoleamine 2,3-dioxygenase (IDO-1 and 2; Gao et al., 2018; Agus et al., 2018). Kynurenine can be converted either into neurotoxic metabolites, namely, 3-hydroxykynurenine, 3-hydroxyanthranilic acid, quinolinic acid, or neuroprotective metabolites, such as kynurenic acid, anthranilic acid, xanthurenic acid, and picolinic acid (Savitz, 2020). Depending on gut microbiota composition, approximately 4–6% of tryptophan is converted into various intermediates. For example, the prevalence of *Clostridium sporogenes* and *Ruminococcus gnavus* spp. originates tryptamine (Williams et al., 2014), whereas the prevalence of *Lactobacillus*, *Bacteroides,* and *Clostridium* genera originates indole derivatives (Roager and Licht, 2018). Among bacterial metabolites, indoles play a crucial role in the regulation of gastrointestinal barrier function and integrity by their binding with the pregnane X receptor (PXR), also known as steroid and xenobiotic receptor (SXR; Venkatesh et al., 2014; Oladimeji and Chen, 2018). Finally, 1–2% tryptophan is converted into serotonin (5-hydroxytryptamine), a neurotransmitter and key regulator of intestinal secretion and motility. More than 90% of serotonin is synthesized by the rate-limiting enzyme tryptophan hydroxylase (Tph/TPH) 1 within enterochromaffin cells of the gut (Stasi et al., 2019); notably, gut microbiota may considerably affect serotonergic regulation *via* microbiota-derived SCFA (Reigstad et al., 2015).

Tryptophan and indole-3-acetic are decreased in the blood of patients with IBD (**Table 3**; Ooi et al., 2011; Kohashi et al., 2014; Nikolaus et al., 2017; Bosch et al., 2018; Lai et al., 2019). Conversely, kynurenine and quinolinic acid are increased (Forrest et al., 2002; Yau et al., 2014; Nikolaus et al., 2017; Whiley et al., 2019). In the urine and stool of patients with IBD, tryptophan is increased (Schicho et al., 2012; Bosch et al., 2018). Kynurenic acid blood levels were increased in IBD (Forrest et al., 2002); similarly, picolinic acid was increased (Yau et al., 2014). More recently, kynurenic acid and picolinic acid were found decreased in patients with IBD (Nikolaus et al., 2017; Whiley et al., 2019). These alterations seem to be closely related to gut dysbiosis (de Meij et al., 2018), promoting the massive activation of pro-inflammatory cytokines (e.g., INF-γ and TNF-α), and the upregulation of the IDO expression (Wu et al., 2018). Gut dysbiosis significantly affects the conversion of tryptophan into indole derivatives, such as indole-3 acetic acid and indole-3-acetaldehyde. Low blood levels of indole-3-acetic acid have been associated with the overgrowth of *Clostridium* and *Lactobacillus* genera; both bacterial genera decarboxylate indole-3-acetic acid in 3-methylindole, also known as skatole. As a result, indole-3-acetic acid is metabolized with the concomitant accumulation of skatole. In patients with IBD, skatole blood levels were significantly increased (Lai et al., 2019). Most indole derivatives are ligands of the Aryl Hydrocarbon Receptor (AHR), a cytosolic ligand-dependent transcription factor widely expressed by cells of the immune system and involved in antimicrobial activity and gut immune homeostasis (Stockinger et al., 2014; Lamas et al., 2018). The indole derivatives-induced AHR activation promotes the local synthesis of the anti-inflammatory cytokine IL-22 by the innate lymphoid cells (Monteleone et al., 2011; Qiu et al., 2012; Yang et al., 2020); IL-22 protects the mucosa integrity against fungal infection by Candida albicans, commonly observed in patients with IBD (Sokol et al., 2017; Gronke et al., 2019). In healthy subjects with gut eubiosis, the activation of AHR modulates local IL-22 production; conversely, in IBD patients, the decrease of indole-3 acetic acid due to the imbalance of gut flora reduces

the AHR activity and hence IL-22 synthesis, as observed in an animal model (Lamas et al., 2016). In patients with COVID-19, most metabolomics-based studies reported a significant decrease in blood tryptophan level in conjunction with the significant increase of kynurenine, kynurenic acid, and downstream metabolites of the kynurenine pathway (Barberis et al., 2020; Fraser et al., 2020; Thomas et al., 2020; Lawler et al., 2021). The decrease of tryptophan was inversely correlated with biomarkers of inflammation, such as IL-6 and CRP (Thomas et al., 2020). The increase of kynurenine is significantly associated with the increase of several cytokines, including interferon γ-induced protein 10 (IP-10), the mitogen-inducible cytokine macrophage inflammatory protein-1 β (MIP-1-β), also known as CCL4, TNF-α, interleukin-1 receptor antagonist (IL-1RA), IL-7, IL-18, and IL-8 (Lawler et al., 2021). Controversial results on indole-3-acetic acid in patients with COVID-19 have been published. On the one hand, it was found decreased (Lawler et al., 2021), and this finding remains unclear. A possible explanation may be the decrease of indole-3 acetic acid produced by the host because of the upregulation of the kynurenine pathway. This assumption is plausible, taking into account that a fraction of indole-3 acetic acid is produced by mammalian cells (Zhang et al., 2020). On the other hand, indole-3 acetic acid has been found increased (Blasco et al., 2020); this finding is coherent with gut dysbiosis marked by the prevalence of bacteria converting tryptophan into indoles.

Glutamine

Glutamine is an L-α gluconeogenic and proteogenic amino acid containing five carbons and two amino groups. L-glutamine is considered a conditionally essential amino acid; it is obtained mainly through the diet, as well as it is synthesized *de novo* from glutamate and ammonia in almost all the human cells by the activity of glutamine synthase (E.C.: 6.3.1.2.). In rapidly diving cells, such as enterocytes of the small intestine (Windmueller and Spaeth, 1974), lymphocytes, neutrophils, macrophages, and tumor cells, as well as under catabolic stressed conditions, including severe infections and sepsis, the endogenous synthesis does not meet the energy cell demand and the amount deriving from digested food absorbed through the small intestine becomes vitally important (Cruzat et al., 2018). Glutamine is involved in many cytoplasmatic and mitochondrial pathways, such as (a) the preservation of the reactive oxygen species (ROS) homeostasis, by contributing to the synthesis of the glutathione (Matés et al., 2002); (b) the biosynthesis of hexosamine, nucleotides, asparagine; and (c) the activation of glutaminolysis (Yoo et al., 2020). In the gut, the role of L-glutamine is crucial (Kim and Kim, 2017). Firstly, glutamine exerts an anti-inflammatory activity, preventing the expression of pro-inflammatory cytokines through the inhibition of both nuclear factor *k* light chain-enhancer of activated B cells (NF-*k*B) and signal transducer and activator of transcription (STAT) proteins (Kretzmann et al., 2008). Secondly, L-glutamine enhances tight junction integrity, as demonstrated in animal models (Wang et al., 2016), triggering the mitogen-activated protein kinase (MAPK) function (Basuroy et al., 2006; Perna et al., 2019); in addition, L-glutamine is pivotal for gut cells proliferation (Rhoads et al., 1997). Finally, L-glutamine modulates NO synthetase expression (Hecker et al., 1990; Swierkosz et al., 1990). Thus, it is not surprising that L-glutamine is the most abundant free amino acid in humans. L-glutamine provides energy as a substitute fuel to the tricarboxylic acid (TCA) cycle to produce adenosine triphosphate (ATP; Curi et al., 2005); recently, it was postulated that L-glutamine is the fuel for the immune system, generating the concept of immunometabolism (Wang et al., 2019).

In IBD, L-glutamine plasma levels are decreased (**Table 4**), especially during the acute exacerbation of CD (Sido et al., 2006; Bjerrum et al., 2010; Ooi et al., 2011; Scoville et al., 2018). Supplementation improves inflammation (Sugihara et al., 2019) and the mucosal barrier integrity in patients with remissive CD (Benjamin et al., 2012b). Glutamine is reduced in the blood of patients with non-severe COVID19 and much more in severe forms (Thomas et al., 2020; Doğan et al., 2021; Meoni et al., 2021; **Table 4**). In patients with COVID-19, it was observed a 19% reduction in glutamine blood level compared to that before the onset of the disease (Bruzzone et al., 2020). This finding may be related to higher consumption of gluconeogenic amino acids, especially in patients with severe forms, because of the significant scarcity of amino acids (Fanos et al., 2021). A further possible explanation may be the increased conversion of L-glutamine into glutamate, supported by the blood Krebs cycle's intermediates elevation and by the increase in oxidative stress (Páez-Franco et al., 2021).

Histidine

Histidine is an essential amino acid playing a crucial role as a ROS scavenger and anti-inflammatory mediator (Son et al., 2005; Holeček, 2020). The decarboxylation of histidine, catalyzed by histidine decarboxylase, originates histamine, a primary mediator in allergic diseases and a neurotransmitter involved in the control of food intake and sleep biorhythm. Histidine decarboxylase is expressed in bacteria of the large intestine and human muscle, liver, lung, and gastric mucosa (Moro et al., 2020). Blood histidine has been found significantly decreased in patients with IBD (Bjerrum et al., 2010; Ooi et al., 2011; Dawiskiba et al., 2014; Kohashi et al., 2014; Bosch et al., 2018; Probert et al., 2018). An early study found that histidine was significantly decreased in a large cohort ($n = 387$) of IBD patients (Hisamatsu et al., 2012); interestingly, plasma histidine was significantly lower in patients with active disease than in those in remission. Furthermore, a significant inverse correlation was observed between plasma histidine and serum CRP in patients with UC and CD. Authors postulated that the decrease in plasma histidine may reflect chronic inflammation in patients with IBD, suggesting supplementation with histidine as a novel therapy. In a subsequent 1-year follow-up of patients with UC in clinical remission, the same research group found that the decrease in histidine plasma level over time was associated with the increased risk of clinical relapse (Hisamatsu et al., 2015). Previous *in vitro* studies and in animal models suggested that low levels of histidine do not allow the effective suppression of NF-*k*B (Andou et al., 2009; Hasegawa et al., 2011); thus, LPS-induced TNF-α expression cannot be inhibited,

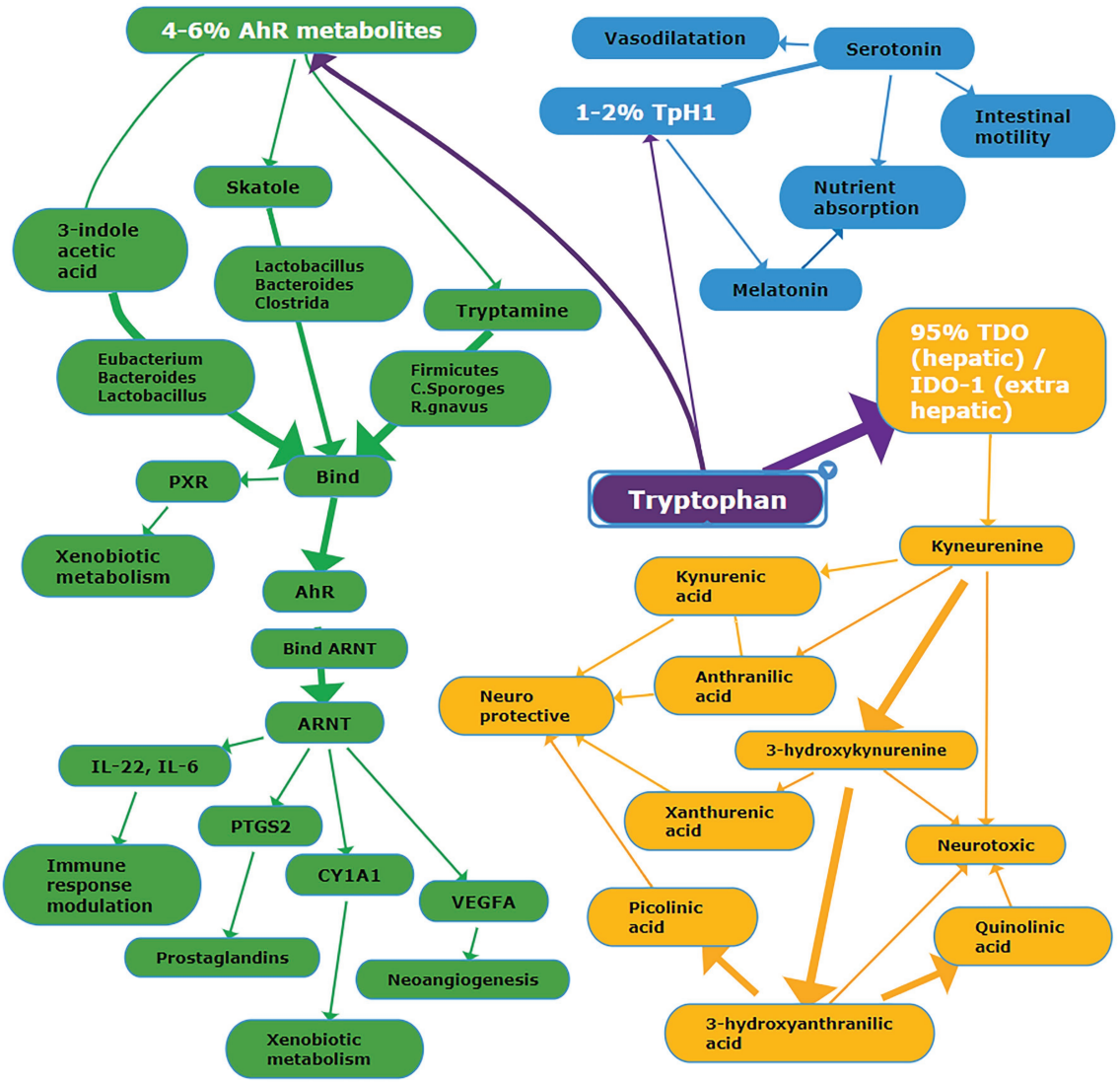


FIGURE 1 | Schema summarizing the main metabolic pathways deriving from tryptophan. Details are reported in the text. Abbreviations: PXR, Pregnane X Receptor; AhR, Aryl Hydrocarbon Receptor; ARNT, AhR Nuclear Translocator protein; PTGS2, Prostaglandin G/H Synthase 2; CYP1A1, Cytochrome P450 1A1; VEGFA, vascular endothelial growth factor A; TpH1, tryptophan hydroxylase-1; TDO, tryptophan dioxygenase; IDO-1, indoleamine 2,3-dioxygenase-1. White font = name of the metabolic pathway; purple background = tryptophan pathway; yellow background = TDO/IDO-1 pathway; green background = AhR metabolic pathway; blue background = TpH1 pathway.

resulting in the exacerbation of inflammation. This was confirmed by the amelioration of intestinal inflammation after oral administration of histidine. A recent metabolomics-based study on serum and stool samples in pediatric patients with IBD found controversial results. Compared with healthy controls, fecal histidine was significantly decreased in children with CD and significantly increased in those with UC (Kolho et al., 2017). Moreover, in the group of children with UC, serum histidine inversely correlated with erythrocyte sedimentation rate (ESR). Histidine fecal levels have been positively associated with more extended disease in UC but not in CD pediatric patients (Jagt et al., 2021). A possible explanation may be either the colonic leakage of histidine and other amino acids or malabsorption. However, the overgrowth of *Bacteroides vulgatus* in patients with UC leads to an increased fecal proteolytic and elastase activity (Galipeau et al., 2021). Therefore, it is reasonable to assume that the increased proteolytic and elastase activity might be the main factor promoting the high concentrations of fecal histidine rather than the protein-losing enteropathy. Several studies found L-histidine significantly decreased in the serum/plasma of patients with COVID-19 (**Table 4**; Barberis et al., 2020; Bruzzone et al., 2020; Thomas et al., 2020); the magnitude of L-histidine decline correlated with the disease severity (Lawler et al., 2021; Meoni et al., 2021). COVID-19 is marked by the activation of gluconeogenesis that is positively correlated with the severity of the disease.

TABLE 3 | Changes in the tryptophan metabolism in IBD and in COVID-19.

Metabolite	Inflammatory bowel disease (IBD)		Coronavirus disease 2019 (COVID-19)	
	Trend	Ref.	Trend	Ref.
Tryptophan (blood)	Decreased	Ooi et al., 2011; Kohashi et al., 2014; Nikolaus et al., 2017; Bosch et al., 2018	Decreased	Barberis et al., 2020; Thomas et al., 2020; Lawler et al., 2021
Tryptophan (urine)	Increased	Schicho et al., 2012		
Tryptophan (stool)	Increased	Bosch et al., 2018		
Kynurenine (blood)	Increased	Forrest et al., 2002; Nikolaus et al., 2017; Whiley et al., 2019	Increased	Fraser et al., 2020; Thomas et al., 2020; Lawler et al., 2021
Quinolinic acid (blood)	Increased	Yau et al., 2014; Nikolaus et al., 2017;	Increased	Lawler et al., 2021
Kynurenic acid (blood)	Increased	Forrest et al., 2002	Increased	Thomas et al., 2020 (only in patients with high levels of interleukine-6)
	Decreased	Nikolaus et al., 2017; Whiley et al., 2019 (statistically not significant)		
Indole-3-acetic acid (blood)	Decreased	Lai et al., 2019 (estimated by the increase in skatole)	Decreased	Lawler et al., 2021
			Increased	Blasco et al., 2020
Picolinic acid (blood)	Increased	Yau et al., 2014	Increased	Thomas et al., 2020
	Decreased	Nikolaus et al., 2017; Whiley et al., 2019		

TABLE 4 | Changes in the concentration of various amino acids in patients with IBD and COVID-19.

Metabolite	Inflammatory bowel disease (IBD)		Coronavirus disease 2019 (COVID-19)	
	Trend	Ref.	Trend	Ref.
L-Glutamine (blood)	Decreased	Sido et al., 2006; Bjerrum et al., 2010; Ooi et al., 2011; Scoville et al., 2018	Decreased	Bruzzone et al., 2020; Thomas et al., 2020; Meoni et al., 2021; Doğan et al., 2021
L-Glutamine (gut mucosa)	Decreased	Ooi et al., 2011		
Histidine (blood)	Decreased	Bjerrum et al., 2010; Ooi et al., 2011; Hisamatsu et al., 2012; Kohashi et al., 2014; Dawiskiba et al., 2014; Bosch et al., 2018; Probert et al., 2018	Decreased	Thomas et al., 2020; Barberis et al., 2020; Bruzzone et al., 2020; Kimhofer et al., 2020; Lawler et al., 2021; Meoni et al., 2021
Histidine (stools)	Decreased	Kolho et al., 2017		
Histidine (stools)	Increased	Kolho et al., 2017		
Phenylalanine (serum)	Increased	Zhang et al., 2013 ; Dawiskiba et al., 2014	Decreased	Kimhofer et al., 2020; Shi et al., 2021
			Increased	Barberis et al., 2020; Bruzzone et al., 2020
Phenylalanine (urine)	Increased	Alonso et al., 2016		
Phenylalanine (stools)	Increased	Jansson et al., 2009; Kolho et al., 2017; Bosch et al., 2018		
Succinate (blood, urine)	Increased	Ooi et al., 2011; Macias-Ceja et al., 2019	Increased	Barberis et al., 2020; Bruzzone et al., 2020
	Decreased	Schicho et al., 2012; Stephens et al., 2013; Dawiskiba et al., 2014	Decreased	Song et al., 2020
Succinate (gut mucosa)	Increased	Macias-Ceja et al., 2019		
	Decreased	Ooi et al., 2011		
Citrate (blood, urine)	Decreased	Schicho et al., 2012; Stephens et al., 2013; Dawiskiba et al., 2014	Decreased	Páez-Franco et al., 2021
			Increased	Bruzzone et al., 2020
Citrate (gut mucosa)	Decreased	Ooi et al., 2011		

In patients with COVID-19, it was observed a 16% reduction in L-histidine blood levels compared to those before the onset of the disease, similarly to glutamine (Bruzzone et al., 2020). In COVID-19, low L-histidine serum levels may be related to the skeletal muscle breakdown (Kimhofer et al., 2020).

Phenylalanine

Phenylalanine is an aromatic, hydrophobic essential amino acid involved in the biosynthesis of catecholamines (Matthews, 2007). Phenylalanine is hydroxylated to tyrosine by phenylalanine 4-hydroxylase; this reaction primarily occurs in the liver and the kidney. In turn, tyrosine is hydroxylated to L-DOPA (3,4-Dihydroxyphenylalanine) by tyrosine-5 hydrolase, and the enzyme DOPA-decarboxylase converts L-DOPA into Dopamine. A restricted number of gut bacteria, mainly belonging to the phylum *Firmicutes,* such as *Clostridium sporogenes* and *C. botulinum* spp., metabolize aromatic amino acids tryptophan, phenylalanine, and tyrosine to their corresponding propionic acid derivatives, namely, phenylpropionic acid and 4-hydroxyphenylpropionic acid (Elsden et al., 1976). In *C. sporogens*, this metabolic pathway can produce nine metabolites; they accumulate in host serum and exhibit specificity by engaging receptors and altering host biology, especially systemic immunity and gut permeability (Dodd et al., 2017). Although an early metabolomics-based study focused on amino acids did not report any data on phenylalanine level in stool samples from adults with IBD (Marchesi et al., 2007), further studies found a significant increase of phenylalanine in serum (Zhang et al., 2013; Dawiskiba et al., 2014), urine (Alonso et al., 2016), and stools (Jansson et al., 2009; Kolho et al., 2017; Bosch et al., 2018). Interestingly, the magnitude of fecal phenylalanine increase differed between CD and UC, and no correlation was found between the disease activity and fecal phenylalanine concentration (Bosch et al., 2018). Given that the activity of phenylalanine-4-hydroxylase is impaired during immune activation and inflammation (Scholl-Bürgi et al., 2013), it is reasonable to assume that the increase of phenylalanine in patients with IBD may originate from the accumulation of this amino acid. A further contribution to the fecal phenylalanine increase in IBD may be related to the microbial biosynthesis of aromatic amino acids *via* the shikimate pathway (Sprenger, 2006). The condensation of phosphoenolpyruvate with erythrose 4-phosphate, deriving from the glycolysis pathway and the non-oxidative branch of the pentose phosphate pathway, respectively, yields 3-deoxy-d-arabino-heptulosonate-7-phosphate (DAHP), which is converted into chorismic acid and then into aromatic amino acids L-phenylalanine, L-tyrosine, and L-tryptophan. Approximately one-third of gut bacteria expresses all the transcripts coded from the genes of the shikimate pathway, including *A. muciniphila* (Mesnage and Antoniou, 2020); the remaining bacteria species do not exhibit a complete shikimate pathway, having lost either one enzyme (nearly 22%) or five or more enzymes (nearly 74%; Zucko et al., 2010).

In COVID-19, the considerable generation of ROS, due to cytokine activation, induces the irreversible non-enzymatic oxidation of 5, 6, 7, 8-tetrahydrobiopterin (BH4), a cofactor of phenylalanine 4-hydroxylase (PAH), the enzyme catalyzing the conversion of phenylalanine to tyrosine. BH4 shortage induces the loss of PAH activity, with the reduced biosynthesis of tyrosine and the accumulation of phenylalanine. Data on phenylalanine emerging from metabolomic studies in patients with COVID-19 are heterogeneous and discordant: both phenylalanine and tyrosine were decreased, especially in severe and fatal outcomes (Barberis et al., 2020) or increased (Kimhofer et al., 2020; Shi et al., 2021). Interestingly, in non-survivors, phenylalanine serum levels are significantly higher than in survivors (Shi et al., 2021). Moreover, the increase in phenylalanine was associated with the decrease in tyrosine (Bruzzone et al., 2020). The role of phenylalanine in SARS-CoV-2 infection seems to be crucial. In a multicenter study on elderly patients with COVID-19, phenylalanine and tyrosine metabolisms were highly upregulated in 132 deceased patients (median age 74 years) compared with 91 survivors (median age 70 years), suggesting that these amino acids contribute as building blocks for the production of internal SARS-CoV-2 protein and its subsequent assembly into viral particles (Mei et al., 2021).

Succinate and Citrate

Succinate and citrate are intermediates of the TCA cycle. Succinate is generated within mitochondria *via* the TCA cycle from succinyl-CoA; then, succinate is oxidized to fumarate by succinate dehydrogenase. Succinate is also a product of the glyoxylate cycle, a pathway active in many bacteria, plants, and fungi. Overall, succinate accumulation is a metabolic signature of anoxia, asphyxia, and ischemia; the dysregulation of succinate metabolism can lead to pathological conditions, such as malignant transformation, inflammation, and tissue injury. Metabolomics-based studies carried out in serum, plasma, and urine of patients with IBD found that succinate and citrate significantly decreased in CD and UC compared to non-IBD individuals (Schicho et al., 2012; Stephens et al., 2013; Dawiskiba et al., 2014). Succinate and citrate are strongly involved in energy metabolism, and their depletion confirms the increased demand and the rapid utilization of cellular energy in IBD. In this context, citrate depletion may be associated with the increase in fatty acids biosynthesis. Indeed, citrate is essential for carrying acetyl-CoA from mitochondria to the cytosol. The significant increase in circulating triglycerides, observed in several studies on IBD, can be considered further evidence (Levy et al., 2000; Sappati Biyyani et al., 2010; Tefas et al., 2020). Beyond the role of succinate as an energy supplier, this amino acid is an inflammation mediator; it selectively binds to and activates the succinate receptor-1 (SUCNR1), which promotes pro-inflammatory signaling pathways (Mills and O'Neill, 2014). In patients with CD, hypoxia, inflammation, and necrosis promote the accumulation of succinate in gut inflamed areas, with a further activation and infiltration of macrophages and fibroblasts, the overexpression of pro-inflammatory cytokines, and ultimately the acceleration of fibrosis (Macias-Ceja et al., 2019).

Data on succinate emerging from metabolomics-based studies in COVID-19 are controversial (**Table 4**); the concentration of this amino acid has been found increased (Barberis et al.,

TABLE 5 | Changes in lipid concentration in IBD and in COVID-19.

Metabolite	Inflammatory bowel disease (IBD)		Coronavirus disease 2019 (COVID-19)	
	Trend	Ref.	Trend	Ref.
3-β-hydroxybutyrate	Increased	Zhang et al., 2013; Dawiskiba et al., 2014; Kohashi et al., 2014; Keshteli et al., 2017	Increased	Barberis et al., 2020; Bruzzone et al., 2020; San Juan et al., 2020; Páez-Franco et al., 2021; Meoni et al., 2021
Acetone	Increased	Keshteli et al., 2017	Increased	Bruzzone et al., 2020; San Juan et al., 2020
Acetoacetate	Increased	Dawiskiba et al., 2014; Keshteli et al., 2017	Increased	Bruzzone et al., 2020; San Juan et al., 2020
Glycerophospholipids	Decreased	Bjerrum et al., 2010, 2017; Fan et al., 2015; Kolho et al., 2017; Scoville et al., 2018; Tefas et al., 2019, 2020	Decreased	Barberis et al., 2020; Wu et al., 2020; Song et al., 2020; Shen et al., 2020; Schwarz et al., 2021
Lysophospholipids	Decreased		Increased	Barberis et al., 2020; Song et al., 2020; Schwarz et al., 2021
Sphingolipids	Decreased	Fan et al., 2015 ; Kolho et al., 2017	Decreased	Barberis et al., 2020; Shen et al., 2020; Schwarz et al., 2021
			Increased	Song et al., 2020
Arachidonic acid (blood)	Decreased	Esteve-Comas et al., 1992; Scoville et al., 2018; Lai et al., 2019; Manfredi et al., 2019	Increased	Barberis et al., 2020; Schwarz et al., 2021; Thomas et al., 2020
Arachidonic acid (stools)	Increased	Jansson et al., 2009		

2020; Bruzzone et al., 2020) decreased in moderate and severe disease (Song et al., 2020), or unchanged (Thomas et al., 2020). Differences between patients normo-oxygenate and patients undergoing intense respiratory treatment may affect results obtained in different studies. In the paper of Bruzzone, the increase in succinate and citrate (156 and 12%, respectively) has been associated with central carbon metabolism dysfunction (Bruzzone et al., 2020). In severe COVID-19, the association of citrate decrease with succinate increase could be related to mitochondrial dysfunction due to hypoxia (Páez-Franco et al., 2021). Hypoxia inhibits oxidative phosphorylation, and thus energy is supplied by the anaerobic glycolysis, which is activated by the accumulation of the hypoxia-inducible factor 1α (HIF 1α; Majmundar et al., 2010). As a result, the TCA is blocked, with the consequent accumulation of succinate, depletion of citrate, and increased demand for glucose. In turn, the latter induces the decreased availability of gluconeogenic amino acids and oxaloacetate, being utilized as substrates for gluconeogenesis. Oxaloacetate is converted into glucose, whereas mitochondrial acetyl-CoA oxidation is drastically reduced, and acetyl-CoA is redirected to the synthesis of ketone bodies.

Ketone Bodies

Ketone bodies are small lipid-derived molecules, namely, 3-β-hydroxybutyrate, acetone, and acetoacetate (Laffel, 1999). During fasting or prolonged exercise, the liver converts fatty acids mobilized from adipocytes into ketone bodies; then, they enter circulation, serving as an energy source. Ketone bodies are crucial regulators of metabolic health and longevity; in fact, they are neuroprotective and cytoprotective (Yang et al., 2019), having the ability to inhibit histone deacetylase activity and thereby epigenetic gene regulation (Newman and Verdin, 2014). On the one hand, ketone bodies were increased in the serum of patients with IBD (**Table 5**), and this finding was correlated to the increased energy demand (Zhang et al., 2013; Dawiskiba et al., 2014; Kohashi et al., 2014; Keshteli et al., 2017). On the other hand, ketone bodies may play a strategic role in IBD when supplemented with diet, thanks to their capacity to protect from toxic effects of chronic inflammation. An experimental study demonstrated that 3-β-hydroxybutyrate suppresses the activation of the NLRP3 inflammasome in response to urate crystals, ATP, and lipotoxic fatty acids (Youm et al., 2015). Then, in a mouse model of NLRP3-mediated diseases, authors observed that 3-β-hydroxybutyrate attenuates caspase-1 activation and the release of pro-inflammatory cytokines IL-1β and IL-18 from macrophages, reducing in definitive the severity of NLRP3-mediated chronic inflammatory diseases. The same effect was obtained by applying a ketogenic diet or supplementing 3-β-hydroxybutyrate (Youm et al., 2015). The accumulation of ketone bodies following a ketogenic diet strongly impacts gut microbiota composition; *in vivo* and *in vitro* experiments demonstrated that ketone bodies selectively inhibit the growth of several *Bifidobacterial* spp., with downstream consequences for gut immune cells, especially Th17, and induce the decrease in the relative abundance of *Actinobacteria* (Ang et al., 2020). In an animal model of inflammatory colitis, it

was observed that the ketogenic diet alters gut microbiota and serum metabolome, alleviating colitis (Kong et al., 2021). In particular, *Akkermansia*, *Roseburia*, and *Ruminococcaceae* genera were enriched. After colitis induction, the ketogenic diet protected intestinal barrier function and reduced the presence of $RORgt^{+}CD3^{-}$ group 3 innate lymphoid cells and related inflammatory cytokines (IL-17α, IL-18, IL-22, and Ccl4). As a result, the ketogenic diet in patients with IBD may substantially contribute to control inflammation and shape gut microbiota.

Ketone bodies accumulate in the serum of patients with COVID-19, mimicking diabetic ketoacidosis (Li et al., 2020). Further studies confirmed the elevation of circulating ketone bodies in COVID-19 (**Table 5**), suggesting their role as an alternative energy source during SARS-CoV-2 replication (Barberis et al., 2020; Bruzzone et al., 2020; San Juan et al., 2020; Meoni et al., 2021; Páez-Franco et al., 2021). The increase in 3-β-hydroxybutyrate could interfere with viral replication by upregulating the expression of antioxidant genes, the cytoplasmic NADPH, and directly scavenging free radicals (Stubbs et al., 2020). In addition, 3-β-hydroxybutyrate may induce the closing of mitochondrial permeability transition pore, apoptosis, and the inhibition of glycolysis. It was postulated that ketone bodies inactivate the extracellular virions (Shaheen, 2021). Ketone bodies have carbonyl groups reacting with the ε-amino group of lysine to form a Schiff base. Given that SARS-CoV-2 spike protein contains approximately 4.5% lysine residues, almost equally distributed between the two subunits, the reaction between the ketone body, mainly acetoacetate, and the lysine residues of the spike protein forms Schiff bases, altering the conformation of the spike protein. This change promotes the separation of the spike protein from the virion, either by the separation between S1 and S2 subunits or promoting its bending. As a result, acetoacetate induces protein conformational changes by altering the secondary structure, namely, reducing the α-helix content (Bohlooli et al., 2016). The sum of these researches has raised the question of whether or not it may be effective to induce ketosis both in asymptomatic individuals infected with SARS-CoV-2 and in patients with COVID-19 (Bradshaw et al., 2020).

Phospholipids

Glycerophospholipids, commonly termed phospholipids, their by-product lysophospholipids, and sphingolipids are key components of the cellular membrane; remarkably, phospholipids are involved in the metabolism of cell signaling. Phospholipids are essential for the biosynthesis of lipoproteins. As reported in **Table 5**, in patients with IBD, circulating phospholipids were found decreased in various studies (Bjerrum et al., 2010, 2017; Fan et al., 2015; Kolho et al., 2017; Scoville et al., 2018; Tefas et al., 2019, 2020). Factors, such as the compromised integrity of the intestinal mucosa, TNF-α, NF-*k*B, mitogen-activated protein kinase (MAPK) pathway, and peroxisome proliferator-activated receptor (PPAR) signaling, seem to be closely implicated in phospholipids depletion. Glycerophosphocholine was significantly decreased in two subsequent studies (Bjerrum et al., 2010, 2017); more recently, in patients with extensive UC and colonic CD, tetracosanoic acid, phosphatidylcholine, (PC) lysophosphatidylcholine (LPC), sphingomyelin (SM), and diacylglycerol were found decreased compared with healthy controls; interestingly, saturated LPC (18:2) was found decreased whereas polyunsaturated LPC (20:4) and LPC (22:6) increased (Tefas et al., 2019, 2020). It is likely that the anti-inflammatory polyunsaturated LPC effectively antagonized the pro-inflammatory activity of saturated LPC. Two studies found sphingolipids significantly reduced in patients with IBD (Fan et al., 2015; Kolho et al., 2017). The significant depletion in sphingolipids may be the result of the increased activity of sphingomyelinases, which are activated in IBD by the combined action of TNF-α, NF-*k*B, and IFN-γ (Schütze et al., 1992).

Phospholipids metabolism is strongly influenced by COVID-19 and by the severity of the disease (Mussap and Fanos, 2021); data emerging from the literature suggest that SARS-CoV-2 infection promotes the downregulation of most phospholipids and sphingolipids (**Table 5**), while various lysophospholipids are either overexpressed or unchanged (Barberis et al., 2020; Shen et al., 2020; Song et al., 2020; Wu et al., 2020; Schwarz et al., 2021). The downregulation of phospholipids may originate from the liver impairment in patients with severe COVID-19, whereas lysophospholipids upregulation is the result of the increased activity of phospholipase A_2. Low-risk infected patients are well discriminated from non-infected individuals by high levels of phosphatidylcholine (PC38:8), phosphatidylethanolamine (PE38:4), and phosphatidylglycerol (PG20:5). In patients with COVID-19, the predominance of inflammation over the macrophage-driven anti-inflammatory response leads to the underexpression of sphingosine 1-phosphate (Shen et al., 2020; Song et al., 2020). The considerable number of lipids belonging to any lipid class gives a limited value to the definition of upregulation and downregulation of phospholipids, lysophospholipids, and sphingolipids in COVID-19; recently, a targeted lipidomic analysis measured a considerable number of phospholipids ($n = 90$), lysophospholipids ($n = 14$), and sphingolipids ($n = 15$), discovering that certain lipids decrease and other increase within the same lipid class (Caterino et al., 2021).

Arachidonic Acid and Phospholipases

Arachidonic acid is a 20-carbon chain belonging to the ω-6 (n-6) polyunsaturated fatty acids (PUFAs); it is obtained from food and then incorporated in phospholipids (Tallima and El Ridi, 2017). Arachidonic acid is the primary precursor for the biosynthesis of eicosanoids, a complex family of lipid signaling mediators including but not limited to leukotrienes, prostaglandins, thromboxane, and prostacyclin A_2 (Calder, 2020). By the cleavage of arachidonic acid from membrane phospholipids, phospholipases A_2-IID, -IIF, -III, and -X initiate the arachidonic acid cascade and eicosanoid production (Murakami et al., 2020). Eicosanoids are generated by three main pathways, namely, cyclooxygenases, lipoxygenases, and cytochrome P-450 epoxygenase pathways (Dennis and Norris, 2015). Beyond their crucial role in a broad range of physiological processes, eicosanoids are involved in the pathogenesis of IBD (Alhouayek et al., 2021). During the acute phase of the disease, their concentration significantly increases within the inflamed

intestinal mucosa (Wallace, 2019). In patients with IBD, arachidonic acid was increased in feces (Jansson et al., 2009) and decreased in the blood (**Table 5**); these variations were often associated with the exacerbation of the disease activity (Esteve-Comas et al., 1992; Scoville et al., 2018; Lai et al., 2019; Manfredi et al., 2019; Gallagher et al., 2021). Based on current knowledge, the decrease of circulating arachidonic acid in IBD might be related to the increased synthesis of eicosanoids in the gut (Shores et al., 2011). SARS-CoV-2 infection induces the overexpression of phospholipase A_2, especially in macrophages, T, and B cells; as a consequence, the biosynthesis of arachidonic acid is upregulated in patients with COVID-19 (**Table 5**). The arachidonic acid upregulation positively correlates with the IL-6 concentration and the disease's severity (Barberis et al., 2020; Thomas et al., 2020; Schwarz et al., 2021). In patients with COVID-19 and severe liver injury, however, arachidonic acid may be downregulated (Shen et al., 2020). Arachidonic acid is a potent antiviral PUFA that can inactivate the enveloped viruses, including SARS-CoV-2 (Hoxha, 2020); human cells infected by HCoV-229E and MERS-CoV are inhibited by arachidonic acid (Yan et al., 2019). The increased activity of the cytosolic phospholipase $A_2\alpha$ ($cPLA_2\alpha$) in cells infected by SARS-CoV-2 generates the overproduction of lysophospholipids that are essential for the viral replication (Casari et al., 2021); the pharmacological inhibition of $cPLA_2\alpha$ in human Huh-7 cells infected with coronavirus 229E drastically reduces the viral RNA synthesis, blocking an early step in the replication cycle (Müller et al., 2018). This finding opens new perspectives on the effective treatment of SARS-CoV-2 infection. A recent study performed a targeted lipidomic analysis of bronchoalveolar lavages fluids (BALs) from patients with severe COVID-19 (Archambault et al., 2021). The most relevant finding was the significant increase in several bioactive lipids, such as thromboxane, leukotrienes, and 15-lipoxygenase metabolites derived from linoleic acid, linolenic acid, and dihomo-γ-linolenic acid (**Table 5**). During the early stage of inflammation, the enzymatic oxygenation of essential fatty acids, including arachidonic acid, eicosapentaenoic acid, docosapentaenoic acid, and docosahexaenoic acid, generates a class of bioactive lipids, the so-called specialized pro-resolving mediators (SPMs; Serhan et al., 2000). This class includes lipoxins, resolvins, maresins, and protectins (Basil and Levy, 2016). Interestingly, BALs from patients with severe COVID-19 were also marked by the increase in SPMs. Increased levels of pro-inflammatory bioactive lipids and anti-inflammatory SPMs have also been reported in serum samples collected from hospital inpatients with a confirmed diagnosis of COVID-19 (Turnbull et al., 2022).

CONCLUSIVE AND PROSPECTIVE REMARKS

IBD and COVID-19 are characterized by gut dysbiosis associated with impaired gut barrier function and immune-mediated chronic inflammation. However, COVID-19 can rapidly evolve to multisystemic organ damage due to a dysregulated, fulminant inflammatory response, the so-called "cytokine storm" (Jain, 2020; Hu et al., 2021). The evident imbalance between the number of studies on gut microbiota in IBD and that in SARS-CoV-2 does not exclude a reliable data analysis; we noticed more similarities than differences in gut microbial alterations between IBD and COVID-19 (**Table 1**). The depletion in *F. prausnitzii*, *E. rectale*, *R, bromii*, *Lachnospiraceae*, *C. leptum* (cluster IV), and the overgrowth of *Enterococcus*, *E. coli*, *Shigella*, *P. mirabilis*, *Fusobacterium*, *Veillonellaceae* are common patterns of dysbiosis, playing a key role in the severity and clinical outcome of IBD and COVID-19. An intriguing dissimilarity between IBD and COVID-19 is the abundance of *A. muciniphila* (**Table 1**), a Gram-negative bacterium belonging to the *Verrucomicrobia* phylum. *A. muciniphila* colonizes the mucus layer close to gut epithelial cells and is able to degrade mucin sugars and the protein backbone by specific enzymes, such as sialidases and fucosidases, providing carbon and nitrogen for bacteria unable to produce these enzymes (van Passel et al., 2011). Mucus degradation by *A. muciniphila* generates SCFAs, which are strongly involved in promoting host metabolic health. Therefore, *A. muciniphila* has several beneficial effects on host health by reducing inflammation, stimulating mucin biosynthesis and mucus thickness, preserving the integrity of the mucosal barrier, increasing the expression of tight junction proteins (e.g., occluding), and modulating the intestinal adaptative immune response (Ottman et al., 2017; Ansaldo et al., 2019; Ashrafian et al., 2019; Liu et al., 2021). The depletion of *A. muciniphila* in IBD (**Table 1**), reported by several studies, confirms the well-known inverse relationship between this bacterium and IBD (Rajilić-Stojanović et al., 2013); on the other hand, the enrichment in *A. muciniphila* in patients with COVID-19 has been associated with that of opportunistic pathogens, such as *Enterococcus*, *Staphylococcus*, *Serratia*, *Collinsella*, *Actinomyces*, and many others (Gaibani et al., 2021). Concerns emerge about results on the probiotic strains *Lactobacilli* and *Bifidobacteria*, especially in patients with COVID-19 (**Table 1**). These strains are Gram-positive, non-spore-forming, lactic acid producer bacteria with the antiviral activity performed by various mechanisms, including the synthesis of antiviral inhibitory metabolites, the upregulation of the protective immune responses, and by competing for nutrients and colonization sites with the virus and, more extensively, with pathogens (Kesika et al., 2021). Five studies reported *Lactobacillus* enrichment in patients with COVID-19 (**Table 1**), confuting the general notion that gut dysbiosis due to severe infections promotes the depletion of these strains (Harper et al., 2021). Very few data support the depletion of *Lactobacillus* and *Bifidobacterium* in patients with COVID-19. Two articles cited in the PubMed library, written in Chinese with the same English abstract and the same digital object identifier (doi), described gut dysbiosis in "some patients" with COVID-19 marked by the decrease in *Lactobacillus* and *Bifidobacterium* abundance (Xu et al., 2020a,b). Clearly, these data are unreliable. Therefore, further studies are required to elucidate the significance of *Lactobacillus* and *Bifidobacterium* enrichment in COVID-19.

Metabolomics reveals considerable similarities in the tryptophan metabolism between IBD and COVID-19 (**Table 3**). In IBD, the increase in quinolinic acid is associated with the

decrease in picolinic acid, a non-selective metal ion chelating agent with antimicrobial, antiviral, and antifungal activity formed by a non-enzymic cyclization of aminomuconic acid semialdehyde. In COVID-19, the increase in quinolinic acid is associated with the increase of kynurenic acid; the biochemical mechanism underlying this unusual association should be clarified. In fact, quinolinic acid and kynurenic acid are closely related to each other by an inverse relationship that becomes imbalanced in various diseases. In COVID-19, the increase in picolinic acid reflects the activation of the enzymatic conversion of 2-amino-3-carboxymuconate-6-semialdehyde (ACMS) to 2-aminomuconic-6-semialdehyde. In turn, the latter undergoes either non-enzymatic cyclization to form picolinic acid or enzymatic transformation to 2-aminomuconic acid, yielding acetyl-CoA (Badawy, 2017). It is unclear why picolinic acid is increased in COVID-19, taking into account the alteration of brain functions during the disease (Chou et al., 2021; Marshall, 2021). It is reasonable to assume that the limited number of metabolic differences between IBD and COVID-19 (**Tables 4**, **5**), for example, blood arachidonic acid, originates from the acute systemic damage and impairment of organs, tissues, and biological systems (e.g., coagulation) in COVID-19, while IBD remains a chronic disease localized in the gastrointestinal tract with a broad spectrum of extraintestinal symptoms and comorbidities (Argollo et al., 2019). Emerging evidence indicates the role of the microbiome in modulating the immune response to vaccination and the role of metabolic profile in predicting vaccination outcome (Hagan et al., 2019; Alexander et al., 2021). Therefore, microbiomics and metabolomics may be considered powerful tools for the early identification and monitoring of individuals at risk of adverse events, for example, fragile individuals or patients with severe chronic diseases. Since the COVID-19 pandemic is far from over, SARS-CoV-2 vaccination is the mainstay for preventing the COVID-19 spread. As a result, there is the need to extend SARS-CoV-2 vaccination to any adult or young subject, even if affected by pre-existing chronic diseases, such as IBD (Alexander et al., 2021). Therefore, deciphering the individual metabolic and microbial fingerprint in IBD is crucial to define an effective strategy for the safe administration of the vaccine (e.g., to discover any metabolic/microbial alteration induced by recent steroid or rituximab therapy), to manage vaccinated individuals, and to avoid any possible adverse effect in individuals at risk (Ferretti et al., 2021). On the other hand, deciphering the individual metabolic and microbial fingerprint in patients with IBD with COVID-19 may considerably improve the therapeutic approach, preventing the risk of adverse outcomes.

AUTHOR CONTRIBUTIONS

GC and FB contributed to the conception and design of the study. VF conceptualized and performed the modeling. GC wrote the first draft of the manuscript. MM supervised the manuscript and wrote the final version of the paper. All authors contributed to the article and approved the submitted version.

REFERENCES

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

Aktas, B., and Aslim, B. (2021). Neuropathy in COVID-19 associated with dysbiosis-related inflammation. *Turk. J. Biol.* 45, 390–403. doi: 10.3906/biy-2105-53

Alam, M. T., Amos, G. C. A., Murphy, A. R. J., Murch, S., Wellington, E. M. H., and Arasaradnam, R. P. (2020). Microbial imbalance in inflammatory bowel disease patients at different taxonomic levels. *Gut Pathog.* 12:1. doi: 10.1186/s13099-019-0341-6

Aldars-García, L., Chaparro, M., and Gisbert, J. P. (2021). Systematic review: the gut microbiome and its potential clinical application in inflammatory bowel disease. *Microorganisms* 9:977. doi: 10.3390/microorganisms9050977

Alexander, J. L., Moran, G. W., Gaya, D. R., Raine, T., Hart, A., Kennedy, N. A., et al. (2021). SARS-CoV-2 vaccination for patients with inflammatory bowel disease: a British Society of Gastroenterology Inflammatory Bowel Disease section and IBD clinical research group position statement. *Lancet Gastroenterol. Hepatol.* 6, 218–224. doi: 10.1016/S2468-1253(21)00024-8

Alhouayek, M., Ameraoui, H., and Muccioli, G. G. (2021). Bioactive lipids in inflammatory bowel diseases - From pathophysiological alterations to therapeutic opportunities. *Biochim. Biophys. Acta Mol. Cell Biol. Lipids* 1866:158854. doi: 10.1016/j.bbalip.2020.158854

Allocca, M., Chaparro, M., Gonzalez, H. A., Bosca-Watts, M. M., Palmela, C., D'Amico, F., et al. (2020). Patients with inflammatory bowel disease are not at increased risk of COVID-19: a large multinational cohort study. *J. Clin. Med.* 9:3533. doi: 10.3390/jcm9113533

Allocca, M., and Craviotto, V. (2021). Low risk of severe COVID-19 in patients with inflammatory bowel disease: keep calm and take stock. *Pol. Arch. Intern. Med.* 131, 222–223. doi: 10.20452/pamw.15903

Alonso, A., Julià, A., Vinaixa, M., Domènech, E., Fernández-Nebro, A., Cañete, J. D., et al. (2016). Urine metabolome profiling of immune-mediated inflammatory diseases. *BMC Med.* 14:133. doi: 10.1186/s12916-016-0681-8

Alshehri, D., Saadah, O., Mosli, M., Edris, S., Alhindi, R., and Bahieldin, A. (2021). Dysbiosis of gut microbiota in inflammatory bowel disease: current therapies and potential for microbiota-modulating therapeutic approaches. *Bosn. J. Basic Med. Sci.* 21, 270–283. doi: 10.17305/bjbms.2020.5016

Andoh, A., Tsujikawa, T., Sasaki, M., Mitsuyama, K., Suzuki, Y., Matsui, T., et al. (2009). Faecal microbiota profile of Crohn's disease determined by terminal restriction fragment length polymorphism analysis. *Aliment. Pharmacol. Ther.* 29, 75–82. doi: 10.1111/j.1365-2036.2008.03860.x

Andou, A., Hisamatsu, T., Okamoto, S., Chinen, H., Kamada, N., Kobayashi, T., et al. (2009). Dietary histidine ameliorates murine colitis by inhibition of proinflammatory cytokine production from macrophages. *Gastroenterology* 136, 564–574.e2. doi: 10.1053/j.gastro.2008.09.062

Ang, Q. Y., Alexander, M., Newman, J. C., Tian, Y., Cai, J., Upadhyay, V., et al. (2020). Ketogenic diets alter the gut microbiome resulting in decreased intestinal Th17 cells. *Cell* 181, 1263–275.e16. doi: 10.1016/j.cell.2020.04.027

Ansaldo, E., Slayden, L. C., Ching, K. L., Koch, M. A., Wolf, N. K., Plichta, D. R., et al. (2019). Akkermansia muciniphila induces intestinal adaptive immune responses during homeostasis. *Science* 364, 1179–1184. doi: 10.1126/science.aaw7479

Archambault, A. S., Zaid, Y., Rakotoarivelo, V., Turcotte, C., Doré, É., Dubuc, I., et al. (2021). High levels of eicosanoids and docosanoids in the lungs of intubated COVID-19 patients. *FASEB J.* 35:e21666. doi: 10.1096/fj.202100540R

Ardizzone, S., Ferretti, F., Monico, M. C., Carvalhas Gabrielli, A. M., Carmagnola, S., Bezzio, C., et al. (2021) Lower incidence of COVID-19 in patients with inflammatory bowel disease treated with non-gut selective biologic therapy. *J. Gastroenterol. Hepatol.* doi: 10.1111/jgh.15591 [Epub ahead of print].

Argollo, M., Gilardi, D., Peyrin-Biroulet, C., Chabot, J. F., Peyrin-Biroulet, L., and Danese, S. (2019). Comorbidities in inflammatory bowel disease: a call

for action. *Lancet Gastroenterol. Hepatol.* 4, 643–654. doi: 10.1016/S2468-1253(19)30173-6

Ashrafian, F., Shahriary, A., Behrouzi, A., Moradi, H. R., Keshavarz Azizi Raftar, S., Lari, A., et al. (2019). Akkermansia muciniphila-derived extracellular vesicles as a mucosal delivery vector for amelioration of obesity in mice. *Front. Microbiol.* 10:2155. doi: 10.3389/fmicb.2019.02155

Ashrafian, H., Sounderajah, V., Glen, R., Ebbels, T., Blaise, B. J., Kalra, D., et al. (2020). Metabolomics: The stethoscope for the twenty-first century. *Med. Princ. Pract.* 30, 301–310. doi: 10.1159/000513545

Aziz, M., Fatima, R., Haghbin, H., Lee-Smith, W., and Nawras, A. (2020). The incidence and outcomes of COVID-19 in IBD patients: A rapid review and meta-analysis. *Inflamm. Bowel Dis.* 26, e132–e133. doi: 10.1093/ibd/izaa170

Badawy, A. A. (2017). Tryptophan availability for kynurenine pathway metabolism across the life span: control mechanisms and focus on aging, exercise, diet and nutritional supplements. *Neuropharmacology* 112, 248–263. doi: 10.1016/j.neuropharm.2015.11.015

Barberis, E., Timo, S., Amede, E., Vanella, V. V., Puricelli, C., Cappellano, G., et al. (2020). Large-scale plasma analysis revealed new mechanisms and molecules associated with the host response to SARS-CoV-2. *Int. J. Mol. Sci.* 21:8623. doi: 10.3390/ijms21228623

Basil, M. C., and Levy, B. D. (2016). Specialized pro-resolving mediators: endogenous regulators of infection and inflammation. *Nat. Rev. Immunol.* 16, 51–67. doi: 10.1038/nri.2015.4

Basuroy, S., Seth, A., Elias, B., Naren, A. P., and Rao, R. (2006). MAPK interacts with occludin and mediates EGF-induced prevention of tight junction disruption by hydrogen peroxide. *Biochem. J.* 393, 69–77. doi: 10.1042/BJ20050959

Benjamin, J. L., Hedin, C. R., Koutsoumpas, A., Ng, S. C., McCarthy, N. E., Prescott, N. J., et al. (2012a). Smokers with active Crohn's disease have a clinically relevant dysbiosis of the gastrointestinal microbiota. *Inflamm. Bowel Dis.* 18, 1092–1100. doi: 10.1002/ibd.21864

Benjamin, J., Makharia, G., Ahuja, V., Anand Rajan, K. D., Kalaivani, M., and Gupta, S. D. (2012b). Glutamine and whey protein improve intestinal permeability and morphology in patients with Crohn's disease: a randomized controlled trial. *Dig. Dis. Sci.* 57, 1000–1012. doi: 10.1007/s10620-011-1947-9

Bertè, R., Mazza, S., Stefanucci, M. R., Noviello, D., Costa, S., Ciafardini, C., et al. (2021). Seroprevalence of SARS-CoV2 in IBD patients treated with biologic therapy. *J. Crohns Colitis* 15, 864–868. doi: 10.1093/ecco-jcc/jjaa237

Bjerrum, J. T., Nielsen, O. H., Hao, F., Tang, H., Nicholson, J. K., Wang, Y., et al. (2010). Metabonomics in ulcerative colitis: diagnostics, biomarker identification, and insight into the pathophysiology. *J. Proteome Res.* 9, 954–962. doi: 10.1021/pr9008223

Bjerrum, J. T., Steenholdt, C., Ainsworth, M., Nielsen, O. H., Reed, M. A., Atkins, K., et al. (2017). Metabonomics uncovers a reversible proatherogenic lipid profile during infliximab therapy of inflammatory bowel disease. *BMC Med.* 15:184. doi: 10.1186/s12916-017-0949-7

Blasco, H., Bessy, C., Plantier, L., Lefevre, A., Piver, E., Bernard, L., et al. (2020). The specific metabolome profiling of patients infected by SARS-COV-2 supports the key role of tryptophan-nicotinamide pathway and cytosine metabolism. *Sci. Rep.* 10:16824. doi: 10.1038/s41598-020-73966-5

Bohlooli, M., Ghaffari-Moghaddam, M., Khajeh, M., Shahraki-Fallah, G., Haghighi-Kekhaiye, B., and Sheibani, N. (2016). The role of acetoacetate in Amadori product formation of human serum albumin. *J. Photochem. Photobiol. B* 163, 345–351. doi: 10.1016/j.jphotobiol.2016.09.004

Bonovas, S., Fiorino, G., Allocca, M., Lytras, T., Nikolopoulos, G. K., Peyrin-Biroulet, L., et al. (2016). Biologic therapies and risk of infection and malignancy in patients with inflammatory bowel disease: a systematic review and network meta-analysis. *Clin. Gastroenterol. Hepatol.* 14, 1385–97.e10. doi: 10.1016/j.cgh.2016.04.039

Bosch, S., Struys, E. A., van Gaal, N., Bakkali, A., Jansen, E. W., Diederen, K., et al. (2018). Fecal amino acid analysis can discriminate De novo treatment-Naïve pediatric inflammatory bowel disease from controls. *J. Pediatr. Gastroenterol. Nutr.* 66, 773–778. doi: 10.1097/MPG.0000000000001812

Bradshaw, P. C., Seeds, W. A., Miller, A. C., Mahajan, V. R., and Curtis, W. M. (2020). COVID-19: proposing a ketone-based metabolic therapy as a treatment to blunt the cytokine storm. *Oxidative Med. Cell. Longev.* 2020:6401341. doi: 10.1155/2020/6401341

Brenner, E. J., Ungaro, R. C., and Colombel, J. F. (2021). SECURE-IBD database public data update. Available at: https://covidibd.org/ (Accessed September 9, 2021).

Brenner, E. J., Ungaro, R. C., Gearry, R. B., Kaplan, G. G., Kissous-Hunt, M., Lewis, J. D., et al. (2020). Corticosteroids, but not TNF antagonists, are associated with adverse COVID-19 outcomes in patients with inflammatory bowel diseases: results from an international registry. *Gastroenterology* 159, 481–91.e3. doi: 10.1053/j.gastro.2020.05.032

Bruzzone, C., Bizkarguenaga, M., Gil-Redondo, R., Diercks, T., Arana, E., García de Vicuña, A., et al. (2020). SARS-CoV-2 infection dysregulates the metabolomic and lipidomic profiles of serum. *iScience* 23:101645. doi: 10.1016/j.isci.2020.101645

Calder, P. C. (2020). Eicosanoids. *Essays Biochem.* 64, 423–441. doi: 10.1042/EBC20190083

Casari, I., Manfredi, M., Metharom, P., and Falasca, M. (2021). Dissecting lipid metabolism alterations in SARS-CoV-2. *Prog. Lipid Res.* 82:101092. doi: 10.1016/j.plipres.2021.101092

Caterino, M., Gelzo, M., Sol, S., Fedele, R., Annunziata, A., Calabrese, C., et al. (2021). Dysregulation of lipid metabolism and pathological inflammation in patients with COVID-19. *Sci. Rep.* 11:2941. doi: 10.1038/s41598-021-82426-7

Chen, G. L., Zhang, Y., Wang, W. Y., Ji, X. L., Meng, F., Xu, P. S., et al. (2017). Partners of patients with ulcerative colitis exhibit a biologically relevant dysbiosis in fecal microbial metacommunities. *World J. Gastroenterol.* 23, 4624–4631. doi: 10.3748/wjg.v23.i25.4624

Chervy, M., Barnich, N., and Denizot, J. (2020). Adherent-invasive *E. coli*: update on the lifestyle of a troublemaker in Crohn's disease. *Int. J. Mol. Sci.* 21, 3734. doi: 10.3390/ijms21103734

Cheung, K. S., Hung, I. F. N., Chan, P. P. Y., Lung, K. C., Tso, E., Liu, R., et al. (2020). Gastrointestinal manifestations of SARS-CoV-2 infection and virus load in fecal samples from a Hong Kong cohort: systematic review and meta-analysis. *Gastroenterology* 159, 81–95. doi: 10.1053/j.gastro.2020.03.065

Chiodini, R. J., Dowd, S. E., Galandiuk, S., Davis, B., and Glassing, A. (2016). The predominant site of bacterial translocation across the intestinal mucosal barrier occurs at the advancing disease margin in Crohn's disease. *Microbiology* 162, 1608–1619. doi: 10.1099/mic.0.000336

Chou, S. H., Beghi, E., Helbok, R., Moro, E., Sampson, J., Altamirano, V., et al. (2021). Global incidence of neurological manifestations Among patients hospitalized With COVID-19 - a report for the GCS-NeuroCOVID consortium and the ENERGY consortium. *JAMA Netw. Open* 4:e2112131. doi: 10.1001/jamanetworkopen.2021.12131

Clooney, A. G., Eckenberger, J., Laserna-Mendieta, E., Sexton, K. A., Bernstein, M. T., Vagianos, K., et al. (2021). Ranking microbiome variance in inflammatory bowel disease: a large longitudinal intercontinental study. *Gut* 70, 499–510. doi: 10.1136/gutjnl-2020-321106

Cruzat, V., Macedo Rogero, M., Noel Keane, K., Curi, R., and Newsholme, P. (2018). Glutamine: metabolism and immune function, supplementation and clinical translation. *Nutrients* 10:1564. doi: 10.3390/nu10111564

Curi, R., Lagranha, C. J., Doi, S. Q., Sellitti, D. F., Procopio, J., Pithon-Curi, T. C., et al. (2005). Molecular mechanisms of glutamine action. *J. Cell. Physiol.* 204, 392–401. doi: 10.1002/jcp.20339

Cyprian, F., Sohail, M. U., Abdelhafez, I., Salman, S., Attique, Z., Kamareddine, L., et al. (2021). SARS-CoV-2 and immune-microbiome interactions: lessons from respiratory viral infections. *Int. J. Infect. Dis.* 105, 540–550. doi: 10.1016/j.ijid.2021.02.071

D'Amico, F., Danese, S., and Peyrin-Biroulet, L. (2020). Systematic review on inflammatory bowel disease patients with coronavirus disease 2019: it is time to take stock. *Clin. Gastroenterol. Hepatol.* 18, 2689–2700. doi: 10.1016/j.cgh.2020.08.003

Dawiskiba, T., Deja, S., Mulak, A., Ząbek, A., Jawień, E., Pawełka, D., et al. (2014). Serum and urine metabolomic fingerprinting in diagnostics of inflammatory bowel diseases. *World J. Gastroenterol.* 20, 163–174. doi: 10.3748/wjg.v20.i1.163

de Meij, T. G. J., de Groot, E. F. J., Peeters, C. F. W., de Boer, N. K. H., Kneepkens, C. M. F., Eck, A., et al. (2018). Variability of core microbiota in newly diagnosed treatment-naïve paediatric inflammatory bowel disease patients. *PLoS One* 13:e0197649. doi: 10.1371/journal.pone.0197649

de Oliveira, G. L. V., Oliveira, C. N. S., Pinzan, C. F., de Salis, L. V. V., and Cardoso, C. R. B. (2021). Microbiota modulation of the gut-lung axis in COVID-19. *Front. Immunol.* 12:635471. doi: 10.3389/fimmu.2021.635471

Dennis, E. A., and Norris, P. C. (2015). Eicosanoid storm in infection and inflammation. *Nat. Rev. Immunol.* 15, 511–523. doi: 10.1038/nri3859

Devaux, C. A., Lagier, J. C., and Raoult, D. (2021). New insights into the physiopathology of COVID-19: SARS-CoV-2-associated gastrointestinal illness. *Front. Med.* 8:640073. doi: 10.3389/fmed.2021.640073

Dicksved, J., Halfvarson, J., Rosenquist, M., Järnerot, G., Tysk, C., Apajalahti, J., et al. (2008). Molecular analysis of the gut microbiota of identical twins with Crohn's disease. *ISME J.* 2, 716–727. doi: 10.1038/ismej.2008.37

Dipasquale, V., Cucchiara, S., Martinelli, M., Miele, E., Aloi, M., and Romano, C. (2020). Challenges in paediatric inflammatory bowel diseases in the COVID-19 time. *Dig. Liver Dis.* 52, 593–594. doi: 10.1016/j.dld.2020.03.015

Dodd, D., Spitzer, M. H., Van Treuren, W., Merrill, B. D., Hryckowian, A. J., Higginbottom, S. K., et al. (2017). A gut bacterial pathway metabolizes aromatic amino acids into nine circulating metabolites. *Nature* 551, 648–652. doi: 10.1038/nature24661

Doğan, H. O., Şenol, O., Bolat, S., Yıldız, Ş. N., Büyüktuna, S. A., Sarıismailoğlu, R., et al. (2021). Understanding the pathophysiological changes via untargeted metabolomics in COVID-19 patients. *J. Med. Virol.* 93, 2340–2349. doi: 10.1002/jmv.26716

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

Ellermann, M., Huh, E. Y., Liu, B., Carroll, I. M., Tamayo, R., and Sartor, R. B. (2015). Adherent-invasive *Escherichia coli* production of cellulose influences iron-induced bacterial aggregation, phagocytosis, and induction of colitis. *Infect. Immun.* 83, 4068–4080. doi: 10.1128/IAI.00904-15

Elsden, S. R., Hilton, M. G., and Waller, J. M. (1976). The end products of the metabolism of aromatic amino acids by clostridia. *Arch. Microbiol.* 107, 283–288. doi: 10.1007/BF00425340

Esteve-Comas, M., Ramírez, M., Fernández-Bañares, F., Abad-Lacruz, A., Gil, A., Cabré, E., et al. (1992). Plasma polyunsaturated fatty acid pattern in active inflammatory bowel disease. *Gut* 33, 1365–1369. doi: 10.1136/gut.33.10.1365

Falandry, C., Malapert, A., Roche, M., Subtil, F., Berthiller, J., Boin, C., et al. (2021). Risk factors associated with day-30 mortality in patients over 60 years old admitted in ICU for severe COVID-19: the senior-COVID-Rea multicentre survey protocol. *BMJ Open* 11:e044449. doi: 10.1136/bmjopen-2020-044449

Fan, F., Mundra, P. A., Fang, L., Galvin, A., Moore, X. L., Weir, J. M., et al. (2015). Lipidomic profiling in inflammatory bowel disease: comparison Between ulcerative colitis and Crohn's disease. *Inflamm. Bowel Dis.* 21, 1511–1518. doi: 10.1097/MIB.0000000000000394

Fanos, V., Pintus, R., Pintus, M. C., Mussap, M., and Marcialis, M. A. (2021). Seven secrets of COVID-19: fever, ACE2 receptors, gut-lung axis, metabolomics, microbiomics, probiotics, diet. *J. Pediatr. Neon Ind. Med.* 10:e100145. doi: 10.7363/100145

Favier, C., Neut, C., Mizon, C., Cortot, A., Colombel, J. F., and Mizon, J. (1997). Fecal beta-D-galactosidase production and bifidobacteria are decreased in Crohn's disease. *Dig. Dis. Sci.* 42, 817–822. doi: 10.1023/A:1018876400528

Ferretti, F., Cannatelli, R., Benucci, M., Carmagnola, S., Clementi, E., Danelli, P., et al. (2021). How to manage COVID-19 vaccination in immune-mediated inflammatory diseases: an expert opinion by IMIDs study group. *Front. Immunol.* 12:656362. doi: 10.3389/fimmu.2021.656362

Fite, A., Macfarlane, S., Furrie, E., Bahrami, B., Cummings, J. H., Steinke, D. T., et al. (2013). Longitudinal analyses of gut mucosal microbiotas in ulcerative colitis in relation to patient age and disease severity and duration. *J. Clin. Microbiol.* 51, 849–856. doi: 10.1128/JCM.02574-12

Flynn, S., and Eisenstein, S. (2019). Inflammatory bowel disease presentation and diagnosis. *Surg. Clin. North Am.* 99, 1051–1062. doi: 10.1016/j.suc.2019.08.001

Forrest, C. M., Youd, P., Kennedy, A., Gould, S. R., Darlington, L. G., and Stone, T. W. (2002). Purine, kynurenine, neopterin and lipid peroxidation levels in inflammatory bowel disease. *J. Biomed. Sci.* 9, 436–442. doi: 10.1007/BF02256538

Frank, D. N., St Amand, A. L., Feldman, R. A., Boedeker, E. C., Harpaz, N., and Pace, N. R. (2007). Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. *Proc. Natl. Acad. Sci. U. S. A.* 104, 13780–13785. doi: 10.1073/pnas.0706625104

Franzosa, E. A., Sirota-Madi, A., Avila-Pacheco, J., Fornelos, N., Haiser, H. J., Reinker, S., et al. (2019). Gut microbiome structure and metabolic activity in inflammatory bowel disease. *Nat. Microbiol.* 4, 293–305. doi: 10.1038/s41564-018-0306-4

Fraser, D. D., Slessarev, M., Martin, C. M., Daley, M., Patel, M. A., Miller, M. R., et al. (2020). Metabolomics profiling of critically ill coronavirus disease 2019 patients: identification of diagnostic and prognostic biomarkers. *Crit. Care Explor.* 2:e0272. doi: 10.1097/CCE.0000000000000272

Fujimoto, T., Imaeda, H., Takahashi, K., Kasumi, E., Bamba, S., Fujiyama, Y., et al. (2013). Decreased abundance of Faecalibacterium prausnitzii in the gut microbiota of Crohn's disease. *J. Gastroenterol. Hepatol.* 28, 613–619. doi: 10.1111/jgh.12073

Gaibani, P., D'Amico, F., Bartoletti, M., Lombardo, D., Rampelli, S., Fornaro, G., et al. (2021). The gut microbiota of critically ill patients With COVID-19. *Front. Cell. Infect. Microbiol.* 11:670424. doi: 10.3389/fcimb.2021.670424

Galipeau, H. J., Caminero, A., Turpin, W., Bermudez-Brito, M., Santiago, A., Libertucci, J., et al. (2021). Novel fecal biomarkers That precede clinical diagnosis of ulcerative colitis. *Gastroenterology* 160, 1532–1545. doi: 10.1053/j.gastro.2020.12.004

Gallagher, K., Catesson, A., Griffin, J. L., Holmes, E., and Williams, H. R. T. (2021). Metabolomic analysis in inflammatory bowel disease: a systematic review. *J. Crohns Colitis* 15, 813–826. doi: 10.1093/ecco-jcc/jjaa227

Gao, J., Xu, K., Liu, H., Liu, G., Bai, M., Peng, C., et al. (2018). Impact of the gut microbiota on intestinal immunity mediated by tryptophan metabolism. *Front. Cell. Infect. Microbiol.* 8:13. doi: 10.3389/fcimb.2018.00013

Gevers, D., Kugathasan, S., Denson, L. A., Vázquez-Baeza, Y., Van Treuren, W., Ren, B., et al. (2014). The treatment-naive microbiome in new-onset Crohn's disease. *Cell Host Microbe* 15, 382–392. doi: 10.1016/j.chom.2014.02.005

Gillevet, P., Sikaroodi, M., Keshavarzian, A., and Mutlu, E. A. (2010). Quantitative assessment of the human gut microbiome using multitag pyrosequencing. *Chem. Biodivers.* 7, 1065–1075. doi: 10.1002/cbdv.200900322

Gophna, U., Sommerfeld, K., Gophna, S., Doolittle, W. F., and Veldhuyzen van Zanten, S. J. (2006). Differences between tissue-associated intestinal microfloras of patients with Crohn's disease and ulcerative colitis. *J. Clin. Microbiol.* 44, 4136–4141. doi: 10.1128/JCM.01004-06

Grasselli, G., Zangrillo, A., Zanella, A., Antonelli, M., Cabrini, L., Castelli, A., et al. (2020). Baseline characteristics and outcomes of 1591 patients infected with SARS-CoV-2 admitted to ICUs of the Lombardy region, Italy. *JAMA* 323, 1574–1581. doi: 10.1001/jama.2020.5394

Gronke, K., Hernández, P. P., Zimmermann, J., Klose, C. S. N., Kofoed-Branzk, M., Guendel, F., et al. (2019). Interleukin-22 protects intestinal stem cells against genotoxic stress. *Nature* 566, 249–253. doi: 10.1038/s41586-019-0899-7

Gu, S., Chen, Y., Wu, Z., Chen, Y., Gao, H., Lv, L., et al. (2020). Alterations of the gut microbiota in patients with coronavirus disease 2019 or H1N1 influenza. *Clin. Infect. Dis.* 71, 2669–2678. doi: 10.1093/cid/ciaa709

Guo, M., Tao, W., Flavell, R. A., and Zhu, S. (2021). Potential intestinal infection and faecal-oral transmission of SARS-CoV-2. *Nat. Rev. Gastroenterol. Hepatol.* 18, 269–283. doi: 10.1038/s41575-021-00416-6

Hagan, T., Cortese, M., Rouphael, N., Boudreau, C., Linde, C., Maddur, M. S., et al. (2019). Antibiotics-driven gut microbiome perturbation alters immunity to vaccines in humans. *Cell* 178, 1313–328.e13. doi: 10.1016/j.cell.2019.08.010

Halfvarson, J., Brislawn, C. J., Lamendella, R., Vázquez-Baeza, Y., Walters, W. A., Bramer, L. M., et al. (2017). Dynamics of the human gut microbiome in inflammatory bowel disease. *Nat. Microbiol.* 2:17004. doi: 10.1038/nmicrobiol.2017.4

Harper, A., Vijayakumar, V., Ouwehand, A. C., Ter Haar, J., Obis, D., Espadaler, J., et al. (2021). Viral infections, the microbiome, and probiotics. *Front. Cell. Infect. Microbiol.* 10:596166. doi: 10.3389/fcimb.2020.596166

Hasegawa, S., Ichiyama, T., Sonaka, I., Ohsaki, A., Hirano, R., Haneda, Y., et al. (2011). Amino acids exhibit antiinflammatory effects in human monocytic leukemia cell line, THP-1 cells. *Inflamm. Res.* 60, 1013–1019. doi: 10.1007/s00011-011-0362-1

He, F., Zhang, T., Xue, K., Fang, Z., Jiang, G., Huang, S., et al. (2021). Fecal multi-omics analysis reveals diverse molecular alterations of gut ecosystem in COVID-19 patients. *Anal. Chim. Acta* 1180:338881. doi: 10.1016/j.aca.2021.338881

Hecker, M., Sessa, W. C., Harris, H. J., Anggård, E. E., and Vane, J. R. (1990). The metabolism of L-arginine and its significance for the biosynthesis of endothelium-derived relaxing factor: cultured endothelial cells recycle L-citrulline to L-arginine. *Proc. Natl. Acad. Sci. U. S. A.* 87, 8612–8616. doi: 10.1073/pnas.87.21.8612

Hedin, C. R., McCarthy, N. E., Louis, P., Farquharson, F. M., McCartney, S., Taylor, K., et al. (2014). Altered intestinal microbiota and blood T cell phenotype are shared by patients with Crohn's disease and their unaffected siblings. *Gut* 63, 1578–1586. doi: 10.1136/gutjnl-2013-306226

Heidarian, F., Alebouyeh, M., Shahrokh, S., Balaii, H., and Zali, M. R. (2019). Altered fecal bacterial composition correlates with disease activity in inflammatory bowel disease and the extent of IL8 induction. *Curr. Res. Transl. Med.* 67, 41–50. doi: 10.1016/j.retram.2019.01.002

Hisamatsu, T., Okamoto, S., Hashimoto, M., Muramatsu, T., Andou, A., Uo, M., et al. (2012). Novel, objective, multivariate biomarkers composed of plasma amino acid profiles for the diagnosis and assessment of inflammatory bowel disease. *PLoS One* 7:e31131. doi: 10.1371/journal.pone.0031131

Hisamatsu, T., Ono, N., Imaizumi, A., Mori, M., Suzuki, H., Uo, M., et al. (2015). Decreased plasma histidine level predicts risk of relapse in patients with ulcerative colitis in remission. *PLoS One* 10:e0140716. doi: 10.1371/journal.pone.0140716

Hoarau, G., Mukherjee, P. K., Gower-Rousseau, C., Hager, C., Chandra, J., Retuerto, M. A., et al. (2016). Bacteriome and Mycobiome interactions underscore microbial dysbiosis in familial Crohn's disease. *MBio* 7, e01250–e01216. doi: 10.1128/mBio.01250-16

Holeček, M. (2020). Histidine in health and disease: metabolism, physiological importance, and use as a supplement. *Nutrients* 12:848. doi: 10.3390/nu12030848

Hoxha, M. (2020). What about COVID-19 and arachidonic acid pathway? *Eur. J. Clin. Pharmacol.* 76, 1501–1504. doi: 10.1007/s00228-020-02941-w

Hu, B., Huang, S., and Yin, L. (2021). The cytokine storm and COVID-19. *J. Med. Virol.* 93, 250–256. doi: 10.1002/jmv.26232

Huang, H., Vangay, P., McKinlay, C. E., and Knights, D. (2014). Multi-omics analysis of inflammatory bowel disease. *Immunol. Lett.* 162, 62–68. doi: 10.1016/j.imlet.2014.07.014

Hussain, I., Cher, G. L. Y., Abid, M. A., and Abid, M. B. (2021). Role of gut microbiome in COVID-19: An insight into pathogenesis and therapeutic potential. *Front. Immunol.* 12:765965. doi: 10.3389/fimmu.2021.765965

Irving, P. M., de Lusignan, S., Tang, D., Nijher, M., and Barrett, K. (2021). Risk of common infections in people with inflammatory bowel disease in primary care: a population-based cohort study. *BMJ Open Gastroenterol.* 8:e000573. doi: 10.1136/bmjgast-2020-000573

Jacobs, J. P., Goudarzi, M., Singh, N., Tong, M., McHardy, I. H., Ruegger, P., et al. (2016). A disease-associated microbial and metabolomics state in relatives of pediatric inflammatory bowel disease patients. *Cell. Mol. Gastroenterol. Hepatol.* 2, 750–766. doi: 10.1016/j.jcmgh.2016.06.004

Jagt, J. Z., Struys, E. A., Ayada, I., Bakkali, A., Jansen, E. E. W., Claesen, J., et al. (2021). Fecal amino acid analysis in newly diagnosed pediatric inflammatory bowel disease: a multicenter case-control study. *Inflamm. Bowel Dis.* izab256. doi: 10.1093/ibd/izab256 [Epub ahead of print].

Jain, U. (2020). Effect of COVID-19 on the organs. *Cureus* 12:e9540. doi: 10.7759/cureus.9540

Jansson, J., Willing, B., Lucio, M., Fekete, A., Dicksved, J., Halfvarson, J., et al. (2009). Metabolomics reveals metabolic biomarkers of Crohn's disease. *PLoS One* 4:e6386. doi: 10.1371/journal.pone.0006386

Joossens, M., Huys, G., Cnockaert, M., De Preter, V., Verbeke, K., Rutgeerts, P., et al. (2011). Dysbiosis of the faecal microbiota in patients with Crohn's disease and their unaffected relatives. *Gut* 60, 631–637. doi: 10.1136/gut.2010.223263

Kabeerdoss, J., Jayakanthan, P., Pugazhendhi, S., and Ramakrishna, B. S. (2015). Alterations of mucosal microbiota in the colon of patients with inflammatory bowel disease revealed by real time polymerase chain reaction amplification of 16S ribosomal ribonucleic acid. *Indian J. Med. Res.* 142, 23–32. doi: 10.4103/0971-5916.162091

Kabeerdoss, J., Sankaran, V., Pugazhendhi, S., and Ramakrishna, B. S. (2013). Clostridium leptum group bacteria abundance and diversity in the fecal microbiota of patients with inflammatory bowel disease: a case-control study in India. *BMC Gastroenterol.* 13:20. doi: 10.1186/1471-230X-13-20

Kang, S., Denman, S. E., Morrison, M., Yu, Z., Dore, J., Leclerc, M., et al. (2010). Dysbiosis of fecal microbiota in Crohn's disease patients as revealed by a custom phylogenetic microarray. *Inflamm. Bowel Dis.* 16, 2034–2042. doi: 10.1002/ibd.21319

Kang, P., Kalloniatis, M., and Doig, G. S. (2021). Using updated PubMed: new features and functions to enhance literature searches. *JAMA* 326, 479–480. doi: 10.1001/jama.2021.12021

Keshteli, A. H., van den Brand, F. F., Madsen, K. L., Mandal, R., Valcheva, R., Kroeker, K. I., et al. (2017). Dietary and metabolomic determinants of relapse in ulcerative colitis patients: a pilot prospective cohort study. *World J. Gastroenterol.* 23, 3890–3899. doi: 10.3748/wjg.v23.i21.3890

Kesika, P., Sivamaruthi, B. S., Thangaleela, S., and Chaiyasut, C. (2021). The antiviral potential of probiotics—a review on scientific outcomes. *Appl. Sci.* 11:8687. doi: 10.3390/app11188687

Khan, I., Ullah, N., Zha, L., Bai, Y., Khan, A., Zhao, T., et al. (2019). Alteration of gut microbiota in inflammatory bowel disease (IBD): cause or consequence? IBD treatment targeting the gut microbiome. *Pathogens* 8:126. doi: 10.3390/pathogens8030126

Kim, M. H., and Kim, H. (2017). The roles of glutamine in the intestine and its implication in intestinal diseases. *Int. J. Mol. Sci.* 18:1051. doi: 10.3390/ijms18051051

Kimhofer, T., Lodge, S., Whiley, L., Gray, N., Loo, R. L., Lawler, N. G., et al. (2020). Integrative modeling of quantitative plasma lipoprotein, metabolic, and amino acid data reveals a multiorgan pathological signature of SARS-CoV-2 infection. *J. Proteome Res.* 19, 4442–4454. doi: 10.1021/acs.jproteome.0c00519

Knights, D., Silverberg, M. S., Weersma, R. K., Gevers, D., Dijkstra, G., Huang, H., et al. (2014). Complex host genetics influence the microbiome in inflammatory bowel disease. *Genome Med.* 6:107. doi: 10.1186/s13073-014-0107-1

Knyazev, E., Nersisyan, S., and Tonevitsky, A. (2021). Endocytosis and Transcytosis of SARS-CoV-2 across the intestinal epithelium and other tissue barriers. *Front. Immunol.* 12:636966. doi: 10.3389/fimmu.2021.636966

Kohashi, M., Nishiumi, S., Ooi, M., Yoshie, T., Matsubara, A., Suzuki, M., et al. (2014). A novel gas chromatography mass spectrometry-based serum diagnostic and assessment approach to ulcerative colitis. *J. Crohns Colitis* 8, 1010–1021. doi: 10.1016/j.crohns.2014.01.024

Kolho, K. L., Pessia, A., Jaakkola, T., de Vos, W. M., and Velagapudi, V. (2017). Faecal and serum metabolomics in paediatric inflammatory bowel disease. *J. Crohns Colitis* 11, 321–334. doi: 10.1093/ecco-jcc/jjw158

Kong, C., Yan, X., Liu, Y., Huang, L., Zhu, Y., He, J., et al. (2021). Ketogenic diet alleviates colitis by reduction of colonic group 3 innate lymphoid cells through altering gut microbiome. *Signal Transduct. Target. Ther.* 6:154. doi: 10.1038/s41392-021-00549-9

Kretzmann, N. A., Fillmann, H., Mauriz, J. L., Marroni, C. A., Marroni, N., González-Gallego, J., et al. (2008). Effects of glutamine on proinflammatory gene expression and activation of nuclear factor kappa B and signal transducers and activators of transcription in TNBS-induced colitis. *Inflamm. Bowel Dis.* 14, 1504–1513. doi: 10.1002/ibd.20543

Kumari, R., Ahuja, V., and Paul, J. (2013). Fluctuations in butyrate-producing bacteria in ulcerative colitis patients of North India. *World J. Gastroenterol.* 19, 3404–3414. doi: 10.3748/wjg.v19.i22.3404

Laffel, L. (1999). Ketone bodies: a review of physiology, pathophysiology and application of monitoring to diabetes. *Diabetes Metab. Res. Rev.* 15, 412–426. doi: 10.1002/(SICI)1520-7560(199911/12)15:6<412::AID-DMRR72>3.0.CO;2-8

Lai, Y., Xue, J., Liu, C. W., Gao, B., Chi, L., Tu, P., et al. (2019). Serum metabolomics identifies altered bioenergetics, signaling cascades in parallel with exposome markers in Crohn's disease. *Molecules* 24:449. doi: 10.3390/molecules24030449

Lamas, B., Natividad, J. M., and Sokol, H. (2018). Aryl hydrocarbon receptor and intestinal immunity. *Mucosal Immunol.* 11, 1024–1038. doi: 10.1038/s41385-018-0019-2

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

Lamers, M. M., Beumer, J., van der Vaart, J., Knoops, K., Puschhof, J., Breugem, T. I., et al. (2020). SARS-CoV-2 productively infects human gut enterocytes. *Science* 369, 50–54. doi: 10.1126/science.abc1669

Laserna-Mendieta, E. J., Clooney, A. G., Carretero-Gomez, J. F., Moran, C., Sheehan, D., Nolan, J. A., et al. (2018). Determinants of reduced genetic capacity for butyrate synthesis by the gut microbiome in Crohn's disease and ulcerative colitis. *J. Crohns Colitis* 12, 204–216. doi: 10.1093/ecco-jcc/jjx137

Lawler, N. G., Gray, N., Kimhofer, T., Boughton, B., Gay, M., Yang, R., et al. (2021). Systemic perturbations in amine and kynurenine metabolism associated

with acute SARS-CoV-2 infection and inflammatory cytokine responses. *J. Proteome Res.* 20, 2796–2811. doi: 10.1021/acs.jproteome.1c00052

Lee, J. G., Han, D. S., Jo, S. V., Lee, A. R., Park, C. H., Eun, C. S., et al. (2019). Characteristics and pathogenic role of adherent-invasive Escherichia coli in inflammatory bowel disease: potential impact on clinical outcomes. *PLoS One* 14:e0216165. doi: 10.1371/journal.pone.0216165

Lehmann, M., Allers, K., Heldt, C., Meinhardt, J., Schmidt, F., Rodriguez-Sillke, Y., et al. (2021). Human small intestinal infection by SARS-CoV-2 is characterized by a mucosal infiltration with activated CD8+ T cells. *Mucosal Immunol.* 14, 1381–1392. doi: 10.1038/s41385-021-00437-z

Lepage, P., Seksik, P., Sutren, M., de la Cochetière, M. F., Jian, R., Marteau, P., et al. (2005). Biodiversity of the mucosa-associated microbiota is stable along the distal digestive tract in healthy individuals and patients with IBD. *Inflamm. Bowel Dis.* 11, 473–480. doi: 10.1097/01.MIB.0000159662.62651.06

Levy, E., Rizwan, Y., Thibault, L., Lepage, G., Brunet, S., Bouthillier, L., et al. (2000). Altered lipid profile, lipoprotein composition, and oxidant and antioxidant status in pediatric Crohn disease. *Am. J. Clin. Nutr.* 71, 807–815. doi: 10.1093/ajcn/71.3.807

Li, J., Wang, X., Chen, J., Zuo, X., Zhang, H., and Deng, A. (2020). COVID-19 infection may cause ketosis and ketoacidosis. *Diabetes Obes. Metab.* 22, 1935–1941. doi: 10.1111/dom.14057

Li, Q., Wang, C., Tang, C., Li, N., and Li, J. (2012). Molecular-phylogenetic characterization of the microbiota in ulcerated and non-ulcerated regions in the patients with Crohn's disease. *PLoS One* 7:e34939. doi: 10.1371/journal.pone.0034939

Liu, X., and Locasale, J. W. (2017). Metabolomics: A primer. *Trends Biochem. Sci.* 42, 274–284. doi: 10.1016/j.tibs.2017.01.004

Liu, X., Zhao, F., Liu, H., Xie, Y., Zhao, D., and Li, C. (2021). Transcriptomics and metabolomics reveal the adaption of Akkermansia muciniphila to high mucin by regulating energy homeostasis. *Sci. Rep.* 11:9073. doi: 10.1038/s41598-021-88397-z

Lloyd-Price, J., Arze, C., Ananthakrishnan, A. N., Schirmer, M., Avila-Pacheco, J., Poon, T. W., et al. (2019). Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. *Nature* 569, 655–662. doi: 10.1038/s41586-019-1237-9

Lo Presti, A., Zorzi, F., Del Chierico, F., Altomare, A., Cocca, S., Avola, A., et al. (2019). Fecal and mucosal microbiota profiling in irritable bowel syndrome and inflammatory bowel disease. *Front. Microbiol.* 10:1655. doi: 10.3389/fmicb.2019.01655

Łodyga, M., Maciejewska, K., Eder, P., Waszak, K., Stawczyk-Eder, K., Michalak, M., et al. (2021). Inflammatory bowel disease is associated with higher seroprevalence rates of antibodies against SARS-CoV-2. *Pol. Arch. Intern. Med.* 131, 226–232. doi: 10.20452/pamw.15796

Macfarlane, S., Furrie, E., Cummings, J. H., and Macfarlane, G. T. (2004). Chemotaxonomic analysis of bacterial populations colonizing the rectal mucosa in patients with ulcerative colitis. *Clin. Infect. Dis.* 38, 1690–1699. doi: 10.1086/420823

Machiels, K., Joossens, M., Sabino, J., De Preter, V., Arijs, I., Eeckhaut, V., et al. (2014). A decrease of the butyrate-producing species *Roseburia hominis* and *Faecalibacterium* prausnitzii defines dysbiosis in patients with ulcerative colitis. *Gut* 63, 1275–1283. doi: 10.1136/gutjnl-2013-304833

Macias-Ceja, D. C., Ortiz-Masiá, D., Salvador, P., Gisbert-Ferrándiz, L., Hernández, C., Hausmann, M., et al. (2019). Succinate receptor mediates intestinal inflammation and fibrosis. *Mucosal Immunol.* 12, 178–187. doi: 10.1038/s41385-018-0087-3

Magalhães, N. S., Savino, W., Silva, P. M. R., Martins, M. A., and Carvalho, V. F. (2021). Gut microbiota dysbiosis is a crucial player for the poor outcomes for COVID-19 in elderly, diabetic and hypertensive patients. *Front. Med.* 8:644751. doi: 10.3389/fmed.2021.644751

Majmundar, A. J., Wong, W. J., and Simon, M. C. (2010). Hypoxia-inducible factors and the response to hypoxic stress. *Mol. Cell* 40, 294–309. doi: 10.1016/j.molcel.2010.09.022

Manfredi, M., Conte, E., Barberis, E., Buzzi, A., Robotti, E., Caneparo, V., et al. (2019). Integrated serum proteins and fatty acids analysis for putative biomarker discovery in inflammatory bowel disease. *J. Proteome* 195, 138–149. doi: 10.1016/j.jprot.2018.10.017

Manichanh, C., Rigottier-Gois, L., Bonnaud, E., Gloux, K., Pelletier, E., Frangeul, L., et al. (2006). Reduced diversity of faecal microbiota in Crohn's disease revealed by a metagenomic approach. *Gut* 55, 205–211. doi: 10.1136/gut.2005.073817

Mar, J. S., LaMere, B. J., Lin, D. L., Levan, S., Nazareth, M., Mahadevan, U., et al. (2016). Disease severity and immune activity relate to distinct Interkingdom gut microbiome states in ethnically distinct ulcerative colitis patients. *MBio* 7, e01072–e01016. doi: 10.1128/mBio.01072-16

Marchesi, J. R., Holmes, E., Khan, F., Kochhar, S., Scanlan, P., Shanahan, F., et al. (2007). Rapid and noninvasive metabonomic characterization of inflammatory bowel disease. *J. Proteome Res.* 6, 546–551. doi: 10.1021/pr060470d

Marshall, M. (2021). COVID and the brain: researchers zero in on how damage occurs. *Nature* 595, 484–485. doi: 10.1038/d41586-021-01693-6

Martin, H. M., Campbell, B. J., Hart, C. A., Mpofu, C., Nayar, M., Singh, R., et al. (2004). Enhanced *Escherichia coli* adherence and invasion in Crohn's disease and colon cancer. *Gastroenterology* 127, 80–93. doi: 10.1053/j.gastro.2004.03.054

Matés, J. M., Pérez-Gómez, C., Núñez de Castro, I., Asenjo, M., and Márquez, J. (2002). Glutamine and its relationship with intracellular redox status, oxidative stress and cell proliferation/death. *Int. J. Biochem. Cell Biol.* 34, 439–458. doi: 10.1016/S1357-2725(01)00143-1

Matthews, D. E. (2007). An overview of phenylalanine and tyrosine kinetics in humans. *J. Nutr.* 137, 1549S–1555S; discussion 1573S-1575S. doi: 10.1093/jn/137.6.1549S

Mazza, S., Sorce, A., Peyvandi, F., Vecchi, M., and Caprioli, F. (2020). A fatal case of COVID-19 pneumonia occurring in a patient with severe acute ulcerative colitis. *Gut* 69, 1148–1149. doi: 10.1136/gutjnl-2020-321183

Mei, Q., Wang, A. Y., Bryant, A., Yang, Y., Li, M., Wang, F., et al. (2021). Survival factors and metabolic pathogenesis in elderly patients (≥65) with COVID-19: A multi-center study. *Front. Med.* 7:595503. doi: 10.3389/fmed.2020.595503

Meoni, G., Ghini, V., Maggi, L., Vignoli, A., Mazzoni, A., Salvati, L., et al. (2021). Metabolomic/lipidomic profiling of COVID-19 and individual response to tocilizumab. *PLoS Pathog.* 17:e1009243. doi: 10.1371/journal.ppat.1009243

Mesnage, R., and Antoniou, M. N. (2020). Computational modelling provides insight into the effects of glyphosate on the shikimate pathway in the human gut microbiome. *Curr. Res. Toxicol.* 1, 25–33. doi: 10.1016/j.crtox.2020.04.001

Michail, S., Durbin, M., Turner, D., Griffiths, A. M., Mack, D. R., Hyams, J., et al. (2012). Alterations in the gut microbiome of children with severe ulcerative colitis. *Inflamm. Bowel Dis.* 18, 1799–1808. doi: 10.1002/ibd.22860

Mills, E., and O'Neill, L. A. (2014). Succinate: a metabolic signal in inflammation. *Trends Cell Biol.* 24, 313–320. doi: 10.1016/j.tcb.2013.11.008

Mondot, S., Kang, S., Furet, J. P., Aguirre de Carcer, D., McSweeney, C., Morrison, M., et al. (2011, 2011). Highlighting new phylogenetic specificities of Crohn's disease microbiota. *Inflamm. Bowel Dis.* 17, 185–192. doi: 10.1002/ibd.21436

Monteleone, G., and Ardizzone, S. (2020). Are patients with inflammatory bowel disease at increased risk for Covid-19 infection? *J. Crohns Colitis* 14, 1334–1336. doi: 10.1093/ecco-jcc/jjaa061

Monteleone, I., Rizzo, A., Sarra, M., Sica, G., Sileri, P., Biancone, L., et al. (2011). Aryl hydrocarbon receptor-induced signals up-regulate IL-22 production and inhibit inflammation in the gastrointestinal tract. *Gastroenterology* 141, 237–248.e1. doi: 10.1053/j.gastro.2011.04.007

Morais, L. H., Schreiber, H. L., and Mazmanian, S. K. (2021). The gut microbiota-brain axis in behaviour and brain disorders. *Nat. Rev. Microbiol.* 19, 241–255. doi: 10.1038/s41579-020-00460-0

Morgan, X. C., Tickle, T. L., Sokol, H., Gevers, D., Devaney, K. L., Ward, D. V., et al. (2012). Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. *Genome Biol.* 13:R79. doi: 10.1186/gb-2012-13-9-r79

Moro, J., Tomé, D., Schmidely, P., Demersay, T. C., and Azzout-Marniche, D. (2020). Histidine: a systematic review on metabolism and physiological effects in human and different animal species. *Nutrients* 12:1414. doi: 10.3390/nu12051414

Morone, G., Palomba, A., Iosa, M., Caporaso, T., De Angelis, D., Venturiero, V., et al. (2020). Incidence and persistence of viral shedding in COVID-19 post-acute patients With Negativized pharyngeal swab: a systematic review. *Front. Med.* 7:562. doi: 10.3389/fmed.2020.00562

Müller, C., Hardt, M., Schwudke, D., Neuman, B. W., Pleschka, S., and Ziebuhr, J. (2018). Inhibition of cytosolic phospholipase A2α impairs an early step of

coronavirus replication in cell culture. *J. Virol.* 92, e01463–e01417. doi: 10.1128/JVI.01463-17

Murakami, M., Sato, H., and Taketomi, Y. (2020). Updating phospholipase A2 biology. *Biomol. Ther.* 10:1457. doi: 10.3390/biom10101457

Mussap, M., and Fanos, V. (2021). Could metabolomics drive the fate of COVID-19 pandemic? A narrative review on lights and shadows. *Clin. Chem. Lab. Med.* 59, 1891–1905. doi: 10.1515/cclm-2021-0414

Nadalian, B., Yadegar, A., Houri, H., Olfatifar, M., Shahrokh, S., Asadzadeh Aghdaei, H., et al. (2021). Prevalence of the pathobiont adherent-invasive Escherichia coli and inflammatory bowel disease: a systematic review and meta-analysis. *J. Gastroenterol. Hepatol.* 36, 852–863. doi: 10.1111/jgh.15260

Naftali, T., Reshef, L., Kovacs, A., Porat, R., Amir, I., Konikoff, F. M., et al. (2016). Distinct microbiotas are associated with ileum-restricted and colon-involving Crohn's disease. *Inflamm. Bowel Dis.* 22, 293–302. doi: 10.1097/MIB.0000000000000662

Nash, J. H., Villegas, A., Kropinski, A. M., Aguilar-Valenzuela, R., Konczy, P., Mascarenhas, M., et al. (2010). Genome sequence of adherent-invasive *Escherichia coli* and comparative genomic analysis with other *E. coli* pathotypes. *BMC Genomics* 11:667. doi: 10.1186/1471-2164-11-667

Nemoto, H., Kataoka, K., Ishikawa, H., Ikata, K., Arimochi, H., Iwasaki, T., et al. (2012). Reduced diversity and imbalance of fecal microbiota in patients with ulcerative colitis. *Dig. Dis. Sci.* 57, 2955–2964. doi: 10.1007/s10620-012-2236-y

Neurath, M. F. (2019). Targeting immune cell circuits and trafficking in inflammatory bowel disease. *Nat. Immunol.* 20, 970–979. doi: 10.1038/s41590-019-0415-0

Neurath, M. F. (2020). COVID-19 and immunomodulation in IBD. *Gut* 69, 1335–1342. doi: 10.1136/gutjnl-2020-321269

Newman, J. C., and Verdin, E. (2014). Ketone bodies as signaling metabolites. *Trends Endocrinol. Metab.* 25, 42–52. doi: 10.1016/j.tem.2013.09.002

Ng, S. C., Shi, H. Y., Hamidi, N., Underwood, F. E., Tang, W., Benchimol, E. I., et al. (2017). Worldwide incidence and prevalence of inflammatory bowel disease in the 21st century: a systematic review of population-based studies. *Lancet* 390, 2769–2778. doi: 10.1016/S0140-6736(17)32448-0

Nikolaus, S., Schulte, B., Al-Massad, N., Thieme, F., Schulte, D. M., Bethge, J., et al. (2017). Increased tryptophan metabolism is associated with activity of inflammatory bowel diseases. *Gastroenterology* 153, 1504–516.e2. doi: 10.1053/j.gastro.2017.08.028

Nishino, K., Nishida, A., Inoue, R., Kawada, Y., Ohno, M., Sakai, S., et al. (2018). Analysis of endoscopic brush samples identified mucosa-associated dysbiosis in inflammatory bowel disease. *J. Gastroenterol.* 53, 95–106. doi: 10.1007/s00535-017-1384-4

Norsa, L., Cosimo, P., Indriolo, A., Sansotta, N., D'Antiga, L., and Callegaro, A. (2020). Asymptomatic severe acute respiratory syndrome coronavirus 2 infection in patients with inflammatory bowel disease under biologic treatment. *Gastroenterology* 159, 2229–31.e2. doi: 10.1053/j.gastro.2020.08.046

Oladimeji, P. O., and Chen, T. (2018). PXR: more than just a master xenobiotic receptor. *Mol. Pharmacol.* 93, 119–127. doi: 10.1124/mol.117.110155

Ooi, M., Nishiumi, S., Yoshie, T., Shiomi, Y., Kohashi, M., Fukunaga, K., et al. (2011). GC/MS-based profiling of amino acids and TCA cycle-related molecules in ulcerative colitis. *Inflamm. Res.* 60, 831–840. doi: 10.1007/s00011-011-0340-7

Ott, S. J., Musfeldt, M., Wenderoth, D. F., Hampe, J., Brant, O., Fölsch, U. R., et al. (2004). Reduction in diversity of the colonic mucosa associated bacterial microflora in patients with active inflammatory bowel disease. *Gut* 53, 685–693. doi: 10.1136/gut.2003.025403

Ottman, N., Reunanen, J., Meijerink, M., Pietilä, T. E., Kainulainen, V., Klievink, J., et al. (2017). Pili-like proteins of Akkermansia muciniphila modulate host immune responses and gut barrier function. *PLoS One* 12:e0173004. doi: 10.1371/journal.pone.0173004

Páez-Franco, J. C., Torres-Ruiz, J., Sosa-Hernández, V. A., Cervantes-Díaz, R., Romero-Ramírez, S., Pérez-Fragoso, A., et al. (2021). Metabolomics analysis reveals a modified amino acid metabolism that correlates with altered oxygen homeostasis in COVID-19 patients. *Sci. Rep.* 11:6350. doi: 10.1038/s41598-021-85788-0

Palmela, C., Chevarin, C., Xu, Z., Torres, J., Sevrin, G., Hirten, R., et al. (2018). Adherent-invasive *Escherichia coli* in inflammatory bowel disease. *Gut* 67, 574–587. doi: 10.1136/gutjnl-2017-314903

Park, A., Kim, S., Jung, I. H., and Byun, J. H. (2020). An immune therapy model for effective treatment on inflammatory bowel disease. *PLoS One* 15:e0238918. doi: 10.1371/journal.pone.0238918

Pascal, V., Pozuelo, M., Borruel, N., Casellas, F., Campos, D., Santiago, A., et al. (2017). A microbial signature for Crohn's disease. *Gut* 66, 813–822. doi: 10.1136/gutjnl-2016-313235

Perna, S., Alalwan, T. A., Alaali, Z., Alnashaba, T., Gasparri, C., Infantino, V., et al. (2019). The role of glutamine in the complex interaction between gut microbiota and health: a narrative review. *Int. J. Mol. Sci.* 20:5232. doi: 10.3390/ijms20205232

Prideaux, L., Kang, S., Wagner, J., Buckley, M., Mahar, J. E., De Cruz, P., et al. (2013). Impact of ethnicity, geography, and disease on the microbiota in health and inflammatory bowel disease. *Inflamm. Bowel Dis.* 19, 2906–2918. doi: 10.1097/01.MIB.0000435759.05577.12

Probert, F., Walsh, A., Jagielowicz, M., Yeo, T., Claridge, T. D. W., Simmons, A., et al. (2018). Plasma nuclear magnetic resonance metabolomics discriminates between high and low endoscopic activity and predicts progression in a prospective cohort of patients with ulcerative colitis. *J. Crohns Colitis* 12, 1326–1337. doi: 10.1093/ecco-jcc/jjy101

Qi, F., Qian, S., Zhang, S., and Zhang, Z. (2020). Single cell RNA sequencing of 13 human tissues identify cell types and receptors of human coronaviruses. *Biochem. Biophys. Res. Commun.* 526, 135–140. doi: 10.1016/j.bbrc.2020.03.044

Qian, Q., Fan, L., Liu, W., Li, J., Yue, J., Wang, M., et al. (2021). Direct evidence of active SARS-CoV-2 replication in the intestine. *Clin. Infect. Dis.* 73, 361–366. doi: 10.1093/cid/ciaa925

Qiu, J., Heller, J. J., Guo, X., Chen, Z. M., Fish, K., Fu, Y. X., et al. (2012). The aryl hydrocarbon receptor regulates gut immunity through modulation of innate lymphoid cells. *Immunity* 36, 92–104. doi: 10.1016/j.immuni.2011.11.011

Rajilić-Stojanović, M., Shanahan, F., Guarner, F., and de Vos, W. M. (2013). Phylogenetic analysis of dysbiosis in ulcerative colitis during remission. *Inflamm. Bowel Dis.* 19, 481–488. doi: 10.1097/MIB.0b013e31827fec6d

Rausch, P., Rehman, A., Künzel, S., Häsler, R., Ott, S. J., Schreiber, S., et al. (2011). Colonic mucosa-associated microbiota is influenced by an interaction of Crohn disease and FUT2 (secretor) genotype. *Proc. Natl. Acad. Sci. U. S. A.* 108, 19030–19035. doi: 10.1073/pnas.1106408108

Rehman, A., Lepage, P., Nolte, A., Hellmig, S., Schreiber, S., and Ott, S. J. (2010). Transcriptional activity of the dominant gut mucosal microbiota in chronic inflammatory bowel disease patients. *J. Med. Microbiol.* 59, 1114–1122. doi: 10.1099/jmm.0.021170-0

Reigstad, C. S., Salmonson, C. E., Rainey, J. F. 3rd, Szurszewski, J. H., Linden, D. R., Sonnenburg, J. L., et al. (2015). Gut microbes promote colonic serotonin production through an effect of short-chain fatty acids on enterochromaffin cells. *FASEB J.* 29, 1395–1403. doi: 10.1096/fj.14-259598

Rhoads, J. M., Argenzio, R. A., Chen, W., Rippe, R. A., Westwick, J. K., Cox, A. D., et al. (1997). L-glutamine stimulates intestinal cell proliferation and activates mitogen-activated protein kinases. *Am. J. Phys.* 272, G943–G953. doi: 10.1152/ajpgi.1997.272.5.G943

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

Ryan, F. J., Ahern, A. M., Fitzgerald, R. S., Laserna-Mendieta, E. J., Power, E. M., Clooney, A. G., et al. (2020). Colonic microbiota is associated with inflammation and host epigenomic alterations in inflammatory bowel disease. *Nat. Commun.* 11:1512. doi: 10.1038/s41467-020-15342-5

San Juan, I., Bruzzone, C., Bizkarguenaga, M., Bernardo-Seisdedos, G., Laín, A., Gil-Redondo, R., et al. (2020). Abnormal concentration of porphyrins in serum from COVID-19 patients. *Br. J. Haematol.* 190, e265–e267. doi: 10.1111/bjh.17060

Santoru, M. L., Piras, C., Murgia, A., Palmas, V., Camboni, T., Liggi, S., et al. (2017). A. Cross sectional evaluation of the gut-microbiome metabolome axis in an Italian cohort of IBD patients. *Sci. Rep.* 7:9523. doi: 10.1038/s41598-017-10034-5

Sappati Biyyani, R. S., Putka, B. S., and Mullen, K. D. (2010). Dyslipidemia and lipoprotein profiles in patients with inflammatory bowel disease. *J. Clin. Lipidol.* 4, 478–482. doi: 10.1016/j.jacl.2010.08.021

Satsangi, J., Silverberg, M. S., Vermeire, S., and Colombel, J. F. (2006). The Montreal classification of inflammatory bowel disease: controversies, consensus, and implications. *Gut* 55, 749–753. doi: 10.1136/gut.2005.082909

Savitz, J. (2020). The kynurenine pathway: a finger in every pie. *Mol. Psychiatry* 25, 131–147. doi: 10.1038/s41380-019-0414-4

Scanlan, P. D., Shanahan, F., O'Mahony, C., and Marchesi, J. R. (2006). Culture-independent analyses of temporal variation of the dominant fecal microbiota and targeted bacterial subgroups in Crohn's disease. *J. Clin. Microbiol.* 44, 3980–3988. doi: 10.1128/JCM.00312-06

Schicho, R., Shaykhutdinov, R., Ngo, J., Nazyrova, A., Schneider, C., Panaccione, R., et al. (2012). Quantitative metabolomic profiling of serum, plasma, and urine by (1)H NMR spectroscopy discriminates between patients with inflammatory bowel disease and healthy individuals. *J. Proteome Res.* 11, 3344–3357. doi: 10.1021/pr300139q

Scholl-Bürgi, S., Schroecksnadel, S., Jenny, M., Karall, D., and Fuchs, D. (2013). Chronic immune stimulation may cause moderate impairment of phenylalanine 4-hydroxylase. *Pteridines* 22, 120–125. doi: 10.1515/pteridines.2011.22.1.120

Schütze, S., Potthoff, K., Machleidt, T., Berkovic, D., Wiegmann, K., and Krönke, M. (1992). TNF activates NF-kappa B by phosphatidylcholine-specific phospholipase C-induced "acidic" sphingomyelin breakdown. *Cell* 71, 765–776. doi: 10.1016/0092-8674(92)90553-O

Schwarz, B., Sharma, L., Roberts, L., Peng, X., Bermejo, S., Leighton, I., et al. (2021). Cutting edge: severe SARS-CoV-2 infection in humans is defined by a shift in the serum lipidome, resulting in dysregulation of eicosanoid immune mediators. *J. Immunol.* 206, 329–334. doi: 10.4049/jimmunol.2001025

Schwiertz, A., Jacobi, M., Frick, J. S., Richter, M., Rusch, K., and Köhler, H. (2010). Microbiota in pediatric inflammatory bowel disease. *J. Pediatr.* 157, 240–244.e1. doi: 10.1016/j.jpeds.2010.02.046

Scoville, E. A., Allaman, M. M., Brown, C. T., Motley, A. K., Horst, S. N., Williams, C. S., et al. (2018). Alterations in lipid, amino acid, and energy metabolism distinguish Crohn's disease from ulcerative colitis and control subjects by serum metabolomic profiling. *Metabolomics* 14:17. doi: 10.1007/s11306-017-1311-y

Seksik, P., Rigottier-Gois, L., Gramet, G., Sutren, M., Pochart, P., Marteau, P., et al. (2003). Alterations of the dominant faecal bacterial groups in patients with Crohn's disease of the colon. *Gut* 52, 237–242. doi: 10.1136/gut.52.2.237

Sencio, V., Machado, M. G., and Trottein, F. (2021). The lung-gut axis during viral respiratory infections: the impact of gut dysbiosis on secondary disease outcomes. *Mucosal Immunol.* 14, 296–304. doi: 10.1038/s41385-020-00361-8

Serhan, C. N., Clish, C. B., Brannon, J., Colgan, S. P., Chiang, N., and Gronert, K. (2000). Novel functional sets of lipid-derived mediators with antiinflammatory actions generated from omega-3 fatty acids via cyclooxygenase 2-nonsteroidal antiinflammatory drugs and transcellular processing. *J. Exp. Med.* 192, 1197–1204. doi: 10.1084/jem.192.8.1197

Sha, S., Xu, B., Wang, X., Zhang, Y., Wang, H., Kong, X., et al. (2013). The biodiversity and composition of the dominant fecal microbiota in patients with inflammatory bowel disease. *Diagn. Microbiol. Infect. Dis.* 75, 245–251. doi: 10.1016/j.diagmicrobio.2012.11.022

Shah, E. D., Farida, J. P., Siegel, C. A., Chong, K., and Melmed, G. Y. (2017). Risk for overall infection with anti-TNF and antiintegrin agents used in IBD: A systematic review and meta-analysis. *Inflamm. Bowel Dis.* 23, 570–577. doi: 10.1097/MIB.0000000000001049

Shaheen, A. (2021). Can ketone bodies inactivate coronavirus spike protein? The potential of biocidal agents against SARS-CoV-2. *BioEssays* 43:e2000312. doi: 10.1002/bies.202000312

Shahir, N. M., Wang, J. R., Wolber, E. A., Schaner, M. S., Frank, D. N., Ir, D., et al. (2020). Crohn's disease differentially affects region-specific composition and aerotolerance profiles of mucosally adherent bacteria. *Inflamm. Bowel Dis.* 26, 1843–1855. doi: 10.1093/ibd/izaa103

Shen, B., Yi, X., Sun, Y., Bi, X., Du, J., Zhang, C., et al. (2020). Proteomic and metabolomic characterization of COVID-19 patient sera. *Cell* 182, 59–72.e15. doi: 10.1016/j.cell.2020.05.032

Shi, D., Yan, R., Lv, L., Jiang, H., Lu, Y., Sheng, J., et al. (2021). The serum metabolome of COVID-19 patients is distinctive and predictive. *Metabolism* 118:154739. doi: 10.1016/j.metabol.2021.154739

Shores, D. R., Binion, D. G., Freeman, B. A., and Baker, P. R. (2011). New insights into the role of fatty acids in the pathogenesis and resolution of inflammatory bowel disease. *Inflamm. Bowel Dis.* 17, 2192–2204. doi: 10.1002/ibd.21560

Sido, B., Seel, C., Hochlehnert, A., Breitkreutz, R., and Dröge, W. (2006). Low intestinal glutamine level and low glutaminase activity in Crohn's disease: a rational for glutamine supplementation? *Dig. Dis. Sci.* 51, 2170–2179. doi: 10.1007/s10620-006-9473-x

Sokol, H., Leducq, V., Aschard, H., Pham, H. P., Jegou, S., Landman, C., et al. (2017). Fungal microbiota dysbiosis in IBD. *Gut* 66, 1039–1048. doi: 10.1136/gutjnl-2015-310746

Son, D. O., Satsu, H., and Shimizu, M. (2005). Histidine inhibits oxidative stress- and TNF-alpha-induced interleukin-8 secretion in intestinal epithelial cells. *FEBS Lett.* 579, 4671–4677. doi: 10.1016/j.febslet.2005.07.038

Song, J. W., Lam, S. M., Fan, X., Cao, W. J., Wang, S. Y., Tian, H., et al. (2020). Omics-driven systems interrogation of metabolic dysregulation in COVID-19 pathogenesis. *Cell Metab.* 32, 188–202.e5. doi: 10.1016/j.cmet.2020.06.016

Sprenger, G. A. (2006). "Aromatic amino acids," in *Amino Acid Biosynthesis - Pathways, Regulation and Metabolic Engineering. Microbiology Monographs. Vol. 5.* ed. V. F. Wendisch (Berlin, Heidelberg: Springer), 93–127.

Stasi, C., Sadalla, S., and Milani, S. (2019). The relationship Between the serotonin metabolism, gut-microbiota and the gut-brain axis. *Curr. Drug Metab.* 20, 646–655. doi: 10.2174/1389200220666190725115503

Stephens, N. S., Siffledeen, J., Su, X., Murdoch, T. B., Fedorak, R. N., and Slupsky, C. M. (2013). Urinary NMR metabolomic profiles discriminate inflammatory bowel disease from healthy. *J. Crohns Colitis* 7, e42–e48. doi: 10.1016/j.crohns.2012.04.019

Stockinger, B., Di Meglio, P., Gialitakis, M., and Duarte, J. H. (2014). The aryl hydrocarbon receptor: multitasking in the immune system. *Annu. Rev. Immunol.* 32, 403–432. doi: 10.1146/annurev-immunol-032713-120245

Stubbs, B. J., Koutnik, A. P., Goldberg, E. L., Upadhyay, V., Turnbaugh, P. J., Verdin, E., et al. (2020). Investigating ketone bodies as immunometabolic countermeasures against respiratory viral infections. *Med* 1, 43–65. doi: 10.1016/j.medj.2020.06.008

Sugihara, K., Morhardt, T. L., and Kamada, N. (2019). The role of dietary nutrients in inflammatory bowel disease. *Front. Immunol.* 9:3183. doi: 10.3389/fimmu.2018.03183

Sultan, K., Mone, A., Durbin, L., Khuwaja, S., and Swaminath, A. (2020). Review of inflammatory bowel disease and COVID-19. *World J. Gastroenterol.* 26, 5534–5542. doi: 10.3748/wjg.v26.i37.5534

Swierkosz, T. A., Mitchell, J. A., Sessa, W. C., Hecker, M., and Vane, J. R. (1990). L-glutamine inhibits the release of endothelium-derived relaxing factor from the rabbit aorta. *Biochem. Biophys. Res. Commun.* 172, 143–148. doi: 10.1016/S0006-291X(05)80184-6

Takahashi, K., Nishida, A., Fujimoto, T., Fujii, M., Shioya, M., Imaeda, H., et al. (2016). Reduced abundance of butyrate-producing bacteria species in the fecal microbial Community in Crohn's disease. *Digestion* 93, 59–65. doi: 10.1159/000441768

Tallima, H., and El Ridi, R. (2017). Arachidonic acid: physiological roles and potential health benefits - a review. *J. Adv. Res.* 11, 33–41. doi: 10.1016/j.jare.2017.11.004

Tang, L., Gu, S., Gong, Y., Li, B., Lu, H., Li, Q., et al. (2020). Clinical significance of the correlation between changes in the major intestinal bacteria species and COVID-19 severity. *Engineering* 6, 1178–1184. doi: 10.1016/j.eng.2020.05.013

Tao, W., Zhang, G., Wang, X., Guo, M., Zeng, W., Xu, Z., et al. (2020). Analysis of the intestinal microbiota in COVID-19 patients and its correlation with the inflammatory factor IL-18. *Med. Microecol.* 5:100023. doi: 10.1016/j.medmic.2020.100023

Taxonera, C., Sagastagoitia, I., Alba, C., Mañas, N., Olivares, D., and Rey, E. (2020). 2019 novel coronavirus disease (COVID-19) in patients with inflammatory bowel diseases. *Aliment. Pharmacol. Ther.* 52, 276–283. doi: 10.1111/apt.15804

Tefas, C., Ciobanu, L., Tanțău, M., Moraru, C., and Socaciu, C. (2020). The potential of metabolic and lipid profiling in inflammatory bowel diseases: A pilot study. *Bosn. J. Basic Med. Sci.* 20, 262–270. doi: 10.17305/bjbms.2019.4235

Tefas, C., Socaciu, C., Moraru, C., and Tanțău, M. (2019). Lipidomic signatures of colonic inflammatory bowel diseases: a pilot study. *J. Gastrointestin. Liver Dis.* 28, 246–247. doi: 10.15403/jgld-188

Thomas, T., Stefanoni, D., Reisz, J. A., Nemkov, T., Bertolone, L., Francis, R. O., et al. (2020). COVID-19 infection alters kynurenine and fatty acid metabolism, correlating with IL-6 levels and renal status. *JCI Insight* 5:e140327. doi: 10.1172/jci.insight.140327

Tong, M., Li, X., Wegener Parfrey, L., Roth, B., Ippoliti, A., Wei, B., et al. (2013). A modular organization of the human intestinal mucosal microbiota and its association with inflammatory bowel disease. *PLoS One* 8:e80702. doi: 10.1371/journal.pone.0080702

Turnbull, J., Jha, R., Ortori, C. A., Lunt, E., Tighe, P. J., Irving, W. L., et al. (2022). Serum levels of pro-inflammatory lipid mediators and specialised pro-resolving molecules are increased in SARS-CoV-2 patients and correlate with markers of the adaptive immune response. *J. Infect. Dis.* jiab632 doi: 10.1093/infdis/jiab632, [Epub ahead of print]

van Passel, M. W., Kant, R., Zoetendal, E. G., Plugge, C. M., Derrien, M., Malfatti, S. A., et al. (2011). The genome of Akkermansia muciniphila, a dedicated intestinal mucin degrader, and its use in exploring intestinal metagenomes. *PLoS One* 6:e16876. doi: 10.1371/journal.pone.0016876

Varela, E., Manichanh, C., Gallart, M., Torrejón, A., Borruel, N., Casellas, F., et al. (2013). Colonisation by Faecalibacterium prausnitzii and maintenance of clinical remission in patients with ulcerative colitis. *Aliment. Pharmacol. Ther.* 38, 151–161. doi: 10.1111/apt.12365

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

Vigsnæs, L. K., Brynskov, J., Steenholdt, C., Wilcks, A., and Licht, T. R. (2012). Gram-negative bacteria account for main differences between faecal microbiota from patients with ulcerative colitis and healthy controls. *Benefic. Microbes* 3, 287–297. doi: 10.3920/BM2012.0018

Villapol, S. (2020). Gastrointestinal symptoms associated with COVID-19: impact on the gut microbiome. *Transl. Res.* 226, 57–69. doi: 10.1016/j.trsl.2020.08.004

Vrakas, S., Mountzouris, K. C., Michalopoulos, G., Karamanolis, G., Papatheodoridis, G., Tzathas, C., et al. (2017). Intestinal bacteria composition and translocation of bacteria in inflammatory bowel disease. *PLoS One* 12:e0170034. doi: 10.1371/journal.pone.0170034

Walczuk, U., Sobieszczańska, B., Turniak, M., Rzeszutko, M., Duda-Madej, A., and Iwańczak, B. (2019). The prevalence of mucosa-associated diffusely adherent *Escherichia coli* in children with inflammatory bowel disease. *Adv. Clin. Exp. Med.* 28, 899–905. doi: 10.17219/acem/94149

Walker, A. W., Sanderson, J. D., Churcher, C., Parkes, G. C., Hudspith, B. N., Rayment, N., et al. (2011). High-throughput clone library analysis of the mucosa-associated microbiota reveals dysbiosis and differences between inflamed and non-inflamed regions of the intestine in inflammatory bowel disease. *BMC Microbiol.* 11:7. doi: 10.1186/1471-2180-11-7

Wallace, J. L. (2019). Eicosanoids in the gastrointestinal tract. *Br. J. Pharmacol.* 176, 1000–1008. doi: 10.1111/bph.14178

Walters, S. S., Quiros, A., Rolston, M., Grishina, I., Li, J., Fenton, A., et al. (2014). Analysis of gut microbiome and diet modification in patients with Crohn's disease. *SOJ Microbiol. Infect. Dis.* 2, 1–13. doi: 10.15226/sojmid/2/3/00122

Wang, W., Chen, L., Zhou, R., Wang, X., Song, L., Huang, S., et al. (2014). Increased proportions of bifidobacterium and the lactobacillus group and loss of butyrate-producing bacteria in inflammatory bowel disease. *J. Clin. Microbiol.* 52, 398–406. doi: 10.1128/JCM.01500-13

Wang, A., Luan, H. H., and Medzhitov, R. (2019). An evolutionary perspective on immunometabolism. *Science* 363:eaar3932. doi: 10.1126/science.aar3932

Wang, B., Wu, Z., Ji, Y., Sun, K., Dai, Z., and Wu, G. (2016). L-glutamine enhances tight junction integrity by activating CaMK kinase 2-AMP-activated protein kinase signaling in intestinal porcine epithelial cells. *J. Nutr.* 146, 501–508. doi: 10.3945/jn.115.224857

Wei, S., Bahl, M. I., Baunwall, S. M. D., Hvas, C. L., and Licht, T. R. (2021). Determining gut microbial dysbiosis: a review of applied indexes for assessment of intestinal microbiota imbalances. *Appl. Environ. Microbiol.* 87, e00395–e00321. doi: 10.1128/AEM.00395-21

Whiley, L., Nye, L. C., Grant, I., Andreas, N., Chappell, K. E., Sarafian, M. H., et al. (2019). Ultrahigh-performance liquid chromatography tandem mass spectrometry with electrospray ionization quantification of tryptophan metabolites and markers of gut health in serum and plasma-application to clinical and epidemiology cohorts. *Anal. Chem.* 91, 5207–5216. doi: 10.1021/acs.analchem.8b05884

Williams, B. B., van Benschoten, A. H., Cimermancic, P., Donia, M. S., Zimmermann, M., Taketani, M., et al. (2014). Discovery and characterization of gut microbiota decarboxylases that can produce the neurotransmitter tryptamine. *Cell Host Microbe* 16, 495–503. doi: 10.1016/j.chom.2014.09.001

Willing, B. P., Dicksved, J., Halfvarson, J., Andersson, A. F., Lucio, M., Zheng, Z., et al. (2010). A pyrosequencing study in twins shows that gastrointestinal microbial profiles vary with inflammatory bowel disease phenotypes. *Gastroenterology* 139, 1844–1854.e1. doi: 10.1053/j.gastro.2010.08.049

Windmueller, H. G., and Spaeth, A. E. (1974). Uptake and metabolism of plasma glutamine by the small intestine. *J. Biol. Chem.* 249, 5070–5079. doi: 10.1016/S0021-9258(19)42329-6

Wu, Y., Cheng, X., Jiang, G., Tang, H., Ming, S., Tang, L., et al. (2021). Altered oral and gut microbiota and its association with SARS-CoV-2 viral load in COVID-19 patients during hospitalization. *NPJ Biofilms Microbiomes* 7:61. doi: 10.1038/s41522-021-00232-5

Wu, H., Gong, J., and Liu, Y. (2018). Indoleamine 2, 3-dioxygenase regulation of immune response (review). *Mol. Med. Rep.* 17, 4867–4873. doi: 10.3892/mmr.2018.8537

Wu, D., Shu, T., Yang, X., Song, J.-X., Zhang, M., Yao, C., et al. (2020). Plasma metabolomic and lipidomic alterations associated with COVID-19. *Natl. Sci. Rev.* 7, 1157–1168. doi: 10.1093/nsr/nwaa086

Xiao, F., Tang, M., Zheng, X., Liu, Y., Li, X., and Shan, H. (2020). Evidence for gastrointestinal infection of SARS-CoV-2. *Gastroenterology* 158, 1831–833.e3. doi: 10.1053/j.gastro.2020.02.055

Xu, K., Cai, H., Shen, Y., Ni, Q., Chen, Y., Hu, S., et al. (2020a). Management of corona virus disease-19 (COVID-19): the Zhejiang experience. *Zhejiang Da Xue Xue Bao Yi Xue Ban* 49, 147–157. doi: 10.3785/j.issn.1008-9292.2020.02.02

Xu, K., Cai, H., Shen, Y., Ni, Q., Chen, Y., Hu, S., et al. (2020b). Management of COVID-19: the Zhejiang experience. *Zhejiang Da Xue Xue Bao Yi Xue Ban* 49, 147–157. doi: 10.3785/j.issn.1008-9292.2020.02.02

Yamamoto, S., Saito, M., Tamura, A., Prawisuda, D., Mizutani, T., and Yotsuyanagi, H. (2021). The human microbiome and COVID-19: A systematic review. *PLoS One* 16:e0253293. doi: 10.1371/journal.pone.0253293

Yan, B., Chu, H., Yang, D., Sze, K. H., Lai, P. M., Yuan, S., et al. (2019). Characterization of the Lipidomic profile of human coronavirus-infected cells: implications for lipid metabolism remodeling upon coronavirus replication. *Viruses* 11:73. doi: 10.3390/v11010073

Yang, H., Shan, W., Zhu, F., Wu, J., and Wang, Q. (2019). Ketone bodies in neurological diseases: focus on neuroprotection and underlying mechanisms. *Front. Neurol.* 10:585. doi: 10.3389/fneur.2019.00585

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

Yau, Y. Y., Leong, R. W., Shin, S., Bustamante, S., Pickford, R., Hejazi, L., et al. (2014). Bimodal plasma metabolomics strategy identifies novel inflammatory metabolites in inflammatory bowel diseases. *Discov. Med.* 18, 113–124.

Yeoh, Y. K., Zuo, T., Lui, G. C., Zhang, F., Liu, Q., Li, A. Y., et al. (2021). Gut microbiota composition reflects disease severity and dysfunctional immune responses in patients with COVID-19. *Gut* 70, 698–706. doi: 10.1136/gutjnl-2020-323020

Yilmaz, B., Juillerat, P., Øyås, O., Ramon, C., Bravo, F. D., Franc, Y., et al. (2019). Microbial network disturbances in relapsing refractory Crohn's disease. *Nat. Med.* 25, 323–336. doi: 10.1038/s41591-018-0308-z

Yoo, H. C., Yu, Y. C., Sung, Y., and Han, J. M. (2020). Glutamine reliance in cell metabolism. *Exp. Mol. Med.* 52, 1496–1516. doi: 10.1038/s12276-020-00504-8

Youm, Y. H., Nguyen, K. Y., Grant, R. W., Goldberg, E. L., Bodogai, M., Kim, D., et al. (2015). The ketone metabolite b-hydroxybutyrate blocks NLRP3 inflammasome-mediated inflammatory disease. *Nat. Med.* 21, 263–269. doi: 10.1038/nm.3804

Zamboni, N., Saghatelian, A., and Patti, G. J. (2015). Defining the metabolome: size, flux, and regulation. *Mol. Cell* 58, 699–706. doi: 10.1016/j.molcel.2015.04.021

Zhang, J., Chen, S. L., and Li, L. B. (2017). Correlation between intestinal flora and serum inflammatory factors in patients with Crohn's disease. *Eur. Rev. Med. Pharmacol. Sci.* 21, 4913–4917.

Zhang, X., Gan, M., Li, J., Li, H., Su, M., Tan, D., et al. (2020). Endogenous Indole pyruvate pathway for tryptophan metabolism mediated by IL4I1. *J. Agric. Food Chem.* 68, 10678–10684. doi: 10.1021/acs.jafc.0c03735

Zhang, J., Hoedt, E. C., Liu, Q., Berendsen, E., The, J. J., Hamilton, A., et al. (2021). Elucidation of Proteus mirabilis as a key bacterium in Crohn's disease inflammation. *Gastroenterology* 160, 317–330.e11. doi: 10.1053/j.gastro.2020.09.036

Zhang, Y., Lin, L., Xu, Y., Lin, Y., Jin, Y., and Zheng, C. (2013). 1H NMR-based spectroscopy detects metabolic alterations in serum of patients with early-stage ulcerative colitis. *Biochem. Biophys. Res. Commun.* 433, 547–551. doi: 10.1016/j.bbrc.2013.03.012

Zhang, M., Liu, B., Zhang, Y., Wei, H., Lei, Y., and Zhao, L. (2007). Structural shifts of mucosa-associated lactobacilli and Clostridium leptum subgroup in patients with ulcerative colitis. *J. Clin. Microbiol.* 45, 496–500. doi: 10.1128/JCM.01720-06

Zheng, D., Liwinski, T., and Elinav, E. (2020). Interaction between microbiota and immunity in health and disease. *Cell Res.* 30, 492–506. doi: 10.1038/s41422-020-0332-7

Zucko, J., Dunlap, W. C., Shick, J. M., Cullum, J., Cercelet, F., Amin, B., et al. (2010). Global genome analysis of the shikimic acid pathway reveals greater gene loss in host-associated than in free-living bacteria. *BMC Genomics* 11:628. doi: 10.1186/1471-2164-11-628

Zuo, T., Liu, Q., Zhang, F., Lui, G. C., Tso, E. Y., Yeoh, Y. K., et al. (2021). Depicting SARS-CoV-2 faecal viral activity in association with gut microbiota composition in patients with COVID-19. *Gut* 70:gutjnl-2020-322294. doi: 10.1136/gutjnl-2020-322294

Zuo, T., and Ng, S. C. (2018). The gut microbiota in the pathogenesis and therapeutics of inflammatory bowel disease. *Front. Microbiol.* 9:2247. doi: 10.3389/fmicb.2018.02247

Zuo, T., Zhang, F., Lui, G. C. Y., Yeoh, Y. K., Li, A. Y. L., Zhan, H., et al. (2020). Alterations in gut microbiota of patients With COVID-19 during time of hospitalization. *Gastroenterology* 159, 944–55.e8. doi: 10.1053/j.gastro.2020.05.048

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

ORIGINAL RESEARCH
published: 21 March 2022
doi: 10.3389/fmicb.2022.826916

Digital Therapeutics Care Utilizing Genetic and Gut Microbiome Signals for the Management of Functional Gastrointestinal Disorders: Results From a Preliminary Retrospective Study

*Shreyas V. Kumbhare[1†], Patricia A. Francis-Lyon[1,2†], Dashyanng Kachru[1,2], Tejaswini Uday[1], Carmel Irudayanathan[1], Karthik M. Muthukumar[1], Roshni R. Ricchetti[1], Simitha Singh-Rambiritch[1], Juan Ugalde[3], Parambir S. Dulai[4], Daniel E. Almonacid[1] and Ranjan Sinha[1]**

[1] Digbi Health, Mountain View, CA, United States, [2] Health Informatics, University of San Francisco, San Francisco, CA, United States, [3] Universidad del Desarrollo, Facultad de Ingeniería, Centro de Investigación en Tecnologías para la Sociedad (C+), Santiago, Chile, [4] Division of Gastroenterology, University of California, San Diego, San Diego, CA, United States

OPEN ACCESS

Edited by:
Roshan Kumar,
Magadh University, India

Reviewed by:
Chandni Talwar,
University of Delhi, India
Charu Dogra Rawat,
University of Delhi, India

***Correspondence:**
Ranjan Sinha
ranjan@digbihealth.com

[†] These authors have contributed equally to this work

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

Received: 01 December 2021
Accepted: 28 January 2022
Published: 21 March 2022

Citation:
Kumbhare SV, Francis-Lyon PA, Kachru D, Uday T, Irudayanathan C, Muthukumar KM, Ricchetti RR, Singh-Rambiritch S, Ugalde J, Dulai PS, Almonacid DE and Sinha R (2022) Digital Therapeutics Care Utilizing Genetic and Gut Microbiome Signals for the Management of Functional Gastrointestinal Disorders: Results From a Preliminary Retrospective Study. Front. Microbiol. 13:826916. doi: 10.3389/fmicb.2022.826916

Diet and lifestyle-related illnesses including functional gastrointestinal disorders (FGIDs) and obesity are rapidly emerging health issues worldwide. Research has focused on addressing FGIDs via in-person cognitive-behavioral therapies, diet modulation and pharmaceutical intervention. Yet, there is paucity of research reporting on digital therapeutics care delivering weight loss and reduction of FGID symptom severity, and on modeling FGID status and symptom severity reduction including personalized genomic SNPs and gut microbiome signals. Our aim for this study was to assess how effective a digital therapeutics intervention personalized on genomic SNPs and gut microbiome signals was at reducing symptomatology of FGIDs on individuals that successfully lost body weight. We also aimed at modeling FGID status and FGID symptom severity reduction using demographics, genomic SNPs, and gut microbiome variables. This study sought to train a logistic regression model to differentiate the FGID status of subjects enrolled in a digital therapeutics care program using demographic, genetic, and baseline microbiome data. We also trained linear regression models to ascertain changes in FGID symptom severity of subjects at the time of achieving 5% or more of body weight loss compared to baseline. For this we utilized a cohort of 177 adults who reached 5% or more weight loss on the Digbi Health personalized digital care program, who were retrospectively surveyed about changes in symptom severity of their FGIDs and other comorbidities before and after the program. Gut microbiome taxa and demographics were the strongest predictors of FGID status. The digital therapeutics program implemented, reduced the summative severity of symptoms for 89.42% (93/104) of users who reported FGIDs. Reduction in summative FGID symptom severity and IBS symptom severity were best modeled by a mixture of genomic and microbiome predictors, whereas reduction in diarrhea and constipation symptom severity were best

modeled by microbiome predictors only. This preliminary retrospective study generated diagnostic models for FGID status as well as therapeutic models for reduction of FGID symptom severity. Moreover, these therapeutic models generate testable hypotheses for associations of a number of biomarkers in the prognosis of FGIDs symptomatology.

Keywords: multi-omic models, functional gastrointestinal disorders (FGIDs), IBS – irritable bowel syndrome, diarrhea, constipation, digital therapeutics, non-pharmacological treatment

INTRODUCTION

Background and Rationale

Diseases of the gastrointestinal tract affect 60–70 million people in the United States (Everhart, 2008), and the total expenditure in 2015 for these illnesses was 135.9 billion dollars – greater than for other common diseases and likely to continue increasing (Peery et al., 2019). GI afflictions and altered bowel habits affect the quality of life, social functioning and can result in considerable loss of productivity (Buono et al., 2017; Peery et al., 2019). Thus, early identification of markers to diagnose functional bowel disorders and markers to personalize therapeutic approaches may help reduce healthcare costs and productivity losses and improve quality of life.

According to the National Center for Health Statistics, 73.6% of the US population aged 20 and over was overweight or obese in 2017–2018 (Fryar et al., 2020). Overweight and obesity are known risk factors for the co-occurrence of functional gastrointestinal disorders (FGIDs), and obesity has been shown to negatively impact clinical outcomes of FGID treatment (Emerenziani et al., 2019). Dietary interventions can produce significant weight loss on overweight and obese subjects (Sinha et al., 2021), and at the same time can deliver reduction of FGID symptomatology (Manning and Biesiekierski, 2018).

There is paucity of data looking at demographics, genomic SNPs and microbiome factors that are associated with successfully reducing symptomatology of FGIDs by means of weight loss. Knowledge of these factors could help tailor interventions for these individuals. Thus, our aim for this study was to assess how effective a digital therapeutics intervention personalized on genomic SNPs and gut microbiome signals were at reducing symptomatology of FGIDs on individuals that successfully lost 5% or more body weight. Additionally, we aimed at building statistical diagnostic models to describe the impact of demographics, genomic SNPs, and gut microbiome predictors on the likelihood of a subject presenting or not FGIDs, and then statistical therapeutic models describing reduction of summative symptom severity [irritable bowel syndrome (IBS), diarrhea, constipation, bloating, gas, and abdominal pain] as well as reduction of symptom severity in subjects with IBS, diarrhea, or constipation, based on the same predictors.

Functional Gastrointestinal Disorders

Functional gastrointestinal disorders are conditions that present as normal upon examination of the GI system but still result in poor GI motility – primarily with symptoms in the middle or lower gastrointestinal tract (American College of Gastroenterology, 2021). These disorders include irritable bowel syndrome (IBS), bloating, constipation, diarrhea, gassiness, and dyspepsia, among others (American College of Gastroenterology, 2021; International Foundation for Gastrointestinal Disorders, 2021). Although the pathophysiology of FGIDs is often complex due to their multifactorial nature, they are frequently encountered in both primary care and gastroenterology settings. FGIDs are thought to encompass combinations of dysregulation of the gut-brain interaction, altered gut microbiota, altered mucosal and immune function, gastrointestinal tract motility disturbance, and visceral hypersensitivity (Black et al., 2020; Sperber et al., 2021), and women appear to be afflicted at least twice as frequently as men (Chang, 2004).

Of the six FGIDs studied here, irritable bowel syndrome (IBS) is the most frequent. It is prevalent in the range of 5–25% of the population and accounts for 36% of visits to a gastroenterologist (Chang, 2004; Buono et al., 2017). It is a gastrointestinal tract disease characterized by abdominal pain, discomfort, altered bowel habits, and affected quality of life, however, the absence of demonstrable organic disease makes IBS to not be considered life-threatening. IBS treatment often requires lifestyle changes and medication. The most common symptoms of IBS are chronic diarrhea – IBS-D (1/3 of IBS patients), chronic constipation – IBS-C, or both - IBS-M. Patients with IBS have a lower reported quality of life than sufferers of gastroesophageal reflux disease (GERD) and asthma (Chang, 2004).

Frequent and recurrent pain in the abdominal region not necessarily attributable to gut function is often referred to as Functional Abdominal Pain (FAP) (Thompson et al., 1999), the second FGID analyzed in this manuscript. Children with FAP were found to be significantly more likely to be obese (Galai et al., 2020). FAP symptoms are not as common as other FGIDs and are not necessarily associated with food intake or passage but rather with psychiatric disorders. Management can involve psychotherapy and pharmaceutical interventions encompassing psychiatric regimens like anti-depressants, anticonvulsants, and treatments for other psychiatric disorders (Clouse et al., 2006).

Functional bloating, the third FGID of focus in this work, on the other hand, is typically unlinked to psychiatric or organic causes (Hasler, 2006). Instead, it is a recurrent sensation of abdominal distention that may or may not be associated with measurable distention. Bloating is 2× more commonly reported in women than men and typically worsens after meals, and up to 96% of IBS patients report bloating (Longstreth et al., 2006; Sullivan, 2012).

Gassiness, the fourth FGID analyzed in this work, is the result of gas produced by bacterial fermentation of carbohydrates and proteins in the large intestine, resulting in changes to the gut microbiome, increased short-chain fatty acids, and

increased gas, diarrhea, abdominal pain, and bloating (Hasler, 2006). Although intestinal gas may contribute to bloating, bloating does not necessarily result from more gas. However, malabsorption of simple and complex carbohydrates and dietary fiber are commonly associated with both gas and bloating (Black et al., 2020). Management of gassiness often involves dietary modification, gut microbiome modulation, and lifestyle alteration (Hasler, 2006).

Functional constipation, the fifth FGID studied in our cohort, is a functional bowel disorder that does not meet IBS criteria, and presents as incomplete, infrequent or difficult defecation (Longstreth et al., 2006). Chronic constipation occurs in up to 11% of the population globally and affects all age groups (Sperber et al., 2021). Depending on specific tests, constipation may be defined as (1) straining, hard stools, unproductive movements, infrequent stools, or incomplete evacuation; (2) less than three bowel movements per week OR daily stool weight less than 35 g/day OR straining for more than 25% of the period; (3) lengthy whole-gut or colonic transit. Surprisingly, stool frequency appears to have just a slender relationship with colonic transit and there are usually no demonstrable physiological abnormalities (Longstreth et al., 2006).

Functional diarrhea (FD), the sixth and final FGID analyzed in this work, is a functional gastrointestinal disorder characterized by chronic or recurrent diarrhea not explained by structural or biochemical abnormalities of the gut. Functional diarrhea is characterized as the passage of loose or watery stools without abdominal pain or discomfort (Longstreth et al., 2006). Diarrhea is one of the most commonly reported symptoms for consulting a gastroenterologist, affecting up to 4.7% of the global population (Sperber et al., 2021). It is also a common presenting issue among many patients in general practice (Longstreth et al., 2006). Treatment of FD depends on establishing a correct diagnosis and needs to be distinguished from diarrhea-predominant irritable bowel syndrome (IBS-D) and other organic causes of chronic diarrhea. Once a physician has established the diagnosis, aggravating factors will need to be identified and eliminated, possibly including physiologic factors (e.g., small bowel bacterial overgrowth), psychological factors (e.g., stress and anxiety), and dietary factors (e.g., carbohydrate malabsorption) (Dellon and Ringel, 2006).

Interestingly, it has been hypothesized that FGIDs and obesity may be mechanistically linked (Ho and Spiegel, 2008) and thus that symptom severity for all FGIDs just reviewed may be reduced by means of weight loss through digital therapeutics interventions.

Association of Diet and Lifestyle With Functional Gastrointestinal Disorders

Diet is considered an important trigger of gut-related symptoms. Poor nutrition, for example, consumption of highly processed and "fast" foods, has been implicated in FGID etiology (Shau et al., 2016; Schnabel et al., 2018). Conversely, a Mediterranean diet has been associated with a lower prevalence of FGIDs (Agakidis et al., 2019). A dietary therapeutics approach, for example a low-FODMAP diet in which rapidly fermentable carbohydrates that are poorly absorbed by the gut are eliminated or avoided as much as possible, is a typical dietary protocol for IBS patients (Henström and D'Amato, 2016; Wilder-Smith et al., 2017). Low-FODMAP diets have been associated with relieving other FGID symptoms, although functional dyspepsia seems least responsive to such a regimen (Marsh et al., 2016; Basnayake et al., 2019).

Dietary fiber, which increases mucosal protein production, is widely used to treat chronic constipation (Derrien et al., 2010). This fiber is then digested in the colon, providing a substrate for microbial fermentation and resulting in byproducts such as Short Chain Fatty Acids (SCFAs), which have pro-motility effects and help with stool bulk and gas transit in the colon (Zhao and Yu, 2016). Moreover, a few reports demonstrate the association of low fiber diet and alterations in gut microbial communities, eventually leading to an increase in FGID symptom severity (Saffouri et al., 2019).

Non-celiac gluten sensitivity also links diet to FGIDs. Despite lacking serological and histological markers for celiac disease, a subset of FGID patients report significant alleviation of symptomatology upon elimination of dietary gluten and re-experience these symptoms upon reintroducing gluten (Elli et al., 2016).

Psychosocial factors have been implicated to varying degrees in FGID etiology, including sleep disturbance, dysfunctional coping, and psychiatric disorders (Longstreth et al., 2006). In particular, sleep disturbance has been demonstrated to be strongly associated with GERD, IBS and functional dyspepsia (Morito et al., 2014). Rotating shift work and poor sleep quality have also been identified as risk factors for IBS (Kim et al., 2013).

A randomized controlled trial by Johannesson et al. (2011) demonstrated that increased physical activity significantly reduced symptom severity in adult IBS patients, leading them to suggest that physical activity should be a primary treatment modality for IBS. Furthermore, a subset of the patients that continued increased physical activity 5 years later, continued to show improvements in IBS symptom severity, as well as in quality of life, fatigue, depression and anxiety in the long term (Johannesson et al., 2015). There is certainly a need for including diet and lifestyle changes in interventions to treat FGIDs.

Association of Biomarkers With Functional Gastrointestinal Disorders

Both gut microbiome and human genetics are likely contributors to FGID etiology (Henström and D'Amato, 2016; Wei et al., 2021). Evidence indicates a vital genetic component to FGIDs as demonstrated by prevalence within families and more substantial concordance between monozygotic versus dizygotic twins (Lembo et al., 2007; Saito et al., 2010). One evidence of a mutation in the SCN5A gene that regulates the sodium channel, linked to IBS, was replicated across two studies, and although the mutation is found in only approximately 2% of IBS patients, this finding indicates the influence genetics may have on IBS symptoms (Saito et al., 2009; Beyder et al.,

2014). Functional constipation has also been associated with specific genes (Locke et al., 2006). However, much of the research around genome-wide association studies and FGIDs is hampered by small sample sizes (Henström and D'Amato, 2016), have not been independently replicated, or are otherwise not robust. Yet, the study of the genetics of FGIDs is rapidly evolving. What seems clear so far is that most IBS sufferers either share several common gene variants that each nominally contribute toward the overall risk of the disease, or, for a subset of sufferers, a few highly penetrant alleles are likely the significant risk factors. Given that IBS spans both complex polygenic conditions and rare single-gene forms, it evidences the need for different strategies to identify these genetic factors.

The gut microbiome is also extensively implicated in FGID and particularly IBS pathogenesis (Agnello et al., 2020; Carco et al., 2020). Both IBS and functional dyspepsia have been shown to arise in susceptible individuals following a course of acute onset gastroenteritis (Simrén et al., 2013). Recent studies have revealed dysbiosis of the gut microbiota in constipated patients compared with healthy controls, associated with suppressed intestinal motility by metabolites produced by intestinal bacteria (Zhao and Yu, 2016; Ohkusa et al., 2019). Other research has elucidated several mechanisms playing important roles in IBS. A dysregulated gut-brain axis has been adopted as a suitable model for IBS, and poor gut microbiome diversity may contribute to the onset and exacerbation of IBS symptoms. Dietary fiber appears to influence the gut microbiota, encouraging the growth of beneficial probiotics while preventing pathogenic and obesogenic bacteria from overgrowing (Chen et al., 2013; Zhao and Yu, 2016). Although clinical trials, which have attempted to characterize the gut microbiota in IBS, do not yet allow for a causal role to be inferred, they do confirm alterations in both community stability and diversity (Kennedy, 2014). This evidence suggests that genomic SNPs and microbiome taxa or functions may help build statistical diagnostic models of the likelihood of a subject presenting or not FGIDs, and therapeutic models describing reduction of FGID symptom severity.

MATERIALS AND METHODS

Subject Enrollment

Subjects were recruited from those who achieved 5% or more body weight loss when enrolled in the Digbi Health personalized digital care program (see Intervention below). Only those subjects who had retrospectively responded to questions about symptomatology of their FGIDs and other comorbidities at the start of the program and after successful weight loss were included in the study (**Figure 1**). The average number of days for participants in the program at the time of the survey was 84.26 days. The final cohort studied in this manuscript included 177 subjects with either genome SNP (n = 169) or baseline gut biome (n = 168) or both analyses performed. Subjects were divided into two groups for comparisons and models: those who reported any of 6 FGIDs (IBS, diarrhea, constipation, bloating, gassiness, and cramping) at baseline or at the time of survey (n = 104), and those who reported no FGIDs (n = 73).

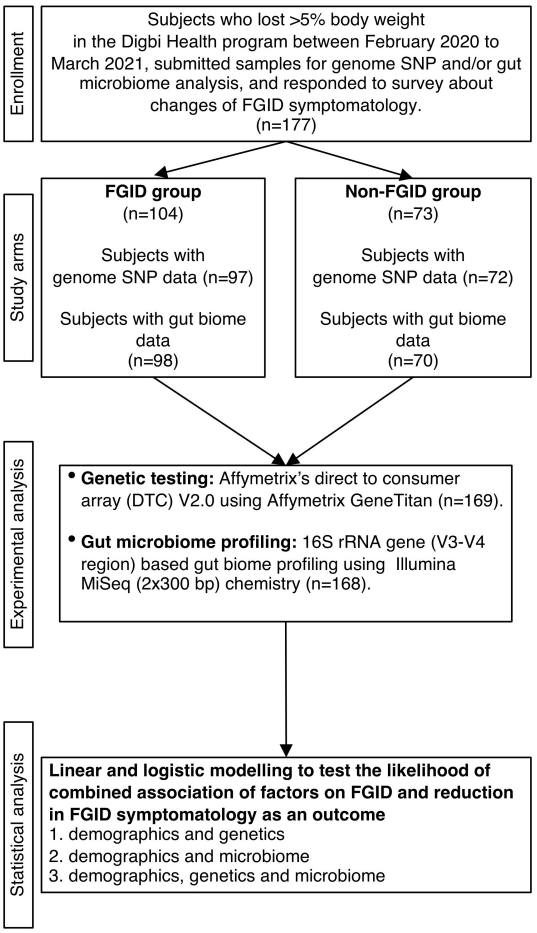


FIGURE 1 | Study design flowchart.

Intervention

Digbi Health is a next-generation, prescription-grade, digital therapeutics platform that analyzes genetics, gut bacteria, lifestyle, and demographics to build evidence-based individualized dietary and lifestyle plans using artificial intelligence (AI). It aims to help subjects reduce at least 5% of their baseline body weight and reduce weight-related inflammatory gut, musculoskeletal, cardiovascular, mental, and insulin-related comorbidities.

Upon enrollment, program participants were provided with online login access to the Digbi Health app and were asked to complete a health questionnaire. A bluetooth compatible digital weighing scale and buccal swab and stool sampling kits were shipped to all participants. The app was used to track subjects' weight, assess dietary intake (via uploaded photographs of food items consumed), and track wellness and lifestyle associated metrics such as sleep quality and quantity, exercise type and duration, stress and meditation, energy levels, cravings, and recommended foods consumed/avoided. Dietary intake was assessed by coaches who assigned a nutrient density score to meals based on their inflammatory, fiber diversity and expected insulin response.

Based on analysis of genetic and gut microbiome profiles, as well as lifestyle vitals, a Digbi Health Wellness Report was generated for subjects. The results were evaluated with the participants one-on-one by a health coach over the course of 4 months at pre-determined weekly and bi-weekly intervals. To achieve its goal, the program sought to nudge participants toward making incremental lifestyle changes focused on reducing sugar consumption and timing meals to optimize insulin sensitivity, reducing systemic inflammation by identifying possibly inflammatory and anti-inflammatory nutrients, and increasing fiber diversity to improve gut health. Most importantly, these behavioral changes were implemented with the help of virtual health coaching and the app to ensure that these changes are habit forming, i.e., long-term sustainable.

Sample Collection and Processing: Genome SNP Array and Gut Microbiome Profiling

Subjects self-collected buccal swab samples (Mawi Technologies iSwab DNA collection kit, Model no. iSWAB-DNA-1200) and fecal swab samples (Mawi Technologies iSWAB Microbiome collection kit, Model no. ISWAB-MBF-1200). Sample collection was completed by following standardized directions provided to all subjects in an instruction manual. DNA extraction, purification, and genotyping from buccal swab samples was performed using Affymetrix's Direct to Consumer Array version 2.0 ("DTC") on the Affymetrix GeneTitan platform at Akesogen Laboratories in Atlanta, GA, United States. Sample processing of baseline (pre-intervention) fecal samples was followed by 16S rRNA gene amplicon sequencing also performed at Akesogen Laboratories in Atlanta, GA, United States. DNA extraction was performed using Qiagen MagAttract Power Microbiome DNA Kit on an automated liquid handling DNA extraction instrument. The V3–V4 region of the 16S rRNA gene was amplified and sequenced on the Illumina MiSeq platform using 2 × 300 bp paired-end sequencing (Illumina, 2013). Sequence reads were demultiplexed, and Amplicon Sequence Variants (ASVs) generated using DADA2 in QIIME2 (version 2020.8) (Bolyen et al., 2019). We trimmed primers off the reads and low-quality bases ($Q < 30$). Taxonomic annotation was performed using the Naive Bayes classifier against the 99% non-redundant Silva database (Silva 138 Reference database, 2021). We excluded hits to Mitochondria, Chloroplast, Eukaryota, and unassigned taxa at the phylum level.

Statistical Analysis

Survey data from 177 respondents (with either genome SNP or gut biome data) over the course of their successful weight loss journey in the Digbi Health personalized digital care program were analyzed retrospectively. Variables for models included demographic (gender, age, weight, and BMI during gut microbiome sample collection and weight loss achieved during the program), genomic SNPs, and baseline gut microbiome data. We used the Wilcoxon sum rank/signed-rank test or Kendall correlation test, as appropriate, to assess: (a) the change in the severity of either of the FGID symptoms and a summative severity change and (b) the effect of variables on the change in severity. Significance results were adjusted for multiple comparisons using the false discovery rate (FDR) correction method for the microbiome data.

Genome SNP Related Statistics

The 16 SNP predictors, associated with lactose intolerance, gluten sensitivity, milk and peanut allergies, caffeine metabolism, and inflammatory markers (TNF and IL10), were from Digbi Health curated panels used for personalized interventions for subjects (**Supplementary Table 1**). Each SNP value was encoded as the number of risk alleles (0, 1, or 2) for each subject.

Microbiome Related Statistics

Bacterial genera abundances were analyzed for 168 survey respondents with baseline gut microbiome data available, using Qiime2 (Bolyen et al., 2019), qiime2R (Jordan E Bisanz, 2018), and phyloseq (McMurdie and Holmes, 2013). The following microbial features were filtered out from downstream analysis: (a) ASVs not classified at the phylum level, (b) phyla that had <25 ASVs (*Elusimicrobiota*, *Nanoarchaeota*, *Bdellovibrionota*, WPS-2), (c) uncultured and Incertae Sedis taxa, (d) genera that had <30 reads in at least 15% of samples, and (e) genera for which >25% of samples had zero read count. In total, 105 genera were kept in downstream analysis. The abundance of these bacterial genera was transformed to centered log-ratio (CLR) using the zCompositions package (Palarea-Albaladejo and Martín-Fernández, 2015) after first replacing zeros with pseudo counts based on a Bayesian-multiplicative replacement from the zCompositions package (Fernandes et al., 2014). Permutational multivariate analysis of variance (PERMANOVA) was performed on the gut microbiome Aitchison distance matrix, using gender, BMI (closest to date of gut biome sampling), and FGID status as variables using the CLR transformed abundances. Additionally, the read counts after zero replacement were transformed using additive log-ratio (ALR) to be utilized in downstream model statistics (see below) (Friedman et al., 2010; Gloor et al., 2017; Quinn et al., 2019).

Model Statistics

Linear and logistic regression models were built, and visualizations generated using the R stats, ggplot2, pscl, car, pROC, Metrics, caret, glmnet, tidyverse, lubridate, imputeTS, and ggpubr packages. In order to utilize lasso regression for variable selection before fitting linear and logistic models, SNPs with >10% missing values were removed, then remaining missing SNPs were imputed to their most frequent value (mode). This resulted in the removal of rs4713586 gluten sensitivity SNP from both reduction in summative symptom severity and reduction in constipation symptom severity models. In order to avoid poor performance in regression models, variables with Pearson correlation to another variable of ≥80% were removed. This excluded two SNPs from all regression models that incorporated DNA: rs182549 of the lactose persistence haplotype was removed while the more highly cited correlated (Pearson correlation, $r = 0.99$) haplotype SNP, rs4988235, was retained.

Additionally, IL10 SNP rs3024496 was removed as being highly correlated (Pearson correlation, $r = -0.95$) with rs1800896, another IL10 SNP, which was preferentially retained as having a higher risk in the population (SNPedia database, 2021).

For regression models, we transformed bacterial abundance data using the additive log-ratio (ALR) (Friedman et al., 2010; Gloor et al., 2017), which maintains sub-compositional coherence, permitting genera to be removed in downstream analysis (for example, removing insignificant predictors from a regression model). For this, the easy Coda package (Greenacre, 2018) was employed to analyze all 105 microbial genera utilizing variances, variances explained, and Procrustes correlations to select candidate references for ALR according to methodology of Greenacre et al. (2021). An additional criterion we employed was prevalence, as zero values are problematic with log-ratios. Based on these criteria, *Blautia* was the selected reference. In all cases of highly correlated microbe pairs, the unclassified microbe was removed, or if both were unclassified, then the microbe with the larger mean absolute correlation in that dataset was removed. This resulted in removing up to 5 microbes (of 105) from regression models incorporating microbial predictors. Gender and SNPs were scaled to a range between −1 and 1. Age and gender were the two demographic variables used in all models; age was used in models with no transformation. Our hypothesis was that age and gender would be associated with bacterial taxa as evident from literature and are important factors considered in our personalization and care interventions. Hence, we decided to include gender and age as well in all the models and test their association in the context of the moderation effect of genetic and microbiome variables.

After the above data preparation, lasso was employed (glmnet package), for variable selection in the modeling of datasets that included the 105 microbial predictors. Optimal lambda was chosen by a fivefold cross-validation grid search, setting standardize = TRUE so that variables would compete fairly in regularization. The lambda resulting in the minimum mean cross-validated error was selected, or if this resulted in a paucity of predictors, then a plateau lambda in error vs. lambda plot having a sufficient number of predictors was selected. Mean squared error was employed in linear lasso regression, while the mean absolute error was used in logistic regression models. Predictors with non-zero coefficients were retained for subsequent best-fit regression to describe the FGID differences of the subjects over the course of treatment.

To arrive at these descriptive models the step function (stats package) was employed, using the Akaike information criterion (AIC) to obtain a high-quality fit. If any insignificant variables remained, these were removed one-by-one beginning with the least significant, until only significant variables remained, resulting in a final interpretable model for each investigation and set of predictors.

The descriptive modeling of FGID vs. non-FGID was conducted by fitting a logistic regression model, as described above, to demographic and 14 SNPs genomic data (D + G model), producing coefficients to describe the impact of each predictor on FGID in this cohort. A second logistic model was fit with demographic plus 105 baseline gut microbiome genera remaining after pre-processing (D + M model). A third logistic model similarly employed lasso for variable selection from demographic variables, 14 genomic SNPs, and 104 baseline genera (D + G + M model).

For those respondents who self-reported any of the 6 FGIDs, change in symptom severity was analyzed with respect to their demographic, genome SNPs, and baseline gut microbiome data. 104 survey respondents rated the severity of their FGID symptoms on a scale of 1–5. A linear regression model was fit to describe the change in summative symptom severity as a function of demographic and genomic variables (D + G model). As above, this model was compared with two additional models: demographic plus microbial predictors (D + M model) and demographic, genomic, plus microbial predictors (D + G + M model). Lasso regression was employed as above for variable selection, followed by the use of the step function. Additionally, changes in IBS, diarrhea and constipation symptom severity were modeled for those subsets of participants who reported them, using lasso with threefold cross-validation and linear regression as above with D + G, D + M, and D + G + M predictors.

Ethics Statement

E&I Review Services, an independent institutional review board, reviewed and approved IRB Study #18053 on 05/22/2018. Additionally, IRB Study #21141 was determined to be exempt from E&I Review Services on 08/06/2021. Research material derived from human participants included self-collected buccal and fecal swabs. Informed consent was obtained electronically from study participants.

RESULTS

Subject Characteristics

In total, 177 subjects who were successful at losing 5% or more body weight and had genetic and/or gut biome data while enrolled in the Digbi Health program were surveyed to assess any changes in the symptomatology of their FGIDs (**Figure 1**). We compared baseline characteristics of those who reported FGIDs vs. those who did not. The distributions of gender, age, and BMI are seen in **Table 1**. In this dataset, a significant difference was found in gender ($X^2_1 = 8.39$, $P = 0.004$) and initial BMI (Wilcoxon sum rank test, $P = 0.009$), between those who reported FGID and those who did not, but no significant difference was found in age (Wilcoxon sum rank test, $P = 0.60$), number of individuals consuming alcohol or using recreational drugs such as cannabinoids and nicotinoids (including tobacco smoking) ($X^2_1 = 2.52$, $P = 0.112$ and $X^2_1 = 0.28$, $P = 0.595$, respectively). Subsequently, we investigated the effect of gender, BMI, and FGID status on the beta diversity of the baseline gut microbiome of subjects. The PERMANOVA analysis (**Supplementary Table 2**) shows that gender had a significant effect on the beta diversity ($R^2 = 0.012$, $P = 0.030$), whereas BMI ($R^2 = 0.007$, $P = 0.279$) and FGID status ($R^2 = 0.005$, $P = 0.532$) did not.

TABLE 1 | Distribution and demographics of FGID and non-FGID groups in the study.

Descriptor variable	FGID group (*n* = 104)	Non-FGID group (*n* = 73)	*P*-value
Gender, *n* (%)			
Male	12 (11.54%)	22 (30.14%)	*0.004*[a]
Female	92 (88.46%)	51 (69.86%)	
Age, in years [median (*SD*)]	49 (11.92)	51 (12.25)	*0.60*
BMI closest to gut biome sampling [median (*SD*)]	31.17 (7.37)	33.51 (7.50)	*0.009*
Alcohol consumption, *n* (%)	62 (59.61%)	52 (71.23%)	0.112
Recreational drugs (including tobacco smoking), *n* (%)	19 (19.2%)	16 (22.5%)	0.595

[a]*Values in italics are significant (P = < 0.05).*

TABLE 2 | Functional gastrointestinal disorders (FGID) vs. non-FGID logistic model: demographics + genomics (D + G).

Variable	OR	2.5% CI	97.5% CI
Gender	3.255	1.421	7.856
Caffeine metabolism (rs2472297), Risk Allele C	0.447	0.211	0.888
Gluten sensitivity (rs2187668), Risk Allele T	2.926	1.140	8.269
Peanut allergy (rs9275596), Risk Allele C	0.556	0.323	0.935

McFadden pseudo R^2: 0.089, OR, odds ratio.

TABLE 3 | Functional gastrointestinal disorders vs. non-FGID logistic model: demographics + microbiome (D + M).

Variable	OR	2.5% CI	97.5% CI
Gender	2.334	1.213	4.686
Ruminococcus torques group	1.193	1.039	1.391
Akkermansia	1.076	1.011	1.151
Holdemanella	0.932	0.878	0.986
Unclassified genus CAG-56 of *Lachnospiraceae* family	1.095	1.024	1.175
Unclassified genus UCG-010 of *Oscillospirales* order	0.904	0.839	0.970
Anaerostipes	0.806	0.680	0.938
Haemophilus	1.115	1.034	1.210
Fusicatenibacter	0.894	0.803	0.985
Terrisporobacter	1.076	1.008	1.153

McFadden pseudo R^2: 0.220, OR, odds ratio.

TABLE 4 | Functional gastrointestinal disorders vs. non-FGID logistic model: demographics + genomics + microbiome (D + G + M).

Variable	OR	2.5% CI	97.5% CI
Gender	2.256	1.151	4.638
Ruminococcus torques group	1.186	1.030	1.388
Akkermansia	1.080	1.013	1.156
Holdemanella	0.928	0.871	0.985
Unclassified genus CAG-56 of *Lachnospiraceae* family	1.100	1.028	1.184
Unclassified genus UCG-010 of *Oscillospirales* order	0.912	0.845	0.978
Anaerostipes	0.797	0.671	0.929
Haemophilus	1.098	1.017	1.193
Fusicatenibacter	0.892	0.801	0.983
Terrisporobacter	1.074	1.003	1.153

McFadden pseudo R^2: 0.227, OR, odds ratio.

Gut Microbiome Taxa Are a Stronger Predictor of Functional Gastrointestinal Disorders Status Than the Genomic SNPs Analyzed in This Study

The SNP values were not significantly different (Welch's two-sample *t*-tests, results not shown) between respondents with FGID and those without. Logistic regression modeled the associations of demographic, genome SNP, and baseline gut microbiome variables with FGID status in this cohort (**Tables 2–4**), fitting separate average effect size for each predictor while controlling for all other model variables. The D + G model described females in this cohort as 3.26 times more likely than males to suffer FGID while controlling for the genomic predictors in the model (**Table 2**), and each risk allele of the rs2187668 gene was associated with a 2.93 times greater likelihood of being an FGID sufferer. Similarly, risk alleles for rs2472297 and rs9275596 were associated with a lowered likelihood of FGID – conferring 0.45 and 0.56 times likelihood of being an FGID sufferer.

In a second logistic model, we studied the associations of D + M together on the likelihood of a subject having an FGID (**Table 3**). The nine taxa identified by lasso regression were used as predictors for a logistic regression model to describe the classification of 168 subjects into their corresponding FGID status, using the genus *Blautia* as the reference denominator for ALR. In this model, the effect of the female gender while controlling for microbial predictors was an average 2.33 times likelihood of FGID compared with the male gender. Genera *Ruminococcus torques* group, *Akkermansia*, unclassified genus CAG-56 of Lachnospiraceae family, *Haemophilus*, and *Terrisporobacter* were all associated with increased FGID. Whereas *Holdemanella*, unclassified genus UCG-010 of *Oscillospirales* order, *Anaerostipes*, and *Fusicatenibacter* were all associated with decreased (protective of) FGID status.

In a third logistic model studying the associations of D + G + M variables together on the likelihood of a subject having FGID status (**Table 4**), no SNPs had a significant association with FGID risk. Variables in **Tables 3**, **4** are identical, with just slight variations in odd ratios. Not surprisingly, pseudo R^2 values from the D + M model (0.220) and the D + G + M model (0.227) were very similar, but most importantly, improved from that of the D + G model (0.089).

Subjects Reported a Reduction in Severity of Functional Gastrointestinal Disorders Related Symptoms Over the Course of Treatment

The proportion of subjects who experienced at least one point improvement in symptom severity ranged from 75.32% (77/104) for constipation to 90.63% (96/104) for bloating (**Figure 2**). Improvement in summative severity across the 6 FGIDs was seen by 89.42% of respondents (93/104), with an average summative reduction of 51.17% (Wilcoxon signed-rank test, $P = <0.001$). The improvement in FGID symptomatology over

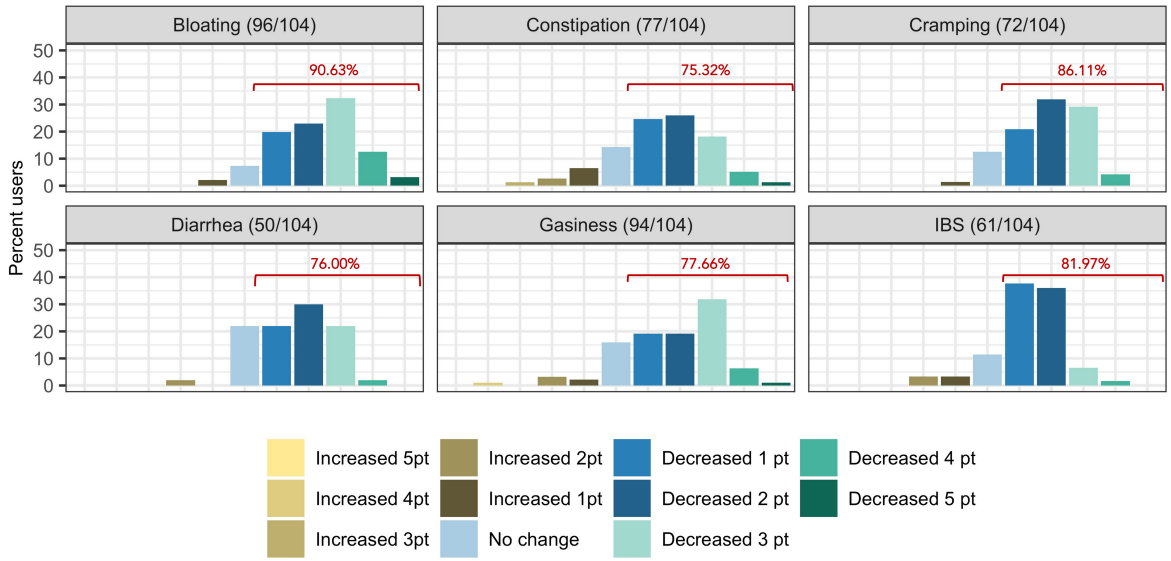


FIGURE 2 | Self-reported symptom severity change (number of levels increased/decreased severity) in functional gastrointestinal disorders from baseline to time of survey. Percentages displayed in red indicate the proportion of users who reported at least one point reduction is symptom severity.

the course of the digital therapeutics intervention (percent summative reduction) was not correlated with percent weight loss (Kendall = 0.12, $P = 0.91$), age (Kendall = −0.66, $P = 0.51$), or gender (Wilcoxon sum rank, $P = 0.809$). Individually, we noted an average 45.93% reduction in the severity of IBS (Wilcoxon signed-rank, $P = <0.001$), 61.01% average reduction in the severity of bloating (Wilcoxon signed-rank test, $P = <0.001$), 38.55% average reduction in the severity of gassiness (Wilcoxon signed-rank, $P =< 0.001$), 61.69% average reduction in the severity of cramping/belly pain (Wilcoxon signed-rank, $P = <0.001$), 37.54% average reduction in the severity of constipation (Wilcoxon signed-rank, $P = <0.001$) and a 48.97% average reduction in the severity of diarrhea (Wilcoxon signed-rank, $P = <0.001$).

Genomic and Microbiome Predictors Can Model Reduction in Summative Functional Gastrointestinal Disorders Symptom Severity as a Result of the Digital Therapeutics Intervention

Supplementary Tables 3, **4** and **Table 5** present linear D + G, D + M, and D + G + M models of reduction in summative FGID symptom severity, respectively. In the D + G model (**Supplementary Table 3**), each additional risk allele for SNPs rs4639334 and rs7775228 was associated with an increase in self-reported summative FGID symptom severity. In the linear D + M model (**Supplementary Table 4**), gut microbial genera Candidatus *Soleaferrea*, *Eubacterium hallii* group, *Alistipes*, and *Desulfovibrio* were all associated with an increase in self-reported summative FGID symptom severity. Microbial genera *Ruminococcus torques* group, *Intestinimonas*, unclassified genus GCA-900066575 of Lachnospiraceae family, and *Megasphaera* were all associated with a self-reported reduction in summative FBD symptom severity. In the linear D + G + M model (**Table 5**), risk alleles of the same two SNPs identified (**Supplementary Table 3**) were again associated with an increase in self-reported summative FGID symptom severity. Similarly, gut microbial genera *Desulfovibrio*, Candidatus *Soleaferrea*, and *Eubacterium ventriosum* group were associated with self-reported increase in summative FGID symptom severity. Microbial genera *Megasphaera*, unclassified genus CAG-352 of Ruminococcaceae family, *Ruminococcus torques* group, *Streptococcus*, and *Intestinimonas* were all associated with self-reported reduction in summative FGID symptom severity. Adjusted R^2 values were 0.124 for the D + G model, 0.318 for the D + M model, and 0.442 for the D + G + M model. This indicates that the fit of the models improved when adding microbiome predictors, with the best fit model for a reduction in summative FGID symptom severity containing a mixture of genomic SNP and microbiome variables.

In the Descriptive Models, Reduction in Irritable Bowel Syndrome Symptom Severity Was Better Explained by a Mixture of Genomic and Microbiome Predictors, Whereas Reduction in Diarrhea and Constipation Symptom Severity Was Better Explained by Microbiome Predictors Only

Tables 6–8 present linear D + G + M models for reduction of symptom severity for IBS, constipation, and diarrhea, respectively. **Supplementary Tables 5**, **7**, and **9**, present linear D + G models, and **Supplementary Tables 6**, **8**, and **10**, present

TABLE 5 | Reduction in summative FGID symptom severity linear model: demographics + genomics + microbiome (D + G + M).

Variable	Estimate	Std. error	*t*-value	*P*-value
Megasphaera	0.495	0.136	3.631	*<0.001*[a]
Desulfovibrio	−0.517	0.105	−4.910	*<0.001*
Unclassified genus *CAG-352* of *Ruminococcaceae* family	0.257	0.104	2.477	*0.015*
Gluten Sensitivity (rs4639334), Risk Allele A	−2.013	0.792	−2.541	*0.013*
Ruminococcus torques group	0.646	0.246	2.626	*0.010*
Gluten Sensitivity (rs7775228), Risk Allele C	−1.936	0.828	−2.338	*0.022*
Streptococcus	0.505	0.218	2.322	*0.023*
Intestinimonas	0.319	0.116	2.756	*0.007*
Candidatus *Soleaferrea*	−0.461	0.150	−3.076	*0.003*
Eubacterium ventriosum group	−0.290	0.134	−2.157	*0.034*

Adjusted R^2: 0.442.
[a]*Values in italics are significant (P = <0.05).*

TABLE 6 | Reduction in IBS symptom severity linear model: demographics + genomics + microbiome (D + G + M).

Variable	Estimate	Std. error	*t*-value	*P*-value
Gluten sensitivity (rs7775228), Risk Allele C	−0.566	0.225	−2.515	*0.016*[a]
Unclassified genus *Clostridia UCG-014*	0.092	0.024	3.844	*<0.001*
Escherichia-Shigella	0.087	0.028	3.122	*0.003*
Fusicatenibacter	−0.096	0.038	−2.501	*0.016*
Megasphaera	0.101	0.034	3.007	*0.004*
Moryella	−0.079	0.028	−2.775	*0.008*

Adjusted R^2: 0.487.
[a]*Values in italics are significant (P = <0.05).*

TABLE 7 | Reduction in constipation symptom severity linear model: demographics + genomics + microbiome (D + G + M).

Variable	Estimate	Std. error	*t*-value	*P*-value
Parabacteroides	0.231	0.073	3.174	*0.003*[a]
Eubacterium coprostanoligenes group	−0.161	0.051	−3.125	*0.003*
Lachnospira	0.123	0.047	2.606	*0.013*
Gluten sensitivity (rs7775228), Risk Allele C	0.757	0.327	2.311	*0.026*

Adjusted R^2: 0.389.
[a]*Values in italics are significant (P = <0.05).*

linear D + M models for reduction of symptom severity for IBS, constipation, and diarrhea, respectively. Adjusted R^2 values for the reduction in IBS symptom severity models were 0.130 for the D + G model, 0.432 for the D + M model, and 0.487 for the D + G + M model. Adjusted R^2 values for the reduction in constipation symptom severity models were 0.038 for the D + G model, 0.413 for the D + M model, and 0.389 for the D + G + M model. Adjusted R^2 values for the reduction in diarrhea symptom severity models were 0.090 for the D + G model, 0.610 for the D + M model and 0.528 for the D + G + M model. This shows that genomic SNP models based on the variables selected for this study performed relatively poorly and that the inclusion of microbiome variables constantly improved the fit

TABLE 8 | Reduction in diarrhea symptom severity linear model: demographics + genomics + microbiome (D + G + M).

Variable	Estimate	Std. error	*z*-value	*P*-value
Ruminococcus torques group	0.203	0.076	2.669	*0.013*[a]
Lactobacillus	0.114	0.042	2.706	*0.012*
Unclassified genus UCG-009 of *Butyricicoccaceae* family	−0.107	0.035	−3.045	*0.005*
Prevotella	0.069	0.031	2.249	*0.034*

Adjusted R^2: 0.528.
[a]*Values in italics are significant (P = <0.05).*

of the models. Interestingly, the D + G + M was the best fit model for IBS symptom reduction, whereas, for constipation and diarrhea symptom reduction, the D + M models were the best fit.

Looking at the best fit models, Unclassified genus *Clostridia UCG-014*, *Escherichia-Shigella*, and *Megasphaera* were associated with reduction of IBS symptom severity, whereas risk alleles of rs7775228 (gluten sensitivity) and genera *Fusicantenibacter* and *Moryella* were associated with an increase. For constipation, there were only microbial taxa in the D + M model: *Parabacteroides*, Unclassified genus of *Anaerovoracaceae* Family XIII AD3011 group, *Lachnospira* and *Terrisporobacter* were associated with a reduction in symptom severity, whereas *Eubacterium coprostanoligenes* group was associated with an increase. For diarrhea change in symptom severity, like for constipation, only microbial taxa were found significant in the D + M model: genera *Intestinimonas*, *Prevotella*, *Lactobacillus*, and *Phascolarctobacterium* were associated with a reduction in symptom severity, while Unclassified genus UCG-009 of the *Butyricicoccaceae* family was associated with an increase.

DISCUSSION

Out of 177 subjects enrolled in this study who successfully lost 5% or more body weight through a digital therapeutics program, 104 presented one or more FGIDs. These FGID sufferers were significantly different from the non-FGID group in terms of gender and BMI at the time of sampling (**Table 1**). Additionally, gender was significantly associated with the composition of baseline gut microbiome samples in these subjects (**Supplementary Table 2**). Logistic regression models trained to differentiate FGID status confirmed that the female gender was associated with a higher prevalence of FGID as seen in the logistic regression D + G model, where females were 3.26 times more likely to be FGID sufferers than males while holding genomic predictors constant (**Table 2**). This is in accordance with what has been reported elsewhere (Narayanan et al., 2021). Additionally, when baseline microbiome was added to the model (D + G + M model), females were on average 2.26 times more likely to suffer from FGID than males while holding constant both genomic and microbial predictors (**Table 4**). Thus, some of the gender association in the D + G model is explained by microbiome variables in the D + G + M model, reinforcing the role of gender in shaping the baseline gut microbiome of subjects (Kim et al., 2020).

Of note, the SNP: gluten sensitivity (rs2187668, risk allele T) was seen to be strongly associated with FGID status in our cohort. This SNP variant is identified as HLA-DQ2.5 and has been reported as one of the most common HLA-DQ2 haplotypes associated with celiac disease (Van Heel et al., 2007). Interestingly, the D + G + M model did not select any genomic variables and was identical to the D + M model (same variables but slightly different odd ratios), and not surprisingly, the pseudo R^2 scores for these models are similar (0.227 for D + G + M vs. 0.220 for D + M). These models have pseudo R^2 scores higher than for the D + G model (0.089), indicating that baseline microbiome better classified participants having FGID than models based on genomic predictors and that the addition of SNPs did not improve classification of FGID by gender plus baseline microbiome. Many of the microbiome taxa variables identified in the models have already been reported in the literature associated with FGIDs. For instance, previous studies show a strong association of the *Ruminococcus torques* group with FGIDs (Lyra et al., 2009).

We then investigated the change in reported symptom severity over the course of the digital therapeutics program. In total, 89.42% (93/104) of subjects experienced improvement in summative symptom severity of their reported FGIDs. This improvement in symptom severity was not correlated with percent weight loss, gender, or age. Reduction in symptom severity was significant for all six FGIDs investigated individually and for summative symptom severity (all of them together). When we modeled the reduction in summative FGID symptom severity over the course of digital therapeutics intervention, we identified the D + G + M linear model as the best fitting, with an adjusted R^2 of 0.442, compared with 0.124 for the D + G model and 0.318 for the D + M model. Two genomic SNPs and eight microbial taxa were the significant predictors in the best model for these participants. Thus, in our cohort, gender, and baseline microbiome best classified subjects into their FGID status, whereas a combination of genomic SNPs and microbiome variables (but not gender) best-modeled reduction in summative FGIDs symptom severity.

We then looked at the reduction of symptom severity for three functional bowel disorders of our interest: IBS, constipation, and diarrhea. Interestingly, for IBS, the best fit model was the one containing D + G + M variables, whereas, for constipation and diarrhea, the best fit models were those only containing D + M variables.

When analyzing the variables found significant to the best fit models for this cohort, we identified several genomic SNPs and microbial taxa that are shared across two or more models. SNP rs7775228 (gluten sensitivity; risk allele C) was associated in the linear D + G + M model with an increase in summative symptom severity (**Table 5**), as well as in the linear D + G + M model with an increase in IBS symptom severity (**Table 6**). Interestingly, despite not being the best fit model, it also was associated in the linear D + G model with an increase in diarrhea symptom severity (**Supplementary Table 9**) and associated in the linear D + G + M model with reduction of constipation symptom severity (**Table 7**). In addition to its association with gluten sensitivity (Monsuur et al., 2008), which is why we included this SNP as part of our care protocols, rs7775228 is involved in seasonal allergic rhinitis and as a protein biomarker for inflammation (Monsuur et al., 2008; Enroth et al., 2014).

In terms of the microbial taxa shared across two or more models, genus *Fusicatenibacter* was associated with decreased likelihood of having FGID status in the D + G + M logistic model of this cohort (**Table 4**) and associated with an increase in IBS symptom severity in the D + G + M linear model (**Table 6**). Genus *Intestinimonas* was associated with a reduction in summative FGID symptom severity in the D + G + M model (**Table 5**) and a reduction in diarrhea symptom severity in the D + M model (**Supplementary Table 10**). Genus *Megasphaera* was noted to be strongly associated with a reduction in summative FGID symptom severity in the D + G + M linear model of **Table 5** and a reduction in IBS symptom severity in the D + G + M linear model of **Table 6**. Interestingly, genus *Lactobacillus* appears associated with a reduction in diarrhea symptom severity in the linear D + M model (**Supplementary Table 10**), an effect that has been amply demonstrated in the literature (McFarland and Goh, 2019), supporting the validity of our methods. Moreover, these bacteria, specifically, *Fusicatenibacter*, *Intestinimonas,* and *Megasphaera* have been previously reported to be Short Chain Fatty Acids (SCFAs) producers (Takeshita et al., 2016; Jin et al., 2019; Bui et al., 2020; Luu et al., 2021). There is ample evidence of the role of SCFAs in improving gut integrity, which plays an essential role in maintaining mucosal homeostasis (Tan et al., 2014), and may explain why these taxa were either negatively associated with FGID status or were associated with a reduction in the severity of FGID symptoms in our cohort.

Our analysis also revealed some bacterial taxa that were associated with FGID status. *Desulfovibrio* was observed to possess a significant association with an increase in FGID summative symptom severity (**Table 5**). Previous reports suggest that bacteria belonging to the genus *Desulfovibrio* generate H2S gas via a dissimilatory sulfate reduction pathway, leading to inflammatory gut disorders (Singh and Lin, 2015; Kushkevych et al., 2019). We also noted the association of *Akkermansia* with increased FGID likelihood (**Table 4**). Although this bacterium is suggested to have beneficial associations with gut health, a few studies have reported its inverse correlation with reduction of abdominal pain (Cruz-Aguliar et al., 2019).

Ruminococcus torques group appears associated with FGID status in the D + G + M logistic model of **Table 4** and associated with a reduction in summative FGID symptom severity in the D + G + M model of **Table 5**. Despite not being the best fit model, this taxon also was associated with a reduction of diarrhea symptom severity in the linear D + G + M model of **Table 8**. Different *Ruminococcus torques* subgroups have been associated in the literature with IBS-D, IBS-M, and Crohn's disease subjects (Lyra et al., 2009). Additionally, genus *Terrisporobacter*, associated with inflammation and gut dysbiosis (Lee et al., 2020), was associated with FGID status in the D + G + M logistic model (**Table 4**). Collectively, these findings demonstrate the potential of gut microbial profiling not only for predicting current gastrointestinal health but also prognosis in FGID related symptoms as a response to personalized dietary intervention.

Limitations

This study has some limitations that are important to note. First, the descriptive modeling exercise performed in this work is the best fit for the cohort analyzed here and is not intended to infer for the larger population. In particular, we aimed to investigate which demographic, genomic, and baseline microbiome predictors improved the fit of the models, along with the magnitude and direction of their association with FGID status or symptom prognosis. Second, whereas we used all microbiome taxa present in the baseline gut microbiome samples (n = 105 after the filters imposed), for genomic SNPs, we selected only markers that are used to inform diet and lifestyle interventions of subjects under the Digbi Health program, specifically those associated with intolerances and allergies. So, the fact that microbiome markers almost always outperformed genomic SNP markers may be due to the markedly different dataset sizes. Third, inclusion criteria in this study did not consider factors known to influence the microbiome composition (probiotic or antibiotic usage) or other comorbidities (musculoskeletal pain, skin conditions, hypothyroidism, diabetes, cholesterol, hypertension, and mental health) that may confound the results presented. Fourth, the survey instrument utilized was an *ad hoc* questionnaire that asked participants to rate their symptom severity for different FGIDs on a scale of 1–5 and was not a validated clinical instrument. The survey was performed retrospectively for both time points after subjects successfully achieved 5% or more body weight loss. And fifth, the findings from this study are derived from a weight loss cohort and thus may be only reflective of the population with FGID that is overweight or obese, or that may benefit from weight loss.

CONCLUSION

Despite the above limitations, the digital therapeutics care provided to subjects, informed by genetic and baseline gut microbiome and their interaction with participant's lifestyle, effectively reduced symptom severity of FGIDs, including IBS, diarrhea, and constipation. One of our earlier studies supported the use of this care as a therapy for insulin resistance (Ricchetti et al., 2020), empowering subjects to manage their inflammation by awareness of the impact of processed foods and foods to which they are sensitive as per their genomic SNPs and gut microbiome results. Dietary fiber coaching also resulted in increased vegetable diversity and quantity. Whereas further research is required to better understand the effect of different components of the care (e.g., fiber types) on modulating the microbial taxa and genomic SNPs identified in the models and their corresponding effect on reduction of FGIDs symptom severity, this preliminary retrospective study generates testable hypotheses for associations of several biomarkers with FGID status and with the prognosis of FGID symptomatology. This study thus provides proof of concept on how a combined genetic and gut microbiome-based dietary intervention can yield biomarkers from human studies. Moreover, the methods presented add to the existing set of tools (e.g., Polster et al., 2021) that can be readily implemented to understand the role that genetics and gut microbiome play on disease etiology. Additionally, FGID and overweight or obesity are common comorbidities, yet concomitant reduction of body weight and reduction of FGID symptom severity is an endpoint which has been poorly studied but it is of high interest to clinicians and patients.

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 below: https://www.ncbi.nlm.nih.gov/bioproject/PRJNA760529.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethical and Independent Review Services (E&I). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

SK: formal analysis, methodology, software, visualization, writing – original draft, and writing – review and editing. PF-L: conceptualization, data curation, formal analysis, methodology, software, validation, visualization, writing – original draft, and writing – review and editing. DK: data curation, formal analysis, methodology, software, visualization, and writing – original draft. TU and KM: data curation and software. CI and RR: writing – original draft and writing – review and editing. SS-R and DA: project administration, writing – original draft, and writing – review and editing. JU: software and writing- reviewing. PD: writing – review and editing. RS: conceptualization, funding acquisition, and writing – review and editing. All authors contributed to the article and approved the submitted version.

FUNDING

This study was funded by Digbi Health, Mountain View, CA, United States. There were no additional external funding sources.

ACKNOWLEDGMENTS

The authors would like to acknowledge Santosh Saravanan for data mining and curation efforts, and Inti Pedroso for useful comments and discussions.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.826916/full#supplementary-material

REFERENCES

Agakidis, C., Kotzakioulafi, E., Petridis, D., Apostolidou, K., and Karagiozoglou-Lampoudi, T. (2019). Mediterranean diet adherence is associated with lower prevalence of functional gastrointestinal disorders in children and adolescents. *Nutrients* 11:1283.

Agnello, M., Carroll, L. N., Imam, N., Pino, R., Palmer, C., Varas, I., et al. (2020). Gut microbiome composition and risk factors in a large cross-sectional IBS cohort. *BMJ Open Gastroenterol.* 7:e000345. doi: 10.1136/bmjgast-2019-000345

American College of Gastroenterology (2021). *10 Tips on Belching, Bloating, and Flatulence*. Available online at: https://gi.org/topics/digestive-health-tips/ (accessed September 21, 2021).

Basnayake, C., Kamm, M. A., Salzberg, M., Stanley, A., Khera, A., Burrell, K., et al. (2019). Outcome of hospital outpatient treatment of functional gastrointestinal disorders: functional gastrointestinal disorders. *Intern. Med. J.* 49, 225–231. doi: 10.1111/imj.14067

Beyder, A., Mazzone, A., Strege, P. R., Tester, D. J., Saito, Y. A., Bernard, C. E., et al. (2014). Loss-of-function of the voltage-gated sodium channel NaV1.5 (Channelopathies) in patients with irritable bowel syndrome. *Gastroenterology* 146, 1659–1668. doi: 10.1053/j.gastro.2014.02.054

Black, C. J., Drossman, D. A., Talley, N. J., Ruddy, J., and Ford, A. C. (2020). Functional gastrointestinal disorders: advances in understanding and management. *Lancet* 396, 1664–1674. doi: 10.1016/S0140-6736(20)32115-2

Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., et al. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat. Biotechnol.* 37, 852–857. doi: 10.1038/s41587-019-0209-9

Bui, T. P. N., Troise, A. D., Nijsse, B., Roviello, G. N., Fogliano, V., and de Vos, W. M. (2020). Intestinimonas-like bacteria are important butyrate producers that utilize Nε-fructosyllysine and lysine in formula-fed infants and adults. *J. Funct. Foods* 70:103974. doi: 10.1016/j.jff.2020.103974

Buono, J. L., Mathur, K., Averitt, A. J., and Andrae, D. A. (2017). Economic burden of irritable bowel syndrome with diarrhea: retrospective analysis of a U.S. Commercially Insured Population. *J. Manag. Care Spec. Pharm.* 23, 453–460. doi: 10.18553/jmcp.2016.16138

Carco, C., Young, W., Gearry, R. B., Talley, N. J., McNabb, W. C., and Roy, N. C. (2020). Increasing evidence that irritable bowel syndrome and functional gastrointestinal disorders have a microbial pathogenesis. *Front. Cell. Infect. Microbiol.* 10:468. doi: 10.3389/fcimb.2020.00468

Chang, L. (2004). Review article: epidemiology and quality of life in functional gastrointestinal disorders. *Aliment. Pharmacol. Ther.* 20, 31–39. doi: 10.1111/j.1365-2036.2004.02183.x

Chen, H., Mao, X., He, J., Yu, B., Huang, Z., Yu, J., et al. (2013). Dietary fibre affects intestinal mucosal barrier function and regulates intestinal bacteria in weaning piglets. *Br. J. Nutr.* 110, 1837–1848. doi: 10.1017/S0007114513001293

Clouse, R. E., Mayer, E. A., Aziz, Q., Drossman, D. A., Dumitrascu, D. L., Mönnikes, H., et al. (2006). Functional abdominal pain syndrome. *Gastroenterology* 130, 1492–1497. doi: 10.1053/j.gastro.2005.11.062

Cruz-Aguliar, R. M., Wantia, N., Clavel, T., Vehreschild, M. J. G. T., Buch, T., Bajbouj, M., et al. (2019). An open-labeled study on fecal microbiota transfer in irritable bowel syndrome patients reveals improvement in abdominal pain associated with the relative abundance of akkermansia muciniphila. *Digestion* 100, 127–138. doi: 10.1159/000494252

Dellon, E. S., and Ringel, Y. (2006). Treatment of functional diarrhea. *Curr. Treat. Options Gastroenterol.* 9, 331–342.

Derrien, M., van Passel, M. W. J., van de Bovenkamp, J. H. B., Schipper, R., de Vos, W., and Dekker, J. (2010). Mucin-bacterial interactions in the human oral cavity and digestive tract. *Gut Microbes* 1, 254–268. doi: 10.4161/gmic.1.4.12778

Elli, L., Tomba, C., Branchi, F., Roncoroni, L., Lombardo, V., Bardella, M., et al. (2016). Evidence for the presence of non-celiac gluten sensitivity in patients with functional gastrointestinal symptoms: results from a multicenter randomized double-blind placebo-controlled gluten challenge. *Nutrients* 8:84. doi: 10.3390/nu8020084

Emerenziani, S., Pier Luca Guarino, M., Trillo Asensio, L., Altomare, A., Ribolsi, M., Balestrieri, P., et al. (2019). Role of overweight and obesity in gastrointestinal disease. *Nutrients* 12:111.

Enroth, S., Johansson, Å, Enroth, S. B., and Gyllensten, U. (2014). Strong effects of genetic and lifestyle factors on biomarker variation and use of personalized cutoffs. *Nat. Commun.* 5:4684. doi: 10.1038/ncomms5684

Everhart, J. E. (2008). "All digestive diseases," in *The Burden of Digestive Diseases in the United States*, ed. J. E. Everhart (Washington, DC: US Government Printing Office), 1–6.

Fernandes, A. D., Reid, J. N., Macklaim, J. M., McMurrough, T. A., Edgell, D. R., and Gloor, G. B. (2014). Unifying the analysis of high-throughput sequencing datasets: characterizing RNA-seq, 16S rRNA gene sequencing and selective growth experiments by compositional data analysis. *Microbiome* 2:15. doi: 10.1186/2049-2618-2-15

Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* 33, 1–22. doi: 10.18637/jss.v033.i01

Fryar, C. D., Carroll, M. D., and Afful, J. (2020). *Prevalence of Overweight, Obesity, and Severe Obesity Among Adults Aged 20 and Over: United States, 1960–1962 Through 2017–2018. NCHS Health E-Stats. 2020*. Available online at: https://www.cdc.gov/nchs/data/hestat/obesity-adult-17-18/obesity-adult.htm (accessed September 10, 2021).

Galai, T., Moran-Lev, H., Cohen, S., Ben-Tov, A., Levy, D., Weintraub, Y., et al. (2020). Higher prevalence of obesity among children with functional abdominal pain disorders. *BMC Pediatr.* 20:193. doi: 10.1186/s12887-020-2106-2109

Gloor, G. B., Macklaim, J. M., Pawlowsky-Glahn, V., and Egozcue, J. J. (2017). Microbiome datasets are compositional: and this is not optional. *Front. Microbiol.* 8:2224. doi: 10.3389/fmicb.2017.02224

Greenacre, M. (2018). *Compositional Data Analysis in Practice*. New York, NY: Chapman and Hall/CRC.

Greenacre, M., Martínez-Álvaro, M., and Blasco, A. (2021). Compositional data analysis of microbiome and any-omics datasets: a validation of the additive logratio transformation. *Front. Microbiol.* 12:727398. doi: 10.3389/fmicb.2021.727398

Hasler, W. L. (2006). Gas and bloating. *Gastroenterol. Hepatol.* 2, 654–662.

Henström, M., and D'Amato, M. (2016). Genetics of irritable bowel syndrome. *Mol. Cell. Pediatr.* 3:7. doi: 10.1186/s40348-016-0038-6

Ho, W., and Spiegel, B. M. (2008). The relationship between obesity and functional gastrointestinal disorders: causation, association, or neither? *Gastroenterol. Hepatol.* 4, 572–578.

Illumina (2013). *Illumina documentation, Part # 15044223 Rev. B*. Available online at: https://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/16s/16s-metagenomic-library-prep-guide-15044223-b.pdf (accessed September 21, 2021).

International Foundation for Gastrointestinal Disorders (2021). *About GI Motility*. Available online at: https://aboutgimotility.org/learn-about-gi-motility/ (accessed September, 2021).

Jin, M., Kalainy, S., Baskota, N., Chiang, D., Deehan, E. C., McDougall, C., et al. (2019). Faecal microbiota from patients with cirrhosis has a low capacity to ferment non-digestible carbohydrates into short-chain fatty acids. *Liver Int.* 39, 1437–1447. doi: 10.1111/liv.14106

Johannesson, E., Ringström, G., Abrahamsson, H., and Sadik, R. (2015). Intervention to increase physical activity in irritable bowel syndrome shows long-term positive effects. *World J. Gastroenterol.* 21, 600–608. doi: 10.3748/wjg.v21.i2.600

Johannesson, E., Simrén, M., Strid, H., Bajor, A., and Sadik, R. (2011). Physical activity improves symptoms in irritable bowel syndrome: a randomized controlled trial. *Am. J. Gastroenterol.* 106, 915–922. doi: 10.1038/ajg.2010.480

Jordan E Bisanz (2018). *qiime2R: Importing QIIME2 Artifacts and Associated Data into R Sessions*. Available online at: https://github.com/jbisanz/qiime2R (accessed July 10, 2021).

Kennedy, P. J. (2014). Irritable bowel syndrome: a microbiome-gut-brain axis disorder? *World J. Gastroenterol.* 20:14105. doi: 10.3748/wjg.v20.i39.14105

Kim, H. I., Jung, S. A., Choi, J. Y., Kim, S. E., Jung, H. K., Shim, K. N., et al. (2013). Impact of shiftwork on irritable bowel syndrome and functional dyspepsia. *J. Korean Med. Sci.* 28, 431–437. doi: 10.3346/jkms.2013.28.3.431

Kim, Y. S., Unno, T., Kim, B.-Y., and Park, M.-S. (2020). Sex differences in gut microbiota. *World J. Mens. Health* 38:48. doi: 10.5534/wjmh.190009

Kushkevych, I., Dordević, D., Kollar, P., Vítězová, M., and Drago, L. (2019). Hydrogen sulfide as a toxic product in the small–large intestine axis and its role in IBD development. *J. Clin. Med.* 8:1054.

Lee, S. H., You, H. S., Kang, H.-G., Kang, S. S., and Hyun, S. H. (2020). Association between altered blood parameters and gut microbiota after synbiotic intake in healthy, elderly korean women. *Nutrients* 12:3112. doi: 10.3390/nu12103112

Lembo, A., Zaman, M., Jones, M., and Talley, N. J. (2007). Influence of genetics on irritable bowel syndrome, gastro-oesophageal reflux and dyspepsia: a twin study: influence of genetics on ibs, gerd and dyspepsia. *Aliment. Pharmacol. Ther.* 25, 1343–1350. doi: 10.1111/j.1365-2036.2007.03326.x

Locke, G. R., Ackerman, M. J., Zinsmeister, A. R., Thapa, P., and Farrugia, G. (2006). Gastrointestinal symptoms in families of patients with an SCN5A-encoded cardiac channelopathy: evidence of an intestinal channelopathy. *Am. J. Gastroenterol.* 101, 1299–1304. doi: 10.1111/j.1572-0241.2006.00507.x

Longstreth, G. F., Thompson, W. G., Chey, W. D., Houghton, L. A., Mearin, F., and Spiller, R. C. (2006). Functional bowel disorders. *Gastroenterology* 130, 1480–1491. doi: 10.1053/j.gastro.2005.11.061

Luu, M., Riester, Z., Baldrich, A., Reichardt, N., Yuille, S., Busetti, A., et al. (2021). Microbial short-chain fatty acids modulate CD8+ T cell responses and improve adoptive immunotherapy for cancer. *Nat. Commun.* 12:4077. doi: 10.1038/s41467-021-24331-1

Lyra, A., Rinttilä, T., Nikkilä, J., Krogius-Kurikka, L., Kajander, K., Malinen, E., et al. (2009). Diarrhoea-predominant irritable bowel syndrome distinguishable by 16S rRNA gene phylotype quantifcation. *World J. Gastroenterol.* 15:5936. doi: 10.3748/wjg.15.5936

Manning, L. P., and Biesiekierski, J. R. (2018). Use of dietary interventions for functional gastrointestinal disorders. *Curr. Opin. Pharmacol.* 43, 132–138. doi: 10.1016/j.coph.2018.09.003

Marsh, A., Eslick, E. M., and Eslick, G. D. (2016). Does a diet low in FODMAPs reduce symptoms associated with functional gastrointestinal disorders? A comprehensive systematic review and meta-analysis. *Eur. J. Nutr.* 55, 897–906. doi: 10.1007/s00394-015-0922-1

McFarland, L. V., and Goh, S. (2019). Are probiotics and prebiotics effective in the prevention of travellers' diarrhea: a systematic review and meta-analysis. *Travel Med. Infect. Dis.* 27, 11–19. doi: 10.1016/j.tmaid.2018.09.007

McMurdie, P. J., and Holmes, S. (2013). phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* 8:e61217. doi: 10.1371/journal.pone.0061217

Monsuur, A. J., de Bakker, P. I. W., Zhernakova, A., Pinto, D., Verduijn, W., Romanos, J., et al. (2008). Effective detection of human leukocyte antigen risk alleles in celiac disease using tag single nucleotide polymorphisms. *PLoS One* 3:e2270. doi: 10.1371/journal.pone.0002270

Morito, Y., Aimi, M., Ishimura, N., Shimura, S., Mikami, H., Okimoto, E., et al. (2014). Association between sleep disturbances and abdominal symptoms. *Intern. Med.* 53, 2179–2183. doi: 10.2169/internalmedicine.53.2591

Narayanan, S. P., Anderson, B., and Bharucha, A. E. (2021). Sex- and gender-related differences in common functional gastroenterologic disorders. *Mayo Clin. Proc.* 96, 1071–1089. doi: 10.1016/j.mayocp.2020.10.004

Ohkusa, T., Koido, S., Nishikawa, Y., and Sato, N. (2019). Gut microbiota and chronic constipation: a review and update. *Front. Med.* 6:19. doi: 10.3389/fmed.2019.00019

Palarea-Albaladejo, J., and Martín-Fernández, J. A. (2015). zCompositions — R package for multivariate imputation of left-censored data under a compositional approach. *Chemom. Intell. Lab. Syst.* 143, 85–96. doi: 10.1016/j.chemolab.2015.02.019

Peery, A. F., Crockett, S. D., Murphy, C. C., Lund, J. L., Dellon, E. S., Williams, J. L., et al. (2019). Burden and cost of gastrointestinal, liver, and pancreatic diseases in the United States: update 2018. *Gastroenterology* 156, 254.e11–272.e11. doi: 10.1053/j.gastro.2018.08.063

Polster, A., Öhman, L., Tap, J., Derrien, M., Le Nevé, B., Sundin, J., et al. (2021). A novel stepwise integrative analysis pipeline reveals distinct microbiota-host interactions and link to symptoms in irritable bowel syndrome. *Sci. Rep.* 11:5521. doi: 10.1038/s41598-021-84686-9

Quinn, T. P., Erb, I., Gloor, G., Notredame, C., Richardson, M. F., and Crowley, T. M. (2019). A field guide for the compositional analysis of any-omics data. *GigaScience* 8:giz107. doi: 10.1093/gigascience/giz107

Ricchetti, R. R., Sinha, R., Muthukumar, K., Singh-Rambiritch, S., Underwood, B., Junaid, I., et al. (2020). Outcomes of a precision digital care program for obesity and associated comorbidities: results in real world clinical practice. *Int. J. Clin. Med. Cases* 7:160.

Saffouri, G. B., Shields-Cutler, R. R., Chen, J., Yang, Y., Lekatz, H. R., Hale, V. L., et al. (2019). Small intestinal microbial dysbiosis underlies symptoms associated with functional gastrointestinal disorders. *Nat. Commun.* 10:2012. doi: 10.1038/s41467-019-9964-9967

Saito, Y. A., Petersen, G. M., Larson, J. J., Atkinson, E. J., Fridley, B. L., de Andrade, M., et al. (2010). Familial aggregation of irritable bowel syndrome: a family case–control study. *Am. J. Gastroenterol.* 105, 833–841. doi: 10.1038/ajg.2010.116

Saito, Y. A., Strege, P. R., Tester, D. J., Locke, G. R., Talley, N. J., Bernard, C. E., et al. (2009). Sodium channel mutation in irritable bowel syndrome: evidence for an ion channelopathy. *Am. J. Physiol. Gastrointest. Liver Physiol.* 296, G211–G218. doi: 10.1152/ajpgi.90571.2008

Schnabel, L., Buscail, C., Sabate, J.-M., Bouchoucha, M., Kesse-Guyot, E., Allès, B., et al. (2018). Association between ultra-processed food consumption and functional gastrointestinal disorders: results from the french nutrinet-santé cohort. *Am. J. Gastroenterol.* 113, 1217–1228. doi: 10.1038/s41395-018-0137-1

Shau, J.-P., Chen, P.-H., Chan, C.-F., Hsu, Y.-C., Wu, T.-C., James, F. E., et al. (2016). Fast foods - are they a risk factor for functional gastrointestinal disorders? *Asia Pac. J. Clin. Nutr.* 25, 393–401.

Silva 138 Reference database (2021). Available online at: https://www.arb-silva.de/documentation/release-138/ (accessed September 10, 2021).

Simrén, M., Barbara, G., Flint, H. J., Spiegel, B. M. R., Spiller, R. C., Vanner, S., et al. (2013). Intestinal microbiota in functional bowel disorders: a Rome foundation report. *Gut* 62, 159–176. doi: 10.1136/gutjnl-2012-302167

Singh, S., and Lin, H. (2015). Hydrogen sulfide in physiology and diseases of the digestive tract. *Microorganisms* 3, 866–889. doi: 10.3390/microorganisms3040866

Sinha, R., Kachru, D., Ricchetti, R. R., Singh-Rambiritch, S., Muthukumar, K. M., Singaravel, V., et al. (2021). Leveraging genomic associations in precision digital care for weight loss: cohort study. *J. Med. Internet Res.* 23:e25401. doi: 10.2196/25401

SNPedia database (2021). Available online at: https://www.snpedia.com/index.php/Rs1800896 (accessed September 10, 2021).

Sperber, A. D., Bangdiwala, S. I., Drossman, D. A., Ghoshal, U. C., Simren, M., Tack, J., et al. (2021). Worldwide prevalence and burden of functional gastrointestinal disorders, results of rome foundation global study. *Gastroenterology* 160, 99.e13–114.e13. doi: 10.1053/j.gastro.2020.04.014

Sullivan, S. N. (2012). Functional abdominal bloating with distention. *ISRN Gastroenterol.* 2012:721820. doi: 10.5402/2012/721820

Takeshita, K., Mizuno, S., Mikami, Y., Sujino, T., Saigusa, K., Matsuoka, K., et al. (2016). A single species of clostridium subcluster XIVa decreased in ulcerative colitis patients. *Inflamm. Bowel Dis.* 22, 2802–2810. doi: 10.1097/MIB.0000000000000972

Tan, J., McKenzie, C., Potamitis, M., Thorburn, A. N., Mackay, C. R., and Macia, L. (2014). "The role of short-chain fatty acids in health and disease," in *Advances in Immunology*, ed. F. W. Alt (Amsterdam: Elsevier), 91–119.

Thompson, W. G., Longstreth, G. F., Drossman, D. A., Heaton, K. W., Irvine, E. J., and Muller-Lissner, S. A. (1999). Functional bowel disorders and functional abdominal pain. *Gut* 45, ii43–ii47. doi: 10.1136/gut.45.2008.ii43

Van Heel, D. A., Franke, L., Hunt, K. A., Gwilliam, R., Zhernakova, A., Inouye, M., et al. (2007). A genome-wide association study for celiac disease identifies risk variants in the region harboring IL2 and IL21. *Nat. Genet.* 39, 827–829. doi: 10.1038/ng2058

Wei, L., Singh, R., Ro, S., and Ghoshal, U. C. (2021). Gut microbiota dysbiosis in functional gastrointestinal disorders: underpinning the symptoms and pathophysiology. *JGH Open* 5, 976–987. doi: 10.1002/jgh3.12528

Wilder-Smith, C. H., Olesen, S. S., Materna, A., and Drewes, A. M. (2017). Predictors of response to a low-FODMAP diet in patients with functional gastrointestinal disorders and lactose or fructose intolerance. *Aliment. Pharmacol. Ther.* 45, 1094–1106. doi: 10.1111/apt.13978

Zhao, Y., and Yu, Y.-B. (2016). Intestinal microbiota and chronic constipation. *SpringerPlus* 5:1130. doi: 10.1186/s40064-016-2821-1

Conflict of Interest: All authors except JU and PD were employees of Digbi Health. JU did contractual work for Digbi Health. PD was an advisor to Digbi Health. The authors declare that this study received funding from Digbi Health. The funder Digbi Health was not involved in the study design, collection, analysis, interpretation of data, the writing of this article, or the decision to submit it for publication. SK, DA, and RS had a patent-pending concerning this work: US Application No. 63/246,348, Methods and systems for multi-omic interventions as diagnostics for personalized care of functional gastrointestinal disorders. The digital therapeutics program provided to study participants in this work is a commercially available program developed and marketed by Digbi Health.

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: 13 April 2022
doi: 10.3389/fmicb.2022.848010

Microbial Ecology of Sulfur Biogeochemical Cycling at a Mesothermal Hot Spring Atop Northern Himalayas, India

*Shekhar Nagar[1], Chandni Talwar[1], Mikael Motelica-Heino[2], Hans-Hermann Richnow[3], Mallikarjun Shakarad[4], Rup Lal[5] and Ram Krishan Negi[1]**

[1] Fish Molecular Biology Laboratory, Department of Zoology, University of Delhi, New Delhi, India, [2] UMR 7327, Centre National de la Recherche Scientifique, Institut des Sciences de la Terre D'Orleans (ISTO), Université d'Orleans-Brgm, Orleans, France, [3] Department of Isotope Biogeochemistry, Helmholtz Centre for Environmental Research – UFZ, Leipzig, Germany, [4] Evolutionary Biology Laboratory, Department of Zoology, University of Delhi, New Delhi, India, [5] NASI Senior Scientist Platinum Jubilee Fellow, The Energy and Resources Institute, New Delhi, India

OPEN ACCESS

Edited by:
Vasvi Chaudhry, University of Tübingen, Germany

Reviewed by:
Digvijay Verma, Babasaheb Bhimrao Ambedkar University, India
Puneet Singh Chauhan, National Botanical Research Institute (CSIR), India

***Correspondence:**
Ram Krishan Negi
rknegi@zoology.du.ac.in; negigurukul@gmail.com

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

Received: *03 January 2022*
Accepted: *09 March 2022*
Published: *13 April 2022*

Citation:
Nagar S, Talwar C, Motelica-Heino M, Richnow H-H, Shakarad M, Lal R and Negi RK (2022) Microbial Ecology of Sulfur Biogeochemical Cycling at a Mesothermal Hot Spring Atop Northern Himalayas, India. Front. Microbiol. 13:848010. doi: 10.3389/fmicb.2022.848010

Sulfur related prokaryotes residing in hot spring present good opportunity for exploring the limitless possibilities of integral ecosystem processes. Metagenomic analysis further expands the phylogenetic breadth of these extraordinary sulfur (S) metabolizing microorganisms as well as their complex metabolic networks and syntrophic interactions in environmental biosystems. Through this study, we explored and expanded the microbial genetic repertoire with focus on S cycling genes through metagenomic analysis of S contaminated hot spring, located at the Northern Himalayas. The analysis revealed rich diversity of microbial consortia with established roles in S cycling such as *Pseudomonas*, *Thioalkalivibrio*, *Desulfovibrio*, and *Desulfobulbaceae* (*Proteobacteria*). The major gene families inferred to be abundant across microbial mat, sediment, and water were assigned to *Proteobacteria* as reflected from the reads per kilobase (RPKs) categorized into translation and ribosomal structure and biogenesis. An analysis of sequence similarity showed conserved pattern of both *dsrAB* genes (n = 178) retrieved from all metagenomes while other S disproportionation proteins were diverged due to different structural and chemical substrates. The diversity of S oxidizing bacteria (SOB) and sulfate reducing bacteria (SRB) with conserved (r)*dsrAB* suggests for it to be an important adaptation for microbial fitness at this site. Here, (i) the oxidative and reductive *dsr* evolutionary time–scale phylogeny proved that the earliest (but not the first) *dsrAB* proteins belong to anaerobic *Thiobacillus* with other (*rdsr*) oxidizers, also we confirm that (ii) SRBs belongs to δ-*Proteobacteria* occurring independent lateral gene transfer (LGT) of *dsr* genes to different and few novel lineages. Further, the structural prediction of unassigned DsrAB proteins confirmed their relatedness with species of *Desulfovibrio* (TM score = 0.86, 0.98, 0.96) and *Archaeoglobus fulgidus* (TM score = 0.97, 0.98). We proposed that the genetic repertoire might provide the basis of studying time–scale evolution and horizontal gene transfer of these genes in biogeochemical S cycling.

Keywords: metagenomics, hot spring, biogeochemical cycle, sulfur spring, evolution

INTRODUCTION

The untapped sulfur (S) compounds oxidizing microorganisms (SOM) and S compounds reducing microorganisms (SRM) microbial communities residing in extreme and contaminated environmental conditions such as hot water, sulfide contaminated springs offer an intriguing opportunity to explore the unique microbial diversity with uncovered metabolic potential (Ghilamicael et al., 2017). The investigations of such microbiota began with focus on identifying and culturing novel thermostable biocatalysts with huge biotechnological applications (Inskeep et al., 2010; Li et al., 2014; Ayangbenro and Babalola, 2017). However, little progress has been made in exploring the correlation between microbiome and geochemistry of hot spring systems particularly that possess mesothermic hot waters with neutral pH and elemental S and sulfate richness (Ghosh et al., 2012; Roy et al., 2020). Moreover, the survival of microbes in these niches is often supported by community dynamics and interactions. Studies of such ecosystems may provide insights into the microbial evolution of specific pathways for microbial biogeochemical cycling of minerals. However, with about 400 thermal hot water springs located in India, less than 15% have been explored for biogeochemical and taxonomical classification using genomics and metagenomics approaches (Cinti et al., 2009; Saxena et al., 2017). Sulfur springs provide harsh physiochemical conditions to sustain the growth of only meso- and hyper-thermophilic microbes which includes S oxidizers and sulfate reducers (Chan et al., 2015; Gonsior et al., 2018). The survival could also be achieved with "microorganism adaptation" by several resistance mechanism such as activity of bioprecipitation, biosorption, extracellular sequestration, and/or chelation (Haferburg and Kothe, 2007). During these changes, the exchange of genetic material by means of horizontal gene transfer (HGT) is prevalent and necessary for the adaptation of microbes through the acquisition of novel genes.

Khirganga, the mesothermal S spring in Northern Himalayas discharging waters rich in sulfate, chlorine, sodium, and magnesium ions has remained uncharted so far (Shirkot and Verma, 2015; Poddar and Das, 2018). High levels of sulfides in the environment accounts for the milky appearance of the hot spring water with white microbial mats predicted to be formed from sulfide reduction by the S-related prokaryotes (SRP) enriched at this site (Sharma et al., 2004; Dong et al., 2019). The microbial S disproportionation one of the oldest (about 3.5 billion years ago; Finster, 2011) biological processes on Earth producing sulfide, sulfite, and sulfate compounds establishes a complex network of pathways in the biogeochemical S cycle. Thus far, it is the very foremost metagenomic investigation of microbial communities in Khirganga (average atmospheric temperature 6.9 ± 0.3) focused on exploring the microbial biogeochemical S cycling with a complex of disproportionation of elemental S conforming intermediary compounds. The current study was carried out *via* microbial mats, sediments, and hot spring water samples in hot spring to decipher the stabilized and diversified genes involved in S cycle intermediary process in anoxygenic, photolithotrophic and chemolithotrophic S-oxidizing and reducing bacteria (Dahl and Truper, 1994; Hipp et al., 1997). The work expands the genetic and evolutionary information for S cycling genes and evaluates the biodiversity and applications for screening of the novel thermostable enzymes from microorganisms. Further, understanding these adaptations *vis-à-vis* the physiological properties and metabolic processes in these springs could be monitored as the engineered SRP consortia could develop into an effective tool in optimizing degradation of sewage waste in industrial processes (Ayangbenro et al., 2018). Also, the sulfate reducing bacteria (SRB) implied to treat various environment contaminants including metals (Mothe et al., 2016; Zhang et al., 2016), metalloids (Battaglia-Brunet et al., 2012; Sahinkaya et al., 2015), various non-methane hydrocarbons (Callaghan et al., 2012), alicyclic hydrocarbons (Jaekel et al., 2015), nitroaromatic compounds (Boopathy, 2014; Mulla et al., 2014), and aromatic hydrocarbons (Stasik et al., 2015; Meckenstock et al., 2016; Kamarisima et al., 2019).

MATERIALS AND METHODS

Sample Collection, Physicochemical Analysis, and Helium Ion Microscopy

Samples of microbial mat deposits (250 g), sediment (250 g) and water (5L) were collected from Khirganga hot water spring (31°59′34″ N, 77°30′35″ E) in February 2017. Sampling was performed in two replicates for each habitat from two closely located primary thermal outlets (31°99′18″ N, 77°50′96″ E) and secondary outlets (31°99′19″ N, 77°50′96″ E). The surface temperature and pH of each habitat were recorded on site.

First, microbial mats and sediment were digested in pure nitric acid and water samples were filtrated to 0.1 μm prior to chemical analysis. All samples were subjected to physicochemical analysis for major elements. Concentrations of major cations (Na^+, K^+, Mg^{2+}, and Ca^{2+}) and anions (SO_4^{2-} and Cl^-) were analyzed by ionic chromatography (Dionex ICS-2000, Sunnyvale, CA) using the columns CS16A for measuring cations and AS17 for anions. An elemental analysis of minor and trace elements through inductively coupled plasma mass spectrometry (ICP-MS) Agilent ICP-MS 7,900 with ultra-high matrix introduction (UHMI). The samples of sediments and microbial mats were desiccated overnight followed by ethanolic dehydration and microstructure was studied using a scanning electron microscope at the Center for Chemical Microscopy (ProVIS). Images were captured using a high efficiency detector.

Metagenomic DNA Extraction, Sequencing, and Assembly

For the extraction of total DNA from microbial mats, 0.25-g samples were processed following a method described by Varin et al. (2010). The total community DNA from 0.25-g sediment samples and 5 L of filtered water (0.45 μm) were extracted using PowerMax Soil DNA isolation kit (MoBio Laboratories Inc., Carlsbad, CA, United States) following the manufacturer's instructions. Sequencing was performed at Beijing Genome Institute (BGI), Hongkong, China using Illumina Hiseq 2,500 platform. Paired end libraries of read length 100 base pairs

(bp) were generated with insert size of 350 bp. The raw sequences were quality filtered using SolexaQA (Cox et al., 2010), and the low-quality sequences below Q_{20} quality cut-off and artificially duplicated reads (ARDs) were castoff using Illumina-Utils (Eren et al., 2013) and duplicate read inferred sequencing error estimation (DRISEE) (Gomez-Alvarez et al., 2009), respectively. Further, the assembly was integrated in IDBA-UD (Peng et al., 2012) with 50-bp insertion length, minimum *k-mer*: 31, maximum *k-mer*: 93 (61 for water) using seed *k-mer* size for alignment 30 bp and the minimum size of contig as 200 bp while allowing minimum multiplicity for filtering *k-mer* while building the graph.

Taxonomic and Functional Assignments

Alpha diversity within each sample was estimated as abundance-weighted average of annotated species from source databases built in MG-RAST v3.0 (Meyer et al., 2008) and expressed as Shannon diversity index transformed based on rarefaction curve using the following formula:

$$\log_{10}(\alpha - diversity)/ln(10)$$

Diversity at phylum level was inferred from MG-RAST (maximum *e*-value, 1×10^5 and minimum percentage identity cutoff, 60%). A paired-sample *t*-test was applied on the phylum determined in any habitat pair to estimate significant similarities based on taxonomic mean abundance using SPSS (SPSS Inc., version 20.0, IBM). Microbial genera were deciphered based on clade-specific markers to identify taxonomy up to species level using MetaPhlAn v2.0 (Truong et al., 2015) and a heatmap was constructed using Bray–Curtis dissimilarity with supporting dendrograms for both species and samples. We used HUMAnN2 (Franzosa et al., 2018) to perform phylum-resolved functional profiling of the communities that maps contigs onto the pangenomes of the known species of the community and quantifies the pathways and UniRef90 gene families database (Suzek et al., 2015). Later, these UniRef90 families were regrouped as clusters of orthologous groups (COGs) annotations based on eggNOG (Huerta-Cepas et al., 2017). Open reading frames (ORFs) of assembled metagenome were predicted using Prodigal v2.6.1 (Hyatt et al., 2010) and annotated at hierarchy levels, namely, subsystems, protein families, and individual enzymes using Prokka v1.12 (Seemann, 2014). The amino acid sequences were mapped against Kyoto Encyclopedia of Genes and Genomes (KEGG) database (Kanehisa et al., 2004) and top 50 metabolic pathways in all of the six samples were compared through heatmap constructed using package *pheatmap* (Kolde and Kolde, 2015) and *ggplot2* (Wickham, 2009) in R (R Development Core Team, 2011). Identification of S substrates disproportionation genes was performed by mapping all predicted ORFs on the HMM databases obtained from TIGRfam v10 (Haft et al., 2012) and Pfam (Finn et al., 2014) using hmmscan v3.1b2 (Eddy, 2011). An abundance of each enzyme was plotted as number of copies annotated within each sample. The sequences with more than 150 amino acids were queried against the National Center for Biotechnology Information (NCBI) Microbial proteins from RefSeq *nr* database (04 April 2020) using BLASTp (Altschul et al., 1990) to identify the sequences producing significant alignments for taxonomic confirmation.

Analysis of Diversity of Sulfate Reduction Proteins

For sequence similarity networks (SSN), amino acid sequences of sulfide oxidation and sulfate reduction proteins annotated in all six samples annotated by KEGG Ids were implied over an empirical measurement of diversity. For this, an all-vs.-all BLAST was performed to define the similarities/variations between sequence pairs of diversifying sulfate reduction proteins. A user defined threshold was optimized according to the alignment score and maximum length of BLAST results in diversifying and stabilized protein sequences. Clustering was performed using CD-HIT (Li and Godzik, 2006) on the scores of BLASTp pairwise alignments at a threshold value (*e*-value of 1*e*-30). The networks were visualized in Cytoscape v3.7.1 (Shannon et al., 2003). The average number of degree and neighbors for a protein sequence or a node was calculated as:

$$k = \frac{2k}{N}$$

where, K is denoted with number of edges and N is denoted with total number of nodes. Also, to determine the divergence/similarity among nodes or protein sequences was calculated as:

$$D = \frac{2k}{N(N-1)}$$

The attributes of node degree distribution, average clustering coefficient, average neighborhood connectivity, and closeness centrality were studied through power law fits to determine their correlation with number of neighbors. Sulfur oxidizing bacteria (SOB) and SRB were identified for the sequences that could be classified up to genus level to study the distribution of S substrates oxidation and reduction genes in the different clusters.

Sequence Alignment, Phylogeny and Structure Prediction of Putative Unidentified Dsr and Asr Enzymes

To elucidate the phylogeny of key sulfite reductases, the DsrA/B and AsrA/B protein sequences (more than 150 amino acids) were individually aligned using MUSCLE v3.8.31 (Edgar, 2004) and clustered using UPGMB (unweighted pair group method with arithmetic mean). All the alignments were end trimmed manually and maximum likelihood (ML) phylogeny was inferred with 500 bootstrap resampling using RAxML v8.0.26 (Stamatakis, 2014). For this, we used standalone version of RAxML which was called as follows:

$$\text{raxmlHPC} - \text{PTHREADS} - \text{s input} - \text{N 500} - \text{n result} - \text{f a}$$
$$-\text{p 12345} - \text{x 12345} - \text{m PROTGAMMAGTR}.$$

The resulting phylogenies were also confirmed using most complex general time-reversible model (CAT-GTR; Tavaré, 1986)

with PhyloBayes v1.7b using CIPRES Science Gateway v 3.3 (Lartillot et al., 2009) that incorporates different rates for every change and different nucleotide frequencies.

For proteins showing similarity with those from uncultured bacteria, we determined the structures using I-TASSER suite (Yang and Zhang, 2015). These predicted structures were then aligned onto their top structural analogs and C-scores, TM-scores, and RMSD were computed and ligand binding sites with conserved residues were identified. The TM-score is to compare two models based on their given residue equivalency (i.e., based on the residue index in the PDB file). It is usually NOT applied to compare two proteins of different sequences. The TM-score predicted from structural alignment of two proteins while comparing them based on residue equivalency such that a score of 0.6 and above denote the two proteins to be fairly aligned (Yang and Zhang, 2015). The TM-align will first find the best equivalent residues of two proteins based on the structure similarity and then output a TM-score.

RESULTS

Description of Sampling Site and Microscopic Analysis of Samples

The Khirganga is a natural hot spring setting that lies in the Parvati Valley in the Northern hemisphere of the great Himalayas (31°59′34″ N, 77°30′35″ E, altitude 2,978 m) at district Kullu, Himachal Pradesh, India (**Figures 1A,B**). For this study, samples from all three habitats, namely, microbial mats, sediments, and water were collected proximal to the major opening (KgM1, KgS1, KgW1) and from a distance of 10 m (KgM2, KgS2, KgW2) as shown in **Figure 1C**.

Using an electron microscopy, we dissected the microstructure of the niches and were able to visualize cellular structures on complex sample matrices. The microbial diversity was visualized as numerous filamentous structures in microbial mats and sediments that resembled Cyanobacteria. In addition, rod and cocci shaped cells of varying sizes were also observed in the sample matrices that providing a visual insight into the microbial diversity at this mesothermic site (**Figure 1D**).

Physicochemical and Elemental Analysis

The *in-situ* measures of water temperature were from 59°C at the outlet to 55°C at 10 m distance (**Table 1**). Microbial mat deposits and sediments had much lower temperature (42–45°C) than water. The pH of the hot spring water was 6.7 while sediments and mats were slightly acidic with pH 6.1 and 6.3, respectively. Thus, all three habitats were recorded to be mesothermic. The physicochemical composition of the hot spring is dominated by anions of chloride (up to 11,024 ug/g) and sulfate (up to 10,079 ug/g) while ions of calcium and potassium were abundant (**Table 1** and **Supplementary File 1**). Importantly, sulfates (SO_4^{2-}) concentration in microbial mats and sediments were higher (9,529 ± 313.29 ug/g; 10,079 ± 863.29 ug/g, respectively) and exceeded the limit of 8,000 ug/g standardized by Environment Protection Act (EPA, 2001) and also found to be exceeded the limit of 53 mg/L in surface waters (79.82 ± 1.85 mg/L) (EPA, 2001; **Table 1**). The chlorides (1,456.77 ± 367.27 mg/L), manganese, sodium, and silicon constituents in the hot spring waters were surpassing the normal average concentrations of 250, 0.05, 200, and 4 mg/L, respectively (EPA, 2001) in surface water samples (**Supplementary File 1**). Among others, the predominant elements and minerals in water samples were aluminum, magnesium, copper, zinc, and arsenic.

Metagenomic DNA Sequencing and Assembly

A large metagenomic dataset was obtained from sequencing having number of reads sized up to ∼18 Gb for each sample. We retrieved a total number of reads ranging between 1.1×10^8 to 1.5×10^8 in all samples which were assembled into 180,849-519,194 (more than 200 bp) contigs. After assembly, the metagenomes sizes varied between 329 and 600 Mbp. A summary of characteristics of the datasets and assembled metagenomes is provided in **Table 2**. The alpha diversity estimated as the Shannon diversity indices ranged between 2.5 and 3 (**Table 2** and **Supplementary File 2**).

Microbial Consortia and Proportionality of S Oxidizing Bacteria and Sulfate Reducing Bacteria

Bacteria belonging to 15 different phyla dominated the microbial communities. The average percentage relative abundances of major phyla in the three habitats shown in parentheses in the order microbial mat, sediment, and water is as follows: *Proteobacteria* (62.1, 50.5, and 58.7%), *Bacteroidetes*, *Firmicutes*, *Cyanobacteria, Planctomycetes*, and *Chloroflexi* (**Figure 2A**). Species belonging to phylum *Proteobacteria* are found in varied temperature ranges which results in their dominance in various hot springs (Bowen de León et al., 2013; Singh and Subudhi, 2016; Saxena et al., 2017) and disproportionation of S compounds is mainly carried out by SRM of *Proteobacteria* (Finster, 2011), Besides, *Actinobacteria*, *Spirochaetes*, *Verrucomicrobia, Acidobacteria, Deinococcus-Thermus, Deferribacteres, Chlorobi, Gemmatimonadetes*, and *Nitrospirae* were also detected in all three habitats with relative abundances less than 3%. Among the three habitats, microbial diversity profiles of water and sediments were more similar compared to those of microbial mats.

The highest genus level diversity was revealed in sediment ($n = 196$) followed by water ($n = 132$) and mat ($n = 63$). The top-50 genera in all habitats were plotted (**Figure 2B**). The microbial mats were dominated by *Pseudomonas* (51.8%) followed by unclassified genera of family *Desulfobulbaceae* (SRB; 5.2%) and *Flavobacterium* (4.3%). Among the abundant genera in the microbial mats were *Thioalkalivibrio* (1.3%, SOB), *Aeromonas* (1.3 %), *Klebsiella* (1%), *Exiguobacterium*, *Enterobacter*, and *Escherichia* (more than 1%) (**Figure 2B**). On the other hand, sediment habitats were found to be enriched in *Thioalkalivibrio* (SOB; 18.9%), *Desulfobulbaceae* (SRB; 9.9%), *Halothiobacillus* (8.3%), *Burkholderia* (7.1%), and unclassifed genera of families *Acetobacteraceae* (6.1%). The hot spring waters with highest diversity of bacterial genera were dominated by *Thioalkalivibrio* (SOB; 20.5%), *Acetobacteracea* (9.6%), and *Desulfovibrio* (SRB;

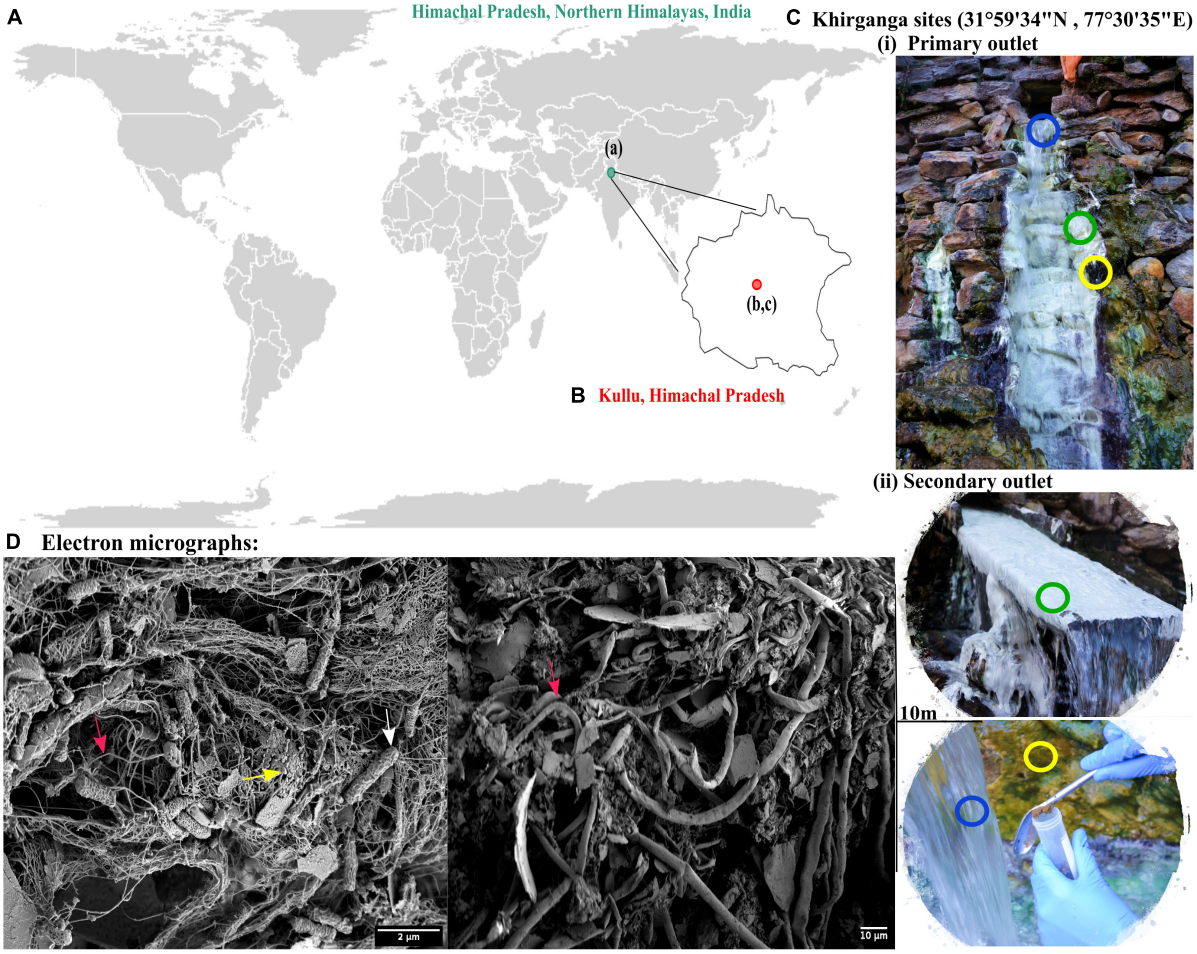


FIGURE 1 | (A) Geographical location of Himachal Pradesh, Northern Himalayas. **(B)** Khirganga hot spring as shown in the maps are located in Parvati valley in Kullu district of Himachal Pradesh, India. **(C)** The different habitats from where the samples were collected are shown: microbial mat (green), sediment (yellow), and water (blue). Samples were collected in replicates from two outlets located 10-m distance apart. **(D)** The scanning electron micrographs of different habitat samples shown with arrow demarcating the filamentous (cyanobacteria), cocci-shaped and rod-shaped bacteria (SRB) in pink, yellow, and white, respectively.

5.9%). Other genera with less than 6% abundances in all three habitats were also detected as shown in **Figure 2B**.

Metabolic Functions of the Community and S Disproportionation Genes

The major gene families inferred to be abundant across all three habitats were assigned to *Proteobacteria* followed by *Chloroflexi*, *Firmicutes*, *Bacteroidetes*, and *Spirochaetes* as reflected from the reads per kilobase (RPKs) in the metagenomes. These gene families were then regrouped as COGs and the top functions were determined to be translation, ribosomal structure and biogenesis (COG: J), amino acid transport and metabolism (E), general function prediction (R), energy production and conversion (C), replication, recombination and repair (L), and carbohydrate transport and metabolism (G). These functions in microbial mats were carried out by *Proteobacteria* (COGs: J, E, C, and G) and unclassified bacteria (COGs: R and L); in sediments by unclassified bacteria (COG: J) and *Proteobacteria* (COGs: E, R, C, L, and G) and in water by *Proteobacteria* (COGs: J and C), *Firmicutes*, and *Chloroflexi* (COGs: E, R, G, and L) (**Figure 2C** and **Supplementary File 3**).

The ORFs that were categorized on the basis of KEGG categories were mapped onto the metabolic functions and the pathways that could be reconstructed with more than 60% completeness were used to define the metabolic potential of the habitats. Based on this criterion, we studied the top-50 functional pathways of each habitat and identified the core functions ($n = 37$ pathways) of the communities that included the common pathways for metabolism of nucleotides, carbohydrates, and amino acids (**Figure 3A** and **Supplementary File 4**). In addition, we determined differentially abundant pathways in each habitat: microbial mats ($n = 7$), sediments ($n = 7$), and water ($n = 2$). The community functional profiles of sediment and water were more similar compared to those of microbial mats which may be due to the stratified layered organization of the mats which are different in sediment and water. The microbial communities in mats were optimized for metabolism of methane specifically,

TABLE 1 | Physicochemical and elemental analysis of the niche samples.

	Value for indicated sampling sites		
Environmental data	**Mat**	**Sediment**	**Water**
Temperature (°C)	42.4 ± 0.56	44.55 ± 1.06	57 ± 2.2
pH	6.2 ± 0.14	6.5 ± 0.14	6.7 ± 0
Major ions/metals	**(μg/g)**	**(μg/g)**	**(mg/L)**
Cl^-	11024.88 ± 983.17	15547.08 ± 204.96	1456.77 ± 367.27
$SO4^{2-}$	9529.06 ± 313.29	10079 ± 863.29	79.82 ± 1.85
Na^+	976.88 ± 64.83	991.47 ± 7.88	1338.52 ± 167.97
Ca^{2+}	7529.71 ± 1034.68	5111.4 ± 588.21	30.27 ± 7.32
K^+	3681.66 ± 886.65	7743.91 ± 412.12	41.56 ± 9.03
Al	7205.41 ± 1685.37	22011.1 ± 2312.9	0.05 ± 0
Mg	3528.62 ± 735.04	8204.28 ± 964.02	8.83 ± 0.02
Si	3.22 ± 0.66	5.28 ± 0.42	31.5 ± 1.36
Mn	1149.7 ± 178.85	1219.29 ± 13.93	0.34 ± 0.03
Cu	14.4 ± 2.62	27.23 ± 1.71	<0.01
Zn	27.44 ± 6.36	67.35 ± 0.47	<0.01
As	3.9 ± 0.64	6.87 ± 0.22	<0.01
Ag	5.51 ± 3.7	5.1 ± 0.03	<0.01
Pb	17.74 ± 2.54	11.4 ± 1.27	<0.00

Values for microbial mats and sediments samples are given in (μg/g) and those of water samples are shown in parts per million (ppm).

members of genus *Methanospirillum*, which were abundant in mats (**Figure 2B**). Other metabolic pathways such as lysine biosynthesis, C5-branched dibasic acid metabolism, thiamine metabolism, pyrimidine metabolism, vitamin B6 metabolism, and other glycan degradation were also found to be abundant in microbial mat.

The S metabolism pathway could be reconstructed within a range of 76.31–84.21% which was maximum in sediment and minimum in microbial mat. A total of 75 genes were responsible for S metabolism present in all the samples with a mean copy number of 1580 ± 249.2. To gain insights into S disproportionation potential across all the habitats, we mapped the TIGRfam and Pfam (**Supplementary File 4**) ids of the 25 associated genes (mean copy number of 980 ± 222.1) on to the ORFs and copy numbers of these genes involved in S oxidation, sulfide oxidation, and sulfate reduction (as described in **Figure 3B**) were estimated. Thiosulfate ions are fused to the carrier complex of S-oxidizing proteins *soxYZ* (*soxY* = 145, *soxZ* = 71), while L-cysteine S-thiosulfotransferase (*soxX* = 76, *soxA* = 86) and S-sulfosulfanyl-L-cysteine sulfohydrolase (*soxB* = 145) mediate the hydrolytic release of reduced S ions from S-bound *soxYZ*. Besides, the dissimilatory sulfite reductase (DsrAB) encoding genes: *dsrA* = 95, *dsrB* = 83 and anaerobic sulfite reductase subunits *asrA* = 44, *asrB* = 30, *asrC* = 9 reduces sulfite to sulfide (**Supplementary File 4**). The other 15 genes for reduction of sulfate ions included solute binding protein (*cysP* = 37), ATP binding protein (*cysA* = 168), transport system permease proteins (*cysU* = 158, *cysW* = 154), ATP sulfurylase (*sat* = 254), transferases (*cysC* = 348, bifunctional *cysNC* = 155, *cysN* = 75), phosphoadenosine phosphosulfate reductase (*cysH* = 314), adenylylsulfate reductase subunit A and B (*aprA* = 88, *aprB* = 104), sulfite reductase flavoprotein alpha-component (*cysJ* = 111), sulfite reductase hemoprotein beta-component (*cysI* = 180), homocysteine desulfhydrase (*mccB* = 6), and cysteine synthase (*ATCYSC1* = 4). **Figure 3C** shows the comparative abundances of these proteins in the three habitats. Here, the results revealed that the microbial S disproportionation occurs largely through the dissimilatory pathway carried out by DsrAB as compared to assimilatory sulfite reductase (AsrABC) mediated reduction. In nature, the dissimilatory pathway is shorter and thus, preferred route of microbial sulfate reduction (**Figure 3B**; Kushkevych et al., 2020).

We identified the sequences of S disproportionation producing significant alignments from the *nr* database for taxonomic confirmation and assigned each sequence that could be classified up to genus level to either SOB or SRB (**Figure 4A** and **Supplementary File 5**). The taxonomy and evolutionary phylogenetic topologies are discussed in detail in the next section.

Diversification and Evolution of S Disproportionation Proteins

To gain insights in the differentiation of S disproportionation genes, we study the diversity and evolution of the key enzymes of SOB and SRB communities in this environmental biosystem. Therefore, we employed a two-step strategy of comparing similarities of all sequences in a pairwise fashion through SSN analysis and further estimated the measures of the rates of non-synonymous to synonymous substitutions in their orthologous proteins between each pair of habitats. SSN effectively resolves the pairwise similarities of each sequence (node) with every other sequence of an enzyme or a group of enzymes for a pathway such that any two nodes are connected by edges only if they share

TABLE 2 | Characteristics of sequenced datasets generated and assembled metagenomes obtained for each sample from different habitats.

Attributes	Microbial mat		Sediment		Water	
Sample name	KgM1	KgM2	KgS1	KgS2	KgW1	KgW2
(Clean reads × 100) bases	110,861,650 × 100	152,895,302 × 100	148,273,248 × 100	145,473,498 × 100	144,915,928 × 100	141,139,772 × 100
No. of contigs	180,849	216,350	404,684	398,499	519,194	506,942
Size of assembly (Mbp)	329	397	592	583	600	591
N_{50}	7,668	8,123	2,672	2,664	3,945	3,985
GC-content (%)	55.55 ± 12.97	54.0 ± 12	59.70 ± 10.74	59.61 ± 11	56.43 ± 12.97	56.34 ± 13
Shannon diversity index	2.6	2.6	2.7	2.7	3.0	3.03
Paired-sample *t*-test (p-value < 0.05)	0.194		0.334		0.416	
Predicted proteins	305,932	369,390	570,660	562,428	545,606	537,498
CRISPRs	452	595	560	570	1,105	1,081

sequence homology above a certain cutoff (here *e*-value of 1*e*-30). Thus, SSN provides for an accurate placement of a sequence among its putative homologs (Talwar et al., 2020).

Here, we examined the diversity among 19 key S substrates oxidizing and reducing proteins determined from the communities as shown in **Figure 3C**, except membrane permeases (CysPUWA) and genes for cysteine synthesis (MccB, ATCYSC1). In total, we retrieved 2,413 protein sequences (mean; M = 254, S = 480, W = 472) denoted by nodes in SSN. The network was organized into 88 connected components including 46 isolated nodes with an average clustering coefficient of 0.84. The connected components were represented by the homologous and heterologous clusters depending on whether they were constituted by the same gene or a number of different genes involved in a pathway, respectively. The number of connected components formed through SSN analysis of S metabolic proteins distributed into gene clusters and the isolated nodes denoted the diverging and highly diverged sequences, respectively (**Figure 4A**). Hence, we looked into these components to study the diversity of each gene that were distributed as shown in **Table 3**. The proteins CysNC, CysH, CysI, and CysJ catalyze important steps and act as cofactors for the AsrABC which were all found to be diverging with many isolated nodes and loosely connected components. The enzymes for S oxidation (Sox) were also found to be diverging as observed from loosely formed clusters. On the other hand, all sequences of the key enzyme of dissimilatory pathway, DsrAB, formed only one connected component, which suggested that they might be under convergent evolution at this site (**Figure 4A**). Further, we compared the distribution of diverged sequences that could be separated as isolated nodes and found that hot spring sediments harbored a high diversity of these enzymes (n = 23) in comparison with microbial mats (n = 12) and water (n = 11).

The node degree distribution estimated to be decreasing with increasing protein quantity (correlation = 0.52, r^2 = 0.29), average neighborhood connectivity within the networks interpreted as function in k was increasing and positively correlated (correlation = 0.90, r^2 = 0.74) (**Figure 4B**). Furthermore, closeness centrality curve that measures closeness between nodes was unable to reach the bench top (correlation = 0.03, r^2 = 0.01), might be due to the maximum number of connected components and less sequence homology. So, we also analyzed each protein cluster individually by using network analysis, 178/2,413 nodes of the network DsrAB protein cluster found to be conserved showed higher clustering coefficient values (0.94), followed by AprA (0.93), AprB, CysNC (0.88), and others (**Table 3**). The evolutionary selection pressures on these genes were studied through estimation of dN/dS values calculated for a subset of conserved gene sequences in all three habitats (**Figure 4C**). The number of core genes and the range of dN/dS values identified for each gene are shown in **Table 3**. The *cysJ* and *cysI* genes were found to be under moderate selection pressures with dN/dS values in the range 0.4–0.7 (**Figure 4C** and **Supplementary File 6**). The results supported the observation as these enzymes code for important co-factors for the AsrABC that were found to be diverging in through SSN analysis. Therefore, the microbial

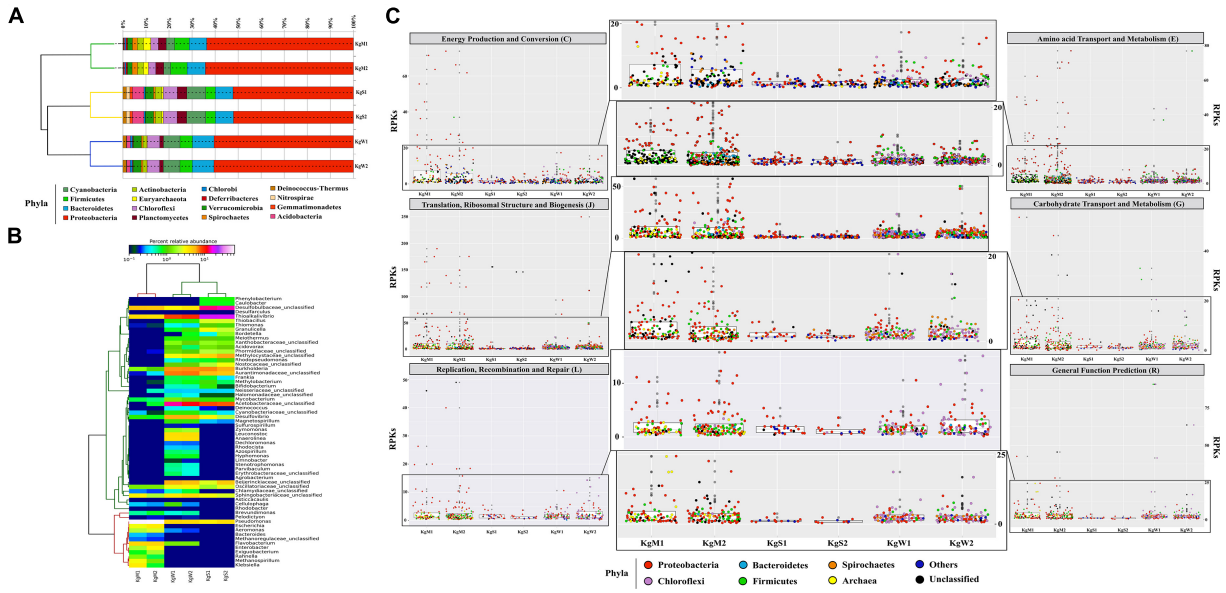


FIGURE 2 | Relative abundance of phylum and genera. **(A)** The stacked bar representation shows the dominating phylum in all three habitats. The dendrograms show hierarchical clustering between species and samples. **(B)** The abundant common genera *Pseudomonas* (20–60%), *Desulfobulbaceae_unclasssifed* (15–20%), *Burkholderia*, *Desulfovibrio*, and *Thioalkalivibrio* (1–20%) are shown here relative abundance × $\log_x$ scale. **(C)** Taxa based functional profiles demonstrating the major phylum contributing toward COG subsystems and percentage of proteins annotated within each COG category for the habitat sites. The distribution of the RPKs were mapped in accordance to the member abundance in the habitats. Dots were colored according to the phylum.

genes for assimilatory reduction pathway are diversifying under moderate selection pressures.

Divergence, Phylogeny, and Structural Relationships of Dissimilatory Sulfite Reductase and Assimilatory Sulfite Reductase

The enzymes catalyzing the reductive (Dsr) or oxidative (rDsr) transformation between sulfite and sulfide appear to be related with respect to their subunit composition and catalytic properties (Loy et al., 2009). The *dsr* genes have been characterized from bacterial as well as archaeal domains (Chang et al., 2001; Grim et al., 2011; Colman et al., 2020). However, their evolution in these domains has long remained a subject of discussion. Our preliminary results showed that both subunits of *dsr* genes (*dsrA* = 79, *dsrB* = 72) corresponds to about 70 newly identified organism for both oxidation and reduction processes (**Supplementary File 8**). Through RAxML phylogenetic analysis, it can be confirmed that the *dsrAB* genes have been introduced in most of the newly identified members by a multiple independent LGT (Anantharaman et al., 2018). Importantly, organisms from *Acidobacteria*, *Candidate division Zixibacteria*, *Chloroflexi*, and β-*Proteobacteria* form completely novel lineages other than known DsrAB clusters identified through RefSeq *nr* protein database with accession numbers (National Center for Biotechnology Information; **Supplementary File 5**). Hence, the dsr from sulfate reducers formed a separate cluster, with sequences from *Desulfarculus, Desulfocarbo, Desulfarcinum, Thermodesulfobacteria, Syntrophobacter, Desulfomonile, Desulfovibrio, Desulfatirhabdium*, and *Desulfobacteriaceae* in both DsrA and DsrB phylogenies and additionally, *Dissulfuribacter* in DsrB (**Supplementary File 8**). We proposed that these organisms with newly identified lineages of *dsr* genes involved in sulfite/sulfate oxidation and reduction likely serves an important control on S cycling on terrestrial subsurface. The divergence of *dsrAB* between unrelated taxa could be driven through combination of speciation, functional diversification, and LGT. Also, there is equal possibility of non-functionality of the genes in these taxa (Loy et al., 2008; Anantharaman et al., 2014).

Through time–scale evolutionary phylogeny of the sequences of DsrAB and AsrAB identified from the metagenomes, we determine the most basal and earliest evolved lineages involved in dsr and *rsdr* pathways (**Figure 5** and **Supplementary File 7**). Our results suggested that both the subunits of the oxidative type reverse-Dsr evolved much earlier than the reductive type Dsr subunits (**Figure 5**). We used the phylogenetic analysis to further assign taxa to the sequences that showed similarity with yet uncultured bacteria and predicted their structures to gain insights into the more common phylogenetic ancestor of the two Dsr subunits. Interestingly, these sequences of the oxidative type Dsr (rDsr) formed monophyletic clade with a more recently identified genus, *Sulfuritortus* in both DsrA and DsrB phylogenies. Our analysis revealed similar tree topologies with these unassigned sequences forming clade with *Sulfuritortus*, *Thiobacillus*, and *Hydrogenophilales bacterium* in both DsrA (n = 1) and DsrB (n = 3). Although the sequences were similar to *Thiobacillus* phylogenetically, prediction of their structures revealed DsrA to be highly similar to that of *Desulfovibrio gigas*

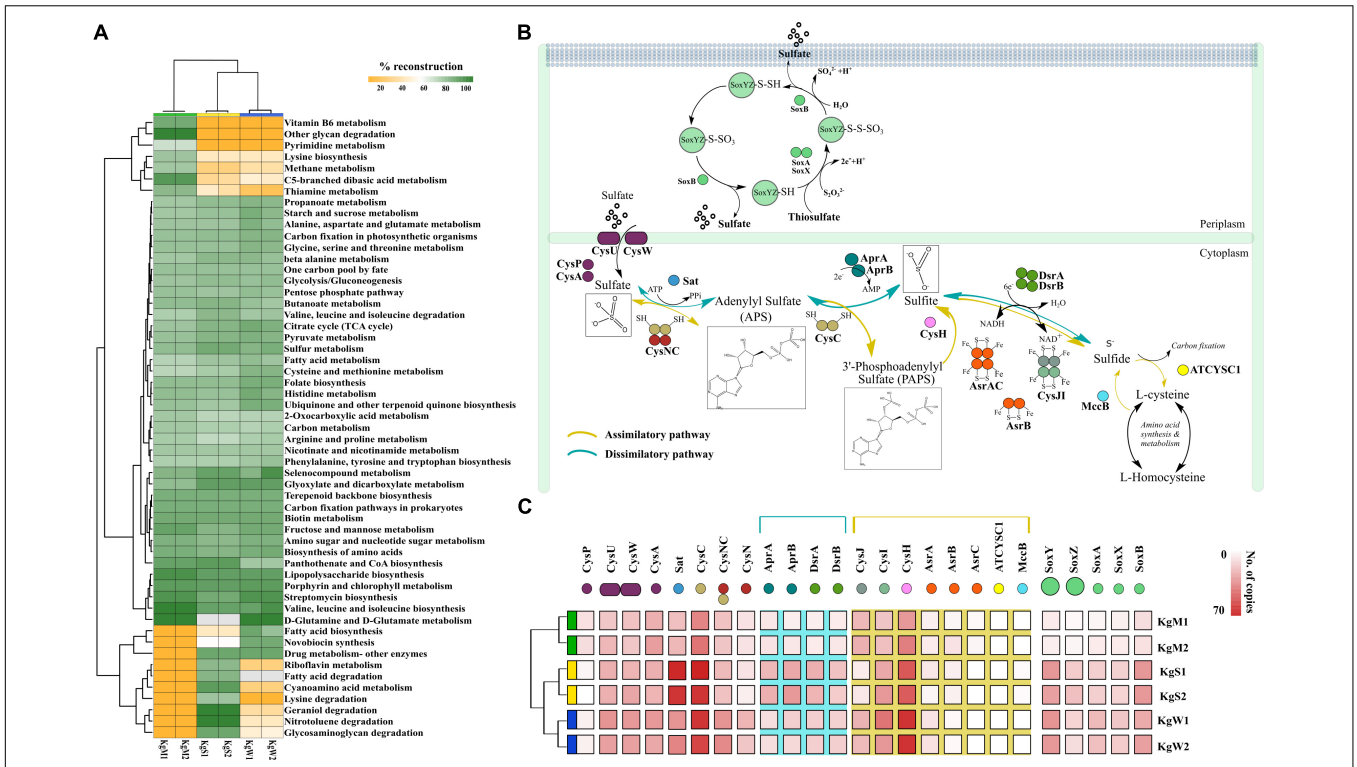


FIGURE 3 | (A) Reconstruction of top 50 pathways annotated using KEGG automatic annotation server. Heatmap matrix representation and clustering was performed by using "pheatmap" package (Kolde and Kolde, 2015) in R (R Development Core Team, 2011). **(B)** The sulfate reduction pathway involved a group of reductases, kinases, and transferases with the product chemical structures generated through chemDraw7 and Inkscape v0.9 (Inkscape Project, 2020). **(C)** The gene copy number of both sulfate reduction and sulfide oxidation pathway that were partitioned in different habitats showed here using ggplot2 in R (R Development Core Team, 2011).

(TM score = 0.75; analog TM = 0.86; C-score = 0.25, 1.6, and 1.58; **Figure 6**) and structures of DsrB proteins aligned most closely with *Archaeoglobus fulgidus* (TM score = 0.94; analog TM = 0.97; C-score = 1.64; **Supplementary File 9**). While the other DsrB subunit (n = 2) formed clades with *Thioalkalivibrio* (β-*Proteobacteria*) also showing similarity to *A. fulgidus* (TM score = 0.94; analog TM = 0.98; C-score = 1.64). In the reductive type Dsr clades, two DsrA sequences from uncultured bacteria formed clades with *Nitrospiraceae* and Chloroflexi bacteria (**Figure 5**). However, a consistent observation was seen in the similarity of all these structures of DsrA proteins with *Desulfovibrio vulgaris* (TM score = 0.94; analog TM = 0.96; C-score = 1.58).

DISCUSSION

Hot water is continually discharged from a major outlet at Khirganga from where it deposits S upon microbial mats over the sediments along its course (Sharma et al., 2004). Microscopic analysis showed that cyanobacteria are widely distributed in mats and sedimentary deposits of thermal springs (van Gemerden, 1993; Podar et al., 2020). The physio-chemical data signified that the hot spring waters emerges from the confluence of rivers Parvati and Beas have high concentrations of chlorides and sulfates that are characteristic of majority of other hot springs in the Himalayan ranges (Cinti et al., 2009; Sangwan et al., 2015). The alpha diversity was higher in the water samples than microbial mat and sediments as has been reported previously (Ghilamicael et al., 2017; Nagar et al., 2021). Species richness as rarefaction curves obtained for all samples attained a plateau indicating optimum metagenomic sequencing data and sampling of a reasonable number of species for all metagenomes. The Bray–Curtis index calculated and plotted using non-metric multidimensional scaling (NMDS) demonstrated a significant difference in the beta-diversity of all three habitats at phyla level (PERMANOVA; $p < 0.01$). Abundance of class γ-*Proteobacteria* in microbial mats significantly distinguished latter from the other two habitats. Similar results have been reported in previous studies (Selvarajan et al., 2018; Pohlner et al., 2019). Communities in sediment and water samples may be varied from each other majorly due to differences in the abundance of δ-*Proteobacteria*. The dominance of aerobic and facultative anaerobic bacteria like *Pseudomonas* in all three habitats could be possible due to mesothermic environment and natural subsurface water hydrodynamics (Nazina et al., 2005). The other taxa that are often found associated with mat deposits are active biofilm producers that use adherence to the surface as a strategy to survive, evolve, and to cope with various abiotic stresses at such extreme habitats (López et al., 2006; Wright et al., 2013).

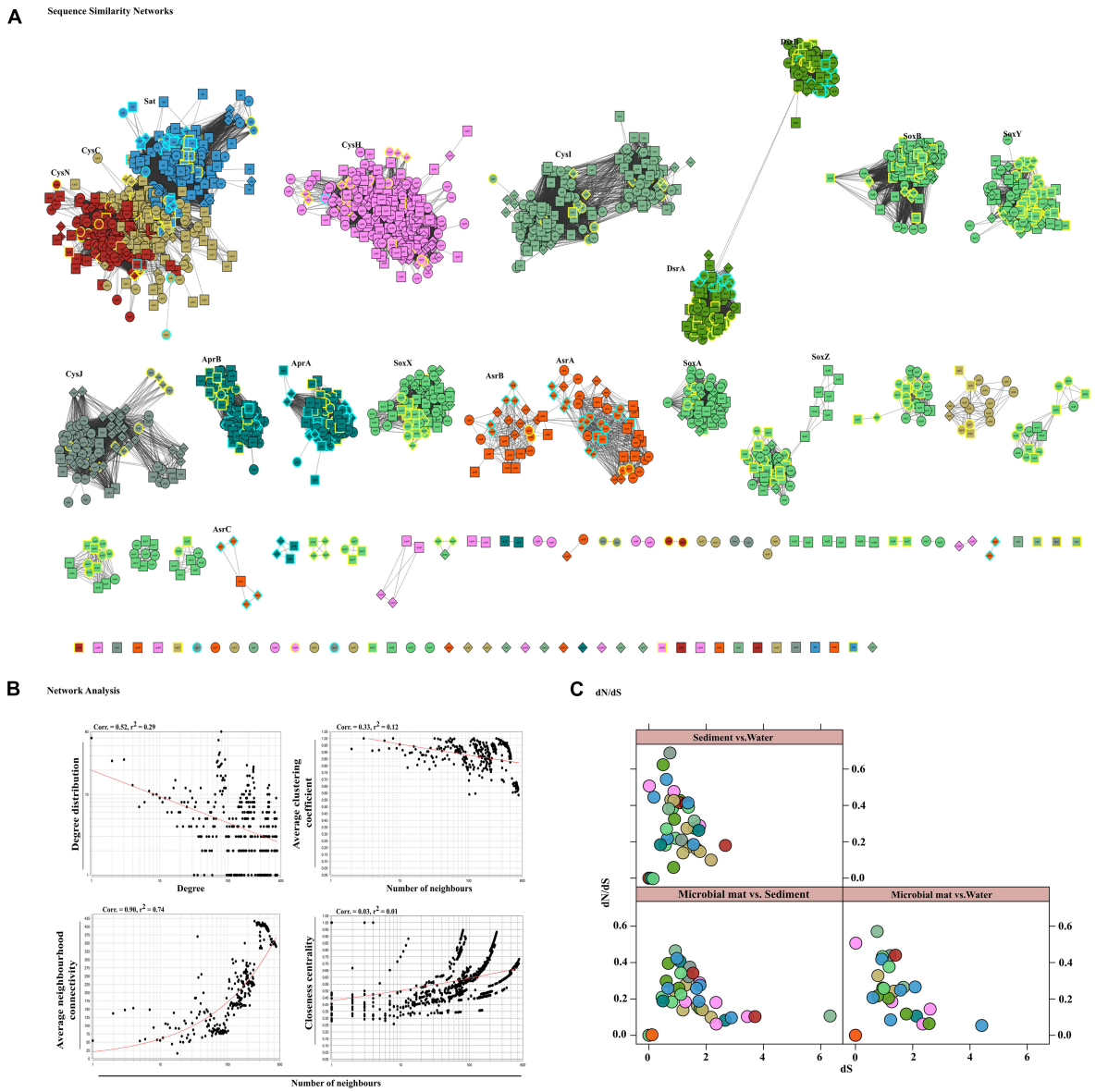


FIGURE 4 | Sequence similarity network analyses. **(A)** Diversity of sulfate reduction genes of both assimilatory and dissimilatory pathways in microbial mat (diamond-shaped), sediment (square), and water (spherical) habitats visualized in cytoscape v3.7.1. Highlighted only the classified taxa, where color cyan belong to SRBs and yellow to SOBs. The network was set at threshold *e*-value cutoff of 1e-30 and nodes represent sequences connected through edges if the similarity exceeds the cutoff score. Here, the clusters and isolated nodes were showing the conserved pattern and diversified pattern of the proteins significantly playing an important role in sulfate reduction. **(B)** Topological properties of the similarity networks: degree distribution, average clustering coefficient, average neighborhood connectivity, and closeness centrality are plotted against the number of neighbors. The power law fit curves are shown within each graph. **(C)** Habitat vs. habitat *dN/dS* values of all S cycle genes were estimated and plotted using *xy*-plot in R (R Development Core Team, 2011).

In contrast, abundance of SOB and SRB in sediment and water was observed (Agostino and Rosenbaum, 2018). These SOB and SRB are usually categorized as lithoautotrophs that play key microbial role in biogeochemical cycling of S in various habitats. In general, hydrogen sulfide account for the S present in the underground geothermal waters originating from pyrites or leaching of other sulfides by deep hypothermal waters (Picard et al., 2016). Sulfide (S^{2-}) is oxidized to sulfate (SO_4^{2-}) as the water rises to the surface and under mild oxidizing conditions, sulfide is only oxidized to sulfate or S dioxide (Picard et al., 2016; Wu et al., 2021). The results provided pertinent information on the geochemical composition of the three habitats to be correlated with the microbial diversity and community functions. More importantly, high concentrations of sulfate ions in microbial mats and sedimentary deposits supported the hypothesis of a key role of the bacterial S cycle in sustaining the microbial community at the hot water spring. Sulfur oxidizing bacteria oxidize the reduced S compounds such as hydrogen sulfide

TABLE 3 | Attributes of the SS N and *dN/dS* analysis of S metabolism genes.

Network parameters	**Network analysis of homologous and heterologous clusters**													
Sulfate reduction genes		**CysNC**	**Sat**	**CysH**	**DsrAB**	**CysI**	**CysJ**	**AprA**	**AprB**	**AsrAB**	**AsrC**	**SoxA**	**SoxB**	**SoxX**	**SoxY**	**SoxZ**
No. of nodes	2,413	578	254	346	178	170	102	87	98	68	9	86	145	76	145	71
No. of edges	1,62,389	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Average degree	134.59	264.2	205.1	121.2	77.8	65.2	57.1	72.4	61.0	17.8	3.4	35.3	119.6	52.1	43.3	21.7
Connected components	88		17	14	1	11	8	2	3	6	4	4	2	2	7	7
Isolated nodes	46		12	8	0	10	5	1	0	4	2	0	0	0	2	2
Network density	0.05	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Characteristic path length	1.89	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Shortest path	14%	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Network centralization	0.16	–	–	–	—-	–	–	–	–	–	–	–	–	–	–	–
Clustering coefficient	0.84	0.88	0.86	0.80	0.94	0.85	0.86	0.93	0.88	0.79	0.76	0.97	0.91	0.89	0.83	0.83
Core genes clusters	–	14	6	9	5	3	2	1	2	1	0	2	4	3	6	2
Range *dN/dS* (≤0)	–	≥0.4	≥ 0.5	≥0.5	≥0.6	≥0.5	≥0.6	≥0.4	≥ 0.1	0	–	0	≥0.4	≥0.1	≥0.3	≥ 0.3

(H_2S), elemental S (S^0), sulfite (SO_3^{2-}), thiosulfate ($S_2O_3^{2-}$), and various polythionates ($S_nO_6^{2-}$ or -S_nO^{6-}) into sulfate (SO_4^{2-}). On the contrary, SO_4^{2-} can serve as an electron acceptor of SRB under anaerobic conditions, and they reduce the SO_4^{2-} and other oxidized S compounds ($S_2O_3^{2-}$, SO_3^{2-}, and S^0) into H_2S (Agostino and Rosenbaum, 2018). The abundance of SOB such as *Thioalkalivibrio* and *Burkholderia* as well as SRB such as *Desulfobulbaceae unclassified* and *Desulfovibrio* in spring is not surprising as high levels of sulfate dominate the site and relative abundance of these bacteria provide evidence of an active S cycling mediated by microbial communities. The enriched diversity for Sox and sulfate reduction as well as the geochemical analysis of sulfide rich habitats compelled us to mine the regulatory genes involved in the different pathways of S cycle. In natural system, the S intermediates are reduced by different bacteria through two different reduction processes, namely, dissimilatory and assimilatory reactions (Vermeij and Kertesz, 1999; Zavarzin, 2008; **Figure 3B**). In dissimilatory reduction, SRB utilize three enzymes [(ATP sulfurylase (sat), APS reductase (apr), and sulfite reductase (dsr)] to reduce sulfate and produce toxic hydrogen sulfide (Agostino and Rosenbaum, 2018; Kushkevych et al., 2020). On contrary, sulfate is assimilated into organic compounds under assimilatory process (Kushkevych et al., 2020).

Based on sequence similarity, majority of the AsrABC genes (assimilatory) that were taxonomically related to SRB could be distinguished from the rest of the sequences that were not identified as either SOB or SRB. Thus, assimilatory reduction was diverged among the SRB communities while those for dissimilatory pathway were rather conserved at the site. Previous reports of sequence comparisons have confirmed that DsrAB, Dsr, to be highly conserved enzyme that could serve as marker gene for SRBs (Loy et al., 2009). The DsrAB and AprAB enzymes were contributed by both SOB and SRB with syntrophic interaction which suggests for the presence of reverse dsr (rdsr) mediated oxidation of S substrates in addition to dissimilatory reduction (Kumar et al., 2017). The inherent complexity of S-based metabolic network revealed that there are controlled mutation rates in dsrAB genes in presence or increased selective pressure of contamination and extreme conditions. A relatively high diversity of the other sulfate disproportionation proteins in all three segments unveiled the high nutritional demands and efficiency of the microbes toward uptake of a wide range of structurally and chemically diverse amino acid side chains from environment (Talwar et al., 2020). The syntropy of SOB and SRB prevailing in anoxic and anaerobic conditions governs the dissimilatory S metabolism (oxidation and reduction parallelly) and indirectly promoting the growth of diverse microbes in this natural ecosystem (Bhatnagar et al., 2020).

It was noted that mutation rates were low in all subunits of Dsr proteins which we readily analyzed to trace down the ancestor among SOBs or SRBs in dissimilatory and reverse dissimilatory (rdsr) pathways. The result of the time–scale phylogeny suggested that an increase in substitution rate in both subunits of DSR might have occurred on the branch connecting δ-*Proteobacteria* to all other taxa as

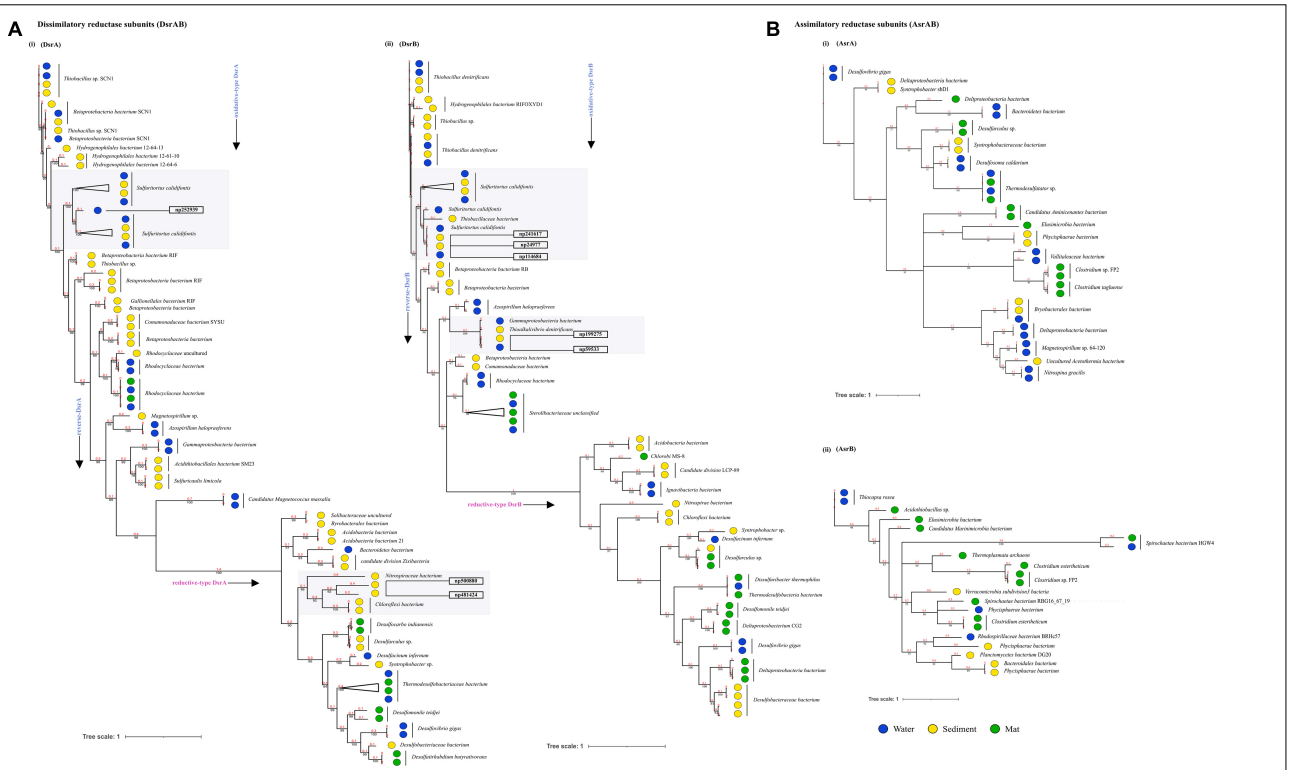


FIGURE 5 | Divergence estimation over time. Reconstruction of the phylogenic tree of optimized full length DsrAB and AsrABC subunits in three habitats using PhyloBayes with the CAT-GTR model. The highlighted squares consist of clades with proteins that were remained unclassified through *nr* database. **(A)** (i) Among, 78 DsrA nodes that showed here the earliest evolution of the rDsr oxidative proteins occurred in *Thiobacillus* sp. (ii) 72 nodes of DsrB proteins with similar results. **(B)** (i) 38 nodes of AsrA and (ii) 21 nodes of AsrB were also compared as a control for branch length shown here.

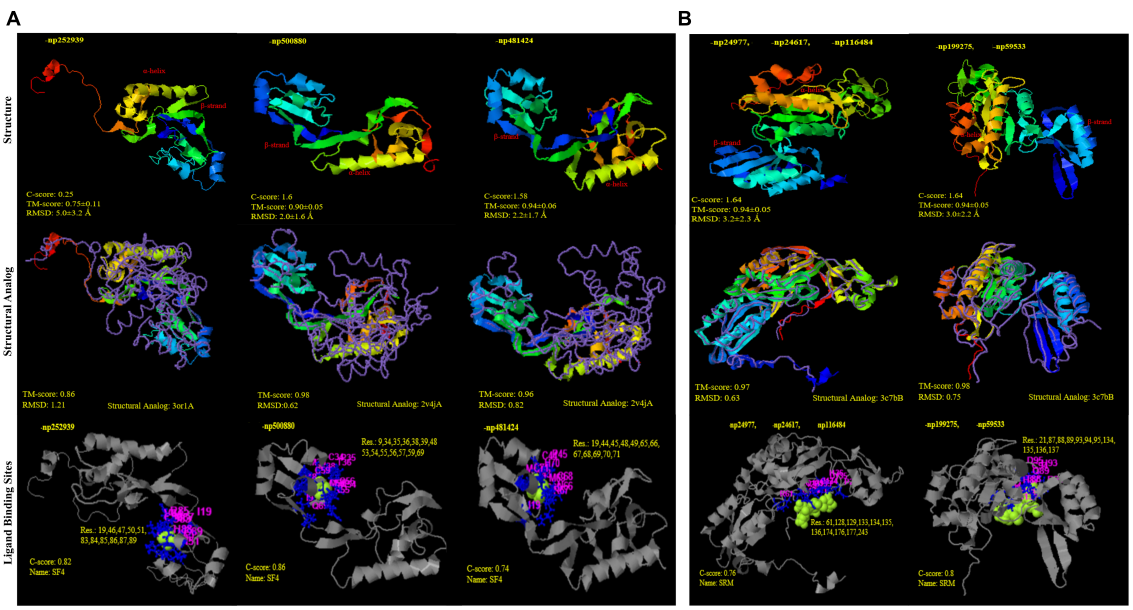


FIGURE 6 | Structural similarity and analogy of unclassified proteins from PDB (Protein data bank) using i-TASSER. **(A)** DsrA protein subunits: (i) np252939 and (ii); (iii) np500880; np481424 showing Tm-align similarity with PDB ids 3or1 and 2v4j (*Desulfovibrio gigas* and *D. vulgaris)*, respectively; with SF4 (iron S cluster) ligand binding sites **(B)** DsrB protein subunits (i) np24977, np24617, np116484 and (ii) np199275, np59533 showing Tm-align similarity with PDB ids 3c7b; with siroheme ligand binding sites.

observed from the branch lengths. The different rates of substitution of the two DSR subunits has this far only been reported in δ-*Proteobacteria* lineage (Wagner et al., 1998). The templates of these *dsrAB* genes have potential to study the genotypic and phenotypic traits in SRPs and the dissimilatory S metabolism processes which will expand the gene-environment interaction mechanism. Also, prior analysis has proved that the evolution of *dsrAB* have been influenced by LGT only among major taxonomic lineages (Klein et al., 2001; Müller et al., 2015) but the findings here provide evidence of independent multiple LGT events distributed throughout the dissimilatory gene clusters. Currently, the time-scale study of this site cannot produce evidence of the progenitor lineages, as the evolutionary history of dissimilatory reduction is complex and yet ascertain. Although it had provided information of the earliest lineage where sulfate/sulfite oxidation and reduction appeared. The genus of SOB currently has few recognized species and is closely related to the members of genus *Thiobacillus* (Kojima et al., 2017). Through structural homology, we predicted that the genus derives its two subunits of Dsr from different ancestors. A plausible explanation for this is observed in the previous reports of high sequence homology between the Dsr of *A. fulgidus* and *D. vulgaris* that suggest a common origin of archaeal and bacterial DSRs or their HGT (Karkhoff-Schweizer et al., 1995). In addition to homology in their sequences, the evolutionary distance separating the enzymes from *A. fulgidus* and *D. vulgaris* was deciphered. For DsrB subunits, the archaeal and bacterial sequences were not particularly distant; such that the branches with structural homology to *A. fulgidus* were approximately the same length as branches leading to bacteria such as *Thiobacillus*, *Thioalkalivibrio*, and β-*Proteobacteria bacterium* (Larsen et al., 1999).

CONCLUSION

The mesothermic hot spring have been composed of a diverse group of microbes (Bacteria and Archaea) and genotypes (*dsrAB*) that could be screened out as novel thermozymes that cannot be underestimated. From the results, it could be concluded that the microbial community functions were distinguished in microbial mat from water and sediment. Here, the genomic repertoire suggested the ongoing specific adaptations to cope up with extreme values of sulfide content in this ecological setting. The S metabolic pathways are completed where inorganic S compounds being the main source for SRB releasing toxicity in the form of sulfides (S^{2-}). The sulfate reduction profiling in all three habitats reveals dissimilatory sulfate reduction process (*dsrAB*) is active than assimilatory sulfate reduction (*asrAB*). Later, the genes involved in S reduction/oxidation were classified and belong mostly to *Proteobacteria* with maximum homologous proteins classified in anoxygenic SOBs. In all S disproportionation proteins, the sulfite reductase DsrAB proteins showed conserved behavior with 0/1 isolated nodes that have been signified as phylogenetic markers for SRBs. The evolutionary phylogenetic analysis showed that the oxidative rDsr were the earliest than the reductive Dsr which may predict that the condition with more sulfides oxidized in more sulfates, directing SRB to perform dissimilatory reduction later. Phylogenetic clades of DsrAB proteins showed unanimous distribution of taxa except the δ-*Proteobacteria* which could be the reason for occurring LGT to other phyla. On the basis of structural alignments, the lineages with unclassified clades have shown different analogy in both Dsr subunits where DsrB derived from Archaea and DsrA are δ-*Proteobacteria* in origin at this mesothermic niche. The stabilization and evolutionary time–scale phylogeny of DsrAB revealing a positive syntrophic relationship between SOB and SRB. These thermophilic microbial inhabitants are very crucial in expanding the metal toxification, ion exchange, and biogeochemical cycling of the elements.

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

SN: conceptualization, methodology, investigation, formal analysis, data curation, methodology, writing—original draft, review, and editing. CT: formal analysis, data curation, methodology, writing—original draft, review, and editing. MM-H and H-HR: formal analysis, writing—review, and editing. MS: writing—review and editing. RL and RN: conceptualization, Writing—review, and editing. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by funds from the Department of Biotechnology (DBT), National Bureau of Agriculturally Important Microorganisms (NBAIM), University Grants Commission-Career Advancement Scheme, and Department of Science and Technology-Purse grant.

ACKNOWLEDGMENTS

The sequence data were produced by the Beijing Genome Institute, China (BGI Tech. Solutions Co., Ltd., Hongkong) in collaboration with the user community. We thank Matthias Schmidt for providing the electron micrographs at the Center for Chemical Microscopy (ProVIS) scanning at UFZ Leipzig, which was supported by European Regional Development Funds (EFRE-Europe Funds Saxony), and the Helmholtz Association, and The Energy and Resources Institute (TERI). SN and CT thank Council

of Scientific and Industrial Research (CSIR) for providing doctoral fellowships.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.848010/full#supplementary-material

Supplementary File 1 | NCBI accession and *in-situ* measures of temperatures, pH, and elemental analysis of each sample from three studied niches.

Supplementary File 2 | (A) Rarefaction curves based on measures of alpha diversity species richness calculated for each of the six samples. **(B)** The interquartile range of alpha diversity measures based on Shannon indices. **(C)** Non-metric multidimensional scaling plots based on Bray–Curtis indices of community dissimilarities in the six samples types ($p < 0.05$, PERMANOVA).

Supplementary File 3 | The list of all metabolic gene families in each sample (KgM1, KgM2, KgS1, KgS2, KgW1, and KgW2) and the RPKs abundances of taxa (*Proteobacteria, Bacteroidetes*, etc.) summarized in details.

Supplementary File 4 | Detailed information of percentage reconstruction of metabolic pathways; Common and unique pathways in all six samples; Identification sources for 19 important proteins involved in S disproportionation using TIGRfam and Pfam conserved domain database, *x* indicates no hits found. Also, the copy number of 25 S disproportionation proteins recorded.

Supplementary File 5 | Details of taxonomic information of 2,413 nodes of 19 S cycle genes identified from nr NCBI microbial proteins database.

Supplementary File 6 | Habitat vs. habitat *dN*/*dS* values of the S cycling genes.

Supplementary File 7 | Sequence alignment of DsrA, DsrB, AsrA, and AsrB proteins identified in this study.

Supplementary File 8 | Phylogenetic analysis of DsrAB proteins: showing percentage similarity with taxa on NCBI *nr* database. Few newly identified lineages were marked in red. In DsrAB, maximum number of LGT events were occurred in Proteobacteria class itself *clade* **(A2,B2)** while the independent events of LGT toward different phyla were occurred through δ-Proteobacteria depicted in *clade* **(A1,B1)**.

Supplementary File 9 | Structural models for other ligand binding sites of unclassified proteins of DsrA and DsrB subunits.

REFERENCES

Agostino, V., and Rosenbaum, M. A. (2018). Sulfate-reducing electro-autotrophs and their applications in bioelectrochemical systems. *Front. Energy Res.* 6:55. doi: 10.3389/fenrg.2018.00055

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. doi: 10.1016/S0022-2836(05)80360-2

Anantharaman, K., Duhaime, M. B., Breier, J. A., Wendt, K. A., Toner, B. M., and Dick, G. J. (2014). Sulfur oxidation genes in diverse deep-sea viruses. *Science* 344, 757–760. doi: 10.1126/science.1252229

Anantharaman, K., Hausmann, B., Jungbluth, S. P., Kantor, R. S., Lavy, A., Warren, L. A., et al. (2018). Expanded diversity of microbial groups that shape the dissimilatory sulfur cycle. *ISME J.* 12, 1715–1728. doi: 10.1038/s41396-018-0078-0

Ayangbenro, A. S., and Babalola, O. O. (2017). A new strategy for heavy metal polluted environments: a review of microbial biosorbents. *Int. J. Environ. Res. Public Health* 14:94. doi: 10.3390/ijerph14010094

Ayangbenro, A. S., Olanrewaju, O. S., and Babalola, O. O. (2018). Sulfate-reducing bacteria as an effective tool for sustainable acid mine bioremediation. *Front. Microbiol.* 9:1986. doi: 10.3389/fmicb.2018.01986

Battaglia-Brunet, F., Crouzet, C., Burnol, A., Coulon, S., Morin, D., and Joulian, C. (2012). Precipitation of arsenic sulphide from acidic water in a fixed-film bioreactor. *Water Res.* 46, 3923–3933. doi: 10.1016/j.watres.2012.04.035

Bhatnagar, S., Cowley, E. S., Kopf, S. H., Pérez Castro, S., Kearney, S., and Dawson, S. C. (2020). Microbial community dynamics and coexistence in a sulfide-driven phototrophic bloom. *Environ. Microb.* 15:3. doi: 10.1186/s40793-019-0348-0

Boopathy, R. (2014). Biodegradation of 2, 4, 6-trinitrotoluene (TNT) under sulfate and nitrate reducing conditions. *Biologia* 69, 1264–1270. doi: 10.2478/s11756-014-0441-1

Bowen de León, K., Gerlach, R., Peyton, B. M., and Fields, M. W. (2013). Archaeal and bacterial communities in three alkaline hot springs in Heart Lake Geyser Basin, Yellowstone National Park. *Front. Microbiol.* 4:330. doi: 10.3389/fmicb.2013.00330

Callaghan, A. V., Morris, B. E., Pereira, I. A., McInerney, M. J., Austin, R. N., Groves, J. T., et al. (2012). The genome sequence of *Desulfatibacillum alkenivorans* AK-01: a blueprint for anaerobic alkane oxidation. *Environ. Microbiol.* 14, 101–113. doi: 10.1111/j.1462-2920.2011.02516.x

Chan, C. S., Chan, K. G., Tay, Y. L., Chua, Y. H., and Goh, K. M. (2015). Diversity of thermophiles in a Malaysian hot spring determined using 16S rRNA and shotgun metagenome sequencing. *Front. Microbiol.* 6:177. doi: 10.3389/fmicb.2015.0017

Chang, Y. J., Peacock, A. D., Long, P. E., Stephen, J. R., McKinley, J. P., Macnaughton, S. J., et al. (2001). Diversity and characterization of sulfate-reducing bacteria in groundwater at a uranium mill tailings site. *Appl. Environ. Microbiol.* 67, 3149–3160. doi: 10.1128/AEM.67.7.3149-3160.2001

Cinti, D., Pizzino, L., Voltattorni, N., Quattrocchi, F., and Walia, V. (2009). Geochemistry of thermal waters along fault segments in the Beas and Parvati valleys (north-west Himalaya, Himachal Pradesh) and in the Sohna town (Haryana), India. *Geochem. J.* 43, 65–76. doi: 10.2343/geochemj.1.0011

Colman, D. R., Lindsay, M. R., Amenabar, M. J., Fernandes-Martins, M. C., Roden, E. R., and Boyd, E. S. (2020). Phylogenomic analysis of novel *Diaforarchaea* is consistent with sulfite but not sulfate reduction in volcanic environments on early Earth. *ISME J.* 14, 1316–1331. doi: 10.1038/s41396-020-0611-9

Cox, M. P., Peterson, D. A., and Biggs, P. J. (2010). SolexaQA: at-a-glance quality assessment of Illumina second-generation sequencing data. *BMC Bioinformatics* 11:485. doi: 10.1186/1471-2105-11-485

Dahl, C., and Truper, H. G. (1994). Enzymes of dissimilatory sulfide oxidation in phototrophic sulfur bacteria. *Methods Enzymol.* 243, 400–421. doi: 10.1016/0076-6879(94)43030-6

Dong, Y., Sanford, R. A., Inskeep, W. P., Srivastava, V., Bulone, V., Fields, C. J., et al. (2019). Physiology, metabolism, and fossilization of hot-spring filamentous microbial mats. *Astrobiology* 19, 1442–1458. doi: 10.1089/ast.2018.1965

Eddy, S. R. (2011). Accelerated profile HMM searches. *PLoS Comput. Biol.* 7:e1002195. doi: 10.1371/journal.pcbi.1002195

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340

EPA (2001). *Parameters of Water Quality: Interpretation and Standards*. Wexford: Environmental Protection Agency.

Eren, A. M., Vineis, J. H., Morrison, H. G., and Sogin, M. L. (2013). A filtering method to generate high quality short reads using Illumina paired-end technology. *PLoS One* 8:e66643. doi: 10.1371/journal.pone.0066643

Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., and Eddy, S. R. (2014). Pfam: the protein families database. *Nucleic Acids Res.* 42, D222–D230. doi: 10.1093/nar/gkt1223

Finster, K. (2011). Microbiological dispropotionation of inorganic sulfur compounds. *J. Sulf. Chem.* 29, 281–292. doi: 10.1080/17415990802105770

Franzosa, E. A., McIver, L. J., Rahnavard, G., Thompson, L. R., Schirmer, M., Weingart, G., et al. (2018). Species-level functional profiling of metagenomes and metatranscriptomes. *Nat. Methods* 15, 962–968. doi: 10.1038/s41592-018-0176-y

Ghilamicael, A. M., Budambula, N. L. M., and Anami, S. E. (2017). Evaluation of prokaryotic diversity of five hot springs in Eritrea. *BMC Microbiol.* 17:203. doi: 10.1186/s12866-017-1113-4

Ghosh, W., Mallick, S., Haldar, P. K., Pal, B., Maikap, S. C., and Gupta, S. K. D. (2012). Molecular and cellular fossils of a mat-like microbial community in geothermal boratic sinters. *Geomicrobiol. J.* 29, 879–885. doi: 10.1080/01490451.2011.635761

Gomez-Alvarez, V., Teal, T. K., and Schmidt, T. M. (2009). Systematic artifacts in metagenomes from complex microbial communities. *ISME J.* 3, 1314–1317. doi: 10.1038/ismej.2009.72

Gonsior, M., Hertkorn, N., and Hinman, N. (2018). Yellowstone hot springs are organic chemodiversity hot spots. *Sci. Rep.* 8:14155. doi: 10.1038/s41598-018-32593-x

Grim, F., Franz, B., and Dahl, C. (2011). Regulation of dissimilatory sulfur oxidation in the purple sulfur bacterium *Allochromatium vinosum*. *Front. Microbiol.* 2:51. doi: 10.3389/fmicb.2011.00051

Haferburg, G., and Kothe, E. (2007). Microbes and metals: interactions in the environment. *J. Basic Microbiol.* 47, 453–467. doi: 10.1002/jobm.200700275

Haft, D. H., Selengut, J. D., Richter, R. A., Harkins, D., Basu, M. K., and Beck, E. (2012). TIGRFAMs and genome properties in 2013. *Nucleic Acids Res.* 41, 387–D395. doi: 10.1093/nar/gks1234

Hipp, W. M., Pott, A. S., Thum-Schmitz, N., Faath, I., Dahl, C., and Truper, H. G. (1997). Towards the phylogeny of APS reductases and sirohaem sulfite reductases in sulfate-reducing and sulfur-oxidizing prokaryotes. *Microbiology* 143, 2891–2902. doi: 10.1099/00221287-143-9-2891

Huerta-Cepas, J., Forslund, K., Coelho, L. P., Szklarczyk, D., Jensen, L. J., and Von Mering, C. (2017). Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Mol. Biol. Evol.* 34, 2115–2122. doi: 10.1093/molbev/msx148

Hyatt, D., Chen, G. L., Locascio, P. F., Land, M. L., Larimer, F. W., and Hauser, L. J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119. doi: 10.1186/1471-2105-11-119

Inkscape Project (2020). Available online at: https://inkscape.org/en/

Inskeep, W. P., Rusch, D. B., Jay, Z. J., Herrgard, M. J., Kozubal, M. A., Richardson, T. H., et al. (2010). Metagenomes from high-temperature chemotrophic systems reveal geochemical controls on microbial community structure and function. *PLoS One* 5:e9773. doi: 10.1371/journal.pone.0009773

Jaekel, U., Zedelius, J., Wilkes, H., and Musat, F. (2015). Anaerobic degradation of cyclohexane by sulfate-reducing bacteria from hydrocarbon-contaminated marine sediments. *Front. Microbiol.* 6:116. doi: 10.3389/fmicb.2015.00116

Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y., and Hattori, M. (2004). The KEGG resource for deciphering the genome. *Nucleic Acids Res.* 32, 277–280. doi: 10.1093/nar/gkh063

Karkhoff-Schweizer, R. R., Huber, D. P., and Voordouw, G. (1995). Conservation of the genes for dissimilatory sulfite reductase from *Desulfovibrio vulgaris* and *Archaeoglobus fulgidus* allows their detection by PCR. *Appl. Environ. Microbiol.* 61, 290–296. doi: 10.1128/aem.61.1.290-296.1995

Klein, M., Friedrich, M., Roger, A. J., Hugenholtz, P., Fishbain, S., and Abicht, H. (2001). Multiple lateral transfers of dissimilatory sulfite reductase genes between major lineages of sulfate-reducing prokaryotes. *J. Bacteriol.* 183, 6028–6035. doi: 10.1128/JB.183.20.6028-6035.2001

Kojima, H., Watanabe, M., and Fukui, M. (2017). *Sulfuritortus calidifontis* gen. nov., sp. nov., a sulfur oxidizer isolated from a hot spring microbial mat. *Int. J. Syst. Evol. microbiol.* 67, 1355–1358. doi: 10.1099/ijsem.0.001813

Kolde, R., and Kolde, M. R. (2015). *Package 'Pheatmap'. R Package* 1, 790.

Kumar, S. S., Malyan, S. K., Basu, S., and Bishnoi, N. R. (2017). Syntrophic association and performance of *Clostridium*, *Desulfovibrio*, *Aeromonas* and *Tetrathiobacter* as anodic biocatalysts for bioelectricity generation in dual chamber microbial fuel cell. *Environ. Sci. Pollut. Res. Int.* 24, 16019–16030. doi: 10.1007/s11356-017-9112-4

Kushkevych, I., Abdulina, D., Kováè, J., Dordeviæ, D., Vítizov, M., Iutynska, G., et al. (2020). Adenosine-5'-Phosphosulfate-and sulfite reductases activities of sulfate-reducing bacteria from various environments. *Biomolecules* 10:921. doi: 10.3390/biom10060921

Larsen, O., Lien, T., and Birkeland, N. K. (1999). Dissimilatory sulfite reductase from *Archaeoglobus profundus* and *Desulfotomaculum thermocisternum*: phylogenetic and structural implications from gene sequences. *Extremophiles* 3, 63–70. doi: 10.1007/s007920050100

Lartillot, N., Rodrigue, N., Stubbs, D., and Richer, J. (2009). Phylobayes-MPI 1.7: a Bayesian software for phylogenetic reconstruction using mixture models. *Syst. Biol.* 62, 611–615. doi: 10.1093/sysbio/syt022

Li, S. J., Hua, Z. S., and Huang, L. N. (2014). Microbial communities evolve faster in extreme environments. *Sci. Rep.* 4:6205. doi: 10.1038/srep06205

Li, W., and Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22, 1658–1659. doi: 10.1093/bioinformatics/btl158

López, M. A., Zavala-Díaz de la Serna, F. J., Jan-Roblero, J., Romero, J. M., and Hernández-Rodríguez, C. (2006). Phylogenetic analysis of a biofilm bacterial population in a water pipeline in the Gulf of Mexico. *FEMS Microbiol. Ecol.* 58, 145–154. doi: 10.1111/j.1574-6941.2006.00137

Loy, A., Duller, S., and Wagner, M. (2008). "Evolution and ecology of microbes dissimilating sulfur compounds: insights from siroheme sulfite reductases," in *Microbial Sulfur Metabolism*, ed. C. Dahl (Berlin: Springer), 46–59. doi: 10.1007/978-3-540-72682-1_5

Loy, A., Duller, S., Baranyi, C., Mussmann, M., Ott, J., Sharon, I., et al. (2009). Reverse dissimilatory sulfite reductase as phylogenetic marker for a subgroup of sulfur-oxidizing prokaryotes. *Environ. Microbiol.* 11, 289–299. doi: 10.1111/j.1462-2920

Meckenstock, R. U., Boll, M., Mouttaki, H., Koelschbach, J. S., Cunha Tarouco, P., Weyrauch, P., et al. (2016). Anaerobic degradation of benzene and polycyclic aromatic hydrocarbons. *J. Mol. Microbiol. Biotechnol.* 26, 92–118. doi: 10.1159/000441358

Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E. M., Kubal, M., et al. (2008). The metagenomics RAST server – a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* 9:386. doi: 10.1186/1471-2105-9-386

Kamarisima, Miyanaga, K., and Tanji, Y. (2019). The utilization of aromatic hydrocarbon by nitrate- and sulfate-reducing bacteria in single and multiple nitrate injection for souring control. *Biochem. Eng. J.* 143, 75–80. doi: 10.1016/j.bej.2018.12.006

Mothe, G. K., Pakshirajan, K., and Das, G. (2016). Heavy metal removal from multicomponent system by sulfate reducing bacteria: mechanism and cell surface characterization. *J. Hazard. Mater.* 324, 62–70. doi: 10.1016/j.jhazmat.2015.12.042

Mulla, S. I., Talwar, M. P., and Ninnekar, H. Z. (2014). "Bioremediation of 2, 4, 6-trinitrotoluene explosive residues," in *Biological Remediation of Explosive Residues*, ed. S. N. Singh (Cham: Springer), 201–233. doi: 10.1007/978-3-319-01083-0

Müller, A. L., Kjeldsen, K. U., Rattei, T., Pester, M., and Loy, A. (2015). Phylogenetic and environmental diversity of DsrAB-type dissimilatory (bi)sulfite reductases. *ISME J.* 9, 1152–1165. doi: 10.1038/ismej.2014.208

Nagar, S., Talwar, C., Bharti, M., Yadav, S., Siwach, S., and Negi, R. K. (2021). Metagenome-assembled genomes recovered from the datasets of a high-altitude Himalayan hot spring Khirganga, Himachal Pradesh, India. *Data Brief.* 39:107551. doi: 10.1016/j.dib.2021.107551

Nazina, T. N., Sokolova, D. Sh, Shestakova, N. M., Grigor'ian, A. A., Mikhailova, E. M., Babich, T. L., et al. (2005). The phylogenetic diversity of aerobic organotrophic bacteria from the Dagan high-temperature oil field. *Mikrobiologiia* 74, 401–409.

Peng, Y., Leung, H. C., Yiu, S. M., and Chin, F. Y. (2012). IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* 28, 1420–1428. doi: 10.1093/bioinformatics/bts174

Picard, A., Gartman, A., and Girguis, P. R. (2016). What do we really know about the role of microorganisms in iron sulfide mineral formation? *Front. Earth Sci.* 4:68. doi: 10.3389/feart.2016.00068

Podar, P. T., Yang, Z., Björnsdóttir, S. H., and Podar, M. (2020). Comparative analysis of microbial diversity across temperature gradients in hot springs from Yellowstone and Iceland. *Front. Microbiol.* 11:1625. doi: 10.3389/fmicb.2020.01625

Poddar, A., and Das, S. K. (2018). Microbiological studies of hot springs in India: a review. *Arch. Microbiol.* 200, 1–18. doi: 10.1007/s00203-017-1429-3

Pohlner, M., Dlugosch, L., Wemheuer, B., Mills, H., Engelen, B., and Reese, B. K. (2019). The majority of active *Rhodobacteraceae* in marine sediments belong to uncultured genera: a molecular approach to link their distribution

to environmental conditions. *Front. Microbiol.* 10:659. doi: 10.3389/fmicb.2019.00659

R Development Core Team (2011). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Roy, C., Rameez, M. J., and Haldar, P. K. (2020). Microbiome and ecology of a hot spring-microbialite system on the Trans-Himalayan Plateau. *Sci. Rep*. 10:5917. doi: 10.1038/s41598-020-62797-z

Sahinkaya, E., Yurtsever, A., Toker, Y., Elcik, H., Cakmaci, M., and Kaksonen, A. H. (2015). Biotreatment of As-containing simulated acid mine drainage using laboratory scale sulfate reducing upflow anaerobic sludge blanket reactor. *Miner. Eng.* 75, 133–139. doi: 10.1016/j.mineng.2014.08.012

Sangwan, N., Lambert, C., Sharma, A., Gupta, V., Khurana, P., Khurana, J. P., et al. (2015). Arsenic rich Himalayan hot spring metagenomics reveal genetically novel predator-prey genotypes. *Environ. Microbiol. Rep.* 7, 812–823. doi: 10.1111/1758-2229.12297

Saxena, R., Dhakan, D. B., Mittal, P., Waiker, P., Chowdhury, A., Ghatak, A., et al. (2017). Metagenomic analysis of hot springs in Central India reveals hydrocarbon degrading thermophiles and pathways essential for survival in extreme environments. *Front. Microbiol.* 7:2123. doi: 10.3389/fmicb.2016.02123

Seemann, T. (2014). Prokka*:* rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069. doi: 10.1093/bioinformatics/btu153

Selvarajan, R., Sibanda, T., and Tekere, M. (2018). Thermophilic bacterial communities inhabiting the microbial mats of "indifferent" and chalybeate (iron-rich) thermal springs: diversity and biotechnological analysis. *Microbiol. Open* 7:e00560. doi: 10.1002/mbo3.560

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi: 10.1101/gr.1239303

Sharma, A. K., Sharma, R., and Dandi, H. R. (2004). *Mineral Resources of Himachal Pradesh*. Available online at: https://inkscape.org/en/

Shirkot, P., and Verma, A. (2015). Assessment of thermophilic bacterial diversity of thermal springs of Himachal Pradesh. *ENVIS Bull. Himal. Ecol.* 23, 27–34. doi: 10.1007/s40009-018-0676-4

Singh, A., and Subudhi, E. (2016). Structural insights of microbial community of Deulajhari (India) hot spring using 16s-rRNA based metagenomic sequencing. *Genom. Data* 7, 101–102. doi: 10.1016/j.gdata.2015.12.004

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033

Stasik, S., Wick, L. Y., and Wendt-Potthoff, K. (2015). Anaerobic BTEX degradation in oil sands tailings ponds: impact of labile organic carbon and sulfate-reducing bacteria. *Chemosphere* 138, 133–139. doi: 10.1016/j.chemosphere.2015.05.068

Suzek, B. E., Wang, Y., Huang, H., McGarvey, P. B., and Wu, C. H. (2015). UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches. *Bioinformatics* 31, 926–932. doi: 10.1093/bioinformatics/btu739

Talwar, C., Nagar, S., Kumar, R., Scaria, J., Lal, R., and Negi, R. K. (2020). Defining the environmental adaptations of genus *Devosia*: insights into its expansive short peptide transport system and positively selected genes. *Sci. Rep.* 10:1151. doi: 10.1038/s41598-020-58163-8

Tavaré, S. (1986). Some probabilistic and statistical problems in the analysis of DNA sequences, lectures on mathematics in the life sciences. *Am. Math. Soc.* 17, 57–86.

Truong, D. T., Franzosa, E. A., Tickle, T. L., Scholz, M., Weingart, G., Pasolli, E., et al. (2015). MetaPhlAn2 for enhanced metagenomic taxonomic profiling. *Nat. Methods* 12, 902–903. doi: 10.1038/nmeth.3589

van Gemerden, H. (1993). Microbial mats: a joint venture. *Mar. Geol.* 113, 3–25. doi: 10.1016/0025-3227(93)90146-M

Varin, T., Lovejoy, C., Jungblut, A. D., Vincent, W. F., and Corbeil, J. (2010). Metagenomic profiling of Arctic microbial mat communities as nutrient scavenging and recycling systems. *Limnol. Oceanogr.* 55, 1901–1911. doi: 10.4319/lo.2010.55.5.1901

Vermeij, P., and Kertesz, M. (1999). Pathways of Assimilative Sulfur Metabolism in *Pseudomonas putida*. *J. Bacteriol.* 181, 5833–5837. doi: 10.1128/JB.181.18.5833-5837.1999

Wagner, M., Roger, A. J., Flax, J. L., Brusseau, G. A., and Stahl, D. A. (1998). Phylogeny of dissimilatory sulfite reductases supports an early origin of sulfate respiration. *J. Bacteriol.* 180, 2975–2982. doi: 10.1128/JB.180.11.2975-2982.1998

Wickham, H. (2009). Elegant graphics for data analysis. *Media* 35, 10–1007.

Wright, K. E., Williamson, C., Grasby, S. E., Spear, J. R., and Templeton, A. S. (2013). Metagenomic evidence for sulfur lithotrophy by Epsilonproteobacteria as the major energy source for primary productivity in a sub-aerial arctic glacial deposit, Borup Fiord Pass. *Front. Microbiol.* 4:63. doi: 10.3389/fmicb.2013.00063

Wu, B., Liu, F., Fang, W., Yang, T., Chen, G. H., He, Z., et al. (2021). Microbial sulfur metabolism and environmental implications. *Sci. Total Environ.* 15:778. doi: 10.1016/j.scitotenv.2021.146085

Yang, J., and Zhang, Y. (2015). Protein structure and function prediction using I-TASSER. *Curr. Protoc. Bioinform.* 52, 5.8.1–5.815. doi: 10.1002/0471250953.bi0508s52

Zavarzin, G. A. (2008). *Microbial Cycles, Encyclopedia of Ecology*. Cambridge, MA: Academic Press, 2335–2341. doi: 10.1016/B978-008045405-4.00745-X

Zhang, J., Tian, Y., Zhang, J., Li, N., Kong, L., Yu, M., et al. (2016). Distribution and risk assessment of heavy metals in sewage sludge after ozonation. *Environ. Sci. Pollut. Res.* 24, 5118–5125. doi: 10.1007/s11356-016-6313-1

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: 03 June 2022
doi: 10.3389/fmicb.2022.872576

Insight Into the Molecular Mechanisms Underpinning the Mycoremediation of Multiple Metals by Proteomic Technique

*Priyadarshini Dey[1,2,3], Anushree Malik[1], Dileep Kumar Singh[4], Sven-Bastiaan Haange[2], Martin von Bergen[2,5,6] and Nico Jehmlich[2]**

[1] Applied Microbiology Lab, Centre for Rural Development and Technology, Indian Institute of Technology Delhi, New Delhi, India, [2] Department of Molecular Systems Biology, Helmholtz Centre for Environmental Research, Helmholtz Association of German Research Centres (HZ), Leipzig, Germany, [3] Department of Biotechnology, MS Ramaiah Institute of Technology, Bengaluru, India, [4] Department of Zoology, Faculty of Science, University of Delhi, New Delhi, India, [5] Institute of Biochemistry, Faculty of Biosciences, Pharmacy and Psychology, University of Leipzig, Leipzig, Germany, [6] German Centre for Integrative Biodiversity, Research (iDiv) Halle-Jena-Leipzig, Leipzig, Germany

OPEN ACCESS

Edited by:
Roshan Kumar, Magadh University, India

Reviewed by:
M. Sudhakara Reddy, Thapar Institute of Engineering & Technology, India
Surajit Das, National Institute of Technology Rourkela, India

***Correspondence:**
Nico Jehmlich
nico.jehmlich@ufz.de

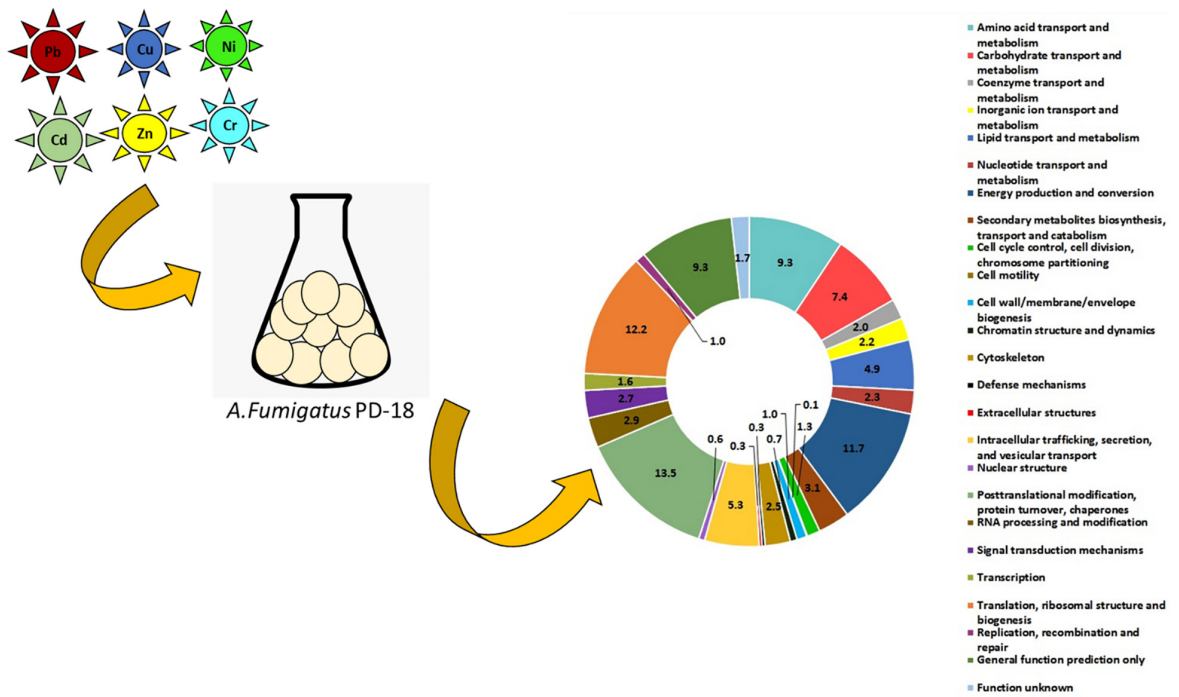
Specialty section:
This article was submitted to Evolutionary and Genomic Microbiology, a section of the journal Frontiers in Microbiology

Received: *09 February 2022*
Accepted: *07 April 2022*
Published: *03 June 2022*

Citation:
Dey P, Malik A, Singh DK, Haange S-B, von Bergen M and Jehmlich N (2022) Insight Into the Molecular Mechanisms Underpinning the Mycoremediation of Multiple Metals by Proteomic Technique. Front. Microbiol. 13:872576. doi: 10.3389/fmicb.2022.872576

We investigated the fungus *Aspergillus fumigatus* PD-18 responses when subjected to the multimetal combination (Total Cr, Cd^{2+}, Cu^{2+}, Ni^{2+}, Pb^{2+}, and Zn^{2+}) in synthetic composite media. To understand how multimetal stress impacts fungal cells at the molecular level, the cellular response of *A. fumigatus* PD-18 to 30 mg/L multimetal stress (5 mg/L of each heavy metal) was determined by proteomics. The comparative fungal proteomics displayed the remarkable inherent intracellular and extracellular mechanism of metal resistance and tolerance potential of *A. fumigatus* PD-18. This study reported 2,238 proteins of which 434 proteins were exclusively expressed in multimetal extracts. The most predominant functional class expressed was for cellular processing and signaling. The type of proteins and the number of proteins that were upregulated due to various stress tolerance mechanisms were post-translational modification, protein turnover, and chaperones (42); translation, ribosomal structure, and biogenesis (60); and intracellular trafficking, secretion, and vesicular transport (18). In addition, free radical scavenging antioxidant proteins, such as superoxide dismutase, were upregulated upto 3.45-fold and transporter systems, such as protein transport (SEC31), upto 3.31-fold to combat the oxidative stress caused by the multiple metals. Also, protein–protein interaction network analysis revealed that cytochrome c oxidase and 60S ribosomal protein played key roles to detoxify the multimetal. To the best of our knowledge, this study of *A. fumigatus* PD-18 provides valuable insights toward the growing research in comprehending the metal microbe interactions in the presence of multimetal. This will facilitate in development of novel molecular markers for contaminant bioremediation.

Keywords: multi-metal, proteomics, mycoremediation, LC-MS/MS, fungi



GRAPHICAL ABSTRACT |

HIGHLIGHTS

- First proteomic study on filamentous fungus *Aspergillus fumigatus* for hexametal removal.
- Most upregulated class posttranslational modification, protein turnover, chaperones.
- Major protein player in protein-protein interaction identified as cytochrome oxidase.
- Expressed proteins are environmental biomarkers in heavy metals-fungus interaction.

INTRODUCTION

Heavy metals, such as copper (Cu), chromium (Cr), cadmium (Cd), zinc (Zn), lead (Pb), and nickel (Ni), occur in the water bodies, such as river water, and drains, in developing countries that are above the permissible mandates as prescribed by the Food and Agriculture Organization (Bhattacharya et al., 2015). These heavy metals are released from numerous small- and medium-scale enterprises (pesticide, textile, electroplating, fertilizer, batteries, etc.) due to lack of or improper sewage treatment plant systems (Bhattacharya et al., 2015; Bhardwaj et al., 2017; Gola et al., 2020). These hazardous contaminants are discharged into various water bodies *via* irrigation during agricultural activities. This results in bioaccumulation of these contaminants that enter the food chain and cause detrimental health ailments, such as cancer in human beings (Kou et al., 2018; Li et al., 2019; Abed et al., 2020; Chauhan et al., 2021; Irawati et al., 2021). Thus, these heavy metals are the environmental priority contaminants threatening the environment and therefore must be remediated before discharge into the environment.

The conventional physico-chemical methods, such as chemical precipitation, ion exchange, adsorption, membrane filtration, coagulation-flocculation, and flotation, are usually utilized to remediate these harmful contaminants. These techniques have the disadvantages of being expensive, having low selectivity, production of additional sludge, and further treatment is required for better results (Dey et al., 2021; Zamora-Ledezma et al., 2021).

On the other hand, living and actively growing microbial cells can be a lucrative option for bioremediation (Malik, 2004). Of the bioremediation techniques, mycoremediation has shown to be a promising technology having the potential to ameliorate these hazardous chemicals (Dey et al., 2020; Sabuda et al., 2020). The superiority of fungi over single-celled microbes, such as bacteria, to remediate these recalcitrant heavy metals is well-documented (Deshmukh et al., 2016). In addition, the fungi are omnipresent, multifarious, and have a wider arsenal to acclimatize to environmental limitations, such as immoderations of temperature, extremes of pH, higher metal concentrations, and low nutrient accessibility, due to their morphological diversity (Anand et al., 2006). Besides, fungal mycelia have enhanced enzymatic and mechanical contact with the pollutant due to a greater cell to the surface ratio (Sagar and Singh, 2011).

Further, in microbial cells, such as fungus, heavy metals are key components in the number of catalytic and structural proteins that are integral to biochemical processes. The outcome of these processes differs depending on the type of metal involved

and its concentration inside the cell (Gadd, 1994). Particularly, the species of *Aspergillus* have a high metal uptake capacity for metals, such as Cu (Dusengemungu et al., 2020). Also, filamentous fungi develop signature metabolic pathways which are species-specific to survival in the harsh environment of heavy metals and other contaminants that are utilized as nutrients and energy sources (Kuhn and Käufer, 2003). Accordingly, proteins expressed in cells under such diverse conditions and at different times are dissimilar (Boopathy, 2000; Keller, 2015; Wisecaver and Rokas, 2015).

Also, there is evidence that fungal resistance toward one element does not necessarily infer resistance to another element even though the elements possess similar valency charges (Høiland, 1995). For example, the toxicity of Pb toward the microbe is less compared to other toxic metals, such as Cd, As, and Hg (Jiang et al., 2020). Further, in fungi, the metallothioneins formation is predominantly induced by the heavy metal Cu (Jaeckel et al., 2005). Thus, understanding the altered heavy metal uptake in the fungus in presence of diverse environmental conditions needs to be evaluated.

To gain such mechanistic insight, high throughput techniques, such as proteomic analysis by LC-MS/MS, can be used to identify and characterize proteins involved in the multimetal resistance mechanism in a filamentous fungus (Pandey and Mann, 2000; Kraut et al., 2009; Zhang et al., 2010). Moreover, proteomics facilitates the development of new and important protein biomarkers that specify and monitor metal contamination in the environments as protein type and also their copy numbers are estimated by the translational regulation (Ohno et al., 2014).

Several researchers have studied the change in proteomes under various metal conditions. Selamoglu et al. (2020) have assessed the change in the differential proteome expression of the prokaryotic organism *Rhodobacter capsulatus* in presence of 5 μM Cu by nano-LC-MS/MS. About 75 proteins were significantly regulated. Most of the proteins present were responsible for maintaining Cu homeostasis. Further, Lotlikar et al. (2020) enumerated the variable expression of proteins in a eukaryotic fungus *Penicillium chrysogenum* in the presence of 100 mg/L and 500 mg/L Cu. Several key proteins related to genetic information, carbohydrate metabolism, glycan biosynthesis and metabolism, amino acid metabolism, and energy metabolism were expressed.

Cherrad et al. (2012) highlighted the overaccumulation of proteins of the oxidoreductase family when exposed to Cd, Cu, and Ni but not when exposed to Zn. Thus, the secretion of proteins in fungus is extremely dynamic and its production depends on different environmental triggers.

However, despite these growing proteomic studies dealing with the bioremediation of single heavy metal, there is a dearth of information on the modulated proteins triggered by the cumulative toxicity of hexametals, namely, Cd, Cr, Cu, Ni, Pb, and Zn, specifically in a filamentous fungus. Interestingly, *Aspergillus fumigatus* has high adaptability when subjected to an altered environment (Bakti et al., 2018). Also, as established in our previous works (Dey et al., 2016, 2020; Bhattacharya et al., 2020), *A. fumigatus* PD-18 is a filamentous fungus capable of removal of 30 mg/L multimetal exceptionally well. Furthermore, there are structural and functional similarities between the numerous genes of lower eukaryotes, such as fungi and mammals (Bae and Chen, 2004).

Thus, this fungus *A. fumigatus* PD-18 would be a good eukaryotic model for helping us understand how cells adopt various cellular strategies and lay a foundation study to decipher the enzymes produced in the presence of simultaneous effects of multimetal cocktail on interaction with a fungus for scale-up process.

MATERIALS AND METHODS

Chemicals and Reagents

The stock solutions (10 g/L) of different individual metals were made by dissolving their respective salts, viz. $K_2Cr_2O_7$, $Cd(NO_3)_2$, $Ni(NO_3)_2$, $Cu(NO_3)_2$, $Zn(NO_3)_2$, and $Pb(CH_3COO)_2$ in double-distilled water and were diluted to the concentrations that were required for the experiments. For preparing the reagents and calibration standards, deionized ultrapure water (RIONS Ultra 370 series) was used. Rest all other chemicals utilized were of analytical grade and were obtained from Merck, Sigma, and Qualigens.

Microorganism and Culture Media Composition

The fungal strain used was *A. fumigatus* PD-18 which was isolated from the polluted banks of the river Yamuna, New Delhi, India, and characterized with accession number KX365202 after depositing the sequence to the Genbank (NCBI) (Dey et al., 2016).

The growth media used was the composite medium with following composition (g/L): (NH_4NO_3, 0.5; K_2HPO_4, 0.5; $MgSO_4.7H_2O$, 0.1; NaCl, 1.0; Yeast extract, 2.5; pH 6.8 ± 0.2). The sterilization of the media was carried out at 121°C for 15 min. Glucose (10.0 g/L) was added to the flasks separately after autoclaving to avoid precipitation.

Methodology

The methodology adopted in this study was to estimate the differential expression of proteins in *A. fumigatus* PD18 brought about by 30 mg/L multimetal (MM) viz. Five milligram per liter of each of the individual Cd, Total Cr, Cu, Ni, Pb, and Zn amended in the composite media in addition to 1% of glucose. The concentration of 5 mg/L of each heavy metal was chosen to take into consideration the permissible mandates for heavy metal occurrence in water that can be utilized for irrigation according to Food and Agriculture Organization (FAO). The prescribed limit for each heavy metal is as follows: Cd, 0.01; Total Cr, 0.1; Cu, 0.2; Ni, 0.2; Pb, 5.0; Zn, 2.0. Also, the typical concentration of heavy metals occurring in a mixture in the Yamuna river was considered (Bhattacharya et al., 2015). The biotic control included composite media added with only 1% of glucose.

Studies were conducted in a series of Erlenmeyer flasks (250 mL) comprising 100 mL of composite growth media. One milliliter of spore suspension (having a concentration of

10^7 spores eluted with sterile distilled water containing 0.01% Tween 80) was inoculated in the flasks and incubated at 30°C and 150 rpm agitation for 72 h to ensure complete uptake of metal ions. The flasks were withdrawn after the late log phase incubation period. Analysis was done in three technical replicates for biotic control and multimetal treated samples separately.

Proteomics Analysis

Cell Lysis and Protein Extraction

The fungal pellets were separated from the media by inversion and followed by centrifugation at 4,000 g for 10 min at 4°C. The mycelia were further snap-freezed in liquid nitrogen and then lyophilized and stored as a dry powder at −20°C before the subsequent steps. The total protein from the dried powdered fungal cells was extracted in a buffer of composition (8M Urea and 2M Thiourea) as per the modified protocol (Rughöft et al., 2020). The supernatant was collected by centrifugation at 14,000 rpm, 4°C for 10 min, and precipitated overnight in five volumes of ice-cold acetone and the pellet was stored at −20°C for further use. The concentration of protein to be measured was estimated according to the Bradford assay of protein quantification (Bradford, 1976). In brief, a 2-mg/mL concentration stock solution of Bovine Serum Albumin (BSA) was prepared in water. BSA standard curve was prepared using the following concentration range (0–2.00 μg/μL) followed by the addition of Coomassie Brilliant Blue dye reagent in 95% ethanol and 100 mL 85% (w/v) phosphoric acid and incubated for 5 min. The absorbance was measured at 595 nm.

Protein Separation by 1D-SDS PAGE and in Gel Tryptic Digestion of the Proteins

The separation of the proteins was carried out by one-dimensional SDS-PAGE and the separated protein bands were visualized by Coomassie brilliant blue staining. The individual gel bands were cut into small pieces and washed with 10 mM ammonium bicarbonate followed by reduction with 10 mM 1,4-Dithiothreitol and 100 mM 2-Iodoacetamide. The proteins lysates were subjected to tryptic digestion and incubated overnight and kept at 37°C. The peptides were extracted from the gel pieces in 5 mM ammonium bicarbonate and further purified by C-18 ZipTip columns according to the protocol (Haange et al., 2012). Thereafter, the peptide lysates were dried in a vacuum centrifuge and stored at −20°C until further analysis.

Mass Spectrometric Analysis

The dried peptides lysates were reconstituted in 0.1% (v/v) formic acid and loaded and separated by reversed-phase chromatography before being measured on mass spectrometer UPLC-coupled (Waters) LTQ Orbitrap Velos MS/MS (Thermo Fisher Scientific).

Data Analysis (Identification and Characterization of the Proteins)

The raw MS-MS ion spectra data from the instrument were processed using the Proteome Discoverer software (v1.0 build 43, Thermo Fisher Scientific) (Orsburn, 2021). The fixed and variable modifications (Carbamidomethylation at cysteines was given as fixed and oxidation of methionines as variable modifications) were taken into consideration. Only ranked one peptide hits with ≥1 high confidence and less than 1% false discovery rate (FDR) were considered as identified and taken up for further analysis. The intensity data were further processed by logarithmic transformation and normalization for the determination of changes in the abundance of proteins (at least 1.5-fold upregulated or downregulated). Only the values of $p < 0.05$ were considered as statistically significant. Triplicate gels were run for biotic control and MM samples. For functional annotation of the proteins, the generated protein lists were analyzed by prophane.[1]

The creation of the VENN diagrams was done by the online tool VENNY.[2] The heatmaps and volcano plots were constructed using the heatmap and volcano programs from R tool version 3.2.0. The WoLF PSORT tool[3] was used for the prediction of the subcellular localization of the proteins. The protein–protein interactions between the differentially expressed proteins were studied by the STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) database.[4]

RESULTS

Fungi usually acclimate by making dynamic changes in the cell structure and composition when exposed to heavy metals. In our previous study on the toxicity of individual metals, that is, Cd, Cr, Cu, Ni, Pb, and Zn (500 mg/L) in *A. fumigatus* PD-18, we determined the highest tolerance index for each metal in solid media. We calculated the cube root growth (k) constant of *A. fumigatus* PD-18 when subjected to 30 mg/L multimetal in liquid composite media. This fungal strain had exceptional multimetal removal ability after ∼72 h. The growth profile of the fungus was altered when multimetal was added to the composite media and substantial variations were observed when compared with individual metals (Dey et al., 2016).

Further, the morphological changes in this fungus in response to the 30 mg/L multimetal were determined by scanning electron microscope, the localization of the heavy metals inside the fungal cell by transmission electron microscopy, and chelation of the heavy metals with the functional groups occurring in the fungus by Fourier-transform infrared spectroscopy (Dey et al., 2020).

Expression of Proteins in Biotic Control and Multimetal Extracts of *Aspergillus fumigatus* PD-18

From the six sample preparations (conditions, including one biological control and one multimetal exposure and their three technical replicates), 434 proteins were uniquely present in multimetal extracts and 400 proteins in biotic control of *A. fumigatus* PD-18.

[1] [1]https://www.prophane.de/

[2]https://bioinfogp.cnb.csic.es/tools/venny/

[3]https://wolfpsort.hgc.jp/

[4]https://string-db.org/

Figure 1 depicts the expression of the proteins isolated from *A. fumigatus* PD-18 mycelia as established with SDS-PAGE. This expression of proteins from the biotic control and multimetal (MM) stressed PD-18 were observed after Coomassie Brilliant Blue staining.

Proteins Present in High Abundances in *Aspergillus fumigatus* PD-18

Figure 2A depicts the heatmap of differently regulated proteins of the multimetal exposure when compared to the biotic control without multimetal. The detected proteins were categorized under three functional KOG groups related to metabolism, cellular process and signaling, and information storage and processing. These categories were associated with fungal physiology and most of the identified proteins were further classified on the basis of different 23 KOG classes which are discussed as follows:

- metabolism [amino acid transport and metabolism (9%), carbohydrate transport and metabolism (7%), coenzyme transport and metabolism (2%), inorganic ion transport and metabolism (2%), lipid transport and metabolism (5%), nucleotide transport and metabolism (3%), energy production and conversion (11%), and secondary metabolites biosynthesis and transport and catabolism (3%)];
- cellular process and signaling [cell cycle control and cell division and chromosome partitioning (1%), cell motility (0.1%), cell wall/membrane/envelope biogenesis (1%), chromatin structure and dynamics (1%), cytoskeleton (3%), defense mechanisms (0.3%), extracellular structures (0.3%), intracellular trafficking, secretion, and vesicular transport (6%), nuclear structure (1%), post-translation modification and protein turnover and chaperones (14%), RNA processing and modification (3%), and signal transduction mechanism (3%)];
- information storage and processing [transcription (2%), translation and ribosomal structure and biogenesis (12%), replication, recombination, and repair (1%), proteins with general function (9%), and function unknown (2%)].

From **Figure 2A**, the proteins that were expressed in higher abundance in *A. fumigatus* when treated with 30 mg/L multimetal belonged to the main metabolic functional classes of RNA processing & modification, post-translational modification & protein turnover & chaperones, nuclear structure, intracellular trafficking & secretion & vesicular transport, translation & ribosomal structure & biogenesis, and nucleotide transport & metabolism. Further, hypothetical proteins with general function prediction were also expressed which could play role in tackling the multimetal stress. The afflicted metabolic functions classes where production of proteins decreased belonged to amino acid transport & metabolism, carbohydrate transport & metabolism, lipid transport & metabolism, and secondary metabolites biosynthesis & transport and catabolism.

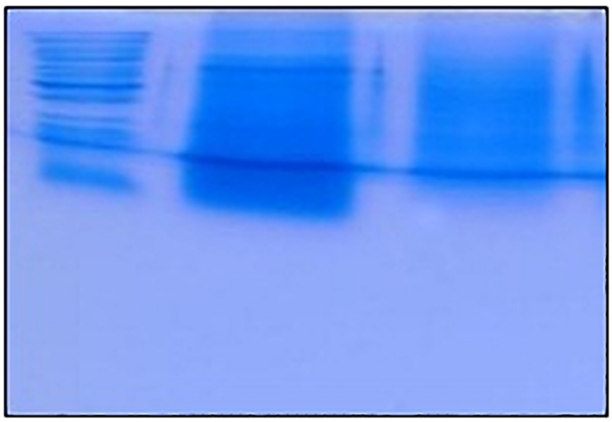


FIGURE 1 | Coomassie brilliant blue stained segment of SDS-PAGE protein expression patterns in heavy metal treated *Aspergillus fumigatus* in absence of multimetal (Biotic control in 2nd lane); 30 mg/L multimetal (3rd lane). Molecular-weight markers (M) are shown on the left (1st lane).

This proteomic study displayed proteins in *A. fumigatus* PD-18 that showed a significant fold increase and decrease of differentially expressed proteins during growth in 30 mg/L multimetal supplemented composite media as depicted in the volcano plot in **Figure 2B**. Among the differentially expressed proteins identified, 399 proteins were differentially upregulated while 266 proteins were differentially downregulated. The highly upregulated protein identified hydroxymethylglutaryl-CoA synthase upregulated upto 5.2-fold times. Other important proteins with general function prediction were upregulation of serine/threonine-protein phosphatase upto 3.42-fold times and G-protein beta subunit SfaD upto 2.98-fold. The highly upregulated and downregulated proteins by fold changes and their locations inside the fungal cell are depicted in **Table 1**. Furthermore, it was found that the majority of the upregulated proteins were located in the cytoplasm (18), mitochondria (12) followed by extracellular (6), nuclear (5), plasma membrane (2), peroxisome (2), nuclear and cytoplasmic (1), and endoplasmic reticulum lumen (1).

Enriched Protein Network and Pathways in *Aspergillus fumigatus* PD-18

The interaction network as constructed by the STRING database of the highly upregulated proteins in *A. fumigatus* is depicted in **Figure 3**. The strength of the association is depicted by the thickness of the line. The number of nodes is 263 that represents the proteins, the number of edges is 467 that represents associations, the average node degree is 3.55, and the average local clustering coefficient is 0.412. The STRING analysis identified three clusters of protein interactions. The first major cluster of proteins (green) was involved in ribosome and preinitiation factors (60S and 40S ribosomal proteins). The predicted partners functionally associated with 60S and 40S ribosomal proteins include translation elongation factor, fibrillarin of class translation, ribosomal structure, and biogenesis. The second cluster

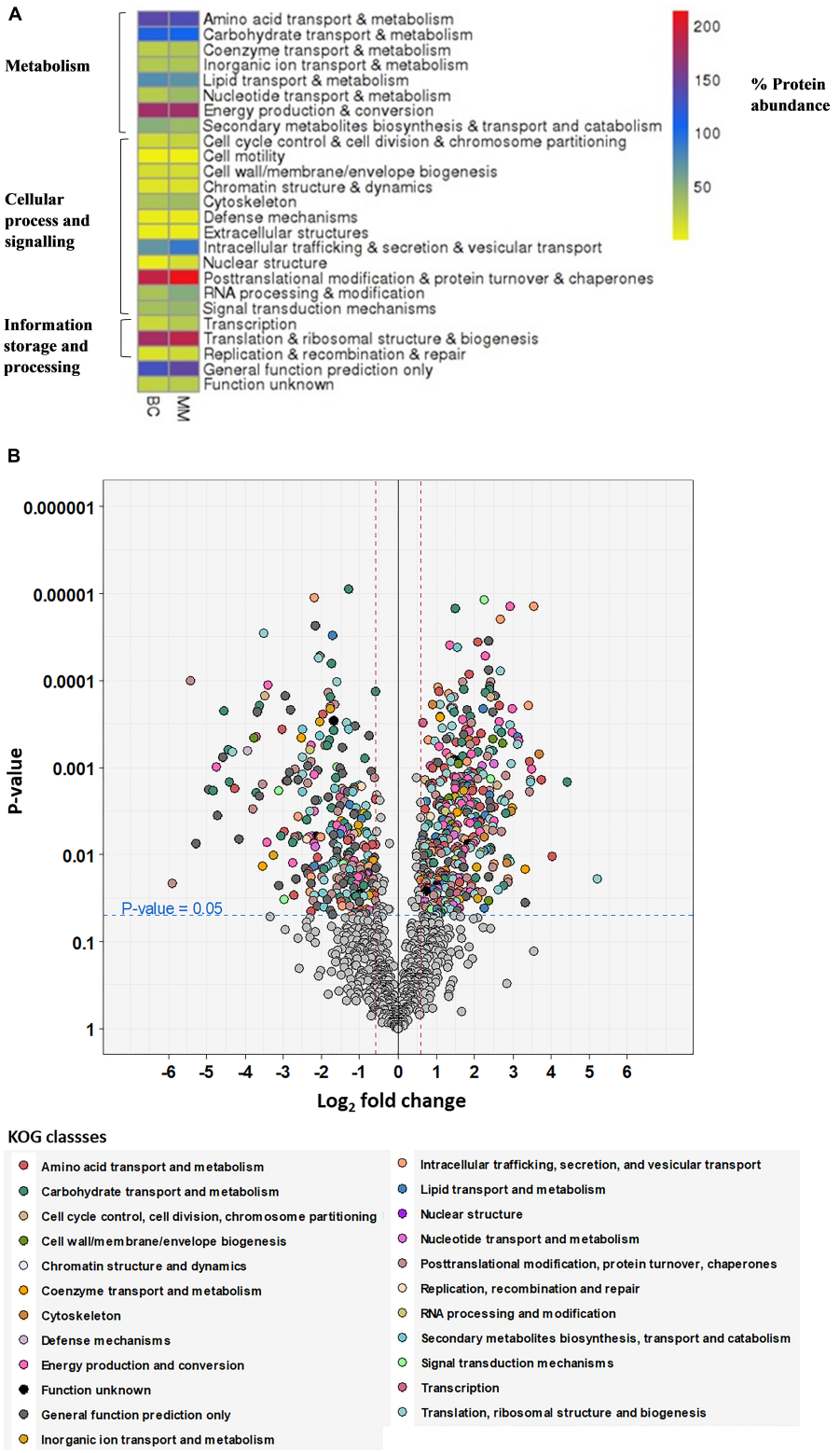


FIGURE 2 | (A) Heatmap visualization of the abundance of the proteins as red = higher values, dark blue = mid values, yellow = lower values depicted on (x-axis) for proteome abundance of 30 mg/L multimetal (MM) and proteome abundance without heavy metal exposure (BC). **(B)** Volcano plot showing the distribution of quantified proteins according to *p*-value and $\log_2$ fold change, indicating significance level at 0.05 (blue dashed line) and fold change at –1.5 and 1.5 (red dashed lines), colored according to their KOG class. Proteins, which were not significant or fold change within –1.5 to 1.5 are colored in gray.

TABLE 1 | Differential abundance of highly regulated selected proteins in *Aspergillus fumigatus* grown in presence 30 mg/L MM.

Protein accession number	Protein name	Fold change	Cellular location of protein	Protein KOG class	*p*-value
Upregulated proteins					
A1CQV3	hydroxymethylglutaryl-CoA synthase	5.19	Cytoplasm	Lipid transport and metabolism	0.019
B0XZB6	aminotransferase family protein	4.03	Mitochondria	Amino acid transport and metabolism	0.010
Q4WV94	proteasome regulatory particle subunit Rpt3	3.74	Peroxisome	Post-translational modification, protein turnover, chaperones	0.001
Q4WXX9	fibrillarin	3.68	Cytoplasm	RNA processing and modification	0.001
Q4X0E9	glutamate dehydrogenase	3.54	Mitochondria	Amino acid transport and metabolism	0.001
Q4WLA8	superoxide dismutase	3.45	Mitochondria	Inorganic ion transport and metabolism	0.001
A0A017SPG0	calcium/calmodulin-dependent protein kinase	3.45	Cytoplasm	Signal transduction mechanisms	0.001
A0A017SD45	ATP synthase	3.44	Mitochondria	Energy production and conversion	0.001
Q4WM81	serine/threonine-protein phosphatase	3.42	Nuclear	Signal transduction mechanisms	0.001
A1C6S1	histone H4	3.40	Mitochondria	Chromatin structure and dynamics	0.001
A0A084BYE7	protein transport protein (SEC31)	3.31	Mitochondria	Intracellular trafficking, secretion, and vesicular transport	0.036
A0A017S654	glutamate decarboxylase	3.16	Peroxisome	Amino acid transport and metabolism	0.001
A2R0I8	26S protease regulatory subunit	3.12	Nuclear and cytoplasmic	Post-translational modification, protein turnover, chaperones	0.001
B0XVU3	RNA binding effector protein Scp160	3.11	Endoplasmic reticulum lumen	Lipid transport and metabolism	0.001
O60022	probable HECT-type ubiquitin ligase	2.99	Extracellular	General function prediction only	0.002
A1CI51	G protein beta subunit SfaD	2.98	Cytoplasm	General function prediction only	0.001
Q4WLV1	26S proteasome regulatory particle subunit Rpn8	2.96	Extracellular	Post-translational modification, protein turnover, chaperones	0.001
Q4WFD3	coatomer subunit gamma	2.94	Cytoplasm	Intracellular trafficking, secretion, and vesicular transport	0.012
A0A017S6B4	phosphoglycerate kinase	2.91	Mitochondria	Carbohydrate transport and metabolism	0.001
Q4WX43	endosomal cargo receptor (P24)	2.86	Nucleus	Intracellular trafficking, secretion, and vesicular transport	0.020
A0A084C1K3	glutamyl-tRNA synthetase	2.86	Nucleus	Translation, ribosomal structure and biogenesis	0.011
A0A017S4K9	pyridoxine biosynthesis protein	2.82	Cytoplasm	Coenzyme transport and metabolism	0.001
B0XXM2	CBS and PB1 domain protein	2.73	Extracellular	Energy production and conversion	0.002
A1CPB0	sulfate adenylyltransferase	2.71	Cytoplasm	Inorganic ion transport and metabolism	0.001
A0A084BD28	importin beta 4 subunit	2.69	Nucleus	Nuclear structure	0.002
Q4WHV3	oxidoreductase short chain dehydrogenase	2.58	Plasma membrane	Secondary metabolites biosynthesis, transport and catabolism	0.001
Q4 × 0N5	T-complex protein 1 subunit gamma	2.5	Nucleus	Post-translational modification, protein turnover, chaperones	0.012
A0A084BP78	acetolactate synthase	2.49	Mitochondria	Amino acid transport and metabolism	0.005
Q0CSN0	Mitochondrial inner membrane translocase subunit TIM44	2.41	Mitochondria	Intracellular trafficking, secretion, and vesicular transport	0.002
Q4WE44	Cytochrome P450 phenylacetate 2-hydroxylase	2.39	Cytoplasm	Secondary metabolites biosynthesis, transport and catabolism	0.001
I7ZRN7	glyceraldehyde-3-phosphate dehydrogenase	2.38	Cytoplasm	Carbohydrate transport and metabolism	0.001
A4FSH5	RAB proteins geranylgeranyltransferase component A	2.37	Extracellular	Post-translational modification, protein turnover, chaperones	0.001
Q5B2V1	Peptidyl-prolyl *cis*-trans isomerase D	2.37	Cytoplasm	Post-translational modification, protein turnover, chaperones	0.001
B8NQD2	mitochondrial enoyl reductase	2.35	Cytoplasm	Transcription	0.033
Q4WGJ9	Tropomyosin	2.26	Mitochondria	Cytoskeleton	0.041
Q4WIT8	Ubiquitin DskB	2.13	Extracellular	Post-translational modification, protein turnover, chaperones	0.004
A0A084BHN1	Importin beta 5 subunit	2.02	Cytoskeleton	Intracellular trafficking, secretion, and vesicular transport	0.018

(Continued)

TABLE 1 | (Continued)

Protein accession number	Protein name	Fold change	Cellular location of protein	Protein KOG class	*p*-value
Upregulated proteins					
Q4WJB0	ribonucleoside-diphosphate reductase	1.93	Mitochondria	Nucleotide transport and metabolism	0.001
Q4W9K3	Mitochondrial outer membrane translocase receptor (TOM70)	1.84	Extracellular	Intracellular trafficking, secretion, and vesicular transport	0.001
G7XPM0	ATP-dependent bile acid permease	1.72	Cytoplasm	Secondary metabolites biosynthesis, transport and catabolism	0.014
Q4 × 237	translation release factor eRF3	1.71	Mitochondria	Translation, ribosomal structure and biogenesis	0.002
Q4WVZ4	isocitrate dehydrogenase LysB	1.69	Nucleus	Amino acid transport and metabolism	0.013
Q96UX3	14-alpha sterol demethylase	1.69	Cytoplasmic and nuclear	Secondary metabolites biosynthesis, transport and catabolism	0.001
A0A084C1V0	DNA damage inducible v SNARE binding protein Ddi1	1.65	Cytoplasm	Replication, recombination and repair	0.018
Q5B7M1	Vesicular-fusion protein sec17	1.65	Cytoplasm	Intracellular trafficking, secretion, and vesicular transport	0.001
A0A084BWQ6	Nuclear pore complex subunit (SEC13)	1.62	Cytoplasmic and nuclear	Intracellular trafficking, secretion, and vesicular transport	0.001
Q0CCZ9	bifunctional purine biosynthetic protein Ade1	1.60	Cytoplasm	Nucleotide transport and metabolism	0.020
A0A017SJP6	protein phosphatase 2a 65kd regulatory subunit	1.54	Plasma membrane	Signal transduction mechanisms	0.044
Q4WN17	septin	1.54	Cytoplasm	Cell cycle control, cell division, chromosome partitioning	0.001
A1CB85	transcription elongation factor spt4	1.49	Cytoplasm	Post-translational modification, protein turnover, chaperones	0.001
P20445	MICOS complex subunit mic60	1.39	Cytoplasm	Cell wall/membrane/envelope biogenesis	0.003
Q6MYF0	plasma membrane H +-ATPase Pma1	1.28	Cytoplasm	Inorganic ion transport and metabolism	0.005
Downregulated proteins					
Q4WHK1	Small nuclear ribonucleoprotein Sm D1	−0.59	Nucleus	Translation, ribosomal structure, and biogenesis	0.034
Q4WHE2	NADH-ubiquinone oxidoreductase	−0.59	Cytoplasm	Energy production and conversion	0.002
B0Y5W4	Pyruvate carboxylase	−0.61	Peroxisome	Post-translational modification, protein turnover, chaperones	0.029
G3XYJ1	GST C-terminal domain-containing protein	−0.63	Cytoplasm and nucleus	Amino acid transport and metabolism	0.011
Q4WCQ8	ATP synthase subunit E	−0.64	Nucleus	Energy production and conversion	0.012
Q4 × 0B7	Arginyl-tRNA synthetase	−0.68	Cytoplasm	Lipid transport and metabolism	0.012
A1CIL4	Histidine–tRNA ligase	−0.75	Mitochondria	Coenzyme transport and metabolism	0.029
B0XS44	Carbamoyl-phosphate synthase	−0.77	Mitochondria	Signal transduction mechanisms	0.013
Q6QNB5	Squalene monooxygenase	−0.78	Plasma membrane	Translation, ribosomal structure and biogenesis	0.010
Q4WY15	Arginase	−0.77	Mitochondria	Energy production and conversion	0.016

proteins (red) have major hub proteins, such as malate dehydrogenase, ATP synthase, pyruvate kinase, fructose biphosphate aldolase, dihydrolipoyl dehydrogenase, glutamate synthase, anthranilate synthase, and isocitrate dehydrogenase. The third cluster proteins (blue) are proteasome regulatory particles, Hsp70 chaperone, protein transport protein (sec31), and V-type proton ATPase.

The significant junction proteins in protein–protein interaction identified were cytochrome c, oxidase, and septin suggest that energy production and conversion and cell cycle control, cell division, and chromosome partitioning that play a pivotal role in the resistance toward multimetal exposure. **Figure 4** summarizes the salient mechanism of the multimetal detoxification process by *A. fumigatus* PD-18.

DISCUSSION

In this study, we identified the chief classes of proteins that are crucial for the resistance and tolerance of *A. fumigatus* PD-18 for multimetals, including intracellular and extracellular mechanisms of metal uptake. These two processes are further elaborated.

Intracellular and Extracellular Mechanism of Metal Uptake by Fungus

The mechanism of the fungal resistance and tolerance could be attributed to its occurrence at the metal-contaminated site (Baker, 1987; Gadd and White, 1993). The fungal detoxification

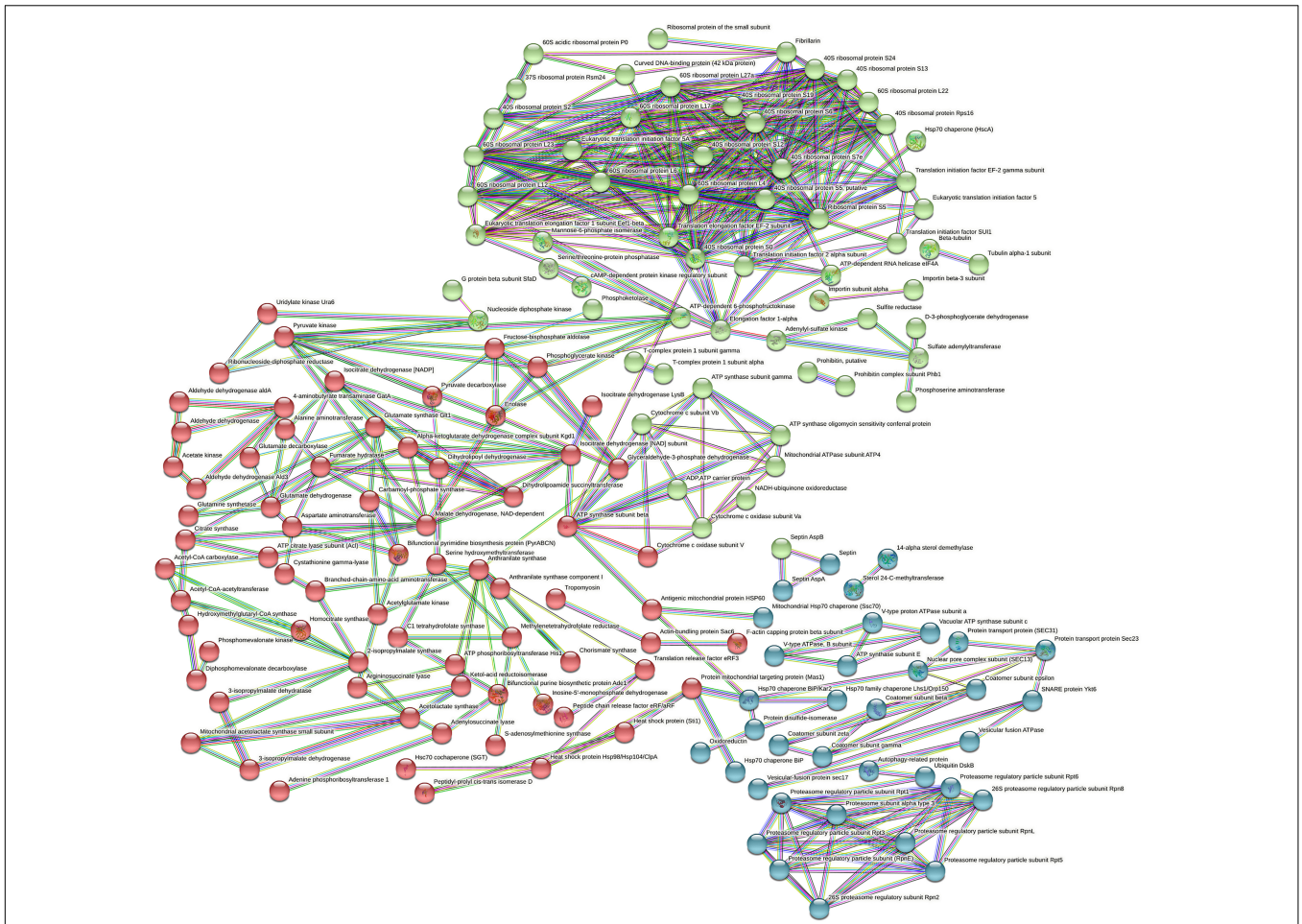


FIGURE 3 | STRING analysis displaying protein–protein interaction. The colored lines represent the edges that show associations among different proteins. Dark blue: connect proteins that are co-occurring phylogenetically. Dark green: connect proteins that are occurring as gene neighborhood. Red: shows proteins with gene fusions. The colored nodes (red, blue, green) are different clusters of proteins.

process involves strategies, such as intracellular bioaccumulation, extracellular precipitation, biotransformation, biomineralization, and biosorption, that involve several signaling pathways (Wang et al., 2019). In addition, the reduction of heavy metals from the cell comprises of activation of different metabolic processes in the cell that are stimulated by heavy metals (Goyal et al., 2003).

Fungi largely counter heavy metals in two ways. The first way involves averting the metal uptake and its passage inside the fungal cell. This happens chiefly by a decrease in metal uptake or increased efflux of metals, metal biosorption to the impermeable cell walls by metal binding polysaccharides, peptides, extracellular formation of complexes, and the release of organic acids that chelates heavy metals outside the cell. Thus, the extracellular mechanism operates in the cell to counter heavy metals by circumventing the entry of heavy metals. Particularly, the secondary metabolites, such as citric acid, oxalic acids, succinate, and fumarate, which are low molecular weight compounds (<900 daltons) are secreted by fungal species namely *A. niger* specifically in response to heavy metals exposure and they bind the heavy metals extracellularly (Kolen, 2013). Also, the secondary metabolite oxalic acid is produced as an intermediate compound in the biochemical tricarboxylic acid cycle (TCA) (Munir et al., 2001). Here, we found the upregulation of the key enzymes that are involved in the TCA cycle viz. phosphoglycerate kinase (2.91-fold), glyceraldehyde-3-phosphate dehydrogenase (2.38-fold), enolase (1.8-fold), and pyruvate kinase (1.86-fold) which mediates the enhanced secretion of oxalic acids.

In the second way, the fungus subsists the high concentration of metals inside the cell by tolerance after the process of detoxification *via* metal chelation by synthesizing ligands, such as metallothioneins and phytochelatins, that bind heavy metals intracellularly or by compartmentalization of heavy metals within the cell organelles of vacuoles by polyphosphates. The three main classes of intracellular peptides binding metal ions are phytochelatins (PCs), metallothioneins (MTs), and glutathione (GSH). MTs are low molecular weight cysteine-rich metal-binding proteins that have high affinity toward both the essential metal ions, such as Cu and Zn, non-essential metal ions, such as Cd, Hg, and Ag, and also have large metal-binding capacities (Reddy et al., 2014). Further, MTs chelate heavy metals by

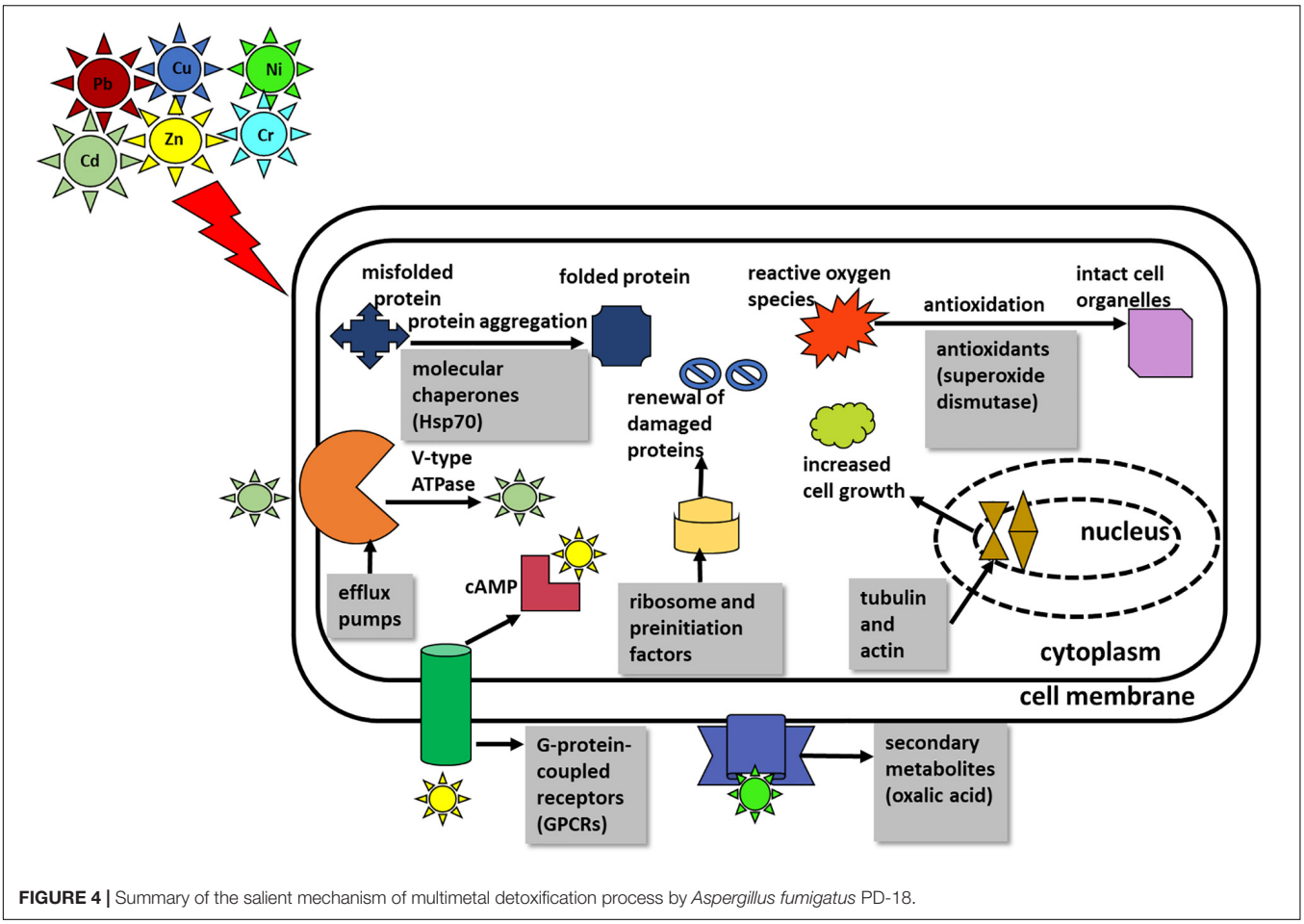


FIGURE 4 | Summary of the salient mechanism of multimetal detoxification process by *Aspergillus fumigatus* PD-18.

forming thiolate bonds with the heavy metals. Glutathione S transferases (GSTs) are enzymes that metabolize heavy metals and other contaminants by catalyzing the binding of glutathione to non-polar compounds comprising of electrophilic nitrogen, carbon, and sulfur atom (Morel et al., 2009). Usually, the metals Cd, Cu, Pb, and Zn are removed *via* glutathione (GSH)-mediated sequestration. However, in this study, there was no evidence of the production of proteins glutathione, metallothioneins, and phytochelatins despite the presence of these heavy metals Cd, Cu, Pb, and Zn in the multimetal mixture. The reason for this phenomenon could be attributed to the dynamics of individual heavy metals when present in a mixture. As the expression of GST is related to the type of heavy metal, its concentration, and the extent of treatment time of the heavy metal (Shen et al., 2015).

Further, it is reported that heavy metals induce oxidative damage to the cell membranes of fungi by the generation of reactive oxygen species (ROS). These ROS are detoxified by the production of antioxidants that are components from the thioredoxin system, such as peroxiredoxins, NADPH dehydrogenases, catalase, superoxide dismutase, and peroxidase, that enables the fungus to confront the reactive-oxygen species that accumulate in the cell on exposure to the metals (Zhang et al., 2015). Thus, in principle, intracellular mechanisms decrease the metal load in the cytosol (Sandau et al., 1996; Gadd, 2000, 2007, 2010). In this study, we found 3.45-fold upregulation of the antioxidant protein Cu-Zn superoxide dismutase. Other important functional groups were detected that expressed amino acid metabolism, lipid metabolism, energy metabolism, and also the proteins involved in signal transduction, transcription, translation, or DNA repair. In general, the upregulated proteins are stimulated to display the fungal resistance against the contaminants' stress, while the downregulated proteins are suppressed by the action of the pollutants' toxicity.

The highly upregulated protein hydroxymethylglutaryl-CoA synthase with 5.19-fold upregulation is responsible for the production of secondary metabolite carotenoid from the precursor molecule of acetyl-CoA when stimulated by the heavy metal stress (Bhosale, 2004). The protein serine/threonine phosphatase with 3.42-fold upregulation is responsible for maintaining the conformation of cell organelles and proteasomes (Dias et al., 2019).

The other proteins, such as G-protein beta subunit *SfaD*, with 2.98-fold upregulation depicted the role of G-protein-coupled receptors (GPCRs) in heavy metal bioremediation. These are the largest transmembrane receptors that aid in communicating the extracellular signals, such as stresses of heavy metals into the intracellular sites. G-protein-coupled receptors (GPCRs) regulate the important effector molecules,

such as adenylate cyclase and phospholipase C, and regulate the function of kinase and ion channel by producing secondary messengers, such as cAMP, thereby inducing signaling cascades (El-Defrawy and Hesham, 2020).

From **Figures 2A**, discussion of the highly regulated selected KOG classes are as follows:

Post-translational Modification, Protein Turnover, Chaperones

Protein homeostasis is crucial for the cell proliferation and viability of all organisms. Further, cellular signaling is greatly affected by the protein homeostasis under different physiological conditions and environmental stresses, such as heavy metals, and therefore they can be suitable biomarkers. Molecular chaperones aid in the delivery of metal ions to the cell organelles and metalloproteins (Lotlikar and Damare, 2018). The two vital regulators of molecular chaperones in the proteostasis network are heat shock transcription factor Hsf1 and heat shock protein Hsp90. Under a stressful environment, heat shock protein enables the folding of newly synthesized proteins and helps in the degradation of damaged/misfolded proteins with the help of the ubiquitin-proteasome system (Hossain et al., 2020). Heat shock proteins bind to the denatured proteins, compelling them to refold into their native conformation and regain their original structure (Feng et al., 2018). In this study, we observed 42 proteins upregulated that belonged to post-translational modification, protein turnover, and chaperones. There was upto 4-fold upregulation of proteasome regulatory particle subunit *Rpt3* (KOG0727), upto 2-fold upregulation of Hsp70 chaperone (*HscA*) (KOG0101), and upto 2.5-fold upregulation of protein geranylgeranyltransferase (KOG1439). The enzyme geranylgeranyltransferase I (GGTase I) aids in the catalysis of the post-translational transfer of lipophilic diterpenoid geranylgeranyl molecule to the cysteine residue of proteins with the termination at CaaX motif (*Rho1p* and *Cdc42p*). This alteration helps in the membrane localization of the protein and thereby rendering it biologically active. *Rho1p* is a regulatory subunit of 1,3-β-D-glucan synthesis and contributes to the cell wall synthesis in fungi which is vital for cell viability under stressful condition of excess metals (Singh et al., 2005).

Translation, Ribosomal Structure, and Biogenesis

Different proteins related to protein translation under multimetal stress were overexpressed. Here, we found 60 proteins of translation, ribosomal structure and biogenesis upregulated. There was upto 3.0-fold increase in glutamyl-tRNA synthetase (KOG1147), upto 2.5-fold increase in eukaryotic translation initiation factor 3 subunit B (KOG2314), upto 2.0-fold increase in 60S ribosomal protein L23 (KOG1751), upto 2.3-fold increase in mitochondrial translation initiation factor IF 2 (KOG1144), and 2.0-fold increase 40S ribosomal protein S10b (KOG3344). Similar elements of protein synthesis, such as translation initiation factor 5A, elongation factor 2, 40S and 60S ribosomal proteins, ATP-dependent RNA helicase, and aspartyl-tRNA synthetase, were overexpressed in *Phanerochaete chrysosporium* under Cu stress as a result of the need for production of new proteins or renewal of the damaged proteins (Okay et al., 2020).

Intracellular Trafficking and Secretion and Vesicular Transport

Proteins such as ion transporters and other solutes are crucial for processes such as detoxification, cell nutrition, cell signaling, cellular homeostasis, and resistance toward metal stress. These polytopic transmembrane proteins are translated altogether and folded in the endoplasmic reticulum (ER) of the eukaryotic cells that are later ultimately arranged to their respective membrane location through vesicular secretion. During any physiological or stressful environment, transporters undertake several regulated turnovers. Thus, in the process, transporters briefly interact dynamically with multiple proteins (Dimou et al., 2021). In this study, we found 18 proteins of intracellular trafficking and secretion and vesicular transport upregulated. The levels of the proteins were expressed in higher amounts (SEC31) (KOG0307) by 3.3-fold, endosomal cargo receptor (P24) (KOG1692) by 2.9-fold, and mitochondrial inner membrane translocase (KOG2580) by 2.4-fold in *A. fumigatus* under the effect of multimetal stress. The vesicle (Ves) are tissues composed of a lipid bilayer whose size varies ∼nanometers to micrometers. The Ves structures fuse with the plasma membrane of the cell and eject the trapped materials either inside or outside of the cytoplasm. There are three types of intracellular Ves viz. protein complex I (COPI)-coated Ves, protein complex II (COPII)-coated Ves, and BAR-domain protein Ves. These Ves proteins aid in physiological processes, such as the exchange of proteins and RNA intercellularly (Jiang et al., 2020).

Energy Production and Conversion

Heavy metals, such as Cd, Cu, Ni, and Zn, function as cofactors in bacteria and fungi. However, excess amounts of these metals are toxic to these cells and also produce reactive oxygen species (Liu et al., 2017). The need for metabolic energy in the fungal increases during abiotic stress, such as exposure to excess heavy metals. Thus, ATPases are responsible for the biochemical and physiological processes by the production of energy. Heavy metal ATPases (HMAs) or P-type ATPases can be categorized into three groups namely, A, B, and C. Further, P-type ATPases are utilized by numerous organisms to facilitate the transport of cations viz. Na^+, K^+, and Ca^{2+}. To eliminate these excess metals, fungal HMA *Saccharomyces cerevisiae* CCC2 (Group A) localizes metals to metal-containing proteins, for example, in the case of copper metal, copper-containing protein *FET3* in trans-golgi compartment transports metals to the cell membrane *via* efflux pumps, such as cadmium efflux pumps, encoded by fungal HMA *Saccharomyces cerevisiae* PCA1 (Group B and Group C) (Adle et al., 2007). Saitoh et al. (2009) studied the CCC2-type *HMA* gene that targets copper-containing proteins from the fungus *Cochliobolus heterostrophus* by cloning. There was upto 0.78-fold upregulation in the production of V-type ATPase and upto 1-fold upregulation in the production of mitochondrial ATPase subunit ATP4. This gene has other multifarious roles, such as in the formation of dark brown colored melanin pigment located in fungal cell walls, that also sequester metals (Chang et al., 2019).

Amino Acid Transport and Metabolism

The nitrogen cycle is essential for nitrogen assimilation and transformation and also for stress tolerance. Heavy metals impact the enzymes that play important role in nitrogen metabolism (Khouja et al., 2014). There was upto 4-fold upregulation in the production of aminotransferase (KOG1549) in response to multimetal. A similar response was observed by Okay et al. (2020), where the production of enzyme aspartate aminotransferase was enhanced in the fungus *Phanerochaete chrysosporium* to tackle the Cu stress. This enzyme has a possible role in the renewal of the mitochondrial NAD/NADH imbalance.

Analysis of Biological Pathways and Protein–Protein Interactions

In addition to these mechanisms, there are contributions from other regulatory systems, such as cross-talks in various pathways, interconnection amongst these different pathways, and regulation of different genes. These regulatory networks of the microbial proteins are intricate and play crucial roles in resistance to metal contaminants by modifying the series of specific functional proteins/non-proteins and altering the different metabolic enzymes at the cellular level. The metabolic processes related to the detoxification of contaminants are usually regulated by the complete set of proteins and their networks instead of a single enzyme (Zhao and Poh, 2008; Feng et al., 2018).

The STRING analysis displayed protein–protein interaction networks that are related to the resistance and tolerance mechanism *A. fumigatus* PD-18 for multimetals. Protein interactions in the first cluster (green) are involved in ribosome and preinitiation factors (60S and 40S ribosomal proteins). Ribosomal proteins form the protein part of ribosomes and participate in protein synthesis in cells in conjunction with rRNA. Thus, the increased expression of the large subunit of ribosome renders resistance against abiotic stresses, such as heavy metals, radiation, cold, and salt (Liu et al., 2014).

The important hub proteins expressed were cytochrome-c oxidase that is mitochondrial proteins and catalyst in the electron transport chain and is responsible for the transport of heavy metals particularly copper (Dias et al., 2019). The second cluster (blue) had hub protein ATP synthase. The important interconnecting protein in the network is the septin protein of KOG class cell cycle control, cell division, and chromosome partitioning, and is involved in vesicular trafficking and countering the apoptotic cell death initiated by the toxicity of heavy metals. This energy-intensive process is mediated by ATP synthase to maintain the cellular structure and function under the lethal environment of heavy metals. Excess to or above the permissible limit of heavy metals exposure can substantially delimit the growth of organisms. Here, more growth is initiated in the fungal cell as a result of the enhanced activity of ATP synthase (Yıldırım et al., 2011). This was evident in our study by the upregulation of tubulin by 1.1-fold and actin proteins by 2.2-fold which are responsible for cellular division and growth (Marks et al., 1986). This enhanced growth in the fungus also corroborated with our previous study where there was an increased dry weight of the fungal biomass under 30 mg/L multimetal (Dey et al., 2016). These results showed that the proteins in this network played important functions in cell functioning under heavy metals stress.

The mechanism of hexametal uptake by *A. fumigatus* PD-18 has been summarized in **Figure 4**.

CONCLUSION

In this study, we found the fungus *A. fumigatus* PD-18 developed stress coping strategies by secreting a suite of proteins that were either unique or upregulated/overexpressed when compared to the control. The proteomics study revealed that maximum proteins that were upregulated belonged to KOG class translation, ribosomal structure, and biogenesis. This study also highlighted the enhanced expression of antioxidants superoxide dismutase, molecular chaperone heat shock proteins, and involvement of proton transporter, such as ATPase. These proteins are involved in the tolerance and detoxification of multimetals by the fungus. *A. fumigatus* PD-18. Therefore, this investigation on the response of cellular proteomes to multimetal stress enabled us to better understand the cellular mechanism regarding the cumulative effect of the inorganic heavy metals stress on microbes. Further, it will be conducive to screening the key genes coding for enzymes that have higher resistance to these inorganic pollutants with enhanced capability to transform pollutants.

DATA AVAILABILITY STATEMENT

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE [1] partner repository with the dataset identifier PXD031741.

AUTHOR CONTRIBUTIONS

PD, AM, and NJ contributed to conception and design of the study. PD organized the database and wrote the first draft of the manuscript. PD and S-BH performed the statistical analysis. PD, AM, NJ, DS, and MB wrote sections of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

FUNDING

PD thankfully acknowledges the fellowship from the Ministry of Human Resource and Development (Government of India) and German Academic Exchange Service (DAAD). MB and NJ are grateful for the funding of the UFZ for the ProMetheus platform for proteomics and metabolomics.

REFERENCES

Abed, R. M. M., Shanti, M., Muthukrishnan, T., Al-Riyami, Z., Pracejus, B., and Moraetis, D. (2020). The role of microbial mats in the removal of hexavalent chromium and associated shifts in their bacterial community composition. *Front. Microbiol.* 11:12. doi: 10.3389/fmicb.2020.00012

Adle, D. J., Sinani, D., Kim, H., and Lee, J. (2007). A cadmium-transporting P1B-type ATPase in yeast *Saccharomyces cerevisiae*. *J. Biol. Chem.* 282, 947–955. doi: 10.1074/jbc.M609535200

Anand, P., Isar, J., Saran, S., and Saxena, R. K. (2006). Bioaccumulation of copper by *Trichoderma viride*. *Bioresour. Technol.* 97, 1018–1025. doi: 10.1016/j.biortech.2005.04.046

Bae, W., and Chen, X. (2004). Proteomic study for the cellular responses to Cd2+ in *Schizosaccharomyces pombe* through amino acid-coded mass tagging and liquid chromatography tandem mass spectrometry. *Mol. Cell. Proteom.* 3, 596–607. doi: 10.1074/mcp.M300122-MCP200

Baker, A. J. M. (1987). Metal Tolerance. *New Phytol.* 106, 93–111. doi: 10.1111/j.1469-8137.1987.tb04685.x

Bakti, F., Sasse, C., Heinekamp, T., Pócsi, I., and Braus, G. H. (2018). Heavy Metal-Induced Expression of PcaA provides cadmium tolerance to *Aspergillus fumigatus* and supports Its virulence in the *Galleria mellonella* Model. *Front. Microbiol.* 9:744. doi: 10.3389/fmicb.2018.00744

Bhardwaj, R., Gupta, A., and Garg, J. K. (2017). Evaluation of heavy metal contamination using environmetrics and indexing approach for River Yamuna. Delhi stretch, India. *Water Sci.* 31, 52–66. doi: 10.1016/j.wsj.2017.02.002

Bhattacharya, A., Dey, P., Gola, D., Mishra, A., Malik, A., and Patel, N. (2015). Assessment of Yamuna and associated drains used for irrigation in rural and peri-urban settings of Delhi NCR. *Environ. Monit. Assess* 187:4146. doi: 10.1007/s10661-014-4146-2

Bhattacharya, A., Gola, D., Dey, P., and Malik, A. (2020). Synergistic and antagonistic effects on metal bioremediation with increasing metal complexity in a hexa-metal environment by *Aspergillus fumigatus*. *Int. J. Environ. Res.* 14, 761–770. doi: 10.1007/s41742-020-00295-w

Bhosale, P. (2004). Environmental and cultural stimulants in the production of carotenoids from microorganisms. *Appl. Microbiol. Biotechnol.* 63, 351–361. doi: 10.1007/s00253-003-1441-1

Boopathy, R. (2000). Factors limiting bioremediation technologies. *Bioresour. Technol.* 74, 63–67. doi: 10.1016/S0960-8524(99)00144-3

Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* 72, 248–254. doi: 10.1006/abio.1976.9999

Chang, P.-K., Scharfenstein, L. L., Mack, B., Wei, Q., Gilbert, M., Lebar, M., et al. (2019). Identification of a copper-transporting ATPase involved in biosynthesis of A. flavus conidial pigment. *Appl. Microbiol. Biotechnol.* 103, 4889–4897. doi: 10.1007/s00253-019-09820-0

Chauhan, N., Gola, S., Surabhi, Sharma, S., Khantwal, S., Mehrotra, R., et al. (2021). "Detrimental effects of industrial wastewater on the environment and health," in *Biological Treatment of Industrial Wastewater*, ed. M. P. Shah (London: Royal Society of Chemistry), 40–52. doi: 10.1039/9781839165399-00040

Cherrad, S., Girard, V., Dieryckx, C., Gonçalves, I. R., Dupuy, J.-W., Bonneu, M., et al. (2012). Proteomic analysis of proteins secreted by Botrytis cinerea in response to heavy metal toxicity†. *Metallomics* 4, 835–846. doi: 10.1039/c2mt20041d

Deshmukh, R., Khardenavis, A. A., and Purohit, H. J. (2016). Diverse metabolic capacities of fungi for bioremediation. *Indian J. Microbiol.* 56, 247–264. doi: 10.1007/s12088-016-0584-6

Dey, P., Gola, D., Chauhan, N., Bharti, R. K., and Malik, A. (2021). "Mechanistic insight to bioremediation of hazardous metals and pesticides from water bodies by microbes," in *Removal of Emerging Contaminants Through Microbial Processes*, ed. M. P. Shah (Berilin: Springer), 467–487. doi: 10.1007/978-981-15-5901-3_23

Dey, P., Gola, D., Mishra, A., Malik, A., Kumar, P., Singh, D. K., et al. (2016). Comparative performance evaluation of multi-metal resistant fungal strains for simultaneous removal of multiple hazardous metals. *J. Hazard. Mater.* 318, 679–685. doi: 10.1016/j.jhazmat.2016.07.025

Dey, P., Malik, A., Mishra, A., Singh, D. K., von Bergen, M., and Jehmlich, N. (2020). Mechanistic insight to mycoremediation potential of a metal resistant fungal strain for removal of hazardous metals from multimetal pesticide matrix. *Environ. Pollut.* 262:114255. doi: 10.1016/j.envpol.2020.114255

Dias, M., Gomes de Lacerda, J. T. J., Perdigão Cota de Almeida, S., de Andrade, L. M., Oller do Nascimento, C. A., Rozas, E. E., et al. (2019). Response mechanism of mine-isolated fungus *Aspergillus niger* IOC 4687 to copper stress determined by proteomics. *Metallomics* 11, 1558–1566. doi: 10.1039/c9mt00137a

Dimou, S., Georgiou, X., Sarantidi, E., Diallinas, G., and Anagnostopoulos, A. K. (2021). Profile of membrane cargo trafficking proteins and transporters expressed under N source derepressing conditions in *Aspergillus nidulans*. *J. Fungi* 7:560. doi: 10.3390/jof7070560

Dusengemungu, L., Kasali, G., Gwanama, C., and Ouma, K. O. (2020). Recent advances in biosorption of copper and cobalt by filamentous fungi. *Front. Microbiol.* 11:3285. doi: 10.3389/fmicb.2020.582016

El-Defrawy, M. M. H., and Hesham, A. E.-L. (2020). G-protein-coupled Receptors in Fungi. *Fungal Biotechnol. Bioeng.* 37–126. doi: 10.1007/978-3-030-41870-0_3

Feng, Y., Zhao, Y., Guo, Y., and Liu, S. (2018). Microbial transcript and metabolome analysis uncover discrepant metabolic pathways in autotrophic and mixotrophic anammox consortia. *Water Res.* 128, 402–411. doi: 10.1016/j.watres.2017.10.069

Gadd, G. M. (1994). "Interactions of fungi with toxic metals," in *The Genus Aspergillus: From Taxonomy and Genetics to Industrial Application Federation of European Microbiological Societies Symposium Series.*, eds K. A. Powell, A. Renwick, and J. F. Peberdy Boston, MA: Springer 361–374. doi: 10.1007/978-1-4899-0981-7_28.

Gadd, G. M. (2000). Bioremedial potential of microbial mechanisms of metal mobilization and immobilization. *Curr. Opin Biotechnol.* 11, 271–279. doi: 10.1016/s0958-1669(00)00095-1

Gadd, G. M. (2007). Geomycology: biogeochemical transformations of rocks, minerals, metals and radionuclides by fungi, bioweathering and bioremediation. *Mycolog. Res.* 111, 3–49. doi: 10.1016/j.mycres.2006.12.001

Gadd, G. M. (2010). Metals, minerals and microbes: geomicrobiology and bioremediation. *Microbiology* 156, 609–643. doi: 10.1099/mic.0.037143-0

Gadd, G. M., and White, C. (1993). Microbial treatment of metal pollution — a working biotechnology? *Trends Biotechnol.* 11, 353–359. doi: 10.1016/0167-7799(93)90158-6

Gola, D., Bhattacharya, A., Dey, P., Malik, A., and Ahammad, S. Z. (2020). Assessment of drain water used for irrigation in the delhi region. *J. Health Pollut.* 10:200610. doi: 10.5696/2156-9614-10.26.200610

Goyal, N., Jain, S. C., and Banerjee, U. C. (2003). Comparative studies on the microbial adsorption of heavy metals. *Adv. Environ. Res.* 7, 311–319. doi: 10.1016/S1093-0191(02)00004-7

Haange, S.-B., Oberbach, A., Schlichting, N., Hugenholtz, F., Smidt, H., von Bergen, M., et al. (2012). Metaproteome analysis and molecular genetics of rat intestinal microbiota reveals section and localization resolved species distribution and enzymatic functionalities. *J. Proteome Res.* 11, 5406–5417. doi: 10.1021/pr3006364

Høiland, K. (1995). Reaction of some decomposer basidiomycetes to toxic elements. *Nordic J. Bot.* 15, 305–318. doi: 10.1111/J.1756-1051.1995.TB00157.X

Hossain, S., Veri, A. O., and Cowen, L. E. (2020). The proteasome governs fungal morphogenesis *via* functional connections with Hsp90 and cAMP-Protein Kinase a Signaling. *mBio* 11:e00290-20. doi: 10.1128/mBio.00290-20

Irawati, W., Djojo, E. S., Kusumawati, L., Yuwono, T., and Pinontoan, R. (2021). Optimizing bioremediation: elucidating copper accumulation mechanisms of *Acinetobacter* sp. IrC2 isolated from an industrial waste treatment center. *Front. Microbiol.* 12:3049. doi: 10.3389/fmicb.2021.713812

Jaeckel, P., Krauss, G., Menge, S., Schierhorn, A., Rücknagel, P., and Krauss, G.-J. (2005). Cadmium induces a novel metallothionein and phytochelatin 2 in an aquatic fungus. *Biochem. Biophys. Res. Commun.* 333, 150–155. doi: 10.1016/j.bbrc.2005.05.083

Jiang, Z., Wang, T., Sun, Y., Nong, Y., Tang, L., Gu, T., et al. (2020). Application of Pb(II) to probe the physiological responses of fungal intracellular vesicles. *Ecotoxicol. Environ. Saf.* 194:110441. doi: 10.1016/j.ecoenv.2020.110441

Keller, N. P. (2015). Translating biosynthetic gene clusters into fungal armor and weaponry. *Nat. Chem. Biol.* 11, 671–677. doi: 10.1038/nchembio.1897

Khouja, H. R., Daghino, S., Abbà, S., Boutaraa, F., Chalot, M., Blaudez, D., et al. (2014). OmGOGAT-disruption in the ericoid mycorrhizal fungus *Oidiodendron* maius induces reorganization of the N pathway and reduces tolerance to heavy-metals. *Fungal Genet. Biol.* 71, 1–8. doi: 10.1016/j.fgb.2014.08.003

Kolen, M. (2013). Leaching of zinc, cadmium, lead and copper from electronic scrap using organic acids and the *Aspergillus niger* strain. *Fresenius Environ. Bull.* 22:8.

Kou, S., Vincent, G., Gonzalez, E., Pitre, F. E., Labrecque, M., and Brereton, N. J. B. (2018). The response of a 16S Ribosomal RNA gene fragment amplified community to lead. zinc, and copper pollution in a shanghai field trial. *Front. Microbiol.* 9:366. doi: 10.3389/fmicb.2018.00366

Kraut, A., Marcellin, M., Adrait, A., Kuhn, L., Louwagie, M., Kieffer-Jaquinod, S., et al. (2009). Peptide storage: are you getting the best return on your investment? Defining optimal storage conditions for proteomics samples. *J. Proteome Res* 8, 3778–3785. doi: 10.1021/pr900095u

Kuhn, A. N., and Käufer, N. F. (2003). Pre-mRNA splicing in *Schizosaccharomyces pombe*. *Curr. Genet* 42, 241–251. doi: 10.1007/s00294-002-0355-2

Li, H., Dong, W., Liu, Y., Zhang, H., and Wang, G. (2019). Enhanced biosorption of nickel ions on immobilized surface-engineered yeast using nickel-binding peptides. *Front. Microbiol.* 10:1254. doi: 10.3389/fmicb.2019.01254

Liu, S.-H., Zeng, G.-M., Niu, Q.-Y., Liu, Y., Zhou, L., Jiang, L.-H., et al. (2017). Bioremediation mechanisms of combined pollution of PAHs and heavy metals by bacteria and fungi: a mini review. *Bioresour. Technol.* 224, 25–33. doi: 10.1016/j.biortech.2016.11.095

Liu, X.-D., Xie, L., Wei, Y., Zhou, X., Jia, B., Liu, J., et al. (2014). Abiotic stress Resistance, a novel moonlighting function of ribosomal protein RPL44 in the halophilic fungus *Aspergillus glaucus*. *Appl. Environ. Microbiol.* 80, 4294–4300. doi: 10.1128/AEM.00292-14

Lotlikar, N., Damare, S., Meena, R. M., and Jayachandran, S. (2020). Variable protein expression in marine-derived filamentous fungus *Penicillium chrysogenum* in response to varying copper concentrations and salinity. *Metallomics.* 12, 1083–1093. doi: 10.1039/C9MT00316A

Lotlikar, N. P., and Damare, S. R. (2018). Variability in protein expression in marine-derived *Purpureocillium lilacinum* subjected to salt and chromium stresses. *Indian J. Microbiol.* 58, 360–371. doi: 10.1007/s12088-018-0733-1

Malik, A. (2004). Metal bioremediation through growing cells. *Environ. Int.* 30, 261–278. doi: 10.1016/j.envint.2003.08.001

Marks, J., Hagan, I. M., and Hyams, J. S. (1986). Growth polarity and cytokinesis in fission yeast: the role of the cytoskeleton. *J. Cell Sci.* 1986, 229–241. doi: 10.1242/jcs.1986.Supplement_5.15

Morel, M., Ngadin, A. A., Droux, M., Jacquot, J.-P., and Gelhaye, E. (2009). The fungal glutathione S-transferase system. Evidence of new classes in the wood-degrading basidiomycete *Phanerochaete chrysosporium*. *Cell Mol. Life Sci.* 66, 3711–3725. doi: 10.1007/s00018-009-0104-5

Munir, E., Yoon, J. J., Tokimatsu, T., Hattori, T., and Shimada, M. (2001). A physiological role for oxalic acid biosynthesis in the wood-rotting basidiomycete *Fomitopsis palustris*. *PNAS* 98, 11126–11130. doi: 10.1073/pnas.191389598

Ohno, M., Karagiannis, P., and Taniguchi, Y. (2014). Protein expression analyses at the single cell level. *Molecules* 19, 13932–13947. doi: 10.3390/molecules190913932

Okay, S., Yildirim, V., Büttner, K., Becher, D., and Özcengiz, G. (2020). Dynamic proteomic analysis of *Phanerochaete chrysosporium* under copper stress. *Ecotoxicol. Environ. Saf.* 198:110694. doi: 10.1016/j.ecoenv.2020.110694

Orsburn, B. C. (2021). Proteome Discoverer—A Community Enhanced Data Processing Suite for Protein Informatics. *Proteomes* 9:15. doi: 10.3390/proteomes9010015

Pandey, A., and Mann, M. (2000). Proteomics to study genes and genomes. *Nature* 405, 837–846. doi: 10.1038/35015709

Reddy, M. S., Prasanna, L., Marmeisse, R., and Fraissinet-Tachet, L. Y. (2014). Differential expression of metallothioneins in response to heavy metals and their involvement in metal tolerance in the symbiotic basidiomycete *Laccaria bicolor*. *Microbiology* 160, 2235–2242. doi: 10.1099/mic.0.080218-0

Rughöft, S., Jehmlich, N., Gutierrez, T., and Kleindienst, S. (2020). Comparative Proteomics of *Marinobacter* sp. TT1 reveals corexit impacts on hydrocarbon metabolism, chemotactic motility, and biofilm formation. *Microorganisms* 9:3. doi: 10.3390/microorganisms9010003

Sabuda, M. C., Rosenfeld, C. E., DeJournett, T. D., Schroeder, K., Wuolo-Journey, K., and Santelli, C. M. (2020). Fungal bioremediation of selenium-contaminated industrial and municipal wastewaters. *Front. Microbiol.* 11:2105. doi: 10.3389/fmicb.2020.02105

Sagar, V., and Singh, D. P. (2011). Biodegradation of lindane pesticide by non white- rots soil fungus *Fusarium* sp. *World J. Microbiol. Biotechnol.* 27, 1747–1754. doi: 10.1007/s11274-010-0628-8

Saitoh, Y., Izumitsu, K., and Tanaka, C. (2009). Phylogenetic analysis of heavy-metal ATPases in fungi and characterization of the copper-transporting ATPase of *Cochliobolus heterostrophus*. *Mycol. Res.* 113, 737–745. doi: 10.1016/j.mycres.2009.02.009

Sandau, E., Sandau, P., and Pulz, O. (1996). Heavy metal sorption by microalgae. *Acta Biotechnologica.* 16, 227–235. doi: 10.1002/abio.370160402

Selamoglu, N., Önder, Ö, Öztürk, Y., Khalfaoui-Hassani, B., Blaby-Haas, C. E., Garcia, B. A., et al. (2020). Comparative differential cuproproteomes of *Rhodobacter capsulatus* reveal novel copper homeostasis related proteins. *Metallomics* 12, 572–591. doi: 10.1039/C9MT00314B

Shen, M., Zhao, D.-K., Qiao, Q., Liu, L., Wang, J.-L., Cao, G.-H., et al. (2015). Identification of glutathione S-transferase (GST) genes from a dark septate endophytic fungus (*Exophiala pisciphila*) and their expression patterns under varied metals stress. *PLoS One* 10:e0123418. doi: 10.1371/journal.pone.0123418

Singh, S. B., Kelly, R., Guan, Z., Polishook, J. D., Dombrowski, A. W., Collado, J., et al. (2005). New fungal metabolite geranylgeranyltransferase inhibitors with antifungal activity. *Nat. Prod. Res.* 19, 739–747. doi: 10.1080/14786410420003347 15

Wang, Y., Yi, B., Sun, X., Yu, L., Wu, L., Liu, W., et al. (2019). Removal and tolerance mechanism of Pb by a filamentous fungus: a case study. *Chemosphere* 225, 200–208. doi: 10.1016/j.chemosphere.2019.03.027

Wisecaver, J. H., and Rokas, A. (2015). Fungal metabolic gene clusters—caravans traveling across genomes and environments. *Front. Microbiol.* 6:161. doi: 10.3389/fmicb.2015.00161

Yıldırım, V., Özcan, S., Becher, D., Büttner, K., Hecker, M., and Özcengiz, G. (2011). Characterization of proteome alterations in *Phanerochaete chrysosporium* in response to lead exposure. *Proteome Sci.* 9:12. doi: 10.1186/1477-5956-9-12

Zamora-Ledezma, C., Negrete-Bolagay, D., Figueroa, F., Zamora-Ledezma, E., Ni, M., Alexis, F., et al. (2021). Heavy metal water pollution: a fresh look about hazards, novel and conventional remediation methods. *Environ. Technol. Innov.* 22:101504. doi: 10.1016/j.eti.2021.101504

Zhang, Q., Zeng, G., Chen, G., Yan, M., Chen, A., Du, J., et al. (2015). The effect of heavy metal-induced oxidative stress on the enzymes in white rot fungus *Phanerochaete chrysosporium*. *Appl. Biochem. Biotechnol.* 175, 1281–1293. doi: 10.1007/s12010-014-1298-z

Zhang, X., Fang, A., Riley, C. P., Wang, M., Regnier, F. E., and Buck, C. (2010). Multi-dimensional Liquid Chromatography in Proteomics. *Anal. Chim. Acta* 664, 101–113. doi: 10.1016/j.aca.2010.02.001

Zhao, B., and Poh, C. L. (2008). Insights into environmental bioremediation by microorganisms through functional genomics and proteomics. *Proteomics* 8, 874–881. doi: 10.1002/pmic.200701005

Conflict of Interest: PD, S-BH, MB, and NJ were employed by Helmholtz-Centre for Environmental Research-UFZ GmbH.

The remaining 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: 06 July 2022
doi: 10.3389/fmicb.2022.928848

Life Within a Contaminated Niche: Comparative Genomic Analyses of an Integrative Conjugative Element ICE*nah*CSV86 and Two Genomic Islands From *Pseudomonas bharatica* CSV86^T Suggest Probable Role in Colonization and Adaptation

*Balaram Mohapatra, Harshit Malhotra and Prashant S. Phale**

Department of Biosciences and Bioengineering, Indian Institute of Technology Bombay, Mumbai, India

OPEN ACCESS

Edited by:
Roshan Kumar,
Magadh University, India

Reviewed by:
François Delavat,
Université de Nantes, France
Olesya Sazonova,
Institute of Biochemistry and Physiology of Microorganisms (RAS), Russia

***Correspondence:**
Prashant S. Phale
pphale@iitb.ac.in

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

Received: *26 April 2022*
Accepted: *08 June 2022*
Published: *06 July 2022*

Citation:
Mohapatra B, Malhotra H and Phale PS (2022) Life Within a Contaminated Niche: Comparative Genomic Analyses of an Integrative Conjugative Element ICEnahCSV86 and Two Genomic Islands From Pseudomonas bharatica CSV86^T Suggest Probable Role in Colonization and Adaptation. Front. Microbiol. 13:928848. doi: 10.3389/fmicb.2022.928848

Comparative genomic and functional analyses revealed the presence of three genomic islands (GIs, >50 Kb size): ICE*nah*CSV86, *Pseudomonas bharatica* genomic island-1 (PBGI-1), and PBGI-2 in the preferentially aromatic-degrading soil bacterium, *Pseudomonas bharatica* CSV86^T. Site-specific genomic integration at or near specific transfer RNAs (tRNAs), near-syntenic structural modules, and phylogenetic relatedness indicated their evolutionary lineage to the type-4 secretion system (T4SS) ICE*clc* family, thus predicting these elements to be integrative conjugative elements (ICEs). These GIs were found to be present as a single copy in the genome and the encoded phenotypic traits were found to be stable, even in the absence of selection pressure. ICE*nah*CSV86 harbors naphthalene catabolic (*nah-sal*) cluster, while PBGI-1 harbors Co-Zn-Cd (*czc*) efflux genes as cargo modules, whereas PBGI-2 was attributed to as a mixed-function element. The ICE*nah*CSV86 has been reported to be conjugatively transferred (frequency of 7×10^{-8}/donor cell) to *Stenotrophomonas maltophilia* CSV89. Genome-wide comparative analyses of aromatic-degrading bacteria revealed *nah-sal* clusters from several *Pseudomonas* spp. as part of probable ICEs, syntenic to conjugatively transferable ICE*nah*CSV86 of strain CSV86^T, suggesting it to be a prototypical element for naphthalene degradation. It was observed that the plasmids harboring *nah-sal* clusters were phylogenetically incongruent with predicted ICEs, suggesting genetic divergence of naphthalene metabolic clusters in the *Pseudomonas* population. Gene synteny, divergence estimates, and codon-based *Z*-test indicated that ICE*nah*CSV86 is probably derived from PBGI-2, while multiple recombination events masked the ancestral lineage of PBGI-1. Diversifying selection pressure (dN-dS = 2.27–4.31) imposed by aromatics and heavy metals implied the modular exchange-fusion of various cargo clusters through events like recombination, rearrangement, domain reshuffling, and active site optimization, thus allowing the strain to evolve, adapt,

and maximize the metabolic efficiency in a contaminated niche. The promoters (P*nah* and P*sal*) of naphthalene cargo modules (*nah*, *sal*) on ICE*nah*CSV86 were proved to be efficient for heterologous protein expression in *Escherichia coli*. GI-based genomic plasticity expands the metabolic spectrum and versatility of CSV86^T, rendering efficient adaptation to the contaminated niche. Such isolate(s) are of utmost importance for their application in bioremediation and are the probable ideal host(s) for metabolic engineering.

Keywords: *Pseudomonas*, integrative conjugative element (ICE), comparative and functional genomics, phylogenomics, naphthalene metabolism, Co-Zn-Cd tolerance, niche adaptation, evolution-speciation

INTRODUCTION

Horizontal transfer of gene(s) "*en masse*" through mobile genetic elements [MGEs: genomic islands (GIs)/integrative conjugative elements (ICEs), prophages, plasmids, etc.] encoding adaptive or niche colonization traits is the decisive factor for species diversification, evolution, and pan-genomic stability in prokaryotes (Burrus et al., 2002; Rodríguez-Beltrán et al., 2020). Amongst MGEs, GIs are strain-specific segments of the bacterial genome, which are often found to be present at limited chromosomal locations. Depending on the functions they encode and the specific lifestyle of the bacterium, GIs can be classified into metabolic, pathogenicity, symbiosis, fitness, or resistance islands (Dobrindt et al., 2004; Juhas et al., 2009). GIs located on chromosomes are typically flanked by direct repeats and are inserted at the 3′ end of a transfer RNA (tRNA) gene. Further, these elements harbor transposase(s) or integrase(s) that are required for chromosomal integration and excision along with other mobile genes, in a strain or clone-specific manner. GIs are usually differentiated from the rest of the core genome by their atypical G + C contents and oligonucleotide compositions, with steep gradients at their boundaries (Reva and Tümmler, 2005). Comparative genetic mapping and evolutionary analyses have uncovered diverse GIs in bacterial taxa, particularly amongst proteobacterial members, with pathogenicity islands of *Enterobacteriaceae* being the most studied GIs (Hacker and Carniel, 2001).

Genomic islands and ICEs have structural similarities, such as the presence of integrase/excisionase or conserved core gene regions that belong to a common group of MGEs, often termed as ICElands (van der Meer and Sentchilo, 2003). While GIs might be self-transmissible and/or unstable, ICEs are distinguished by their conjugative, integrative, and stable/consistent behavior (Wozniak and Waldor, 2010; Delavat et al., 2017). Classically, the core module(s) of ICEs comprise (i) an integrase (*xer*C, responsible for ICE excision to form circular intermediate), often assisted by an excisionase or recombination directionality factors (RDFs), (ii) type-4 secretion system (T4SS) (for conjugative transfer), and (iii) partition (*par*)- maintenance cluster, ensuring successful horizontal transmission amongst near-relatives (Burrus, 2017; Liu et al., 2019). Most importantly, the "cargo gene array(s)" which differ from ICE to ICE, encode diverse metabolic traits which provide survival fitness to the host in a particular niche (Wozniak and Waldor, 2010). Although both GIs and ICEs are synonymous, there are no clear demarcations on their evolution and/or mode of mobilization.

Genomic islands or ICEs harboring genetic modules for the degradation of aromatics and tolerance/resistance to heavy metals (as functional cargo) are of immense scientific interest due to their potential application in bioremediation and metal detoxification/recovery. With large-scale discharge, resonance-stabilized structures, hydrophobicity, and persistent nature of aromatics and toxicity of metals, selection pressure is exerted on the microbial communities, leading to the evolution of efficient metabolic pathways or tolerance mechanisms (Nojiri et al., 2009; Fuentes et al., 2014; Phale et al., 2019). This requires a high degree of genetic exchange and recombination events within the community mediated by mobile genetic elements through horizontal gene transfer (HGT) process. Till date, few GIs and ICEs, such as *phn* island (phenanthrene metabolism), *clc* GI (later classified as ICE*clc*B13; chlorocatechol metabolism), ICE*clc*LB400 (chlorinated biphenyl, chlorocatechol), ICE*clc*JB2 (*o*-halobenzoates and *o*-hydroxybenzoate), ICE*XTD* (xylene-toluene), ICE*bph-sal* (biphenyl-salicylate), and ICE*Tn4371* family (Co, Zn, Cd, Cr, As, and Hg resistance) have been characterized (ICEBerg 2.0 database[1]; Liu et al., 2019). ICE*clc*B13 (105 Kb) of *Pseudomonas knackmussii* B13 encoding *ortho*-cleavage of chlorocatechol (*clc*) and aminophenol (*amp*) is a functionally characterized prototypical element (Gaillard et al., 2006; Miyazaki et al., 2015). However, the coexistence of multiple GIs (either of similar or hybrid nature) encoding degradation/tolerance traits in a single bacterium has been rarely observed (Badhai and Das, 2016; Liu et al., 2019).

An Indian isolate, strain CSV86^T (formerly *Pseudomonas putida*, recently effectively described as *Pseudomonas bharatica*), isolated from petrol-station soil contaminated with petroleum products preferentially metabolizes various aromatic compounds over glucose and displays the following carbon source utilization hierarchy: aromatics $\geq$ organic acid $>$ glucose (Basu et al., 2006; Phale et al., 2021; Mohapatra et al., 2022). This unique preferential naphthalene degradation trait is stable, chromosomally encoded, and was found to be transferred to *Stenotrophomonas maltophilia* CSV89 through conjugation at a very low frequency (7×10^{-8}/donor cell) (Basu and Phale, 2008). Besides, strain CSV86^T is able to tolerate up to 2 mM concentration of heavy metals (Cd, Zn, Co, Cu, and As) during

[1]https://bioinfo-mml.sjtu.edu.cn/ICEberg2/index.php

growth on aromatics (Paliwal et al., 2014). In addition to these unique traits, the absence of plasmid and other pathogenicity and virulence factors renders strain CSV86^T an ideal candidate for pollutant bioremediation and host/*chassis* for metabolic engineering applications (Phale et al., 2021).

Initial genomic analyses of strain CSV86^T revealed the presence of a large extent of MGEs (GIs, prophages, probable ICEs, and insertion sequences), amounting to ∼19% of its total genomic content (6.79 Mb), responsible for the genomic plasticity of this isolate (Mohapatra et al., 2022). To delineate the extent, functionality, and evolutionary significance of MGEs in strain CSV86^T, combined comparative- and functional-genome minings and molecular analyses were performed. Analyses revealed the role and significance of MGEs in adaptation, niche colonization, and species competitive behavior of strain CSV86^T at the contaminated site.

MATERIALS AND METHODS

Growth and Culture Conditions of *Pseudomonas bharatica* CSV86^T

Pseudomonas bharatica CSV86^T (= MTCC 2445^T = ICMP 13424^T = JCM 34937^T) which metabolizes benzoate, naphthalene, and phenylpropanoids, amongst others was used in this study (Mahajan et al., 1994; Mohapatra et al., 2022). The strain was routinely grown on a minimal salt medium (MSM, 150 mL, pH 7.5; Basu et al., 2003) supplemented aseptically with naphthalene or benzoate (0.1%, w/v) or glucose (0.25%, v/v) as the sole source of carbon and energy at 30°C for 12 h at 200 rpm.

Identification and Characterization of Genomic Islands From *Pseudomonas bharatica* CSV86^T

The whole-genome shotgun sequencing project of strain CSV86^T was previously deposited at DDBJ/EMBL/GenBank under the accession number AMWJ00000000 with the first/standard version as AMWJ01000000 (Paliwal et al., 2014). The hybrid PacBio-GS FLX hybrid assembly based genome sequencing project has been deposited as the updated advanced version: AMWJ02000000 with BioProject: PRJNA65243 and BioSample: SAMN02472262 (Mohapatra et al., 2022). The GI replicons (corresponding nucleotide positions) are represented in the genome AMWJ02000000.

To identify the regions of genomic plasticity (RGPs) in strain CSV86^T, advanced version of the genomic draft (AMWJ02000000) of CSV86^T was analyzed by Microscope-Genoscope (MaGe[2]) and IslandViewer 4.0[3] using IslandPick, IslandPath-DIMOB, SIGI-HMM, and Islander prediction methods. The RGPs were manually curated to filter biases (false-positive and negative) using the following criteria: (i) presence of mobility genes (integrases/recombinases, and transposases); (ii) structural RNAs (tRNAs); (iii) atypical genomic content (GC mol% content, GC skewness, and codon usage); and (iv) presence of phenotypic trait cassettes (antibiotic/toxic compounds resistance, virulence, or substrate metabolism). To categorize RGPs, different methods were employed: (i) BLASTn searches against the ICEBerg database 2.0[4] and PHASTER[5] with the *E*-value $\leq 10^{-5}$ and sequence coverage $\geq$50%; (ii) homology search against nr- and RefSeq-database; (iii) evaluation of t-RNAs (tRNA Scan-SE) followed by site-specific integrase/relaxase; and (iv) presence of mobile element and hypothetical proteins. Besides, repeats (direct/inverted) were identified to define the boundaries of these elements. RGPs were identified as GIs and suspected to be ICEs, thus annotated using the RAST-tk pipeline and compared with the reported reference ICEs in ICEBerg 2.0 database (**Table 1**). Genes located on these suspected ICEs were categorized into orthologous groups using OrthoVenn2[6] with an *E*-value cut-off of 10^{-5}. Localized reciprocal best-hit of ICE genes was performed using the orthoANI tool (OAT) of the Ez-BioCloud server.[7] Other genomic features (gene synteny, GC skewness, GC content mol%, and pairwise comparisons) were analyzed using Artemis Comparison Tool (ACT, release 18.1.0) and Mauve progressive alignment in MAUVE [version 20150226 build 10 (c)]. To confirm the integrative nature of the predicted ICEs, the origin of replication/transfer (*ori*C/T) was identified using both the DoriC database[8] and manual curation using *Pseudomonas* spp. and *Escherichia coli* consensus sequences. To validate the genomic prediction and appropriate organization of the suspected ICEs, genome drafts of strain CSV86^T (AMWJ01 and AMWJ02) were aligned, gaps were amplified using custom-designed primers (**Supplementary Table 1** and **Supplementary Figure 1**), and sequenced. For details of the methodology, refer to "**Supplementary Material**."

Phylogenomic Analyses and Diversifying Selection of Core and Cargo Modules

The three suspected ICEs and their similar/syntenic homologs were subjected to multiple alignment using CLUSTALW (without gaps and including all positions) in MEGA 7.0 (Kumar et al., 2016). Phylogenetic reconstruction was performed using the neighbor-joining (NJ) method based on 1,000 bootstrap iterations using Jukes–Cantor distance model, where the robustness of the trees was assessed using maximum-likelihood (ML) and minimum-evolution (ME) algorithms. Owing to high synteny and conservedness of core genes within ICE classes, amino acid sequences were subjected to phylogenetic reconstruction. Multiple alignment of concatenated amino acid sequences of T4SS-T4CP was performed using MUSCLE, and phylogeny was deciphered through the ML method involving Jones–Thornton–Taylor (JTT) substitution matrix with 1,000 bootstrap revaluation in MEGA 7.0. Additionally, individual ICE core proteins like integrase, transposase as well as cargo proteins (*nah-sal* cluster, NahR regulator, and *czc* heavy metal

[2]https://mage.genoscope.cns.fr/microscope/home/index.php

[3]https://www.pathogenomics.sfu.ca/islandviewer/

[4]https://db-mml.sjtu.edu.cn/ICEberg/

[5]https://phaster.ca/

[6]https://orthovenn2.bioinfotoolkits.net/home

[7]https://www.ezbiocloud.net/

[8]http://tubic.org/doric/public/index.php/index/browse/bacteria

tolerance cluster) were used for phylogenetic interpretations using Poisson's correction and Kimura-2-parameter models involving ML algorithm in MEGA 7.0. The robustness and reliability of phylogenetic trees were evaluated by 1,000 bootstrap replications. Further, overall ICE-related indices (average nucleotide-/amino acid identity, pattern disparity index, and substitution estimation) were calculated using the OrthoANI (OAT) tool of the EzBiocloud server,[9] GGDC tool of Type Strain Genome Server (TYGS[10]) with recommended settings, and MEGA 7.0, respectively. Codon-based *Z*-Test of selection (diversifying/purifying-/neutral- selection) was performed by codon-by-codon alignment in MEGA 7.0 followed by dN-dS and dN/dS vs. dS computation to obtain the mean divergence rate.

[9]https://www.ezbiocloud.net/tools/orthoani
[10]https://tygs.dsmz.de/

To validate the genomic divergence, the total GC mol% of the genomes and the mean deviation (ΔG + C mol%) were calculated using the G + C calculator. Nucleotide sequences of OriC and associated elements (*par*B, *par*A, *gid*B, *yid*C, *rnp*A, *dna*A, *dna*N, *rec*F, and *gyr*B) were concatenated, aligned using CLUSTALW, and phylogeny was deciphered in MEGA 7.0.

Stability of ICE*nah*CSV86 and *Pseudomonas bharatica* Genomic Island-1 in Strain CSV86^T

The genetic and phenotypic stability of these elements in strain CSV86^T was evaluated by performing growth experiments in the absence of selection pressure for 60 generations [corresponding to ∼6 generations in a single growth cycle in batch culture (t_d = 1.11 h), thus a total of 10 batch transfers]. Strain CSV86^T

TABLE 1 | Comparative analysis of three GIs, suspected to be ICEs, from *Pseudomonas bharatica* CSV86^T and syntenic members of ICE*clc* family used in this study.

ICE name	Organism harboring	Genome accession*	Size (Kb)	GC mol%	Δ GC%	Function encoding	References
ICE*nah*CSV86	*Pseudomonas bharatica* CSV86^T	AMWJ02.1	105	62.5	0.7	Naphthalene degradation	This study
PBGI-1	*Pseudomonas bharatica* CSV86^T	AMWJ02.2	89.8	53.9	9.3	Heavy metal (Co, Zn, Cd) resistance	This study
PBGI-2	*Pseudomonas bharatica* CSV86^T	AMWJ02.2	91.9	56.07	7.13	Mixed function	This study
ICE*clc*B13	*Pseudomonas knackmussii* B13	HG322950	102.7	62.51	3.49	3-Chloro benzoate	Sentchilo et al., 2009
ICE*clc*JB2	*Pseudomonas aeruginosa* JB2	CP028917.1	123.2	61.5	3.5	*o*-Chloro benzoate degradation	Obi et al., 2018
ICE*Pae*LES	*Pseudomonas aeruginosa* LESB58	NC_011770	111.1	64.42	2.1	Mercuric resistance	Liu et al., 2019
PAGI-2	*Pseudomonas aeruginosa* C	NC_009539	104.9	64.71	1.8	Heavy metal (Hg) resistance, cytochrome biogenesis	Klockgether et al., 2007
ICE*Pae*690	*Pseudomonas aeruginosa* FFUP_PS_690	AP014646	86.2	63.05	3.05	Carbapenem resistance, virulence	Botelho et al., 2018
ICE*Xca*Ve85-1	*Xanthomonas campestris* 85-10	NC_007508	109.2	64.06	0.7	Streptomycin resistance	Liu et al., 2019
ICE*Xca*Ca8004-1	*Xanthomonas campestris* 8004	NC_007086	98.9	60.48	4.48	Unknown	Qian et al., 2005
ICE*Bxe*LB400-1	*Burkholderia xenovorans* LB400	NC_007951	122.8	61.71	1.05	Drug resistance	Chain et al., 2006
ICE*Cme*CH34-1	*Cupriavidus metallidurans* CH34	NC_007973	109.5	64.67	−0.85	Mercuric resistance	Janssen et al., 2010
ICE*Hae*ULPAs-1	*Herminiimonas arsenicoxydans*	NC_009138	88.1	62.85	−8.56	Arsenic resistance	Liu et al., 2019
ICE*Aci*JS42-1	*Acidovorax* sp. JS42	NC_008782	119.9	65.22	0.83	Unknown	Ryan et al., 2009
ICE*XTD*	*Azoarcus* sp. CIB	NC_006513	173.7	59	6.5	Xylene, toluene degradation	Zamarro et al., 2016
ICE*Axy*A8-1	*Achromobacter xylosoxidans* A8	NC_014640	83.1	64.78	1.21	Heavy metals (Mercuric, Copper, Arsenic)	Strnad et al., 2011
ICE-GI1	*Bordetella petrii* DSM 12804	NC_010170	255.5	61.73	3.75	Heavy metals, antibiotic resistance/efflux	Lechner et al., 2009

**These are predicted as ICEs in the current study and are reported to be the part of the genomes of corresponding organisms.*

was grown on MSM (150 mL, pH 7.5) amended with benzoate (0.1%, w/v) as the sole source of carbon and energy. Culture grown on benzoate (growth $OD_{540\ nm}$ = 1.5–1.7) for 0th (control condition), 30th (5 batch transfers), and 60th (10 batch transfers) generations was transferred to MSM (150 mL, pH 7.5) containing: (i) naphthalene (0.1%, w/v) as the sole source of carbon, and (ii) benzoate plus heavy metals (CZC: Co^{2+}, Zn^{2+}, and Cd^{2+}, each at 1 mM). Culture growth was assessed by measuring the optical density ($OD_{540\ nm}$) at periodic intervals for naphthalene-amended conditions. While for heavy metals, growth was monitored at 12 and 24 h intervals. Minimum inhibitory concentration (MIC) was determined by tube dilution method, where cells of strain CSV86^T were inoculated into Luria broth (g/L: peptone, 10.0; yeast extract, 5.0, NaCl, 10.0, pH 7.2) containing increasing concentration of Co^{2+}, Zn^{2+}, or Cd^{2+} (0, 0.5, 1, 1.5, 2, 2.5, and 3 mM) and incubated for 48 h at 30°C (Wiegand et al., 2008). The lowest concentration of metals that inhibited growth ($OD_{540\ nm}$) completely was considered as MIC. The activity (at 10 h) of catechol 2,3-dioxygenase (C23DO, an enzyme catalyzing *meta*-ring cleavage of catechol in naphthalene metabolism) was monitored in the cell-free supernatant spectrophotometrically by measuring an increase in the absorbance at 375 nm due to the formation of 2-hydroxymuconic semialdehyde (ε_{375} = 33,000 M^{-1} cm^{-1}) as described by Basu et al. (2006). Protein was estimated by the Bradford method using BSA as the standard (Bradford, 1976). All experiments were performed thrice in duplicates and the mean with the standard deviations are tabulated.

Functional Analyses of Cargo Operons of ICE*nah*CSV86

In order to assess the functionality of the cargo operons of ICE*nah*CSV86, the promoter regions present upstream of the *nah* and *sal* operons were predicted using BPROM and SAPPHIRE, whereas the NahR binding site was predicted using multiple sequence alignment (for details, refer to "**Supplementary Material**"). The polycistronic nature of the *nah* and *sal* operons was validated using co-transcription analysis (refer to "**Supplementary Material**" and **Supplementary Figure 2**). In order to validate the functionality and strength of the predicted promoters and the NahR regulator; P*nah*, P*sal*, and P*sal*/*nah*R promoters were fused to gene (mcbC) encoding 1-naphthol 2-hydroxylase (1NH) as a reporter gene in pSEVA and the expression of 1NH was tested in *E. coli* BL21 (DE3) (**Supplementary Figure 3**). The details of the construct strategy and methodology are presented in the "**Supplementary Material**."

RESULTS AND DISCUSSION

Pseudomonas bharatica CSV86^T Genome Harbors Three Genomic Islands

An advanced version of the draft genome of *P. bharatica* CSV86^T (AMWJ02000000: 6.79 Mb, 99.9% completeness, 0.03% heterogeneity) revealed the presence of three GIs of greater than 50 kb size on two of the largest contigs (Contig_001 and Contig_008, **Figure 1**). BLAST-based bidirectional hits of GIs against experimentally validated and genomically predicted ICE families in the ICEBerg 2.0 database predicted these GIs to be discrete, complete, and intact ICEs (**Figure 1**). Presence of genes encoding: (i) tRNA at either of the extreme ends of ICEs, (ii) integrase, and (iii) VirB4 homologs (hallmark gene of T4SS) having the highest similarity (72–95%) to ICE*clc* (ICE*clc*B13, ICE*XTD*, ICE*Bxe*LB400-1) confirmed affiliation of these elements to mating-pair type T4SS-ICEs (Wozniak et al., 2009; Poulin-Laprade et al., 2015). Genome analyses indicated that all three predicted ICEs in strain CSV86^T were present as a single copy. These elements were predicted to have no origin of replication (*ori*C). A single *ori*C site (831 bp with eight 9-mer DnaA boxes, constituting 48.01% AT content = 6,355,999 AT disparity, on contig_001) (**Supplementary Figure 4**) was found elsewhere in the genome of the strain CSV86^T. Phylogenetic analysis of *ori*C and its associated elements indicated its lineage to non-ICE bearing members (*Pseudomonas japonica* WL, *Pseudomonas parafulva* CRS01-1) (**Supplementary Figure 4**). Further, the absence of any plasmid in strain CSV86^T confirms the chromosomal nature of these GIs (Basu and Phale, 2008). Similar to T4SS-ICEs, these elements were found to encode the main core gene modules harboring integrases, T4SS-T4CP, and DNA partitioning cluster (*par*, PFG-1) (**Figure 2A**). These three elements: ICE*nah*CSV86 (∼105 Kb), *Pseudomonas bharatica* genomic island-1 (PBGI-1) (∼89.8 Kb), and PBGI-2 (∼91.9 Kb) were found to encode 127, 112, and 136 RAST subsystem features, respectively. Amongst these, ICE*nah*CSV86 has been reported to be successfully transferred through conjugation to a recipient, *S. maltophilia* CSV89 with a frequency of 7×10^{-8} (Basu and Phale, 2008). Thus, this element has been classified as a probable ICE (ICE*nah*CSV86), based on the nomenclature system proposed (Burrus et al., 2002; Liu et al., 2019). A minor average G + C skew was observed in ICE*nah*CSV86 (62.5 mol%), i.e., close to chromosomal G + C content (63.2 mol%) of strain CSV86^T. While significant deviations (ΔG + C = 9.3 and 7.13 mol%) were observed for PBGI-1 (G + C content = 53.9 mol%) and PBGI-2 (G + C content = 56.07 mol%), respectively (**Table 1** and **Figure 2A**). A low synteny at the core gene modules and no synteny for cargo modules were noted for ICE*nah*CSV86 vs. PBGI-1 and PBGI-1 vs. PBGI-2 (**Supplementary Figure 5A**). Whereas ICE*nah*CSV86 and PBGI-2 shared moderate synteny at the core gene level (**Supplementary Figure 5B**). It has been established that genes (encoded on any MGEs, like GIs, including ICEs) originate from a common ancestor during specification events and are usually syntenic between closely related species, hence orthologous in nature. Sequences that show greater divergence from other species are more likely to perform distinct functions (Fang et al., 2010; Xu et al., 2019). Therefore, genes encoded on ICEs were analyzed for orthologous clusters using the OrthoVenn2 platform and annotated through the COG database with the WebMGA server (data not shown). A total of 83 orthologous clusters for ICE*nah*CSV86 were observed, while ICE*clc*B13 formed the maximum, i.e., 126 clusters (**Supplementary Figure 6**). Both PBGI-1 and PBGI-2 showed the highest number of

singletons/unique clusters (n = 94), which were categorized into COG classes of unknown function/hypothetical proteins (I), general function (S), transcription (T), and defense mechanisms (V). Overall, a higher number of integrases (n = 3), hypothetical proteins (77), and conjugative proteins (9) were observed in PBGI-2, while a higher number of repeats (5) and transposases (8) were observed in PBGI-1.

Genomic Islands in Strain CSV86^T Encode Metabolic and Adaptive Traits for Its Survival

ICE*nah*CSV86

ICE*nah*CSV86, a probable ICE, is located on the contig_001 (length 105,221 bp; nucleotide positions: 274,076 to 379,297) and has tRNA-Gly^{CCC} gene (74 bp) at 5′-end. The 18 bp sequence at the 3′-end of tRNA-Gly (TTCCCTTCGCCCGCTCCA) forms the *att*L site (direct repeat), while the *att*R site (18 bp long, TTCCCTTCGCCCACTCCA) is located at 3′-end of ICE*nah*CSV86 with a single base substitution (A with G), defining the structural borders (**Figure 2**). The gene annotation of ICE*nah*CSV86 is summarized in **Supplementary Table 3** and **Supplementary Figure 7A**. This element harbors typical T4SS core modules and cargo genes responsible for naphthalene metabolism. Strain CSV86^T utilizes naphthalene as the sole source of carbon and energy (Mahajan et al., 1994). Genes encoding enzymes for naphthalene metabolism were found to be a part of this element only and not elsewhere on the genome, hence designated as ICE*nah*CSV86. It showed the highest synteny with ICE*clc* elements [ICE*clc*B13 (102.7 Kb) and ICE*XTD* (173.7 Kb)]. ICE*nah*CSV86 exhibits a dual structure, i.e., characteristic of both ICE*clc*B13 and ICE*XTD*. However, unlike ICE*clc*B13, which is present in dual copies (Ravatn et al., 1998b), ICE*nah*CSV86 is present as a single copy, based on the genome analyses. ICE*nah*CSV86 displayed 76–83% nucleotide identity with core modules of various ICE*clc* members and displayed nearly similar gene synteny as observed in ICE*XTD* (**Figure 2B**). ICE*nah*CSV86 possesses a single *ori*T at position 41,464 to 41,838 Kb (375 bp), adjacent to an open reading frame (ORF) encoding SNF2 superfamily II DNA/RNA helicase (**Supplementary Table 3**). The *ori*T showed the highest homology with *ori*T of ICE*XTD* (92.8% identity) and ICE*clc*B13 (67.5%). Signature repeat motifs of 6 bp long (CCCTTT or GCCTTT, each 2 motifs) in *ori*T of ICE*nah*CSV86 were observed and found to be identical to ICE*clc*B13 (**Supplementary Figure 8**). The frequencies of repeats at nucleotide positions from 41,464 to 41,838 bp were 20.8- and 45.4-fold higher, respectively, than their random distribution on whole ICE*nah*CSV86, thus implying their recognition by helicase and *ori*T-processing proteins for DNA unwinding and conjugative transfer (Lawley et al., 2004). The presence of *ori*T and associated elements as a part of ICE*nah*CSV86 supports the successful conjugative transfer of naphthalene degradation property to *S. maltophilia* CSV89 (Basu and Phale, 2008).

Synteny alignment showed the occurrence of substantial rearrangement events (deletions and insertions) in ICE*nah*CSV86 as compared to ICE*clc* relatives (**Figure 2B**). Adjacent to tRNA-Gly^{CCC}, gene encoding integrase (*int*NAH) responsible for catalyzing the site-specific integration and excision of ICE*nah*CSV86 was present (**Supplementary Table 3**). Similar to ICE*clc* (ICE*clc*B13), where integrase and its regulatory module are located at the two extreme ends of the ICE (Sentchilo et al., 2009), in ICE*nah*CSV86, InrR, an integrase regulator along with a DNA binding protein was found to be present ∼100.8 Kb downstream of *int*NAH and is likely to be involved in the regulation of *int*NAH expression (**Figure 2A**, **Supplementary Table 3**, and **Supplementary Figure 7A**). In strain CSV86^T, naphthalene is metabolized to salicylate (upper pathway; *nah* operon: *nah*ABFCED), which is further channeled to TCA cycle intermediates *via* catechol (lower pathway; *sal* cluster: *nah*GHINLOMKX; **Figure 2A**). These cargo genes [*nah* (9.5 Kb) and *sal* (10.7 Kb) clusters] were found to be located on the coding strand of ICE*nah*CSV86, while *nah*R (encoding LysR type transcription regulator, which regulates both clusters) is located on the complementary strand. Most of the core genes (∼68 Kb) of ICE*nah*CSV86 are present on the complementary strand, except for some hypothetical proteins (**Figure 2A** and **Supplementary Figure 7A**). ICE*nah*CSV86 harbors core genes: VirB4, ∼19 Kb downstream to *nah*-cluster, murein hydrolase, and VirD4 at ∼42 Kb downstream of *nah* cluster (**Supplementary Table 3**). These three key proteins: murein hydrolase (degradation of the peptidoglycan for forming mating channel), VirB4-ATPase (DNA transfer by T4SS), and VirD4 coupling protein (DNA transfer by mating channel) are reported to be the essential components for ICE transfer (Christie et al., 2014). The ParAB locus in the PFG-1 core gene cluster was found to be present ∼66 Kb downstream of the *att*L site in ICE*nah*CSV86. This locus is reported to be involved in stabilization and the maintenance of ICEs (Bignell and Thomas, 2001). The entire core module (T4SS, T4CP, and PFG-1) of ICE*nah*CSV86 formed a large synteny with ICE*XTD*, while no synteny was observed for cargo modules. In ICE*nah*CSV86, insertion of the repeat (2,990 bp), transposases (1.12 Kb), and the stretch of hypothetical proteins (5 Kb) segregated the core module (67.2 Kb) into two sub-modules: Core-I (27.1 Kb) and Core-II (31.0 Kb). Similar insertion (of 2,990 bp repeat and 1.12 Kb transposase) along with an additional region comprising mobile element proteins and 593 bp repeat (in alternate) segregated cargo modules into "*nah*" and "*sal*" clusters (**Figure 2A**). BLASTn search of 2,990 bp repeats (of both cargo and core module) showed 99–100% identity with repeats of *Cycloclasticus zancles* 78-ME (obligate marine PAH degrading isolate), *Acinetobacter baumannii* (antibiotic resistant strains 7835 and 9201), and *Pseudomonas aeruginosa* strain AK6U. Interestingly, repeats found in both ICE*nah*CSV86 and *C. zancles* 78-ME genome encodes an ISL3-TnpA family transposase, guarded by Type-4 pilus assembly proteins. The 593 bp repeat harboring IS5 family transposase showed the highest similarity (99–100%) to probable ICE-bearing naphthalene degrading *Pseudomonas citronellolis* strain SJTE-3, isolated from waste sludge. Other transposases (IS66 and IS5) and mobile element proteins located on ICE*nah*CSV86 are affiliated to *Pseudomonas* spp.

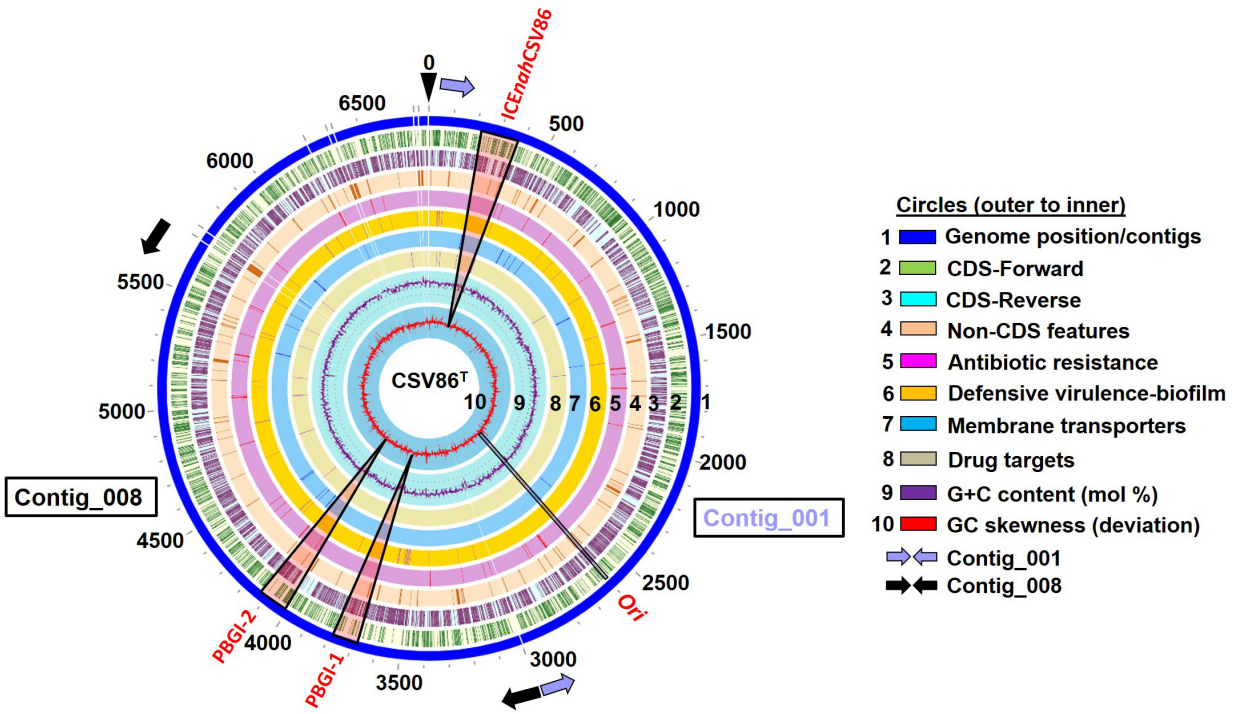


FIGURE 1 | Circular genomic map (0–6,975 Kb) depicting the location of three genomic islands (GIs), predicted to be integrative conjugative elements (ICEs: ICE*nah*CSV86, PBGI-1, and PBGI-2 in the genome of *Pseudomonas bharatica* CSV86^T along with various genomic attributes (GC mol%, skewness, CDS features). The genome map was reconstructed using PATRIC 3.7.2.2. Elongated triangular boxes (light red color) indicate the positions of three GIs on the genome. *Ori*, the origin of replication and replicative gene catamers are depicted by black rectangle. The start and the end positions of Contig_001 and Contig_008 are depicted by light blue and black filled arrows, respectively.

(95–99% similarity). An abundance of repeats and transposases within core and cargo modules serves as hotspots for possible recombination events and implied their acquisition from other organisms (Burrus, 2017). A significant deviation in GC skewness (**Figure 2A**, 12–13 GC mol%) of this stretch further supported the observation of the possibility of horizontal gene transfer events.

Pseudomonas bharatica Genomic Island-1

Pseudomonas bharatica genomic island-1 is located on the contig_008 (length 89,869 bp; nucleotide positions: 6,34,598 to 7,25,367), and has tRNA-Arg^{CCT} gene (75 bp) at 3′-end followed by a repeat of 38 bp sequence that forms the *att*L site. An identical repeat is located at 5′-end (*att*R), thus defining the borders of this suspected ICE (**Figure 2A**). PBGI-1 is present as a single copy in the genome and its annotation is depicted in **Supplementary Table 4** and **Supplementary Figure 7B**. Strain CSV86^T tolerates 2 mM concentration of various heavy metals (Co^{2+}, Zn^{2+}, Cd^{2+}: CZC) supplemented during the growth of aromatics (Paliwal et al., 2014). Genes encoding efflux pump proteins/transporters for heavy metal cations (CZC) are found to be part of this GI as intact clusters, hence designated as PBGI-1. Compared to ICE*nah*CSV86, PBGI-1 exhibited a mixed structure, i.e., characteristic of more than one ICEs [ICE*clc*LB400 (122.2 Kb responsible for polychlorinated biphenyl metabolism), ICE*clc*CH34 (109.5 Kb for mercuric resistance), and ICE(Tn*4371*)6035 (50 Kb for acriflavine resistance)]. Synteny analysis of PBGI-1 showed linear conserved block with respect to a short stretch of core module (∼10 Kb) against ICE*clc*LB400 and ICE(Tn*4371*)6035 (**Figure 2B**). This element showed significantly poor similarity (<30%) with previously reported ICEs [ICE*Pae*LESB58-1 (*P. aeruginosa*), ICE*Cme*CH34-1 (*Cupriavidus metallidurans*), ICE*Hae*ULPAs1-1 (*Herminiimonas arsenicoxydans*), ICE*Axy*A8-1 (*Achromobacter arsenitoxydans*)] encoding heavy metal resistance (Cd, Zn, Co, Hg, As, and Cu). Overall, a different integration site (proximity to tRNA-Arg), the presence of hypothetical proteins within core modules, and Co-Zn-Cd resistance (*czc* operon) as a cargo function demarcated the uniqueness of this suspected ICE amongst reported ICEs for the heavy metal tolerance (**Figure 2**).

Downstream of tRNA-Arg, two core modules (conjugation and partition: T4SS, T4CP: a total of 29.2 Kb) were distributed conjointly on both the strands. Insertion of alternating repeats (sizes: 254 and 5,728 bp) and transposases (two Il-IS-2, size: 1,163 bp) separated the core modules from *czc* cargo modules (**Figure 2A**). Two membrane proteins (YraQ family) and ArsR transcription regulator were present upstream of these repeats and transposases (**Supplementary Figure 7B**). BLASTn searches (against RefSeq) of these observed repeats indicated 99–100% identity with repeats of *Cycloclasticus zancles* 78-ME (obligate marine PAH degrading isolate) and *A. baumannii*

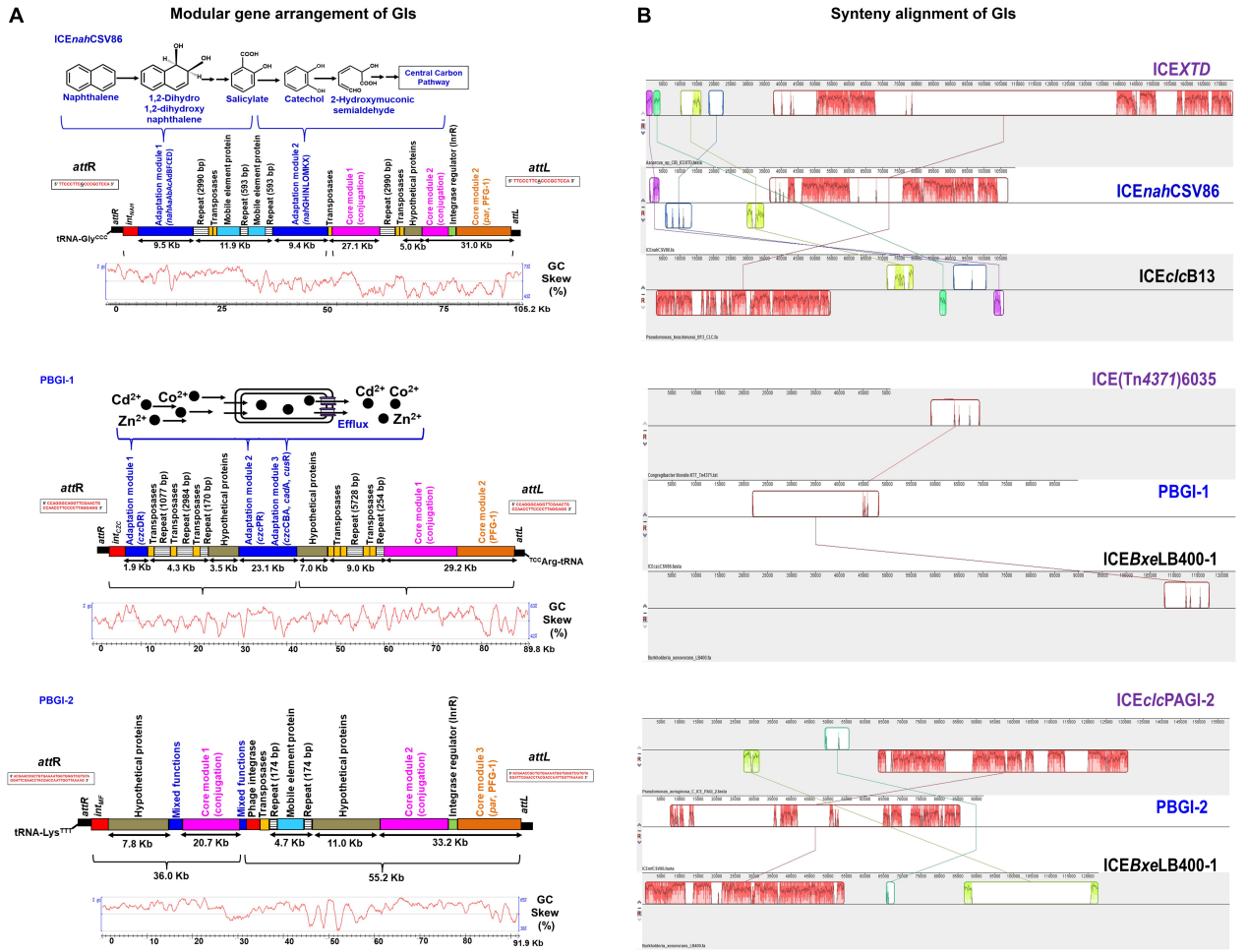


FIGURE 2 | Comparative analyses of the three full-length GIs: ICE*nah*CSV86, PBGI-1 and PBGI-2 from *Pseudomonas bharatica* CSV86^{T}. **(A)** Structural/modular organization of the three GIs, suspected to be ICEs. The color codes and depiction are as follows: cargo gene modules (blue), integration/excision (red), integrase regulator (green), conjugation (purple), partition modules (orange), regions containing repeats (lined boxes), genes encoding transposases (yellow), hypothetical proteins (gray), and mobile element proteins (light blue). Flanking *att*L and *att*R sites are indicated at the extreme ends (black solid rectangles) of the GIs and its sequences are depicted within the rectangular boxes. The functions encoded by respective cargo modules are depicted above the gene arrangement. The GC disparity plots (mol% skewness/deviations) for respective GIs are indicated below the gene arrangement. **(B)** Synteny alignment plots of predicted ICEs with the nearest similar ICEs from other bacteria obtained using progressive MAUVE alignment method. The blocks in different colors indicate linear conserved blocks (LCBs) and criss-cross connecting lines indicate the homologous LCBs amongst ICEs. For gene annotation and ORF details, refer to **Supplementary Tables 3–5**.

7835 (antibiotic resistant), similar to ICE*nah*CSV86. These repeats were found to encode transposases (ISL*3*, TnpA family). Genes encoding Tn*7* transposition proteins (TsnBC) are located next to the repeats. These proteins catalyze transposition by promoting DNA bending to form a highly organized protein-DNA complex (Arciszewska and Craig, 1991). In the vicinity of this region, a stretch of hypothetical proteins (homologous to conserved hypothetical proteins of *Pseudomonas* spp.) was observed (**Figure 2A**). Like ICE*nah*CSV86, the cargo module (total ∼25 Kb) harboring genes for the tolerance of Co-Zn-Cd is located at the 5′-end (**Figure 2A**). Strikingly, PBGI-1 harbors all the reported systems (and the associated mechanisms) for Co^{2+}, Zn^{2+}, and Cd^{2+} resistance-efflux, which are localized into three sub-modules: (I) cation diffusion facilitator (CDF) CzcD based efflux cluster, (II) PIB4-type ATPase CzcP, and (III) HME-RND (Heavy MEtal-Resistance-Nodulation-Division)-driven CzcCBA cluster. Three alternating Il-IS-2 transposases (size: 560, 716, and 1,163 bp) followed by repeats (sizes: 1.077, 2.984, 170 bp), and stretch of hypothetical proteins segregated *czc* cargo clusters into two sub-modules (I and II). Module I comprising of *czc* regulator (*czc*R) and efflux resistance (*czc*D) showed the maximum homology (99%) with marine bacterial taxa, Module II comprising a single component transporting ATPase (*czc*P), and multi-component RND-resistance efflux pumps (*czc*ABC: transmembrane, membrane fusion, and outer membrane) showed affiliation with *Pseudomonas* spp. (**Supplementary Table 4**). This contrasting observation indicated probable differential gene acquisition events mediated through different sets of mobile elements. It is interesting to note that all of the mobile genes on PBGI-1, including repeats and

transposases showed lineage to both aromatic degrading and metal-transforming halophilic marine bacteria (*Cycloclasticus zancles* 78-ME*, Halomonas axialensis, Salinicola peritrichatus*, and *Marinomonas* spp.) and *Pseudomonas* spp. (**Supplementary Table 4**). This signifies possible horizontal acquisition of mobile genes from these taxa. Codon-based *Z*-test for the selection also displayed diversifying selection (dN-dS vs. dS = 1.397, at 0.05% level) for *czc*DR. This observation indicates that in the presence of heavy metals (Co, Zn, and Cd), the tolerance pathway is likely to get stabilized with more non-synonymous substitutions than synonymous substitutions and tends to be retained in the population (Verma et al., 2014). A sigma factor RpoS and an integrase (*int*CZC) were found to be present upstream (5′-end) of the *czc* cluster (**Supplementary Figure 7B**). The RpoS is reported to be involved in making bistability decisions by the stationary phase cells to proceed for ICE transfer in ICE*clc*B13 (Miyazaki et al., 2012). It is also involved in activating promoters for regulated expression of integrase (*int*) and its regulator (InrR), thus mediating the high rate of ICE transfer to the recipient (Miyazaki et al., 2012). The presence of RpoS as a part of PBGI-1 could be attributed to regulating the expression of integrase in the absence of an integrase regulator (InrR).

Pseudomonas bharatica Genomic Island-2

The GI, *P. bharatica* genomic island-2 (PBGI-2) is also located on the contig_008 (length 91,907 bp; nucleotide positions: 947,073 to 1,038979) with tRNA-Lys^{TTT} gene (75 bp) at 3′-end followed by a repeat of 61 bp sequence forming *att*L border and identical sequence repeat at 5′-end forms *att*R borders (**Figure 2A**). The PBGI-2 is present as a single copy in the genome and its gene annotation is summarized in **Supplementary Table 5** and **Supplementary Figure 7C**. It showed structural similarity with ICE*clc* (ICE*clc*LB400), but maximum homology with GIs of *P. aeruginosa* (PAGI-5, PAGI-2 of *P. aeruginosa* C) encoding virulence, metal-resistance, and cytochrome biogenesis traits (Klockgether et al., 2007; Battle et al., 2009). Based on the maximum overall similarity observed with PAGI of *P. aeruginosa*, this GI is designated as PBGI-2 (**Figure 2B**). The core module (53.9 Kb) of PBGI-2 was found to be intact and segregated into three sub-modules (conjugation, *par*, and PFG-1) by two successive repeats of 174 bp flanked by mobile element proteins. Additionally, two long stretches of hypothetical proteins (7.8 and 11 Kb) are present upstream of the core modules (**Figure 2A**). Genes for DNA replication repair, membrane protein-TonB, short-chain dehydrogenase, methyl-accepting chemotaxis protein, disulfide isomerase, DNA-binding regulators, phage proteins, etc., were found to be present on PBGI-2, indicating mixed function nature of its cargo module. It was also found to harbor an additional P4 phage-type integrase and the associated regulator (InrR) near the 5′-end of the core module for regulating *int*PBGI-2 expression. BLASTN/P similarity searches showed a maximum homology of all cargo and core genes with *P. aeruginosa* strains (**Supplementary Table 5**).

The National Center for Biotechnology Information (NCBI) RefSeq-based homology search of PBGI-2 showed the presence of multiple copies of syntenic GIs in *Beta-* and *Gammaproteobacteria* genomes. Whereas PAGI-2 type islands were reported in the genome of various (total of 31) *P. aeruginosa* and *P. putida* strains, of which 12 strains harbor one, 11 strains two, 7 strains three, and 1 strain with four PAGI elements. These are further grouped into 10 sub-types based on differential DNA-hybridization patterns (Klockgether et al., 2007). The presence of these ICE-like GIs in numerous taxa suggests that they probably form a family with a deep evolutionary origin. Based on the structural synteny with ICE*clc* members and higher homology with PAGI-2 of *P. aeruginosa* strain C, it can be hypothesized that PBGI-2 is an intermediate between an ICE and GI. Many PAGI elements are reported to excise spontaneously at a frequency of 10^{-5} to 10^{-4} (Rajanna et al., 2003; Lesic et al., 2004), although mutations, deletions, and genome rearrangements are likely to be responsible for the inability of such elements to achieve precise excision and mobilization. On the other hand, PAGI-2 does not spontaneously excise from its host chromosome during its growth *in vitro* and has been shown to transfer to other proteobacterial members at a low frequency (10^{-5} to 10^{-4}) (Ravatn et al., 1998a; Klockgether et al., 2007). For other similar elements, like pKLC102 and PAPI-1, the frequencies of excision are of 1–3 orders of magnitude higher. Similarly, PBGI-2 harboring the phage-type integrase, plasmid-related genes, a type IV pilus biogenesis gene cassette, and a syntenic set of conserved ORFs similar to those detected in PAGI-2 and PAGI-3 indicate its ability to excise from the chromosome and form an extra-chromosomal circular intermediate during transfer.

Selection Pressure Renders Integrative Conjugative Element Acquisition and Phylogenetic Divergence

Bidirectional alignment and phylogenetic analyses of ICE*nah*CSV86, PBGI-1, and PBGI-2 of strain CSV86^T, as well as its modular components, were performed to delineate their ancestral root/origin and phylogenetic relatedness. ICE*nah*CSV86 formed a coherent cluster with ICE*XTD* as its nearest neighbor and corroborated the pairwise comparison data (**Figure 3A**). Interestingly, PBGI-2 was found to be a near-distant relative of ICE*clc* members, including ICE*nah*CSV86, whereas PBGI-1 deviated from all and formed a separate clade, as an out group [**Figure 3A** (i)]. Bidirectional hit-based DNA-homology (% similarity) data also demarcated ICE*XTD* to be the overall nearest relative of ICE*nah*CSV86, whereas, PBGI-1 and PBGI-2 were congruent with ICE*clc*LB400 (51% similarity) and PAGI-2 (48%), respectively [**Figure 3A** (ii)]. Phylogenetic analysis of integrases revealed their non-coherent nature, forming three distinct clusters (**Supplementary Figure 9A**). The *int*NAH formed a coherent association with integrases of *Pseudomonas, Betaproteobacteria*, and *int*bph-sal (cluster-I), whereas *int*PAGI-2 of *P. aeruginosa* C served to be a distant member. The *Int*PBGI-2 formed a monophyletic grouping with integrases of non-ICE Pseudomonads, whereas *intXTD* and *int*B13 of ICE*clc* were found to be the nearest ICE relatives. Unlike the other two, *int*CZC was a distant relative of all ICE-encoded integrases reported so far (**Supplementary Figure 9A**).

Phylogeny of concatenated core proteins (T4SS: VirB4, VirD4, TraG, TraI, and ParAB) showed two distinct branches.

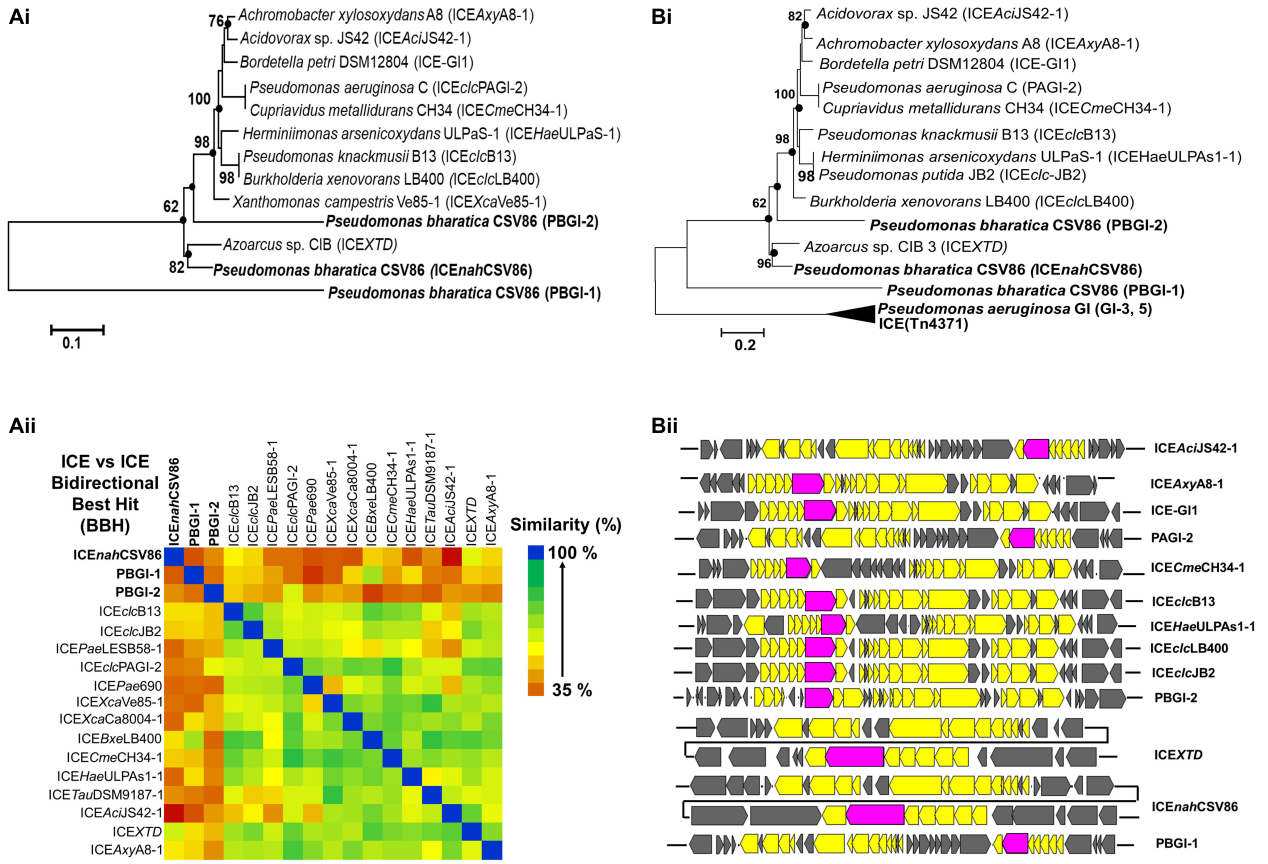


FIGURE 3 | Evolutionary inferences and overall ICE related indices for three GIs, predicted to be ICEs from *Pseudomonas bharatica* CSV86^T. **[A (i)]** Neighbor-joining (NJ) phylogenetic tree reconstructed using full-length nucleotide sequences of predicted ICEs and its nearest similar ICEs, and **[A (ii)]** Heat-map depicting bidirectional best hit-based average nucleotide identity (DNA–DNA homology level) of compared ICEs. **[B (i)]** Maximum-likelihood (ML)-based phylogeny of concatenated amino acid sequences of core gene modules (T4SS, T4CP), and **[B (ii)]** structural arrangement of core gene modules of various compared ICEs shown in **[B (i)]**. Trees were generated with 1,000 bootstrap iterations and the values at each node are represented as percentage replicates. The black solid circles represent generation of trees from all tree-making algorithms (NJ, ML, and ME).

ICE*nah*CSV86 and PBGI-2 were very closely related and clustered with ICE*XTD*, while PBGI-1 was distantly related to ICE*clc* and others (Tn*4371*), thus forming a discrete branch [**Figure 3B** (i)]. The structural arrangement of T4SS core modules from various ICEs also corroborated the observation indicating a genetic coherence between ICE*nah*CSV86 and ICE*XTD*, while a mixed pattern was observed for PBGI-2 and PBGI-1 [**Figure 3B** (ii)]. Similar results were obtained for relaxases, where PBGI-1 remained out-cladded as a distant relative of all compared ICEs (data not shown). All three predicted ICEs from strain CSV86^T harbor a variable number of transposases, which are reported to be a key component for exchanging DNA under specific environmental conditions (Jurka et al., 2007). All these transposases were found to be distinct and showed a poor sequence homology (<30% similarity at 50% coverage) amongst each other. The transposases formed a mixed clustering (three distinct groups) and showed lineages with marine bacterial taxa (*Halomonas* spp., *Marinomonas*, and *Shewanella fodinae*) reported for encoding complex organic carbon metabolism and heavy metal resistance traits (**Supplementary Figure 9B**). Overall, phylogenomics and synteny mapping revealed that both ICE*nah*CSV86 and PBGI-2 might be speciated from a common ancestral lineage, i.e., ICE*XTD/LB4001-1* like ICE*clc* members. Codon-based *Z*-test and the rate of diversifying selection (dN-dS, dN/dS vs. dS) between syntenic core gene modules of ICE*nah*CSV86 and PBGI-2 showed a very high positive dN-dS values (4.07–8.85) (**Supplementary Figure 5B**), indicating a diversifying selection (non-synonymous substitution vs. synonymous substitution) of ICE*nah*CSV86 with respect to PBGI-2. With an intact and complete set of core modules and a conserved synteny of PBGI-2 with PAGI, we hypothesize that this element might be the precursor from which ICE*nah*CSV86 has been diverged after acquiring and stabilizing cargo genes (*nah-sal*) under selection pressure. The PBGI-1 might have undergone high degrees of recombination events, that have masked its evolution from a common ancestor. The presence of similar intermediary ICEs with higher recombination events has also been postulated (Wozniak et al., 2009; Baharoglu et al., 2013).

The presence of functionally active naphthalene degradation and Co^{2+}, Zn^{2+}, and Cd^{2+} transport-efflux genes on two

different GIs in a single bacterium is a novel observation. Many naphthalene-degrading bacterial spp. are reported to harbor naphthalene metabolic clusters either on the plasmid (conjugative/non-conjugative) or transposon (Mohapatra and Phale, 2021). To date, no ICEs have been functionally characterized and reported for encoding naphthalene degradation traits. It is well established that aromatics including naphthalene and its derivatives have become ubiquitous in the environment (Duttagupta et al., 2020; Mohapatra and Phale, 2021). It has also been observed that many aromatic ring-hydroxylating dioxygenases like naphthalene dioxygenase (NDO, the first enzyme of the naphthalene degradation pathway) display "catalytic promiscuity" toward structurally related yet distinct compounds (Verma et al., 2019; Phale et al., 2020). Noticeably, such catalytic promiscuity of metabolic enzymes has been reported for a number of enzymes like biphenyl dioxygenase, benzoate dioxygenase, and toluene monooxygenase involved in the degradation of aromatics. Based on this information, it can be hypothesized that the ubiquitous presence of aromatics like naphthalene has imposed a substantial positive selection pressure on strain CSV86^T or similar spp. (from diverse biogeography) to acquire and evolve genes and enzymes for naphthalene metabolism through horizontal gene transfer events (by MGEs). Further, the genomic integration of these elements, re-organization, domain shuffling, and active-site optimization (for structurally related xenobiotic compounds) might have occurred to fine-tune the pathway, thus providing survival benefits to the recipient.

TABLE 2 | GC content and deviation (mol%, from genomic GC) of *nah* and *sal* cluster from different predicted ICEs and plasmids of various naphthalene degrading *Pseudomonas* spp. isolated from diverse biogeographic regions.

***Pseudomonas* spp.**	**Location of *nah-sal***	***nah* cluster GC mol% (Δ GC)***	***sal* cluster GC mol% (Δ GC)***	**Isolation habitat**	**Country**
P. bharatica CSV86^T	ICE	52.06 (10.66)	62.03 (0.6)	Petroleum-spillage soil	India
P. citronellolis SJTE-3	ICE	52.05 (15.35)	62.30 (5.1)	Wastewater sludge	China
P. stutzeri ATCC 17588	ICE	52.7 (9.8)	62.03 (0.47)	Clinical	United States
P. benzenivorans DSM 8628	ICE	51.8 (13.4)	61.9 (3.3)	Oil-impacted groundwater	Golf Cost
P. balearica DSM 6083	ICE	52.18 (12.52)	62.11 (2.56)	Wastewater lagoon	Spain
Pseudomonas sp. MPDS	ICE	52.0 (8.7)	61.9 (−1.2)	PAH impacted soil	China
P. stutzeri 19SMN4	Plasmid	52.1 (10.9)	61.9 (1.7)	Wastewater lagoon	Spain
P. veronii Pvy	Plasmid	52.6 (7.4)	62 (−2.0)	Oil refinery sediment	Romania
P. bauzanensis W13Z2	ICE	53.85 (8.1)	61.94 (0.01)	Oil-impacted Bohai sea	China
P. fluorescens HK44	Plasmid	52.67 (5.18)	63.47 (5.12)	Gas manufacturing unit soil	United States
P. stutzeri AN10	Plasmid	52.09 (10.61)	60.83 (1.86)	Crude oil enrichment sediment	Mediterranean
P. frederiksbergensis AS1	Plasmid	53.02 (7.49)	62.91 (2.38)	As contaminated waste site	South Korea
P. putida AK5	Plasmid	53.6 (7.56)	–	Oil factory slime pit	Russia
P. putida NCIB 9816-4	Plasmid	53.11 (5.79)	63.59 (4.68)	Oil refinery soil	United States
P. putida ND6	Plasmid	53.11 (8.89)	63.41 (1.39)	Industrial wastewater	China
P. putida G7	Plasmid	53.7 (7.52)	63.30 (2.07)	PAH impacted soil	United States
P. stutzeri KOS6	ICE	53.88 (8.67)	62.82 (0.27)	Hydrocarbon sludge	Russia
Pseudomonas sp. K35	Plasmid	53.8 (8.45)	63.42 (1.15)	Coal-bed methane water	United States
P. indoloxydans JCM 14246	NIDN	–	64.81 (2.58)	Lindane impacted soil	India
P. alcaliphila JAB-1	ICE	–	65.4 (−2.9)	Impacted soil	Czechia
P. saponiphila DSM 9751	ICE	–	63.2 (1.0)	DSM collection	United States
P. furukawaii KF707	Plasmid	–	62.89 (2.51)	Impacted soil	Japan
P. putida B6-2	ICE	–	64.3 (−2.4)	Impacted soil	China

**ΔGC values in bracket with red font denote >5% G + C content deviation from respective genomic GC content indicating acquaintance of module from distant taxa, and those in blue denote G + C content deviation near-by/borderline of 5% deviation, and black denotes less than 5% G + C deviation indicating acquaintance of module from near-similar or similar species members. Positive values of ΔGC denote the net G + C mol% of the genome is higher than the cluster, while negative values indicate net G + C mol% of the genome is lower than the cluster. NIDN indicates Not Yet Identified.*

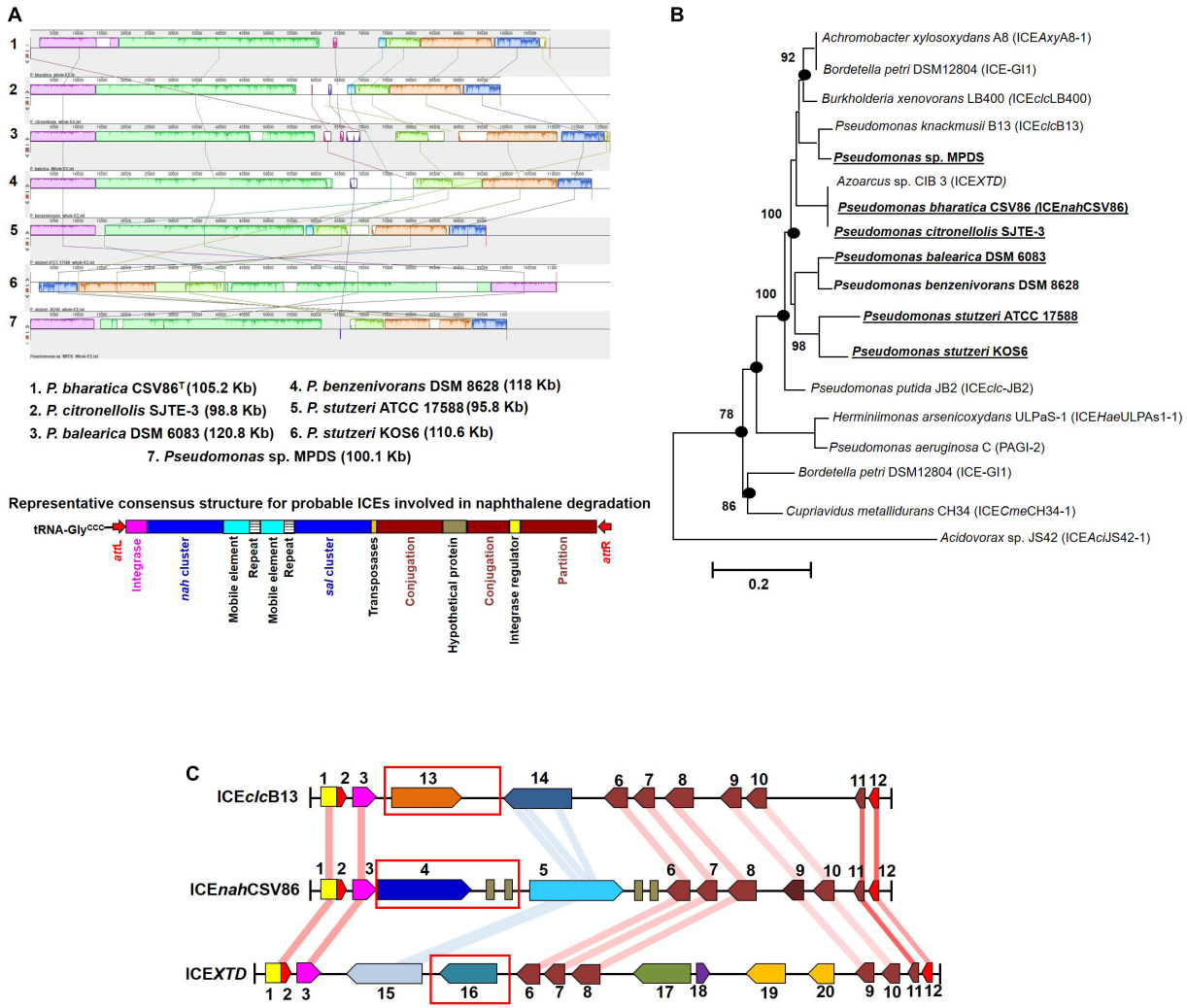


FIGURE 4 | Comparative-functional genomic analyses of *nah-sal* clusters including adjacent up/down-stream sequences from various naphthalene/aromatic degrading *Pseudomonas* spp. **(A)** Synteny mapping of probable ICEs harboring *nah-sal* clusters and its comparison with ICE*nah*CSV86. The representative consensus structure for all naphthalene degradation encoding probable ICEs has been depicted at the bottom of the synteny plot. **(B)** ML phylogenetic analysis of ICE*nah*CSV86 with probable ICEs, belonging to the ICE*clc* family, from naphthalene and aromatic degrading bacteria. Taxa in bold indicate the probable naphthalene degrading bacteria while bold with underline indicate reported naphthalene utilization/degradation phenotype. Black dots at each node represent robustness of clade formed in all tree-making algorithms. Values near each branch of the phylogenetic tree indicate percent bootstrap re-iteration and bar below the trees indicate substitution (0.1–0.2%). **(C)** Artemis comparison-based synteny mapping of prototypical ICE*nah*CSV86 and its phylogenetically related neighbors (ICE*clc*B13 and ICE*XTD*). The connecting lines between LCBs indicate percent similarity between various modules: blue (<60%), light red (60–70%), and deep red (>70%). Rectangular red boxes represent the upper pathway cargo modules of the respective ICEs.

To validate this hypothesis, detailed analyses were performed for *nah* and *sal* clusters from genomes of naphthalene or aromatic degrading *Pseudomonas* spp. reported from diverse biogeographic origins (**Table 2**). Mega-BLAST analysis of *nah-sal* clusters of ICE*nah*CSV86 revealed similar and syntenic regions (of size 100–150 Kb, >70% coverage, 95% similar) in *P. citronellolis* SJTE-3, *Pseudomonas balearica* DSM 6083, *Pseudomonas benzenivorans* DSM 8628, *Pseudomonas bauzanensis* W13Z2, *Pseudomonas stutzeri* KOS6, etc. and some plasmids (pNAH7, pND6, pAK5, AN10, and pAS1) of *Pseudomonas* (**Table 2**).

Interestingly, RAST annotation identified the presence of 5' tRNA-Gly^{CCC}, border repeats (*att*L/R), integrase as well as cargo genes (*nah-sal* clusters), and core modules, suggesting these genomic regions to be probable ICEs, similar to that observed in strain CSV86^{T} (**Figure 4**). The synteny alignment for all these hypothesized ICEs and ICE*nah*CSV86 indicated highly conserved cargo and core gene modules (as linearly conserved blocks, LCBs) (**Figure 4A**), thus denoting identical structural features, including *ori*T (homologous to ICE*nah*CSV86 and ICE*XTD*). Barring strain CSV86^{T} (Basu and Phale, 2008), none of them have been experimentally validated

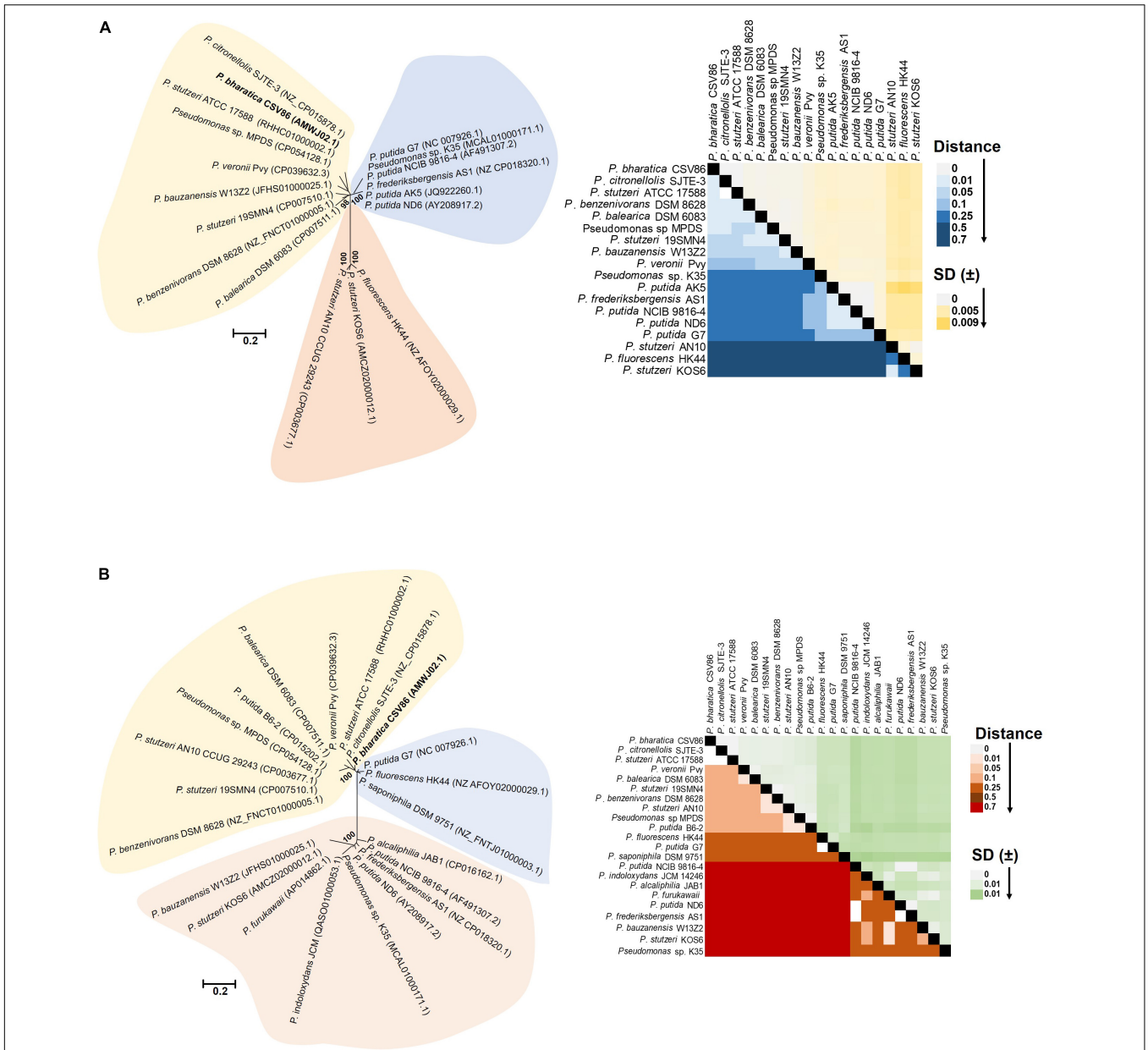


FIGURE 5 | Phylogenomic analyses of cargo gene modules (*nah-sal* cluster) of ICE*nah*CSV86 and other mobile elements (probable ICEs, plasmids, transposons) from naphthalene degrading *Pseudomonas* spp. Phylogenetic trees of **(A)** *nah* cluster and **(B)** *sal* cluster are reconstructed using ML-based algorithm and validated using NJ, ME methods with 1,000 bootstrap iterations. The tree is optimized using interactive Tree of Life (iTOL), and the best optimum tree is depicted here. The color-shaded areas around the clades depict demarcation of differential cladding of clusters present on probable ICEs (light yellow), plasmids (light blue), and mixed elements (light pink, which include conjugative/non-conjugative transposons and intermediate forms). The heat-maps depict the *p*-distance based pairwise similarity of **(A)** *nah* clusters and **(B)** *sal* clusters located on various mobile genetic elements.

for conjugative transfer. Based on this analysis, we propose ICE*nah*CSV86 to be a prototypical element for naphthalene degradation with consensus features. A monophyletic and coherent clustering was observed for ICE*nah*CSV86 and all probable ICEs (**Figure 4B**). Further, a highly conserved consensus *att*L (except substitution of A with G in the last 13 bp in two members), coherent tRNA-Gly, and integrase (syntenic to ICE*nah*CSV86) supported ICE*nah*CSV86 to represent a prototypical element (**Supplementary Figure 10**). A modular synteny plot confirmed ICE*XTD* and ICE*clc*B13 as the nearest relatives, thus classifying these probable ICEs as members of ICE*clc* (**Figure 4C**). It is observed that the typical core and structural elements were highly syntenic (60–76% identical with LCBs) in all three ICEs. Out of two cargo modules (representing upper pathway and lower pathway), a moderate synteny (45–52% identical) was observed in the lower pathway of

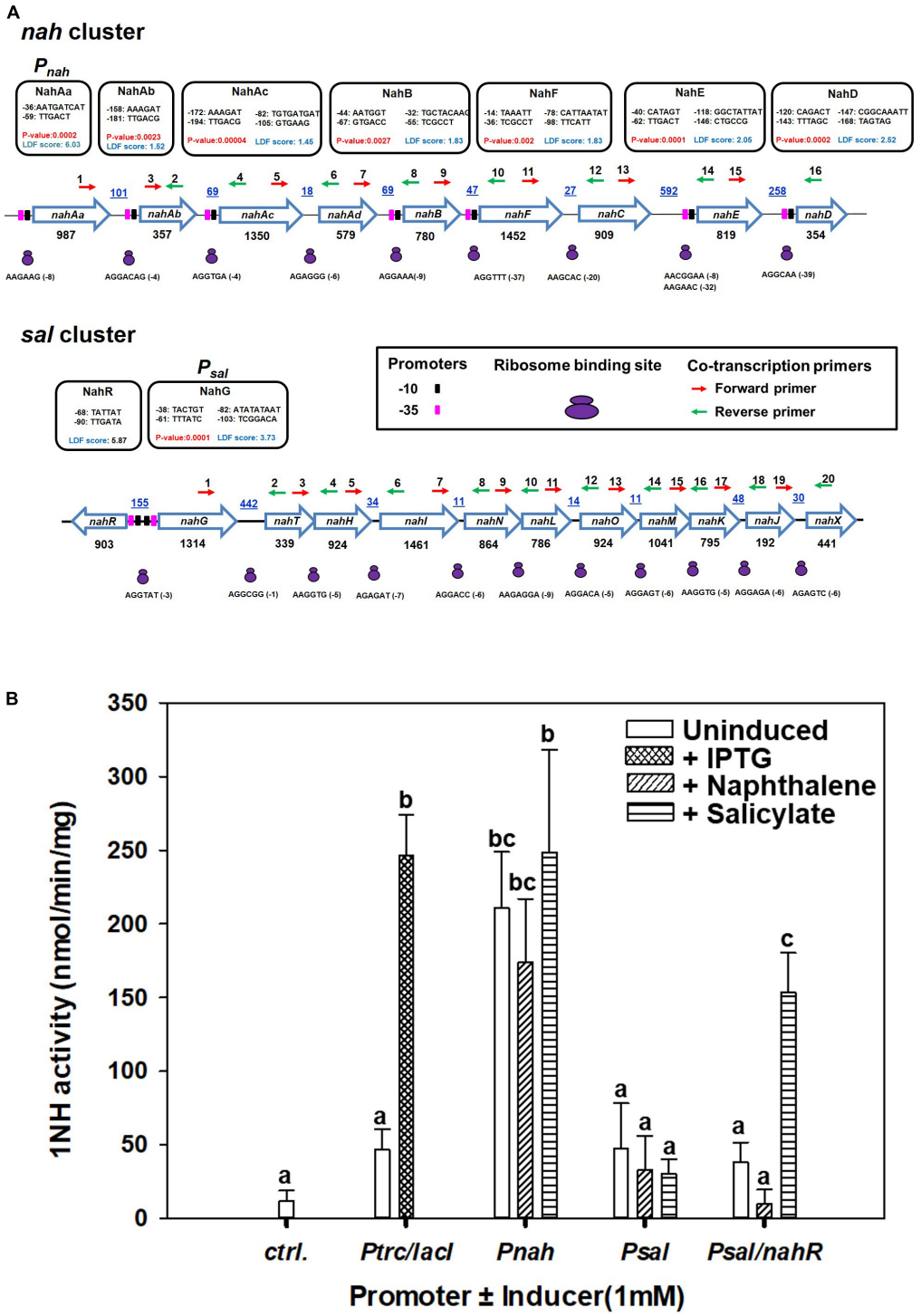


FIGURE 6 | Analysis of the upstream regions of *nah* and *sal* operons present on ICE*nah*CSV86 of *Pseudomonas bharatica* CSV86^T. **(A)** Operonic arrangement of upper (*nah*) and lower pathway *(sal)* genes and identification of putative promoter sequences [-10 (black filled rectangles) and -35 (pink filled rectangles) boxes], ribosome binding sites (RBS, as small and large purple ovals), and the location of primers [forward (red) and reverse (green) arrows] used for co-transcription analysis. The putative promoter sequences are depicted in the black box. The promoters with LDF scores have been predicted using BPROM and *P*-values were predicted using SAPPHIRE. The intergenic distances (bp) are indicated by blue color with underline, while gene lengths (in bp) are denoted below each gene. **(B)** Specific activity of 1-naphthol 2-hydroxylase (1NH) from various constructs expressed in *E. coli* BL21 (DE3) using different promoters and inducers. The bar represents the mean value with standard deviation (n = 3). Statistically significant differences (P = 0.05) among different promoter systems are denoted by a, b, c (Tukey's *post hoc* test).

ICE*nah*CSV86, ICE*clc*B13, and ICE*XTD*. Interestingly, the upper pathway in all three ICEs: *nah* cluster in ICE*nah*CSV86, *clc* cluster in ICE*clc*B13, and *tod* cluster in ICE*XTD* were non-syntenic (<20% similar). Guarding these upper pathways, the presence of various mobile elements (multiple repeats and transposases) suggested a separate modular fusion of the upper cluster/pathway within the ICE frame.

The *p*-distance based pairwise divergence estimate showed that *nah* cluster of ICE*nah*CSV86 has the highest homology with *P. citronellolis* SJTE-3, *P. stutzeri* ATCC 17588 (0.000 ± 0.001 distance), *P. benzenivorans* DSM 8628 (0.002 ± 0.001 distance), and *P. balearica* DSM 6083 (0.008 ± 0.002 distance) (**Figure 5A**). ML-based phylogenetic reconstruction corroborated this observation, where *nah* cluster of ICE*nah*CSV86 cladded with other probable ICE-bearing members and deviated from plasmid-encoded *nah* clusters as distant relatives (**Figure 5A**). Similar results were observed for *sal* cluster (**Figure 5B**). It is interesting to note that, a congruency/commonality was observed in terms of their habitat/ecotype, where all the probable naphthalene-ICE bearing members are isolated from oil, petroleum, or waste-impacted habitat, implying the role of the selection pressure; thus, helping strains to acquire and evolve these elements (**Table 2**). It has been reported that a broad GC content (>5%) change (distinct GC content of a cluster from the host genome) in stretches of prokaryotic genomes is indicative of the exchange of DNA/acquisition of gene(s) through HGT (Bohlin et al., 2010; Hildebrand et al., 2010). GC skewness analysis suggests that *nah* cluster displays higher order of mol% deviation as compared to *sal* cluster (**Table 2**). For *nah* cluster, the lowest deviation (5.18 mol%) was observed for *Pseudomonas fluorescens* HK44 and the highest (15.35 mol%) for *P. citronellolis* SJTE-3, whereas ICE*nah*CSV86 (10.66%) lied close to probable ICE of *P. stutzeri* AN10 (10.61%). On the contrary, the G + C mol% of *sal* cluster was significantly close to genomic G + C content, except for *P. fluorescens* HK44 (5.12% deviation). We have observed that the core genetic module of ICE*nah*CSV86 shows a GC content (62.95%) close to the genome (63.2% GC), whereas adaptation modules deviated sharply, i.e., module-1 (*nah*, 59.2% GC), module-2 (*sal*, 58.9% GC). Insertion of mobile elements (repeats and transposases) flanking *nah-sal* clusters further suggested the acquisition of cargo modules by different horizontal transfer events (Bosch et al., 1999; Hirose et al., 2021). To further validate these observations, dN-dS vs. dS (rate of non-synonymous over synonymous substitution) was performed. Positive dN-dS value (5.439–5.731) for *nah* cluster indicated higher non-synonymous substitutions than synonymous substitutions. This observation indicated its diversifying selection within the population and its still-evolving nature, which can be attributed to differential substrate selection pressure at impacted niches, i.e., naphthalene/structurally related compounds (oil-derived). Whereas lower dN-dS (2.139–2.247) of *sal* cluster indicated its intermediary/near-stable nature within the population. Additionally, the regulatory gene (*nah*R: LysR type regulators) of strain CSV86^T showed 100% similarity with *P. stutzeri* AN10 and phylogenetic closeness to other ICE-encoded NahR, out-cladding it from plasmid encoded members (**Supplementary Figure 11**). Interestingly, the presence of transposases upstream of *nah*R in both CSV86^T and AN10 and the absence of such gene in other plasmid-bearing members further confirms similar horizontal acquisition behavior.

It has been hypothesized that catabolic pathways for aromatic metabolism have evolved through the modular fusion of at least two to three components (upper, middle, and lower pathways routes). For example, catechol *meta*-cleavage route is predominant and homologous in many bacterial taxa and is predicted to fuse with lower operon gene module (generating TCA intermediates). Whereas upper (peripheral) pathway operon might have been acquired as a separate module at a later stage, thus increasing the catabolic versatility of the recipient strain. Further to fine-tune metabolism, regulatory elements were recruited from the evolved bacterial genome(s) and were modified subsequently to optimize the expression of the catabolic enzymes (Harayama, 1994; Gerischer, 2002). The G + C content and codon usage analyses of the *nah* operon of *P. stutzeri* AN10 support this modular fusion and recombination-based evolutionary theory. It has been shown that G + C content and the codon usage patterns of *P. stutzeri* AN10 are different for both the upper and lower naphthalene pathway genes. A similar pattern is also observed in strain CSV86^T. The occurrence of non-coding (repeat-transposases) DNA region, flanking the conserved *nah* upper pathway has been suggested to be involved in the mobilization and further rearrangements of this entire catabolic module (Eaton, 1994). Overall, these observations strengthen our hypothesis of modular fusion of *nah* cluster in ICE*nah*CSV86 upstream to *sal* operon. The modular regulation of enzymes involved in aromatic degradation (benzyl alcohol, hydroxybenzyl alcohol, and other aromatics) has been demonstrated at the biochemical level in CSV86^T (Basu et al., 2003). Carbon source dependent metabolic studies also indicate that the degradation of aromatic alcohol and naphthalene involves two regulons (Basu et al., 2003; Mohan and Phale, 2017). Cellular respiration (nmol O_2 consumed min^{-1} mg^{-1}) studies of strain CSV86^T grown on naphthalene, salicylate, and glucose on substrate naphthalene, salicylate, and catechol showed differential induction of modules/metabolic regulons. Similarly, we observed the metabolite-dependent induction of three regulons for the metabolism of naphthalene-based pesticide Carbaryl in *Pseudomonas* spp. (Singh et al., 2013). These results indicate that the end metabolite generated from the upper operon induces the enzymes involved in the middle or lower operons in a cascading manner, achieving metabolic efficiency.

Integrative conjugative elements-encoded degradation or resistance traits in bacteria provide advantages over plasmid-encoded traits. ICEs are able to integrate into the host chromosome in a site-specific manner, resulting in stable phenotypic property. Further, this eliminates the need for constant replication and maintenance as well as inherent variability in the copy number, which is observed in most plasmids. In addition, ICEs are reported to excise and transfer to new host(s) in response to a variety of

signals like RecA-dependent SOS response to DNA damage, secretion of signaling molecules from recipients, the growth phase of the host, and additional mechanisms related to the expression of cargo genes (Johnson and Grossman, 2015). Interestingly, ICEs are also reported to facilitate the transfer of other mobile elements, thus providing genetic bioaugmentation/plasticity in the community for better adaptability and evolution (Delavat et al., 2017; Liu et al., 2019).

Functional Analyses of P*nah*/P*sal* Promoters From ICE*nah*CSV86

A strong promoter (designated P*nah*, LDF = 6.03, *P*-value = 0.0023) in the upstream and several weaker promoters throughout the *nah* cluster were predicted (**Figure 6A**). A single putative strong promoter (LDF = 3.73) could also be detected upstream of the *sal* cluster (**Figure 6A**). NahR, a LysR family transcription regulator (LTTR), is found to be located in between *nah* and *sal* cluster on ICE*nah*CSV86 with its own promoter (LDF = 5.87) and transcribed in the opposite direction of *nah* cluster. NahR binding site is found to be located 60 bp upstream of the transcriptional start site of *nah* and *sal* clusters (**Supplementary Table 6**). The promoters (-10 box and -35 box) and NahR binding site were found to be similar to other naphthalene degradation operons (**Supplementary Figure 12**). Co-transcription analysis confirmed the polycistronic nature of the *nah* and *sal* operons, thus validating the presence of strong promoters (P*nah* and P*sal*).

Amongst the promoters (P*nah*, P*sal*, and P*sal*/NahR) fused with 1NH, as a reporter gene, the activity from P*nah* was significantly higher than that of P*sal*, which showed expression only in the presence of NahR regulator and the inducer, salicylate (**Figure 6B**), while the expression from IPTG-induced P*trc* was similar to P*nah*. The *Psal*/*nahR* system has also been studied in both *Pseudomonas* and *E. coli* systems and has been found to be inducible (de Lorenzo et al., 1993; Calero et al., 2016),[11] while the P*nah* has been reported to show leaky/constitutive expression (Burlage et al., 1990; Neilson et al., 1999). These observations provide an essential insight into the evolutionary logic driving the regulation of both operons. The presence of a strong but leaky promoter upstream to the upper *nah* operon is essential to generate the inducer (salicylate; the end product of the *nah* enzymes) at a sufficient concentration from the substrate (naphthalene). The lower operon harbors a relatively weak promoter which is tightly regulated and gets subsequently induced by a specific inducer metabolite generated by the action of upper pathway (*nah*) enzymes.

Genomic Islands in *Pseudomonas bharatica* CSV86^T Confer Stable Phenotypic Traits

Strain CSV86^T grown in the absence of GI-encoded selection pressure for 60 generations did not depict a significant change in the naphthalene utilization property and heavy metal tolerance phenotype as compared to a culture grown under selection pressure. Specifically, the growth rate, cell biomass yield, or enzyme activity (catechol 2,3-dioxygenase) in naphthalene amended conditions were similar (**Supplementary Table 7**). The growth in the presence of heavy metals (Co^{2+}, Zn^{2+}, Cd^{2+}) and MIC values also remained unaltered (**Supplementary Table 7**). Therefore, these GIs are stably integrated/maintained in the genome of strain CSV86^T and are not lost even in the absence of selection pressure. As a consequence, such elements have established themselves in diverse populations, playing an important role in bacterial evolution (Miyazaki et al., 2015; Delavat et al., 2017). Moreover, aromatic impacted habitats often contain a variety of toxic compounds like heavy metals, thus hindering the bioremediation process. Therefore, acquiring heavy metal tolerance trait is highly beneficial for a bacterium to survive in such contaminated niches (Jiménez et al., 2004; Phale et al., 2021). ICE*nah*CSV86 and PBGI-1 provide a survival advantage to strain CSV86^T in the presence of naphthalene and/or Co^{2+}-Zn^{2+}-Cd^{2+}.

CONCLUSION

Pseudomonas bharatica CSV86^T harbors three GIs (>50 Kb), suspected to be ICEs as single copies on the genome, belonging to the T4SS-ICE family: ICE*nah*CSV86 (105.2 Kb), PBGI-1 (89.8 Kb), and PBGI-2 (91.9 Kb), and are syntenic to ICE*clc* (ICE*XTD*), ICE*Tn4371*, and PAGI-2, respectively. Based on the cargo genes present on these GIs and phenotypic traits observed in strain CSV86^T, ICE*nah*CSV86 was attributed for naphthalene degradation, PBGI-1 for heavy metal (Co^{2+}, Zn^{2+}, Cd^{2+}) resistance-efflux, and PBGI-2 for mixed/unknown functions. ICE*nah*CSV86 is the only functionally and genomically characterized prototypical element (predicted to be an ICE) for naphthalene degradation with a low conjugative transfer frequency. PBGI-2 is hypothesized to be an intermediate element between ICE and GI and proposed to be an ancestor for ICE*nah*CSV86. The *nah* cluster is predicted to be integrated through modular fusion upstream to the *sal* cluster followed by rearrangement and domain reshuffling to achieve metabolic efficiency. These elements are stably integrated into the genome and are not lost even in the absence of selection pressure, rendering advantageous survival benefits and adaptive lifestyle to the strain CSV86^T under varied environmental conditions. The promoters of *nah-sal* on ICE*nah*CSV86 were transcriptionally active and found to be suitable for heterologous protein expression in *E. coli*. Functional genomic and genome-wide comparative analyses of multiple GIs in a single organism provide crucial insights into the evolution and distribution of metabolic traits, such as naphthalene degradation and metal tolerance through horizontal gene transfer, contributing to genomic plasticity, metabolic diversity, and niche colonization abilities to bacterial species. Further, the role of core modules in the acquisition of diverse cargo functions (as suited to the habitat) mediated through modular exchange and fusion has been hypothesized.

[11] https://parts.igem.org/Part:BBa_J61051

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

PP, BM, and HM designed the study, performed data curation, validation, and interpretation, and agreed for the submission of the final version of the manuscript. BM performed the genome finishing, genomic data mining, comparative genomics, ICE-related experiments, and promoter prediction. HM conducted the promoter analyses, gene construct designing, co-transcription, and expression analyses. BM and HM wrote the manuscript drafts. PP managed the research funds and finalized the manuscript draft. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

BM acknowledges IIT Bombay for providing a Post-Doctoral Fellowship. HM thank CSIR, Government of India for the Senior Research Fellowship. PP thank DST, Government of India for the research funding.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.928848/full#supplementary-material

REFERENCES

Arciszewska, L. K., and Craig, N. L. (1991). Interaction of the Tn7-encoded transposition protein TnsB with the ends of the transposon. *Nucleic Acids Res.* 19, 5021–5029.

Badhai, J., and Das, S. K. (2016). Characterization of three novel SXT/R391 integrating conjugative elements ICE*MfuInd1a* and ICE*MfuInd1b*, and ICE*MprChn1* identified in the genomes of *Marinomonas fungiae* JCM 18476^T and *Marinomonas profundimaris* strain D104. *Front. Microbiol.* 7:1896. doi: 10.3389/fmicb.2016.01896

Baharoglu, Z., Garriss, G., and Mazel, D. (2013). Multiple pathways of genome plasticity leading to development of antibiotic resistance. *Antibiotics* 2, 288–315.

Basu, A., Apte, S. K., and Phale, P. S. (2006). Preferential utilization of aromatic compounds over glucose by *Pseudomonas putida* CSV86. *Appl. Environ. Microbiol.* 72, 2226–2230. doi: 10.1128/AEM.72.3.2226-2230.2006

Basu, A., Dixit, S. S., and Phale, P. S. (2003). Metabolism of benzyl alcohol *via* catechol *ortho*-pathway in methylnaphthalene-degrading *Pseudomonas putida* CSV86. *Appl. Microbiol. Biotechnol.* 62, 579–585. doi: 10.1007/s00253-003-1305-8

Basu, A., and Phale, P. S. (2008). Conjugative transfer of preferential utilization of aromatic compounds from *Pseudomonas putida* CSV86. *Biodegradation* 19, 83–92. doi: 10.1007/s10532-007-9117-7

Battle, S. E., Rello, J., and Hauser, A. R. (2009). Genomic islands of *Pseudomonas aeruginosa*. *FEMS Microbiol. Lett.* 290, 70–78.

Bignell, C., and Thomas, C. M. (2001). The bacterial ParA-ParB partitioning proteins. *J. Biotechnol.* 91, 1–34.

Bohlin, J., Snipen, L., Hardy, S. P., Kristoffersen, A. B., Lagesen, K., Dønsvik, T., et al. (2010). Analysis of intra-genomic GC content homogeneity within prokaryotes. *BMC Genomics* 11:464. doi: 10.1186/1471-2164-11-464

Bosch, R., Garcıìa-Valdés, E., and Moore, E. R. (1999). Genetic characterization and evolutionary implications of a chromosomally encoded naphthalene-degradation upper pathway from *Pseudomonas stutzeri* AN10. *Gene* 236, 149–157. doi: 10.1016/s0378-1119(99)00241-3

Botelho, J., Grosso, F., and Peixe, L. (2018). Unravelling the genome of a *Pseudomonas aeruginosa* isolate belonging to the high-risk clone ST235 reveals an integrative conjugative element housing a bla_{GES-6} carbapenemase. *J. Antimicrob. Chemother.* 73, 77–83. doi: 10.1093/jac/dkx337

Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* 72, 248–254.

Burlage, R. S., Sayler, G. S., and Larimer, F. (1990). Monitoring of naphthalene catabolism by bioluminescence with *nah-lux* transcriptional fusions. *J. Bacteriol.* 172, 4749–4757. doi: 10.1128/jb.172.9.4749-4757.1990

Burrus, V. (2017). Mechanisms of stabilization of integrative and conjugative elements. *Curr. Opin. Microbiol.* 38, 44–50.

Burrus, V., Pavlovic, G., Decaris, B., and Guédon, G. (2002). Conjugative transposons: the tip of the iceberg. *Mol. Microbiol.* 46, 601–610. doi: 10.1046/j.1365-2958.2002.03191.x

Calero, P., Jensen, S. I., and Nielsen, A. T. (2016). Broad-host-range ProUSER vectors enable fast characterization of inducible promoters and optimization of *p*-coumaric acid production in *Pseudomonas putida* KT2440. *ACS Synth. Biol.* 5, 741–753. doi: 10.1021/acssynbio.6b00081

Chain, P. S., Denef, V. J., Konstantinidis, K. T., Vergez, L. M., Agulló, L., Reyes, V. L., et al. (2006). *Burkholderia xenovorans* LB400 harbors a multi-replicon, 9.73-Mbp genome shaped for versatility. *Proc. Natl. Acad. Sci. U.S.A.* 103, 15280–15287. doi: 10.1073/pnas.0606924103

Christie, P. J., Whitaker, N., and González-Rivera, C. (2014). Mechanism and structure of the bacterial type IV secretion systems. *Biochim. Biophys. Acta. Mol. Cell Res.* 1843, 1578–1591.

de Lorenzo, V., Eltis, L., Kessler, B., and Timmis, K. N. (1993). Analysis of *Pseudomonas* gene products using *lac*Iq/*P*trp-lac plasmids and transposons that confer conditional phenotypes. *Gene* 123, 17–24. doi: 10.1016/0378-1119(93)90533-9

Delavat, F., Miyazaki, R., Carraro, N., Pradervand, N., and van der Meer, J. R. (2017). The hidden life of integrative and conjugative elements. *FEMS Microbiol. Rev.* 41, 512–537.

Dobrindt, U., Hochhut, B., Hentschel, U., and Hacker, J. (2004). Genomic islands in pathogenic and environmental microorganisms. *Nat. Rev. Microbiol.* 2, 414–424.

Duttagupta, S., Mukherjee, A., Bhattacharya, A., and Bhattacharya, J. (2020). Wide exposure of persistent organic pollutants (PoPs) in natural waters and sediments of the densely populated Western Bengal basin, India. *Sci. Total Environ.* 717:137187. doi: 10.1016/j.scitotenv.2020.137187

Eaton, R. W. (1994). Organization and evolution of naphthalene catabolic pathways: sequence of the DNA encoding 2-hydroxychromene-2-carboxylate isomerase and *trans-o*-hydroxybenzylidenepyruvate hydratase-aldolase from the NAH7 plasmid. *J. Bacteriol.* 176, 7757–7762. doi: 10.1128/jb.176.24.7757-7762.1994

Fang, G., Bhardwaj, N., Robilotto, R., and Gerstein, M. B. (2010). Getting started in gene orthology and functional analysis. *PLoS Comput. Biol.* 6:e1000703. doi: 10.1371/journal.pcbi.1000703

Fuentes, S., Méndez, V., Aguila, P., and Seeger, M. (2014). Bioremediation of petroleum hydrocarbons: catabolic genes, microbial communities, and applications. *Appl. Microbiol. Biotechnol.* 98, 4781–4794. doi: 10.1007/s00253-014-5684-9

Gaillard, M., Vallaeys, T., Vorhölter, F. J., Minoia, M., Werlen, C., Sentchilo, V., et al. (2006). The *clc* element of *Pseudomonas* sp. strain B13, a genomic island

with various catabolic properties. *J. Bacteriol.* 188, 1999–2013. doi: 10.1128/JB.188.5.1999-2013.2006

Gerischer, U. (2002). Specific and global regulation of genes associated with the degradation of aromatic compounds in bacteria. *J. Mol. Microbiol. Biotechnol.* 4, 111–121.

Hacker, J., and Carniel, E. (2001). Ecological fitness, genomic islands and bacterial pathogenicity. *EMBO Rep.* 2, 376–381.

Harayama, S. (1994). Codon usage patterns suggest independent evolution of two catabolic operons on toluene-degradative plasmid TOL pWW0 of *Pseudomonas putida*. *J. Mol. Evol.* 38, 328–335. doi: 10.1007/BF00163150

Hildebrand, F., Meyer, A., and Eyre-Walker, A. (2010). Evidence of selection upon genomic GC-content in bacteria. *PLoS Genet.* 6:e1001107. doi: 10.1371/journal.pgen.1001107

Hirose, J., Watanabe, T., Futagami, T., Fujihara, H., Kimura, N., Suenaga, H., et al. (2021). A new ICE*clc* subfamily integrative and conjugative element responsible for horizontal transfer of biphenyl and salicylic acid catabolic pathway in the PCB-degrading strain *Pseudomonas stutzeri* KF716. *Microorganisms* 9:2462. doi: 10.3390/microorganisms9122462

Janssen, P. J., Van Houdt, R., Moors, H., Monsieurs, P., Morin, N., Michaux, A., et al. (2010). The complete genome sequence of *Cupriavidus metallidurans* strain CH34, a master survivalist in harsh and anthropogenic environments. *PLoS One* 5:e10433. doi: 10.1371/journal.pone.0010433

Jiménez, J. I., Minambres, B., García, J. L., and Díaz, E. (2004). "Genomic insights in the metabolism of aromatic compounds in *Pseudomonas*," in *Pseudomonas*, ed. J. L. Ramos (Boston, MA: Springer), 425–462.

Johnson, C. M., and Grossman, A. D. (2015). Integrative and conjugative elements (ICEs): what they do and how they work. *Ann. Rev. Genet.* 49, 577–601. doi: 10.1146/annurev-genet-112414-055018

Juhas, M., Van Der Meer, J. R., Gaillard, M., Harding, R. M., Hood, D. W., and Crook, D. W. (2009). Genomic islands: tools of bacterial horizontal gene transfer and evolution. *FEMS Microbiol. Rev.* 33, 376–393.

Jurka, J., Kapitonov, V. V., Kohany, O., and Jurka, M. V. (2007). Repetitive sequences in complex genomes: structure and evolution. *Annu. Rev. Genomics Hum. Genet.* 8, 241–259.

Klockgether, J., Würdemann, D., Reva, O., Wiehlmann, L., and Tummler, B. (2007). Diversity of the abundant pKLC102/PAGI-2 family of genomic islands in *Pseudomonas aeruginosa*. *J. Bacteriol.* 189, 2443–2459. doi: 10.1128/JB.0168806

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33, 1870–1874. doi: 10.1093/molbev/msw054

Lawley, T., Wilkins, B. M., and Frost, L. S. (2004). Bacterial conjugation in gram-negative bacteria. *Plasmid Biol.* 13, 203–226.

Lechner, M., Schmitt, K., Bauer, S., Hot, D., Hubans, C., Levillain, E., et al. (2009). Genomic island excisions in *Bordetella petrii*. *BMC Microbiol.* 9:141. doi: 10.1186/1471-2180-9-141

Lesic, B., Bach, S., Ghigo, J. M., Dobrindt, U., Hacker, J., and Carniel, E. (2004). Excision of the high-pathogenicity island of *Yersinia pseudotuberculosis* requires the combined actions of its cognate integrase and Hef, a new recombination directionality factor. *Mol. Microbiol.* 52, 1337–1348. doi: 10.1111/j.1365-2958.2004.04073.x

Liu, M., Li, X., Xie, Y., Bi, D., Sun, J., Li, J., et al. (2019). ICEberg 2.0: an updated database of bacterial integrative and conjugative elements. *Nucleic Acids Res.* 47, D660–D665. doi: 10.1093/nar/gky1123

Mahajan, M. C., Phale, P. S., and Vaidyanathan, C. S. (1994). Evidence for the involvement of multiple pathways in the biodegradation of 1-and 2-methylnaphthalene by *Pseudomonas putida* CSV86. *Arch. Microbiol.* 161, 425–433. doi: 10.1007/BF00288954

Miyazaki, R., Bertelli, C., Benaglio, P., Canton, J., De Coi, N., Gharib, W. H., et al. (2015). Comparative genome analysis of *Pseudomonas knackmussii* B13, the first bacterium known to degrade chloroaromatic compounds. *Environ. Microbiol.* 17, 91–104. doi: 10.1111/1462-2920.12498

Miyazaki, R., Minoia, M., Pradervand, N., Sulser, S., Reinhard, F., and Van Der Meer, J. R. (2012). Cellular variability of RpoS expression underlies subpopulation activation of an integrative and conjugative element. *PLoS Genet.* 8:e1002818. doi: 10.1371/journal.pgen.1002818

Mohan, K., and Phale, P. S. (2017). Carbon source-dependent inducible metabolism of veratryl alcohol and ferulic acid in *Pseudomonas putida* CSV86. *Appl. Environ. Microbiol.* 83:e03326-16. doi: 10.1128/AEM.03326-16

Mohapatra, B., Nain, S., Sharma, R., and Phale, P. S. (2022). Functional genome mining and taxono-genomics reveal eco-physiological traits and species distinctiveness of aromatic-degrading *Pseudomonas bharatica* sp. nov. *Environ. Microbiol. Rep.* 14, 464–474. doi: 10.1111/1758-2229.13066

Mohapatra, B., and Phale, P. S. (2021). Microbial degradation of naphthalene and substituted naphthalenes: metabolic diversity and genomic insight for bioremediation. *Front. Bioeng. Biotechnol.* 9:602445. doi: 10.3389/fbioe.2021.602445

Neilson, J. W., Pierce, S. A., and Maier, R. M. (1999). Factors influencing expression of *lux*CDABE and *nah* genes in *Pseudomonas putida* RB1353 (NAH7, pUTK9) in dynamic systems. *Appl. Environ. Microbiol.* 65, 3473–3482. doi: 10.1128/AEM.65.8.3473-3482.1999

Nojiri, H., Sota, M., and Shintani, M. (2009). "Catabolic plasmids involved in the degradation of polycyclic aromatic hydrocarbons and heteroaromatic compounds," in *Microbial Megaplasmids*, ed. E. Schwartz (Berlin: Springer), 55–87.

Obi, C. C., Vayla, S., De Gannes, V., Berres, M. E., Walker, J., Pavelec, D., et al. (2018). The integrative conjugative element *clc* (ICE*clc*) of *Pseudomonas aeruginosa* JB2. *Front. Microbiol.* 9:1532. doi: 10.3389/fmicb.2018.01532

Paliwal, V., Raju, S. C., Modak, A., Phale, P. S., and Purohit, H. J. (2014). *Pseudomonas putida* CSV86: a candidate genome for genetic bioaugmentation. *PLoS One* 9:e84000. doi: 10.1371/journal.pone.0084000

Phale, P. S., Malhotra, H., and Shah, B. A. (2020). Degradation strategies and associated regulatory mechanisms/features for aromatic compound metabolism in bacteria. *Adv. Appl. Microbiol.* 112, 1–65. doi: 10.1016/bs.aambs.2020.02.002

Phale, P. S., Mohapatra, B., Malhotra, H., and Shah, B. A. (2021). Eco-physiological portrait of a novel *Pseudomonas* sp. CSV86: an ideal host/candidate for metabolic engineering and bioremediation. *Environ. Microbiol.* 1:15694. doi: 10.1111/1462-2920.15694

Phale, P. S., Shah, B. A., and Malhotra, H. (2019). Variability in assembly of degradation operons for naphthalene and its derivative, carbaryl, suggests mobilization through horizontal gene transfer. *Genes* 10:569. doi: 10.3390/genes10080569

Poulin-Laprade, D., Carraro, N., and Burrus, V. (2015). The extended regulatory networks of SXT/R391 integrative and conjugative elements and IncA/C conjugative plasmids. *Front. Microbiol.* 6:837. doi: 10.3389/fmicb.2015.00837

Qian, W., Jia, Y., Ren, S. X., He, Y. Q., Feng, J. X., Lu, L. F., et al. (2005). Comparative and functional genomic analyses of the pathogenicity of phytopathogen *Xanthomonas campestris* pv. *campestris*. *Genome Res.* 15, 757–767. doi: 10.1101/gr.3378705

Rajanna, C., Wang, J., Zhang, D., Xu, Z., Ali, A., Hou, Y. M., et al. (2003). The *Vibrio* pathogenicity island of epidemic *Vibrio cholerae* forms precise extrachromosomal circular excision products. *J. Bacteriol.* 185, 6893–6901. doi: 10.1128/JB.185.23.6893-6901.2003

Ravatn, R., Zehnder, A. J., and van der Meer, J. R. (1998b). Low-frequency horizontal transfer of an element containing the chlorocatechol degradation genes from *Pseudomonas* sp. strain B13 to *Pseudomonas putida* F1 and to indigenous bacteria in laboratory-scale activated-sludge microcosms. *Appl. Environ. Microbiol.* 64, 2126–2132. doi: 10.1128/AEM.64.6.2126-2132.1998

Ravatn, R., Studer, S., Springael, D., Zehnder, A. J., and van der Meer, J. R. (1998a). Chromosomal integration, tandem amplification, and deamplification in *Pseudomonas putida* F1 of a 105-kilobase genetic element containing the chlorocatechol degradative genes from *Pseudomonas* sp. strain B13. *J. Bacteriol.* 180, 4360–4369. doi: 10.1128/JB.180.17.4360-4369.1998

Reva, O. N., and Tümmler, B. (2005). Differentiation of regions with atypical oligonucleotide composition in bacterial genomes. *BMC Bioinformatics* 6:251. doi: 10.1186/1471-2105-6-251

Rodríguez-Beltrán, J., Sørum, V., Toll-Riera, M., de la Vega, C., Peña-Miller, R., and San Millán, Á (2020). Genetic dominance governs the evolution and spread of mobile genetic elements in bacteria. *Proc. Natl. Acad. Sci. U.S.A.* 117, 15755–15762. doi: 10.1073/pnas.2001240117

Ryan, M. P., Pembroke, J. T., and Adley, C. C. (2009). Novel Tn*4371*-ICE like element in *Ralstonia pickettii* and genome mining for comparative elements. *BMC Microbiol.* 9:242. doi: 10.1186/1471-2180-9-242

Sentchilo, V., Czechowska, K., Pradervand, N., Minoia, M., Miyazaki, R., and van der Meer, J. R. (2009). Intracellular excision and reintegration dynamics of the ICE*clc* genomic island of *Pseudomonas knackmussii* sp. strain B13. *Mol. Microbiol.* 72, 1293–1306. doi: 10.1111/j.1365-2958.2009.06726.x

Singh, R., Trivedi, V. D., and Phale, P. S. (2013). Metabolic regulation and chromosomal localization of carbaryl degradation pathway in *Pseudomonas* sp. strains C4, C5 and C6. *Arch. Microbiol.* 195, 521–535. doi: 10.1007/s00203-013-0903-9

Strnad, H., Ridl, J., Paces, J., Kolar, M., Vlcek, C., and Paces, V. (2011). Complete genome sequence of the haloaromatic acid-degrading bacterium *Achromobacter xylosoxidans* A8. *J. Bacteriol.* 193, 791–792. doi: 10.1128/JB.01299-10

van der Meer, J. R., and Sentchilo, V. (2003). Genomic islands and the evolution of catabolic pathways in bacteria. *Curr. Opin. Biotechnol.* 14, 248–254.

Verma, H., Kumar, R., Oldach, P., Sangwan, N., Khurana, J. P., Gilbert, J. A., et al. (2014). Comparative genomic analysis of nine *Sphingobium* strains: insights into their evolution and hexachlorocyclohexane (HCH) degradation pathways. *BMC Genomics* 15:1014. doi: 10.1186/1471-2164-15-1014

Verma, N., Kantiwal, U., Yadav, Y. K., Teli, S., Goyal, D., and Pandey, J. (2019). "Catalytic promiscuity of aromatic ring-hydroxylating dioxygenases and their role in the plasticity of xenobiotic compound degradation," in *Microbial Metabolism of Xenobiotic Compounds*, ed. P. Arora (Singapore: Springer), 123–143.

Wiegand, I., Hilpert, K., and Hancock, R. E. (2008). Agar and broth dilution methods to determine the minimal inhibitory concentration (MIC) of antimicrobial substances. *Nat. Protoc.* 3, 163–175.

Wozniak, R. A., Fouts, D. E., Spagnoletti, M., Colombo, M. M., Ceccarelli, D., Garriss, G., et al. (2009). Comparative ICE genomics: insights into the evolution of the SXT/R391 family of ICEs. *PLoS Genet.* 5:e1000786. doi: 10.1371/journal.pgen.1000786

Wozniak, R. A., and Waldor, M. K. (2010). Integrative and conjugative elements: mosaic mobile genetic elements enabling dynamic lateral gene flow. *Nat. Rev. Microbiol.* 8, 552–563. doi: 10.1038/nrmicro2382

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

Zamarro, M. T., Martín-Moldes, Z., and Díaz, E. (2016). The ICE*XTD* of *Azoarcus* sp. CIB, an integrative and conjugative element with aerobic and anaerobic catabolic properties. *Environ. Microbiol.* 18, 5018–5031. doi: 10.1111/1462-2920.13465

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: 18 July 2022
doi: 10.3389/fmicb.2022.933373

Metagenomics: An Approach for Unraveling the Community Structure and Functional Potential of Activated Sludge of a Common Effluent Treatment Plant

Gunjan Vasudeva†, *Harpreet Singh*†, *Sakshi Paliwal and Anil Kumar Pinnaka**

MTCC-Microbial Type Culture Collection and Gene Bank, CSIR-Institute of Microbial Technology, Chandigarh, India

OPEN ACCESS

Edited by:
Om Prakash,
National Centre for Cell Science, India

Reviewed by:
Clelton A. Santos,
Brazilian Biorenewables National Laboratory - National Center for Research in Energy and Materials, Brazil
Bhaskar Reddy,
Indian Agricultural Research Institute (ICAR), India
Saisai Zhou,
Huazhong Agricultural University, China

***Correspondence:**
Anil Kumar Pinnaka
apinnaka@imtech.res.in

†*These authors have contributed equally to this work*

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

Received: *30 April 2022*
Accepted: *17 June 2022*
Published: *18 July 2022*

Citation:
Vasudeva G, Singh H, Paliwal S and Pinnaka AK (2022) Metagenomics: An Approach for Unraveling the Community Structure and Functional Potential of Activated Sludge of a Common Effluent Treatment Plant. Front. Microbiol. 13:933373. doi: 10.3389/fmicb.2022.933373

The common effluent treatment plant (CETP) located at Baddi treats the industrial effluent from various industries, leading to the pooling of a diverse range of substrates and metabolites. The nutrient loading and its availability decide the balance of the microbial community and its diversity. The samples thus collected from the activated sludge (BS14) of CETP and Sirsa river (SR1) from the vicinity of CETP effluent discharge were processed for the whole metagenome analysis to reveal the microbial community and its functional potential. The taxonomic classification of the BS14 sample showed the dominance of the bacterial community with 96% of abundance, whereas the SR1 was populated by eukaryotes representing 50.4% of the community of SR1. The bacterial community of SR1 was constituted of 47.2%. The functional analysis of BS14 and SR1 with GhostKOALA against the KEGG database assigned 43.7% and 27.8% of the open reading frames (ORFs) with functions. It revealed the xenobiotic degradation modules with complete pathways along with resistance against the beta-lactams. The analysis with the comprehensive antibiotic resistance database (CARD) revealed 33 and 32 unique types of antimicrobial resistance in BS14 and SR1, respectively. Both the samples were dominated by the beta-lactam resistance genes. The carbohydrate-active enzyme (CAZy) database assigned a total of 6,611 and 2,941 active enzymes to BS14 and SR1, respectively. In contrast, the glycosyl hydrolases (GH) and glycosyltransferases (GT) class of enzymes were found to be abundant in both the samples as compared with polysaccharide lyases (PL), auxiliary activities (AA), carbohydrate esterases (CE), and carbohydrate-binding module (CBM).

Keywords: metagenomics, antimicrobial resistance, effluent treatment plant (ETP), metagenomic binning, functional annotation

INTRODUCTION

Multi-omics, also known as integrated omics, is a modern field of biology that combines more than one biology-based omics data to explore the structure and interactions of the complex biological system at the individual and community level (Krassowski et al., 2020). The different applications of omics studies include genomics (DNA-based), metagenomics (community DNA),

transcriptomics (all RNA-based), epigenomics, pan genomics, proteomics (protein-based), and metabolomics (metabolites content) (Bersanelli et al., 2016; Bock et al., 2016; Vilanova and Porcar, 2016). The different approaches to omics and the identification of biological markers obtained from DNA, RNA, proteins, and metabolites also helped to understand the various biological processes, health diseases, microbiology, and other environment-related processes (Marvasi et al., 2021).

The multi-omics profiling of unique niches for unraveling the microbial community and their metabolite composition involves a variety of high-throughput data, including 16S, shotgun metagenomics, metatranscriptomics, and metabolomics (Shaffer et al., 2021). The microbial communities are a crucial part of all lives on earth, and they maintain all the biogeochemical cycles on the earth by cycling the element between the lithosphere, atmosphere, hydrosphere, and biosphere (Madsen, 2011; Griggs et al., 2013). The advancement of next-generation sequencing (NGS) and omics technologies has generated high throughput data and analysis platforms to study the structure of microbial communities on-site, bypassing the need for pure culture isolation (Unamba et al., 2015; Costessi et al., 2018; Ben Khedher et al., 2022).

Wastewater treatment plants (WWTPs) are specialized systems that collects and treats wastewater for downstream usage to improve the quality of human and aquatic life (Waldrop, 2021). The increased urbanization and industrialization pose a severe risk to all living forms by increasing the accumulation of toxic pollutants (McMichael, 2000; Satterthwaite et al., 2010) and can also disseminate a load of various pathogens and antibiotic resistance genes, thus should be subjected for biological and chemical treatment (Mukherjee et al., 2021; Nguyen et al., 2021). Common effluent treatment plants (CETPs) offer a combined system to treat the wastewater collected from various small and medium-scale local industries (Padalkar and Kumar, 2018). ∼193 CETPs are established in India for primary to secondary treatment to increase water reusability for nonpotable purposes (Ali et al., 2021). The secondary treatment of wastewater involves the application of activated sludge microbiome (bacteria, archaea, protists, and fungi) to remove the dissolved and suspended organic matter, measurable by biological oxygen demand by 90%, both aerobically and anaerobically (Narayanan and Narayan, 2019).

The primary aim of CETPs and ETPs is to remove toxic, hazardous compounds before being discharged into the environment, primarily aquatic systems. Along with hazardous compounds, the CETP incoming water has a high concentration of pharmaceutical products, antibiotics, and heavy metals (Hubeny et al., 2021; Zieliński et al., 2021). The high presence of antibiotics in wastewater also increases the selective pressure to exchange antibiotic resistance genes among the microbes (Kraemer et al., 2019). Microbiome profiling of wastewater, activated sludge of effluent treatment plants, and receiving aquatic environment have already been reported in many studies through culture-dependent, quantitative PCR (qPCR), and culture-independent methods, e.g., 16S and shotgun metagenomics (Chu et al., 2018). Due to the limitation of the culturability of microbes, the whole spectrum of microbial diversity and their functions cannot be captured; however, using shotgun metagenomics, it is possible to untap the hidden microbial community and their interactions in any given environmental sample (Handelsman, 2004; Bodor et al., 2020). Using shotgun metagenomics analysis, the influence of wastewater on the genetic composition of sediment microflora was observed in terms of the dissemination of antibiotic resistance gene mobile genetic elements into the receiving aquatic system (Matviichuk et al., 2022).

The primary objective of this study was microbiome analysis of activated sludge of CETP, Baddi, and its receiving freshwater Sirsa river (SR1). To fulfill our objective, the high throughput shotgun metagenomic sequencing data using the Illumina NextSeq500 platform was obtained. The taxonomic and functional profiling was performed for both samples to elucidate microbial community composition and study their metabolic capabilities. This study could help understand the similarities and differences between the microbial and functional potential of both sites and shed light on wastewater's influence on the microbial and genetic composition of its receiving freshwater.

MATERIALS AND METHODS

Site Description and Sample Collection

Samples were collected from CETP established at Baddi Himachal Pradesh, India. Baddi-Barotiwala-Nalagarh (BBN) is one of the India's most extensive industrial belts and is the world's third-largest pharma hub. Baddi was declared a severely polluted area by Central Pollution Control Board in 2006 (Cluster, 2018). In this regard, the CETP was established in 2005–2006 in Kenduwal village (30.95°N 76.79°E) to treat the effluent load of 25 MLD (Million Liters per Day) from the various industries (e.g., textile, food, paper, detergent, pharmaceutical, and dye industries) in the BBN chain. Due to the massive variety of waste received from the BBN industrial belt, the CETP provides an artificially enriched environment for developing a highly efficient and dynamic microbial community.

The other sample collection was performed from the CETP water receiving fresh river SR1, Himachal Pradesh, Baddi. The SR1 arises in the Shivalik foothill of southern Himachal Pradesh and flows from Solan to the BBN area (29° 32'5.60 "N 75°01'44.33" E), and it meets the Sutlej River in Punjab. The various reports by the Centre of Pollution Control Board and other channels note that the extensive industrial waste is being dumped into the SR1 in the BBN area by CETP and other industries. We have collected the samples from the SR1 ∼100 m away from the drainage site of CETP.

DNA Extraction and Metagenomic Sequencing

The samples were collected in sterile containers, transferred to the laboratory as soon as possible, and processed further for high-molecular-weight metagenomic DNA isolation using the Meta-G-Nome^{TM} DNA Isolation Kit, Epicenter. The BS14 and SR1 samples were submitted for whole metagenomic sequencing to generate paired-end type reads (2×150) in NextSeq500 (Illumina).

Read Processing

The quality of the raw reads was assessed through FastQC (version 0.11.9) (Andrews, 2010), and it was further processed with trimmomatic (version 0.39) (Bolger et al., 2014) to remove the adapters (Leading:33, Trailing:33, without the sliding-window feature). The raw reads were trimmed with the force trim modulo feature of BBDuk (BBMap version 38.93) (Bushnell, 2014) to correct the read length from 151 bps to 150 bps. The read quality was improved with the parameters—rtrim, qtrim = 30, minlen = 51 *via* BBDuk. FastQC analysis was used at every step to assess the sample-specific read processing.

Metagenomic Assembly

The high-quality reads were assembled with a de-Bruijn graph-based assembler, i.e., metaSPAdes (version 3.15.3) (Nurk et al., 2017) at default parameters with a kmer value of 21, 33, 51, and 75 for both samples BS14 and SR-1. Using Bioawk (version 20110810), the contigs shorter than 1,000 bps were removed (Li, 2017).

Taxonomic Profiling

The metagenome assembly of BS14 and SR-1 were analyzed using Kaiju (version 1.8.1) (Menzel et al., 2016) for the taxonomic classification of the whole metagenome. Kaiju generates the Burrows–Wheeler transform (bwt) and Ferragina-Manzini index (fmi) of the nr database (version 26-03-2022) for the alignment. It translates the contigs into six possible reading frames and aligns them against the fm-index of the nr database. The taxonomic classification of BS14 and SR1 was run at default parameters with greedy mode.

Functional Profiling

The functional annotation of BS14 and SR1 assemblies was performed with the help of SqueezeMeta pipeline version 1.5.1 (Tamames and Puente-Sánchez, 2019). SqueezeMeta accepts the reads and does the automated processing and assembly followed by taxonomical and functional annotation of the genes. SqueezeMeta uses nr (Sayers et al., 2022), Pfam (Mistry et al., 2020), Clusters of Orthologous Genes (COG) (Tatusov et al., 2000), and Kyoto Encyclopedia of Genes and Genomes (KEGG) database for functional and taxonomic classification of the ORFs. The COG functions were manually extracted and annotated using DIAMOND (Buchfink et al., 2015) and visualized using ggplot2 (Wickham, 2016). The samples were also screened for the virulence factor through the virulence factor database (VFDB) analysis. The ORFs were aligned with VFDB to extract the relevant functions and visualized through ggplot2.

The amino acid sequences obtained from Prodigal (Hyatt et al., 2010) were further annotated using GhostKOALA (KEGG Orthology And Links Annotation), freely available on (https://www.kegg.jp/ghostkoala/) for assignment of KO numbers to infer functional categories and pathways for both metagenomes (Kanehisa et al., 2016).

Antimicrobial Resistance Profiling

The antimicrobial resistance (AMR) profiling of both BS14 and SR1 assemblies was performed by mapping the predicted ORFs obtained from Prodigal against the CARD database (homologs) (Alcock et al., 2020) using blastx mode of DIAMOND with a minimum identity of 50, query coverage of 90, and minimum bit score of 50 in fast mode. The AMR profile of both metagenomes was linked to the genus level of taxonomy obtained by Kaiju by comparing the first common fields of two files to combine the contents with the help of awk. The total AMR content of both metagenomes BS14 and SR1 was also compared to determine the shared resistome.

Carbohydrate Active Enzyme Profiling

The presence of carbohydrate-active enzymes in metagenome was determined by aligning the ORFs obtained by Prodigal to dbCAN database of carbohydrate-active enzyme (CAZy) annotation (Drula et al., 2022) using blastx mode of DIAMOND with 50% of minimum identity and 90% of query coverage with a minimum bit score of 50, in fast mode. Similarly, the polysaccharide utilization locus database (PULDB) was used to align against the ORFs of BS14 and SR1 (with the same parameters as used for the CAZy database) for the analysis of the polysaccharide metabolizing enzymes (Terrapon et al., 2018).

Binning and Bin Refinement

The contigs of BS14 and SR1 were binned using maxbin2, metabat, and concoct at default parameters (Alneberg et al., 2014; Wu et al., 2016; Kang et al., 2019). The metagenome-assembled genomes (MAGs) were further refined using the DAS Tool (Sieber et al., 2018). The refined bins were checked for completeness, contamination, and taxonomy lineage using checkm version 1.1.3 (Parks et al., 2015) with its lineage-specific mode. The bins with completeness >90% and contamination <5% were selected for further analysis.

Taxonomic Assignment and Functional Annotation of MAGs

The phylogenetic analysis and taxonomic novelty of MAGs were carried out using Genome Taxonomy Database Toolkit (GTDB-TK) for bacterial and archaeal genomes (Chaumeil et al., 2020) by placing the MAGs into a domain-specific reference tree. The functional metabolic potential of MAGs obtained from both SR1 and BS14 metagenomes was annotated using Distilled and Refined Annotation of Metabolism (DRAM). DRAM annotates the MAGs using different databases, including UniRef90, PFAM, dbCAN, RefSeq, VOGDB, and MEROPS peptide database (Shaffer et al., 2020).

RESULTS AND DISCUSSION

Taxonomic Profiling and Biodiversity Analysis of BS14 and SR1

The taxonomic profiling of BS14 and SR1 was performed by aligning the contigs using DIAMOND against the nr-database in fast mode and further analyzed using Kaiju. In this study, we observed the dominance of the bacterial domain in the BS14 sample with a prevalence of 96%. We also observed the

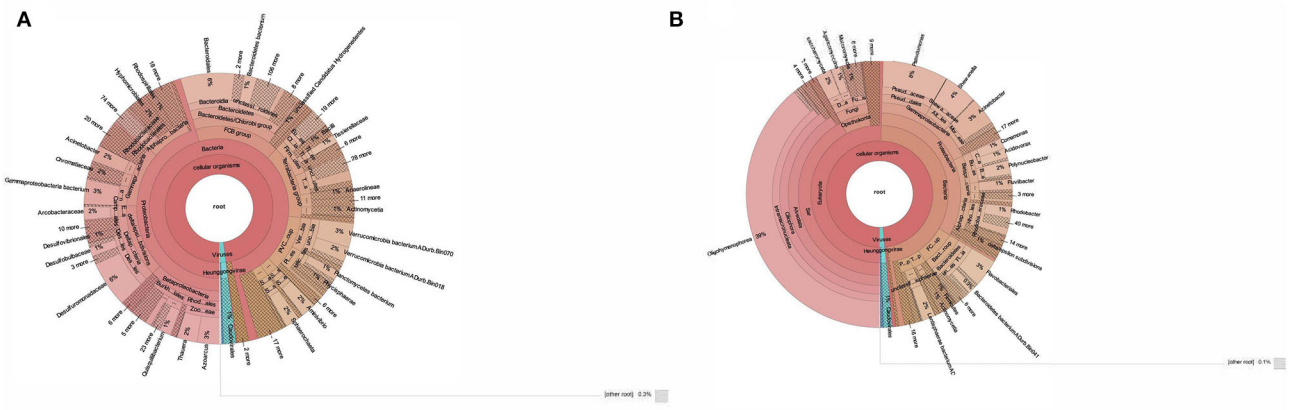


FIGURE 1 | (A) Taxonomic classification of BS14 at the genus level in krona plot. **(B)** Taxonomic classification of SR1 at the genus level in krona plot.

presence of other domains with an abundance of 0.94%, 0.48%, and 1.7% for archaea, eukarya, and viruses, respectively. In BS14, the predominant phyla observed were Proteobacteria, with an abundance of 47%, and the second predominant phylum was Bacteroidetes (15.3%). The other phyla observed in significant numbers belong to Verrucomicrobia, Firmicutes, Tenericutes, Planctomycetes, Acinetobacter, and others, with an abundance of 6.1%, 6.3%, 3.9%, 3.8%, 1.4%, and 8.4%, respectively. We have also observed the presence of Lentisphaerae, Planctomycetes, Spirochaetes, Synergistetes, Chloroflexi, Thermotogae, and Euryarchaeota (**Figure 1A**). In the case of BS14, among β-Proteobacteria, Rhodocyclales and Burkholderiales were dominant orders present with a prevalence of 46% and 42.2%, respectively, and other orders, i.e., Neisseriales (5%) and Nitrosomonadales (2.6%), respectively. Desulfuromonadales (59.4%), Desulfobacterales (15.2%), Desulfovibrionales (9.4%), Syntrophobacterales (4.38%), and Myxococcales (1.38%) were abundant order of δ-proteobacteria in BS14 metagenome. Among γ-proteobacteria, most prevalent orders were Moraxellales, Chromatiales, Pseudomonadales, Xanthomonadales, Enterobacteriales, Methylococcales, and Oceanospirillales with an abundance of 18.9%, 18%, 6.16%, 5.1%, 3.3%, 2.4%, and 2.2%, respectively. In the case of α-proteobacteria, Rhodobacterales (29.5%), Hyphomicrobiales (29.3%), Rhodospirillales (12.7%), Sphingomonadales (10%), and Caulobacterales (7%) were observed. Campylobacteriales (98.4%) were the most prevalent order observed in the case of ε-proteobacteria in BS14.

In contrast, SR1 has a high prevalence in the Eukarya domain, with an abundance of 50.4%. The other domains, i.e., bacteria, archaea, and viruses, were observed at 47.2%, 0.36%, and 1.4%, respectively (**Figure 1B**). Ciliophora (eukarya), with 79.3%, was observed as the predominant phyla in SR1, and other dominant phyla (eukarya) observed in SR1 were Ascomycota, Basidiomycota, Mucormycota, and Chlorophyta ∼2.7%, 2.0%, 1.16%, and 0.95%, respectively. Among Ciliophora, the Oligohymenophora (97%) was observed as the most predominant class, and the other classes observed in Ciliophora involved Spirotrichea (1.3%), Heterotrichea (1.2%), Litostomatea (0.18%), and Armphora (0.025%).

In BS14, 0.94% of the total sequences were assigned to archaea, with 65% and 25% of the sequences assigned to Euryarchaeota and Candidatus phyla. Other archaeal phyla in BS14 are Thaumarchaeota, Crenarchaeota, and Nanoarchaeota, with an abundance of 2.1%, 1.59%, and 0.7%, respectively. In comparison, 0.36% of total sequences were assigned to archaea in SR1, among which Euryarchaeota and Candidatus phyla were also found with a prevalence of 38.3% and 24.3%, respectively. The Crenarchaeota was also found in SR1 at ∼7.5%.

The eukarya domain of BS14 indicated the presence of Ascomycota and Basidiomycota as dominant phyla with an abundance of 32% and 21.8%, and the other eukaryotic phyla, i.e., Chlorophyta (9.2%) and Mucormycota (7.1%), were also present. Among viruses, 85% of the sequences were assigned to Uroviricota.

In SR1, the taxonomic distribution at the phylum level of the bacterial kingdom indicated the dominance of proteobacteria with an abundance of 69.8% and other phyla, Bacteroidetes, Lentisphaerae, Firmicutes, and Verrucomicrobia, with a prevalence of 14.5%, 3.9%, 2.38%, and 0.94%, respectively (**Figure 1B**).

The SR-1 was rich in gamma-proteobacteria with an abundance of 53.6%, and the other classes of proteobacteria, i.e., alpha, beta, delta, and epsilon proteobacteria, were observed at 14.6%, 25.4%, 0.198%, and 2.51%, respectively. In contrast, beta-proteobacteria (28%) was found abundant in BS14, and delta (24%), gamma (23%), alpha (17.5), epsilon (3.9%) proteobacteria were observed. Bacteroidias dominated BS14 (51.6%), while SR1 had only 8.9% of the population.

For γ-proteobacteria class of SR1, Pseudomonadales (45%), Alteromonadales (23.6%), and Moraxellales (22.9%) were the predominant orders. For β-proteobacteria, the most prevalent orders obtained were Burkholderiales (59.8%), Rhodocyclales (19.7%), and Neisseriales (14.29). For α-proteobacteria, the Rhodobacterales (76.4%), Rickettsiales

(64%), and Hyphomicrobiales (8.5%) were observed as the most abundant orders.

The taxonomic distribution at the species level indicated 7,973 and 3,588 unique assigned species, and 931 and 611 unique classifications were found unassigned in BS14 and SR1, respectively. BS14 was found to be enriched in *Desulfuromonas acetexigens, Synergistetes bacterium*, *Macellibacteroides fermentans*, *Arcobacter ellisii*, *Desulfobulbus propionicus*, *Desulfovibrio desulfuricans*, *Azoarcus communis*, *Mesotoga infera*, *Parabacteroides chartae*, *Thauera propionica*, *Quisquillibacterium transsilvanicum*, *Bacteroidetes bacterium*, *Lentimicrobium saccharophilum*, *Gammaproteobacterium bacterium*, *Verrucomicrobia bacterium*, and *Aminivibrio pyruvatiphilus.* In SR1, *Pseudomonas fragi*, *Pseudomonas aeruginosa*, *Lentisphaerae bacterium*, *Acinetobacter kyonggiensis*, *Acinetobacter bohemicus*, *Fluviibacter phosphoraccumulans*, *Acidovorax temperans*, *Paenirhodobacter* sp. MME-103, *Aliarcobacter cryaerophilus*, *Comamonas aquatica*, *Shewanella baltica*, *Arcobacter cryaerophilus*, *Rhodocyclaceae bacterium*, *Neisseriaceae bacterium*, *Polynucleobacter yangtzensis*, *Rickettsiales bacterium*, and *Rhodobacter* spp. were observed as most abundant species. As SR1, metagenome was predominated by eukarya; the most abundant species were *Tetrahymena thermophila*, *Ichthyophthirius multifiliis*, *Pseudocohnilembus persalinus*, *Stentor coeruleus*, *Paramecium sonneborni,* and *Stylonychia lemnae*.

The biodiversity analysis of the BS14 and SR1 metagenome was obtained manually. The contigs were assigned with taxonomy *via* Kaiju, and the abundance and proportions were extracted manually for the estimation of the diversity. The BS14 and SR1 samples showed a Simpson's index (Simpson, 1949) of 0.989 and 0.838, respectively. It indicates that both of the sampling sites exhibited a diverse microbial community. Although the SR1 sample was found to be abundant in eukaryotes, given it is a natural river habitat as compared with the sludge environment of BS14. The Shannon-Weiner index (Shannon, 1948) of 6.53 and 4.44 again supports the diverse microbial profile of the BS14 and SR1 sites, and the maximum allowed diversity for both was found to be 8.98 and 8.18, respectively. The evenness of 0.72 and 0.54 suggests that the population of BS14 was evenly distributed and that of SR1 was found to be quite low. Given the diversified nutritional value of the sludge at the BS14 site, the population was getting a constant supply of all basic nutrients to sustain an even and diverse growth, which was found to be lacking in the natural river (SR1).

The alpha diversity of the BS14 and SR1 was manually estimated from the taxonomic units of the metagenome assigned through Kaiju. The species-level diversity analysis revealed an alpha diversity of 7,973 and 3,588 unique species out of 63,471 and 35,952 contigs in BS14 and SR1 samples, respectively. The beta diversity and beta diversity index were calculated to be 8,793 and 0.239, with 1,384 common species among both the samples, respectively. The predicted gamma diversity was depicted as around 10,177 unique species in both samples.

Functional Analysis of BS14 and SR1

To determine the functional potential of BS14 and SR1, we annotated their predicted gene sequences against the KEGG database using the freely available webserver GhostKOALA. KEGG database is a repository of metabolic pathways to unravel the metabolism, biological pathways, disease, and drugs. For BS14 and SR1, 43.7% and 29.2% of total ORFs were annotated against the KEGG database and assigned to different functional pathways (**Supplementary Figure S1**). Genetic information processing was the most abundant functional category observed for BS14 and SR1 (15,780 and 7,843 ORFs, respectively) (**Supplementary Figure S2**).

The metabolic potential of BS14 and SR1 was also analyzed for biodegradation of xenobiotic compounds (**Supplementary Figure S3**). Most ORFs for BS14 (831) and SR1 (330) were assigned to different KO IDs (70 and 55 for BS14 and SR1, respectively) belonging to benzoate degradation, which also indicated the presence of benzoate as the most prevalent aromatic pollutant of CETP. The complete metabolic pathway was also observed for benzoate degradation in the BS14 and SR1 metagenome (**Figure 2**). The benzoate undergoes complete degradation with the help of ring hydroxylating and ring cleavage enzymes (Valderrama et al., 2012). The key enzymes responsible for benzoate degradation, i.e., benzoate 1,2 dioxygenase subunit alpha and beta (EC: 1.14.12.10), benzoate 1,2 dioxygenase reductase (EC: 1.18.1), dihydroxycyclohexadiene carboxylate dehydrogenase (EC: 1.3.1.25), and catechol 1,2 dioxygenase (EC: 1.13.11.1) were also observed in both metagenomes.

The functional annotation through KEGG also revealed the abundance of rTCA and chemolithotropism, which involves the energy generation using inorganic compounds, especially sulfur or nitrogen, as electron donors where the level of molecular O_2 is deficient (Reddy et al., 2019). During analysis, the equal distribution of sulfur and nitrogen metabolism was observed in both BS14 and SR1. The complete sulfur metabolism was observed, including both assimilatory and dissimilatory sulfate reduction, thiosulfate oxidation by SOX complex and amino acid metabolism. The critical enzymes for sulfur metabolism, i.e., dissimilatory sulfite reductase (EC: 1.8.99.5), adenylylsulfate reductase, subunit A (EC:1.8.99.2), 3′-phosphoadenosine 5′-phosphosulfate synthase (EC: 2.7.7.4), phosphoadenosine phosphosulfate reductase (EC: 1.8.4.8), and sulfate reductase (EC: 1.8.1.2;1.8.1.7) were found in sulfur metabolism pathway. Similarly, the complete nitrogen metabolism, i.e., nitrate reduction and denitrification, were observed in both metagenomes with the presence of key enzymes, i.e., nitric oxide reductase subunit B (EC: 1.7.2.5), nitrous oxide reductase (EC: 1.7.2.4), nitrogenous molybdenum iron protein alpha chain (EC: 1.18.6.1), nitrate reductase (EC: 1.7.5.1), and nitrite reductase (EC: 1.7.2.1).

The functional analysis of both metagenomes also revealed the presence of necessary CO_2 assimilatory enzymes involved in rTCA cycle, i.e., aconitate hydratase (EC: 4.1.2.3), isocitrate dehydrogenase (EC: 1.1.1.42), 2-oxoglutarate/2-oxo acid ferredoxin oxidoreductase subunit alpha (EC: 1.2.7.3), fumarate hydratase (EC: 4.2.1.2), fumarate reductase (EC: 1.3.5.4; 1.3.5.1),

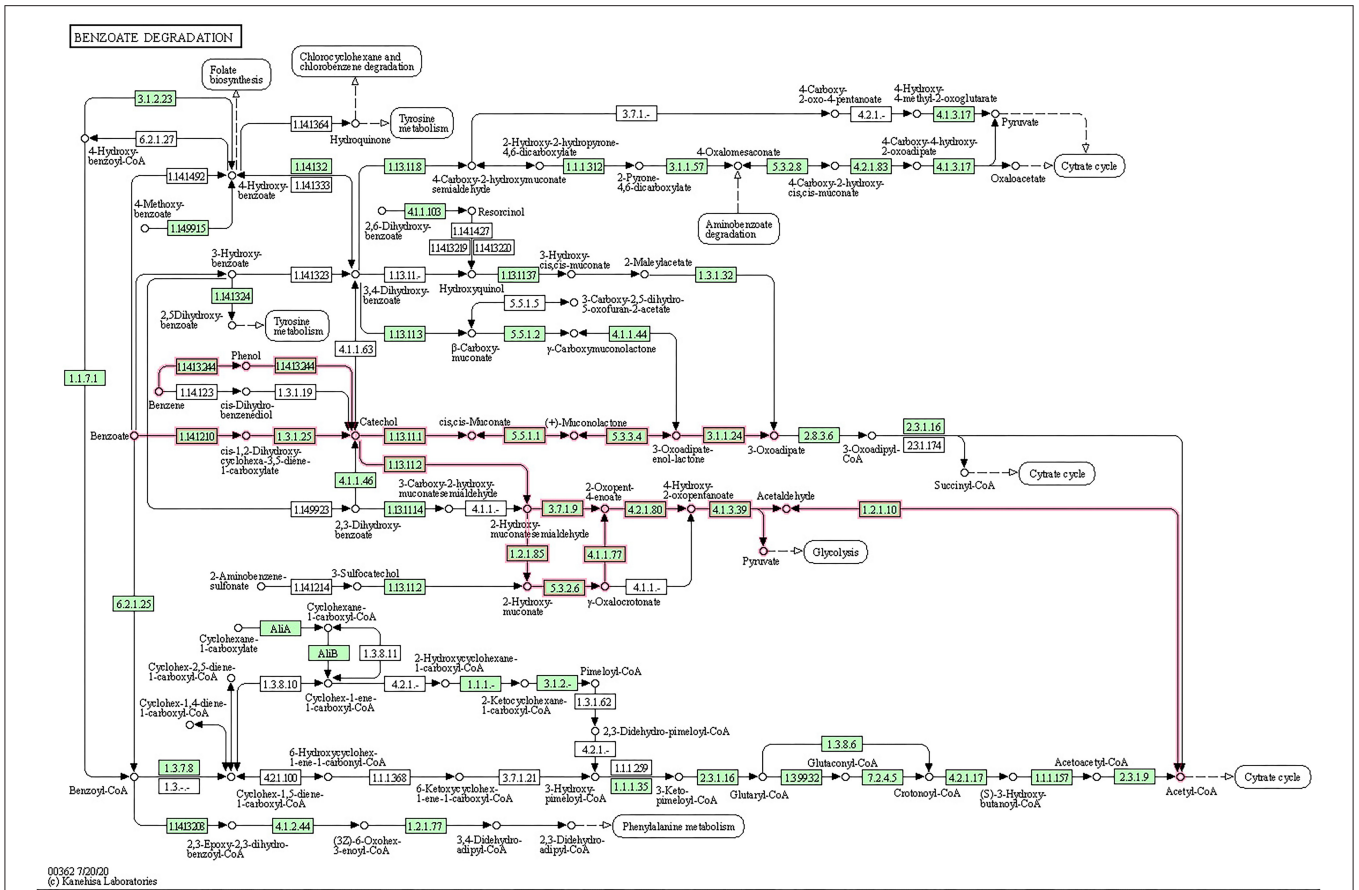


FIGURE 2 | Open reading frames (ORFs) of BS14 and SR1 mapped to benzoate biodegradation pathway.

pyruvate carboxylase (EC: 6.4.1.1), malate dehydrogenase (EC: 1.1.1.37), succinyl CoA synthetase (EC: 6.2.1.5), pyruvate-ferredoxin oxidoreductase (EC: 1.2.7.1), and most important enzyme ATP citrate lyase (EC: 2.3.3.8).

Various studies have shown sewage and effluent treatment plants as sources of AMR genes. Hendriksen et al. have shown the impact of municipal sewage treatment and effluent plants on disseminating AMR genes in its nearby freshwater system (Hendriksen et al., 2019). The KEGG analysis of BS14 and SR1 by GhostKOALA also indicated the prevalence of unique KO terms representing antimicrobial resistance. The resistance against the beta-lactams, cationic antimicrobial peptides (CAMP), and vancomycin was rich in BS14 and SR1. During the KEGG analysis, we observed that 996 ORFs of BS14 metagenome were mapped against 31 unique KO IDs belonging to beta-lactam resistance profile, 786 ORFs against 23 KEGG orthology terms related to CAMP resistance genes, and 457 ORFs related to 9 unique KO IDs of vancomycin resistance genes. Similarly, for SR1, the most prevalent resistance modules were observed against beta-lactam antibiotics, CAMP, and vancomycin. For SR1, 358, 336, and 117 ORFs were assigned to beta-lactam, CAMP, and vancomycin, respectively.

The complete metabolic module for beta-lactam resistance is shown in (**Supplementary Figure S4**). Various resistance modes were observed in BS14 and SR1 for beta-lactam resistance. The resistance modes include the loss or severe reduction of porins of outer membrane encoded by ompU, ampG; coding for MES transporter, nagZ; beta-N-acetylhexosaminidase (EC: 3.2.1.52), ampR; regulator of gene expression of beta-lactamase, ampC; beta-lactamase class C, oppA; oligopeptide transport system substrate-binding proteins, blaI; penicillinase repressor, bla2; beta-lactamase classA, penicillin-binding proteins mrcA, mrcB, pbp2A, mrdA, and fts1. The multidrug efflux pumps, i.e., arcA, mexA, ade1, smeD, mtrC, cmeA, adeA, and adeB for beta-lactam antibiotics, were also observed in BS14 and SR1. The beta-lactamase class A, class B, class C, and class D (EC; 3.5.2.6) were also observed.

The predicted ORFs of BS14 and SR1 samples were aligned using DIAMOND against the COG database, and a total of 95,907 and 42,017 functions were identified from the BS14 and SR1 metagenome, respectively. The COG functions were categorized as shown in (**Figure 3A**), which made it easy to understand that the bacterial community

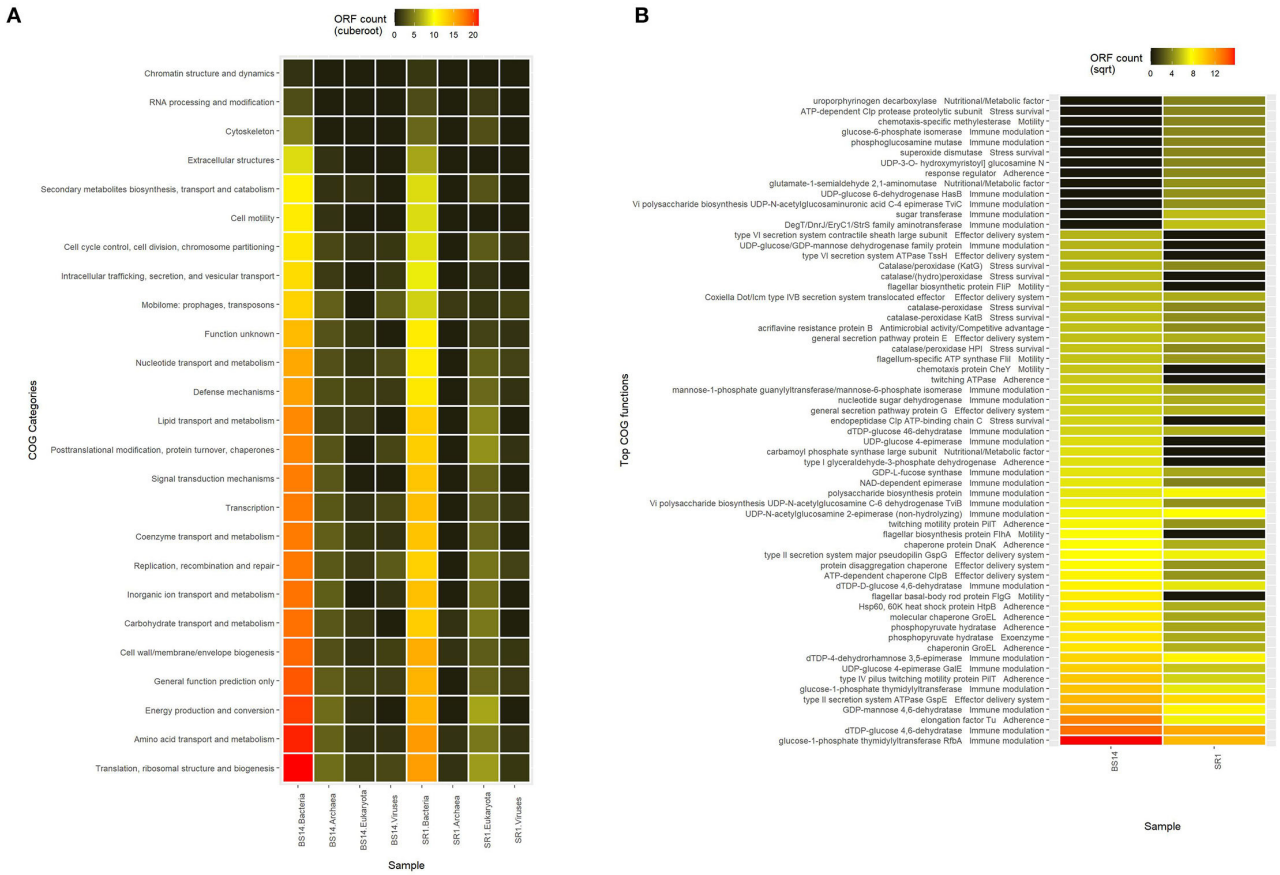


FIGURE 3 | (A) Heatmap showing the Cluster of Orthologous Genes (COG) functional assignment of BS14 and SR1 at the Kingdom level and **(B)** heatmap showing the top 50 COG functional assignments of BS14 and SR1.

majorly contributed to the functional potential of the system, followed by eukaryotes in SR1. The top 50 COG functions of BS14 and SR1 (**Figure 3B**) indicate the abundance of properties such as lipid metabolism, defense mechanism, transcription and translation regulation, cell wall biogenesis, signal transduction, etc. Certain enzymes such as transposase (0.56%), DNA binding response regulator or osmoregulatory protein (OmpR family; 0.44%), glycosyltransferase (0.49%) for cell wall synthesis, short-chain alcohol dehydrogenase (0.49%), and multidrug efflux pump (AcrB; 0.33%) were dominant in BS14 metagenome. The features of SR1 (**Figure 3B**) were similar to the BS14 metagenome. Although the predicted taxonomical dominance of SR1 was covered by the eukaryotes (especially Tetrahymena), it exhibited equal or higher relative abundance of COG functions with respect to BS14. SR1 was also prevalent in DNA binding transcription regulator (LysR family; 0.94%), arabinose efflux permease (0.58%), short-chain alcohol dehydrogenase (0.53%), DNA binding response regulator or osmoregulatory protein (OmpR family; 0.54%), multidrug efflux pump (AcrB; 0.37%), etc. The COG functions of BS14 and SR1 were additionally categorized into carbohydrate metabolism (6.37, 5.55%), inorganic ion transport and metabolism (6.22, 6.53%), secondary metabolism (1.36, 1.57%), and defense mechanism (4.06, 3.69%), and the top 10 enzymes are shown in (**Figure 4**). In BS14 and SR1, the carbohydrate metabolic category exhibited a high count of arabinose efflux permease (0.28, 0.58%), permease of drug/metabolite transporter (DMT) superfamily (0.21, 0.32%), phosphoenolpyruvate synthase (0.16, 0%: zero indicates low abundance), and enzymes for other carbohydrates (**Figure 4A**). The inorganic ion transport and metabolism showed the presence of outer membrane receptor for Fe-transport (0.19, 0.19%) and for ferrienterochelin and colicin (0.16, 0.19%), cation transporting P-type ATPase (0.15, 0.13%), and other metal/inorganic ion transporters (such as copper, potassium, nickel, sulfur, magnesium, sulfite, and nitrate/nitrite) (**Figure 4B**). The secondary metabolic category showed the prevalence of catechol 2,3-dioxygenase (0.09, 0.1%), 3-oxoacyl synthase (0.1, 0.09%), acyl-CoA thioesterase (0.077, 0.095%), and other enzymes (**Figure 4C**) of central aromatic metabolic pathways for benzoate and aminobenzoate utilization as shown in (**Figure 2**). The category of defense mechanism was abundant in multidrug efflux pump (AcrB; 0.33, 0.37%), ABC-type multidrug transport system-ATPase component (0.24, 0.18%), permease component

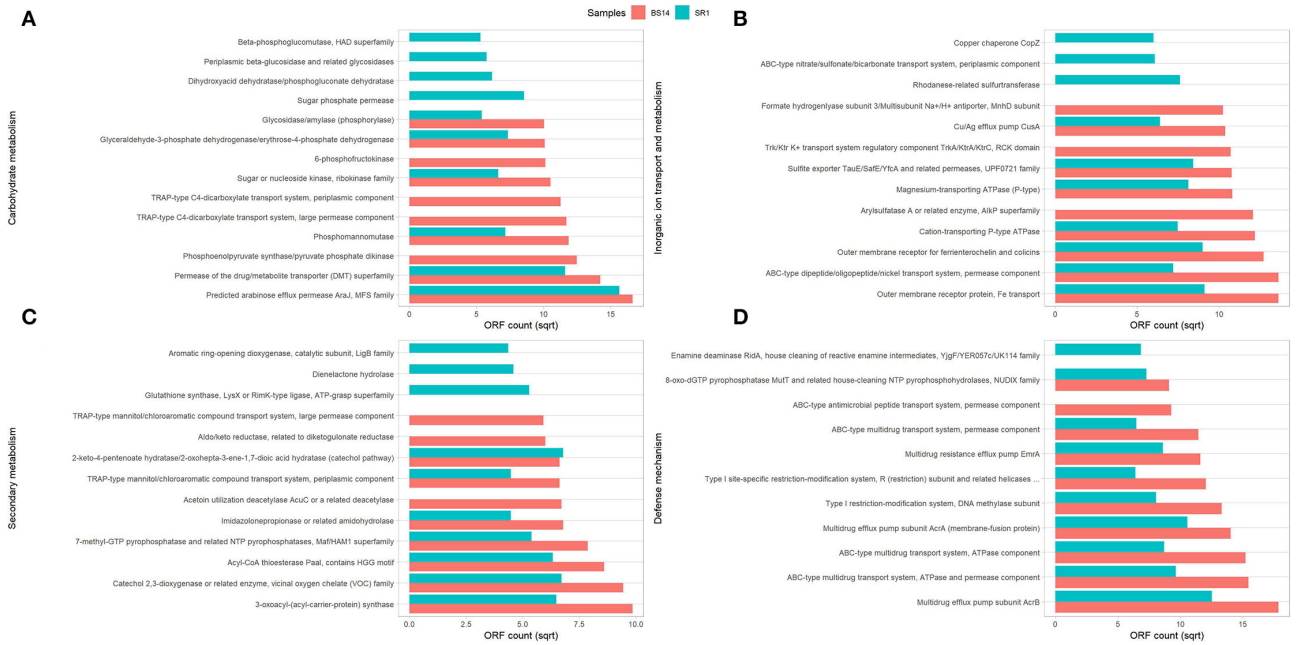


FIGURE 4 | Representation of top 10 abundant features of Clusters of Orthologous Genes (COG) category for BS14 and SR1 involved in **(A)** carbohydrate metabolism; **(B)** inorganic ion transport and metabolism; **(C)** secondary metabolism; and **(D)** defense mechanism.

(0.24, 0.22%), and other efflux system for peptides, etc. (**Figure 4D**).

The ORFs of BS14 and SR1 were also aligned against the PFAM database (Mistry et al., 2021) with the help of the SqueezeMeta pipeline, and the total complete functional profile was extracted with the inbuilt SQM tool of the software. An average of 7,292 and 6,870 unique functions were assigned to the BS14 and SR1 samples. In the BS14 sample, 23% and 25% of the ORFs were unmapped and unclassified respectively, whereas in SR1, 24% and 21% of the ORFs were unmapped and unclassified respectively. The samples were prevalent in the proteins responsible for the ATP binding, phosphorelay signal transduction system, membrane integrity, transmembrane transport, biosynthetic process, catalytic activity, oxidoreductase activity, carbohydrate metabolic process, etc. The top 50 functional assignment is shown in a heatmap in (**Supplementary Figure S5**).

Antimicrobial Resistance Profiling of BS14 and SR1

The alignment of BS14 and SR1 ORFs against the CARD database (Alcock et al., 2020) yielded 33 and 32 unique resistance profiles against different types of antibiotics. Some of them exhibited multidrug resistance properties. In BS14, the majority (11%) of ORFs showed resistance to the cephalosporin class of antibiotics. Second major (7.2%) resistance was found against the penam and several other classes of beta-lactam antibiotics such as carbapenem (5.6%), penem (2.84%), and monobactam (3.51%). In total, 3,872 copies of AMR genes were found in the BS14 metagenome, where most of them targeted beta-lactams and few against fluoroquinolones (8.96%), macrolides (8.13%), tetracycline (8.08%), phenicols (7.6%), aminoglycosides (6.19%), etc. In SR1, the beta-lactam class of cephalosporin and penam constitutes 13% and 11% of the AMR content, respectively, followed by tetracycline (8.7%), fluoroquinolone (7.7%), carbapenem (7.6%), macrolide (7.2%), etc. In SR1, 2,447 copies of AMRs were found and shared a similar profile with BS14. The shared resistome profile of both BS14 and SR1 indicated the dissemination of AMR genes from CETP to its receiving water body.

Out of 35 unique resistomes, SR1 was found to lack the gene for mupirocin, glycopeptide, and bicyclomycin, whereas BS14 lacked the benzalkonium chloride and fosfomycin.

The taxonomic annotation of the CARD data was performed manually, and the classification was assigned from Kaiju to analyze AMRs at the genus level. It was found that in the BS14 sample, the majority of AMRs were prevalent in *Azoarcus* (9.2%), *Acinetobacter* (6.9%), *Desulfuromonas* (6.3%), *Desulfobulbus* (6.4%), *Candidatus* (3.5%), and *Arcobacter* (3.1%). The rest of the AMRs in BS14 was originating from *Thauera* (2.8%), *Geoalkalibacter* (2.9%), *Quisquiliibacterium* (2.4%), *Sphingobium* (2.3%), *Aromatoleum* (2.3%), *Desulfococcus* (2.1%), *Trichlorobacter* (1.9%), *Formosimonas* (1.6%), *Sulfurifustis* (1.5%), *Desulfovibrio* (1.3%), *Escherichia* (0.67%), *Pseudogemmobacter* (0.4%), *Cupriavidus* (0.33%), *Snodgrassella* (0.26%), and *Pseudomonas* (0.39%). Approximately 40.05% of the AMR was not assigned to a genus (**Figure 5A**). Comparatively in SR1, *Pseudomonas* harbored ∼45.8% of the total AMR content, and the rest was distributed among *Acinetobacter*

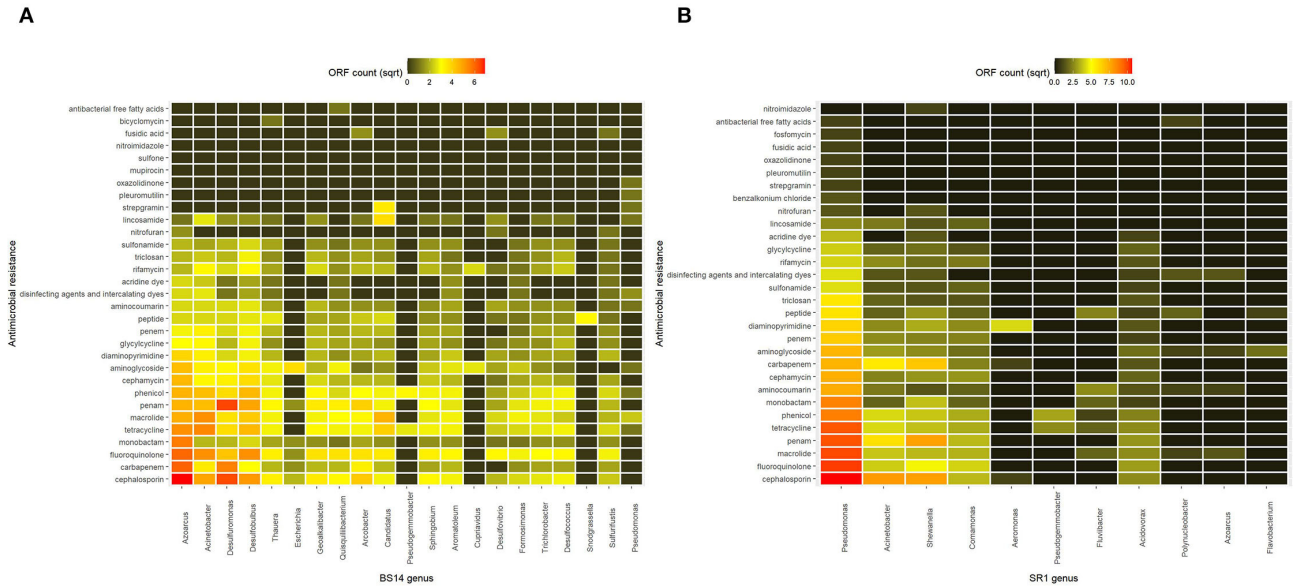


FIGURE 5 | Heatmap showing the antimicrobial resistance (AMR) prevalence of **(A)** BS14 sample at the genus level and **(B)** SR1 sample at the genus level.

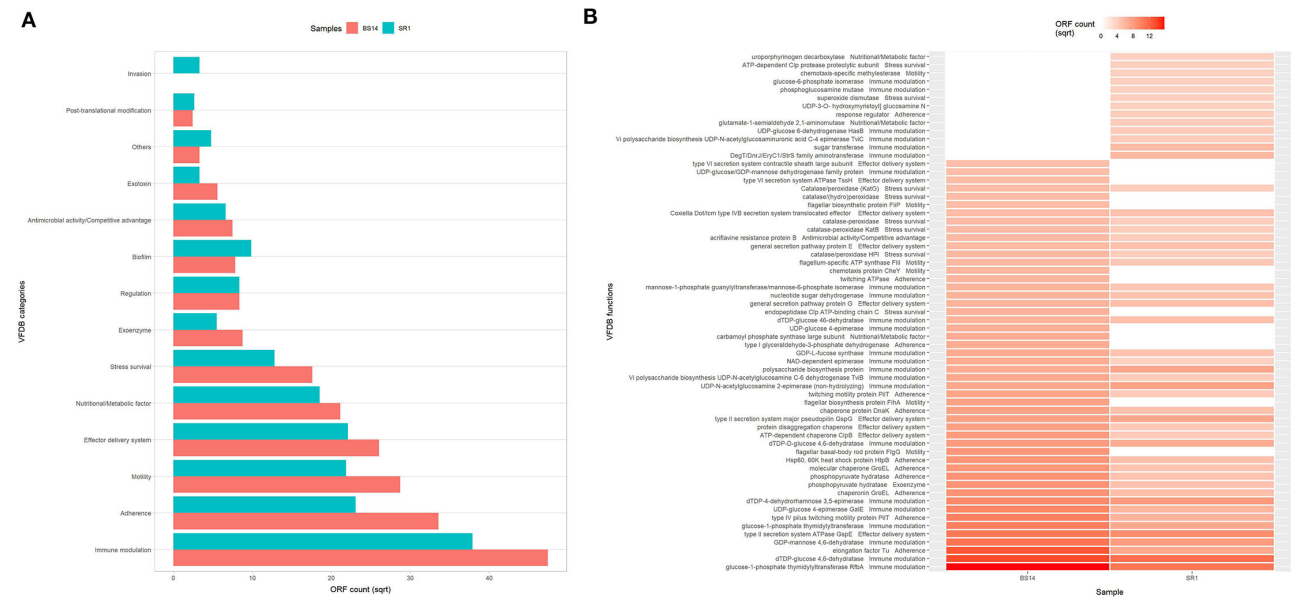


FIGURE 6 | (A) Distribution of virulent categories among the BS14 and SR1 samples from virulence factor database (VFDB) alignment and **(B)** heatmap showing the top 50 abundant virulent factors of BS14 and SR1 sample.

(10.6%), *Shewanella* (12.6%), *Comamonas* (5.3%), *Aeromonas* (0.81%), *Pseudogemmobacter* (0.69%), *Fluviibacter* (0.93%), *Acidovorax* (2.8%), *Polynucleobacter* (0.36%), *Azoarcus* (0.3%), and *Flavobacterium* (0.25%) (**Figure 5B**).

Virulence Profiling of BS14 and SR1

The alignment of BS14 and SR1 ORFs with the virulence factor database (VFDB) yielded 5,953 and 3,736 functions after removing the redundancy. Each ORFs of an individual node were assigned with the taxonomic annotation to determine the dominant class and genus contributing to the extent and type of virulence. In the BS14 class level, the majority of virulence was found to be distributed among the Betaproteobacteria (21%), Gammaproteobacteria (19.62%), Deltaproteobacteria (18.52%), Alphaproteobacteria (11.01%), Bacteroidia (3.61%), Clostridia (1.42%), etc. At the genus level, the abundance of *Desulfuromonas*

(6.81%), *Thauera* (6.38%), *Azoarcus* (5.35%), *Acinetobacter* (4.48%), *Geoalkalibacter* (3.25%), and *Desulfobulbus* (1.86%) was observed, whereas the major fraction was not assigned (38.65%) with a taxonomy. At the SR1 class level, the majority belonged to Gammaproteobacteria (51.04%), Betaproteobacteria (27.7%), Flavobacteriia (5.72), Alphaproteobacteria (5.46%), Oligohymenophorea (0.91%), etc. It is observable in SR1 that despite having a 50% population of eukaryotes, the majority of virulence is sourced from the prokaryotes. On exploring at the genus level, *Pseudomonas* (30.91%) contributed to more virulent factors followed by *Shewanella* (9.71%), *Acinetobacter* (9.28%), *Polynucleobacter* (7.36), *Flavobacteriia* (3.88%), *Fluviibacter* (3.37%), etc. The virulent factors of BS14 and SR1 were further classified into categories (**Figure 6A**) as immune modulation (37.79, 38.46%), adherence (18.96, 14.23%), motility (13.87, 12.79%), effector delivery system (11.4, 13.08%), nutritional metabolic factor (7.5, 9.18%), stress survival (5.2, 4.38%), exoenzyme (1.29, 0.8%), regulation (1.17, 1.87%), biofilm (1.02, 2.59%), posttranslational modification (0.1, 0.18%), invasion (0, 0.29%), exotoxin (0.52, 0.29%), antimicrobial activity/competitive advantage (0.94, 1.17%), and others (0.18, 0.61%), and some top 50 functions of VFDB as shown in (**Figure 6B**) indicates the presence of factors promoting the cell membrane modulation, adherence, motility, secretion system, antimicrobial resistance, radical scavenging, etc. The mining of exotoxin revealed the presence the of hemolysin A and B in both the samples, hemolysin III in BS14, and colibactin (clbI; a genotoxic metabolite) (Balskus, 2015) in SR1 but BS14 had colibactin self-protection protein (clbS) (Tripathi et al., 2017) and a gene for cyclolysin secretion protein (Cya) in both samples as shown in (**Figure 7**). The antibacterial activity/competitive advantage category contained the genes for hydrogen cyanide production (HcnABC; in SR1 only), resistance against acriflavine (AcrAB; in BS14 and SR1), and efflux pumps as shown in (**Figure 7**). The category of biofilm was also explored to reveal the genes responsible for localization and movement (flagellar), adherence, and mucus secretion (**Figure 7**).

Carbohydrate Active Enzyme Profiling of BS14 and SR1

The ORFs of BS14 and SR1 metagenomes were aligned against the CAZy database (Drula et al., 2022) with a minimum identity of 50% and query coverage of 90%. A total hit of 6,611 and 2,941 active enzymes was identified in BS14 and SR1 samples with a minimum bit score >50. The classes of enzymes in the CAZy database are distributed as glycosyl hydrolases (GH), glycosyltransferases (GT), polysaccharide lyases (PL), carbohydrate esterases (CE), auxiliary activities (AAs), and carbohydrate-binding module (CBM). In the BS14 sample, most enzymes belonged to the GH and GT classes having an abundance of 51.5% and 33.6%, respectively. The other enzymes, i.e., PL, CE, AA, and CBM, had an abundance of 1.1%, 3.5%, 1.1%, and 8.9%, respectively. The SR1 sample was abundant, with GH and GT covering 40.9% and 40.5% of the total active enzymes. The other classes, i.e., PL, CE, AA, and CBM occupied 0.4%, 2.8%, 3%, 12.1%, respectively. The taxonomy assignment by Kaiju showed that around 98.65% and 77.52% of active enzymes in BS14 and SR1 belonged to the bacterial kingdom, respectively. Since the SR1 sample had a higher abundance of eukaryotes, it represented 22.2% of the active enzymes. The circos plot of BS14 and SR1 shows the distribution of carbohydrate-active enzymes of BS14 and SR1 samples at kingdom level classification (**Figures 8, 9**). The phyla of Proteobacteria and Bacteroidetes held 37.2% and 48.3% and 23.6% and 18.8% of the active enzymes in the BS14 and SR1 samples, respectively.

The top 50 specific carbohydrate-active enzymes were also extracted from the SqueezeMeta tool, where the BS14 and SR1 ORFs were assigned the functional annotation from the KEGG database. BS14 and SR1 showed a common prevalence of lipid, formate, transketolase, glycogen, glucose, citric acid cycle intermediates, and polyhydroxyalkanoate (PHA) metabolism with different relative abundances. SR1 was also metabolically active for chitin. (**Supplementary Figures S6A,B**) shows the distribution of carbohydrate-active enzymes in the BS14 and SR1 samples.

The ORFs extracted from BS14 and SR1 were also aligned against the dbCAN-PUL database, and it yielded a total of 8,838 and 3,700 assignments after removing the redundancy. The total PULdb functions (as shown in a Sankey plot) reveal the distribution of polysaccharide biosynthesis and degradation potential among the BS14 and SR1 metagenomes (**Figure 10**). BS14 and SR1 were abundant in capsule polysaccharide biosynthesis (21.82, 30.78%), O-antigen biosynthesis (19.66, 25.54%), xylan, beta-glucan, lichenan degradation (7.79, 5.86%), O-glycan, N-glycan degradation (6.53, 4.08%), capsule polysaccharide degradation (2.8, 4.08%), and metabolism of some other polysaccharides. The ORFs with PULdb assignments were filtered to extract the polysaccharide degradative potential of BS14 and SR1. These ORFs were also assigned with the taxonomic profile at the phylum level manually using awk. Proteobacteria and Bacteroidetes represented the majority of the functions in both the BS14 and SR1 metagenome but the former was more prevalent in SR1. In BS14, the proteobacteria were highly active in xylan, beta-glucan, lichenan degradation (2.63%), and capsule polysaccharide degradation (1.98%), whereas Bacteroidetes was active in O-glycan, N-glycan degradation (4.88%), xylan, beta-glucan, lichenan degradation (3%), and some other polysaccharides as shown in (**Figure 11A**). In SR1, the Proteobacterium was a major player in polysaccharide utilization, it was rich in capsule polysaccharide degradation (3.29%), xylan, beta-glucan, and lichenan degradation (2.37%) enzymes, and Bacteroidetes was rich in xylan, beta-glucan, and lichenan degradation (2.45%), and O-glycan and N-glycan degradation (2.72%) (**Figure 11B**). Several features were found to be in common between the BS14 and SR1 samples, which could indicate the carryover of the functional properties from sludge (BS14) to discharge water (SR1).

Taxonomic Identification and Functional Annotation of MAGs

During binning, sixteen and thirty MAGs for SR1 and BS14, respectively, were obtained. For downstream

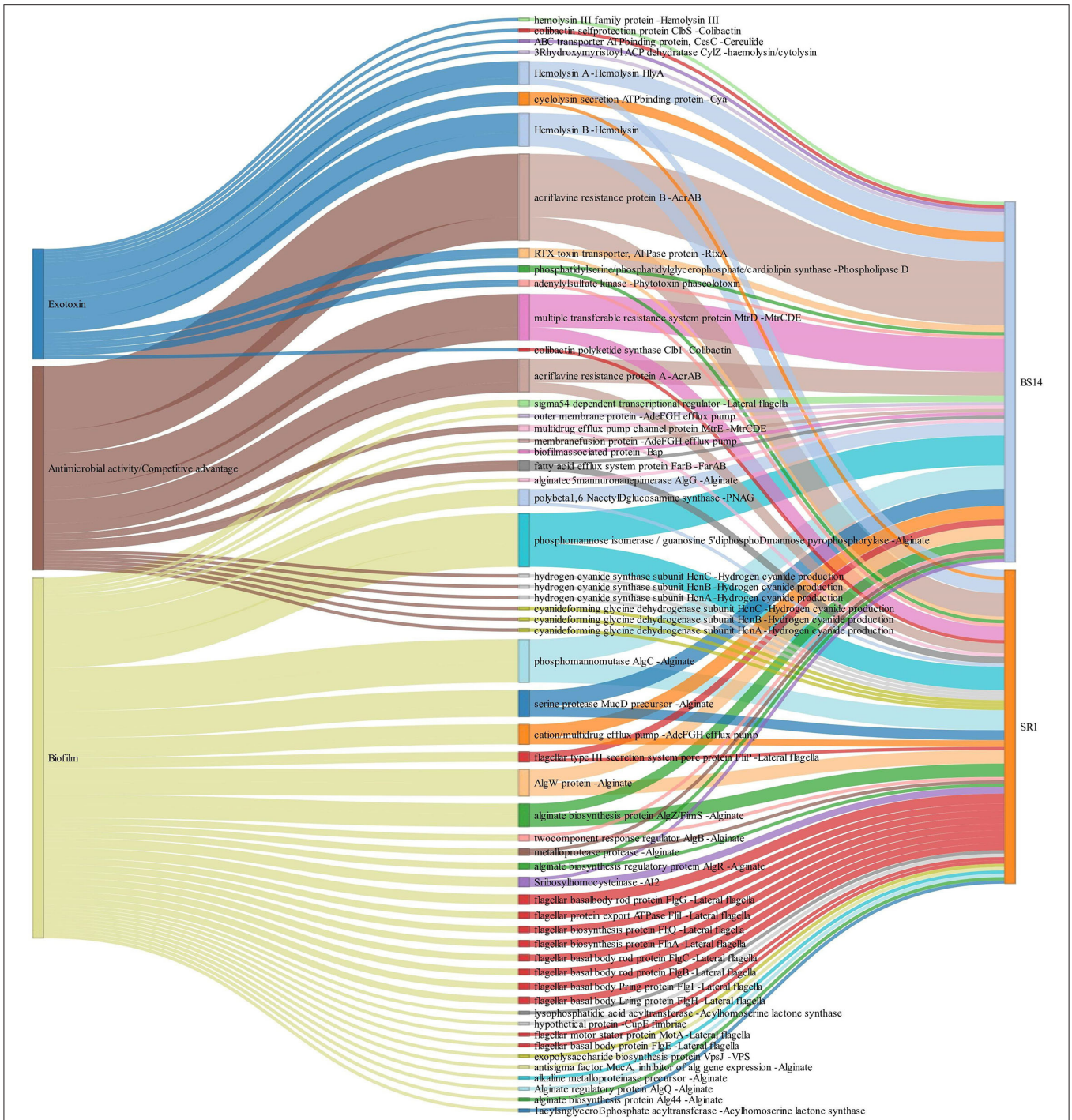


FIGURE 7 | Exploration of the biofilm, antimicrobial activity/competitive advantage, and exotoxin category of VFDB in BS14 and SR1 samples.

analysis, using checkm, the good quality MAGs with >90% completeness and <5% contamination were selected. A total of ten and fifteen good-quality MAGs for SR1 and BS14 were obtained. The checkm profile of MAGs for SR1 and BS14 is given in (**Supplementary Tables S1, S2**) respectively.

The taxonomy of MAGs for bacterial and archaeal genomes using GTDB-Tk workflow was assigned (**Supplementary Tables S3, S4**). The results obtained by GTDB classification demonstrated that the MAGs of BS14 were distributed in five different phyla. The eight MAGs of BS14 belong to phylum Proteobacteria, two MAGs were

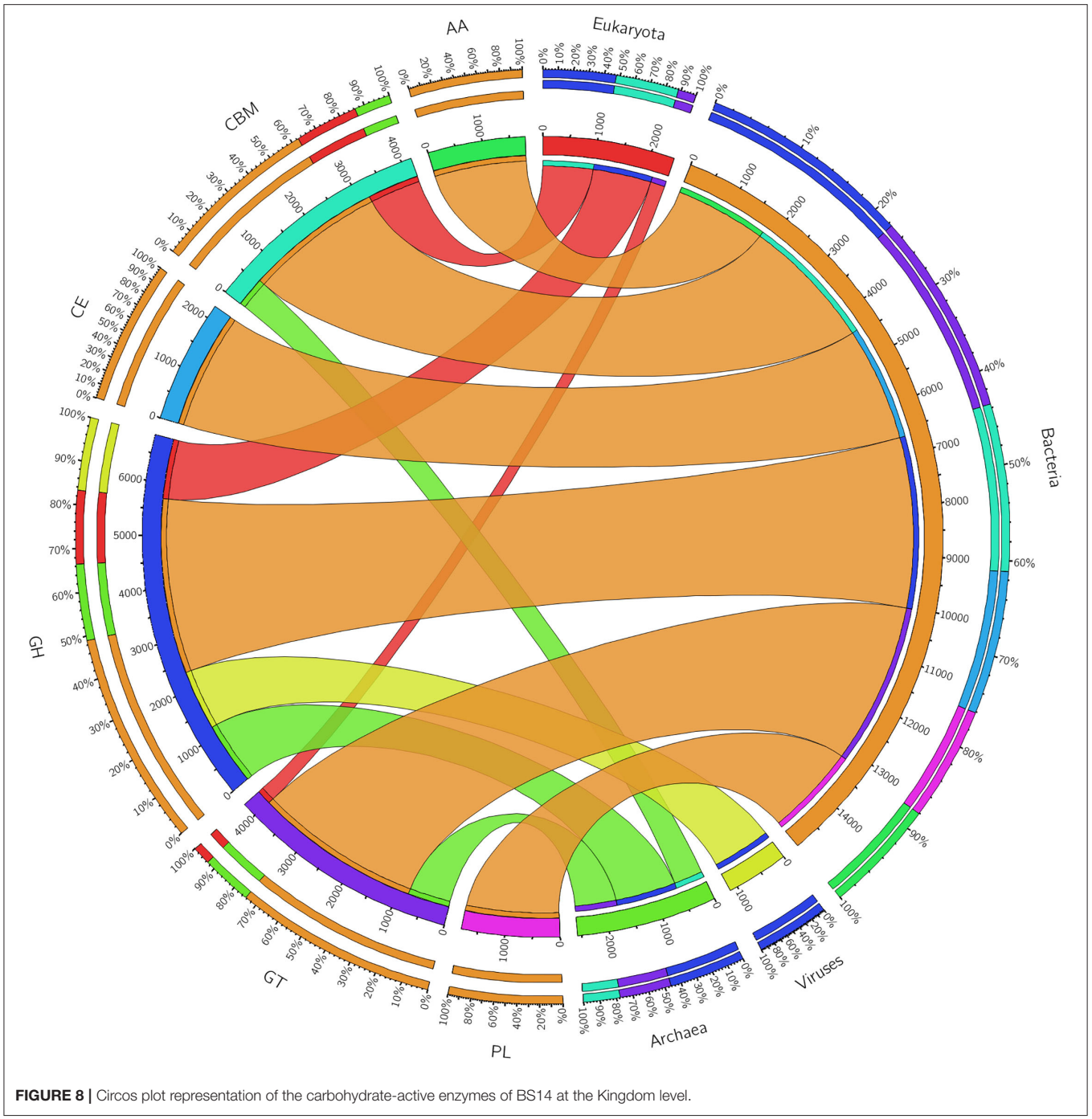


FIGURE 8 | Circos plot representation of the carbohydrate-active enzymes of BS14 at the Kingdom level.

assigned to phylum Bacteroidota, and three MAGs were assigned to Desulfobacterota. Additionally, the phylum Verrucomicrobiota and Firmicutes was also observed. The taxonomic novelty of the MAGs is explained by ANI (average nucleotide identity) value to the closest placed reference as computed by GTDB-Tk workflow (**Supplementary Table S5**). The MAG, BS14_bin1, and BS14_bin7 were thought to be novel taxa belonging to the family *Syntrophobacteraceae* and *Sphingomonadaceae* with ANI values of 78.49 and 78.5, respectively, with the closest reference genome. Similarly, BS14_bin25 and BS14_bin21 were also formally proposed as novel species belonging to the family *Steroidobacteraceae* and *Geoalkalibacteraceae* with ANI values of 88.91 and 79.08, respectively.

Similarly, in SR1, the major recovered phylum was Proteobacteria, whereas two MAGs were assigned to phylum Bacteroidota, and one MAG was assigned to Verrucomicrobiota and Campylobacter

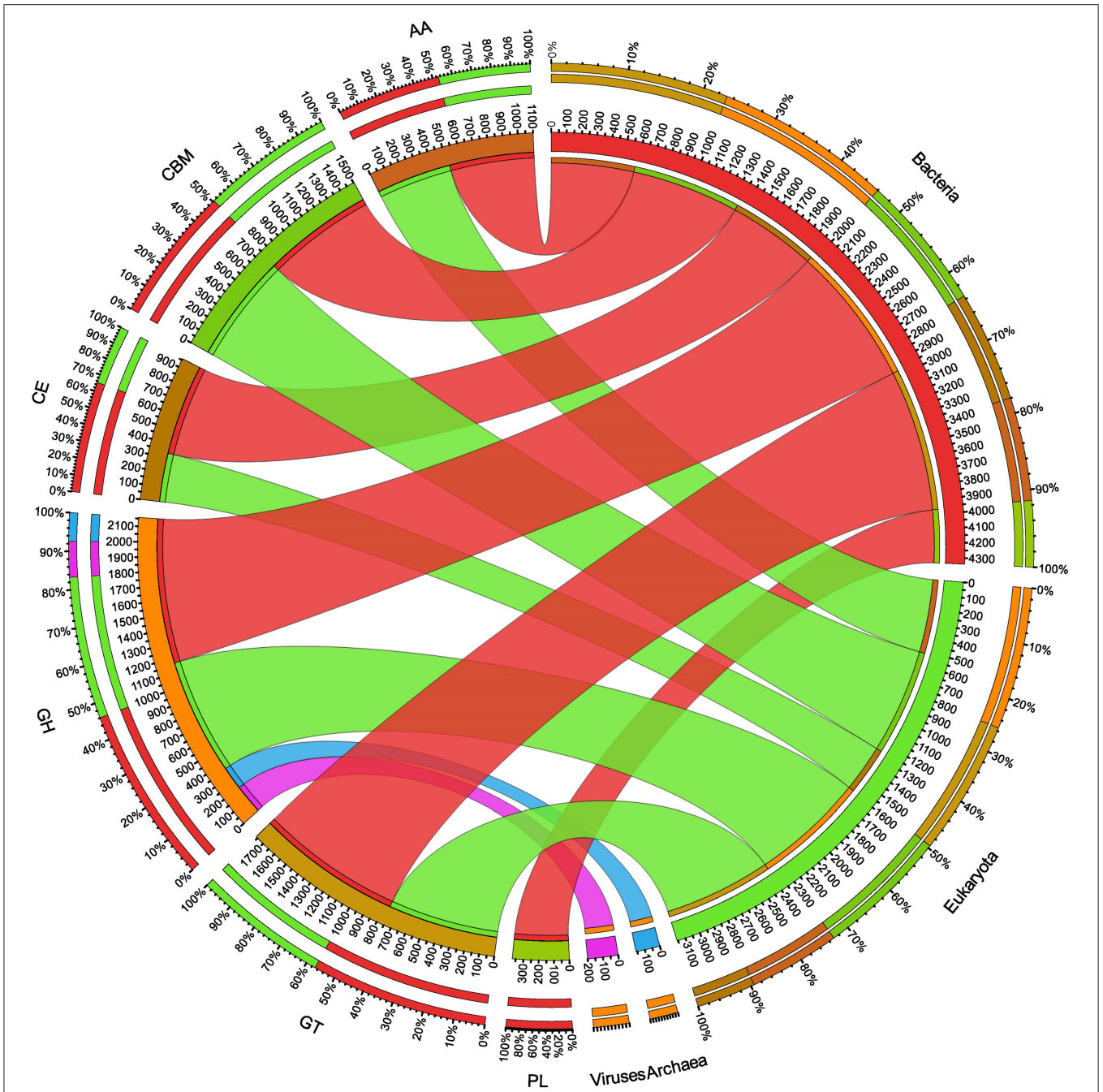


FIGURE 9 | Circos plot representation of the carbohydrate-active enzymes of SR1 at the Kingdom level.

(**Supplementary Table S6**). In SR1, the ANI value of all the MAGs was higher than the species delineation demarcation cutoff, thus no novel taxa were proposed.

The MAGs were annotated for a functional assignment using DRAM. DRAM is used to profile the MAGs for different metabolism, e.g., carbon, nitrogen, sulfur, and methanotrophy. We compared the MAGs obtained from SR1 and BS14 to analyze their metabolic potential in the microbiome (**Supplementary Figures S7A,B**).

The functional annotation using DRAM indicated the presence of various complete energy metabolism pathways in both SR1 and BS14. Most of the MAGs obtained from the SR1 and BS14 showed complete energy metabolism for glycolysis, pentose phosphate pathway, citrate cycle, reductive pentose phosphate pathway, and reductive citrate cycle, whereas

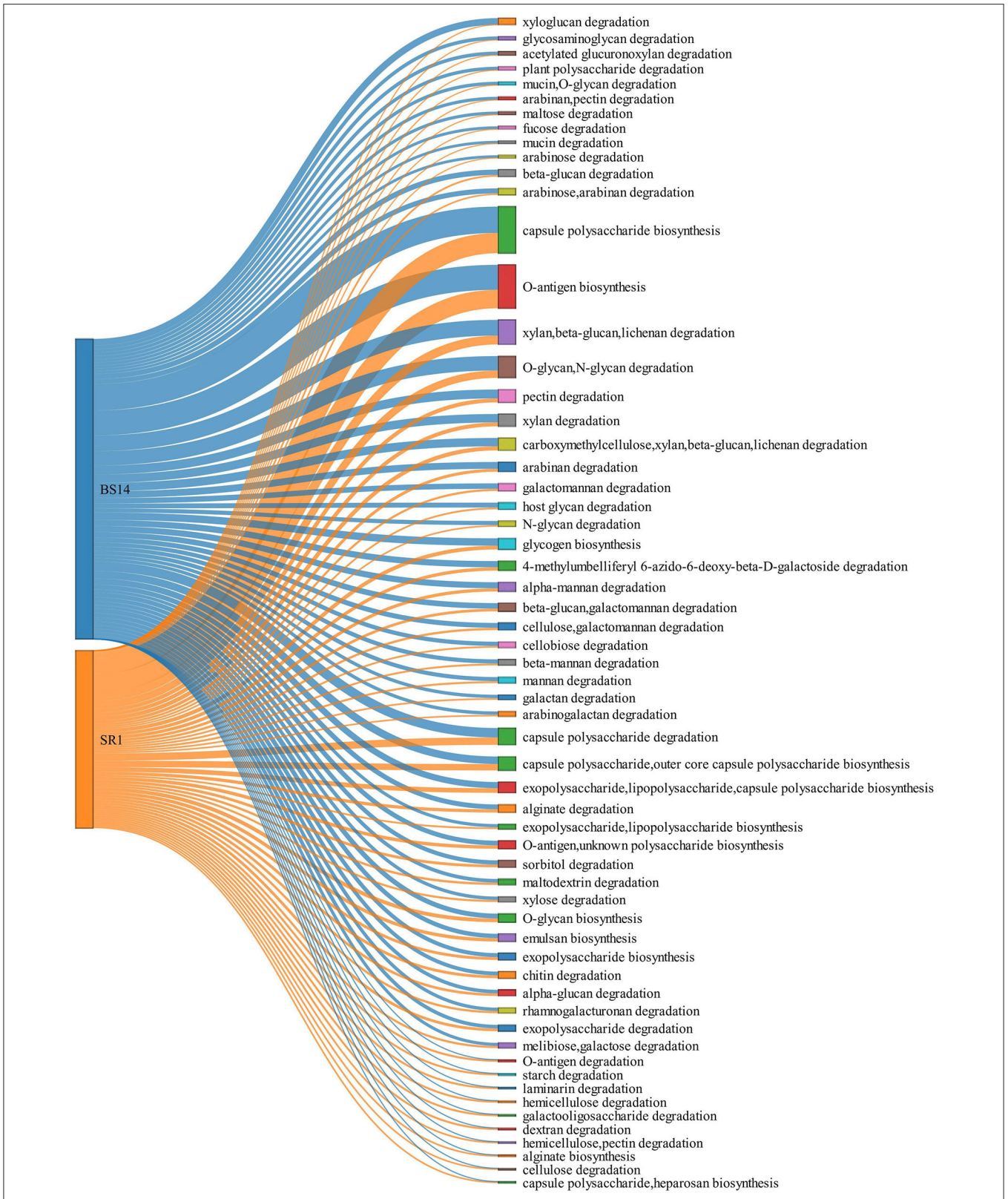


FIGURE 10 | Sankey plot elaborating the total PULDB functions of BS14 and SR1.

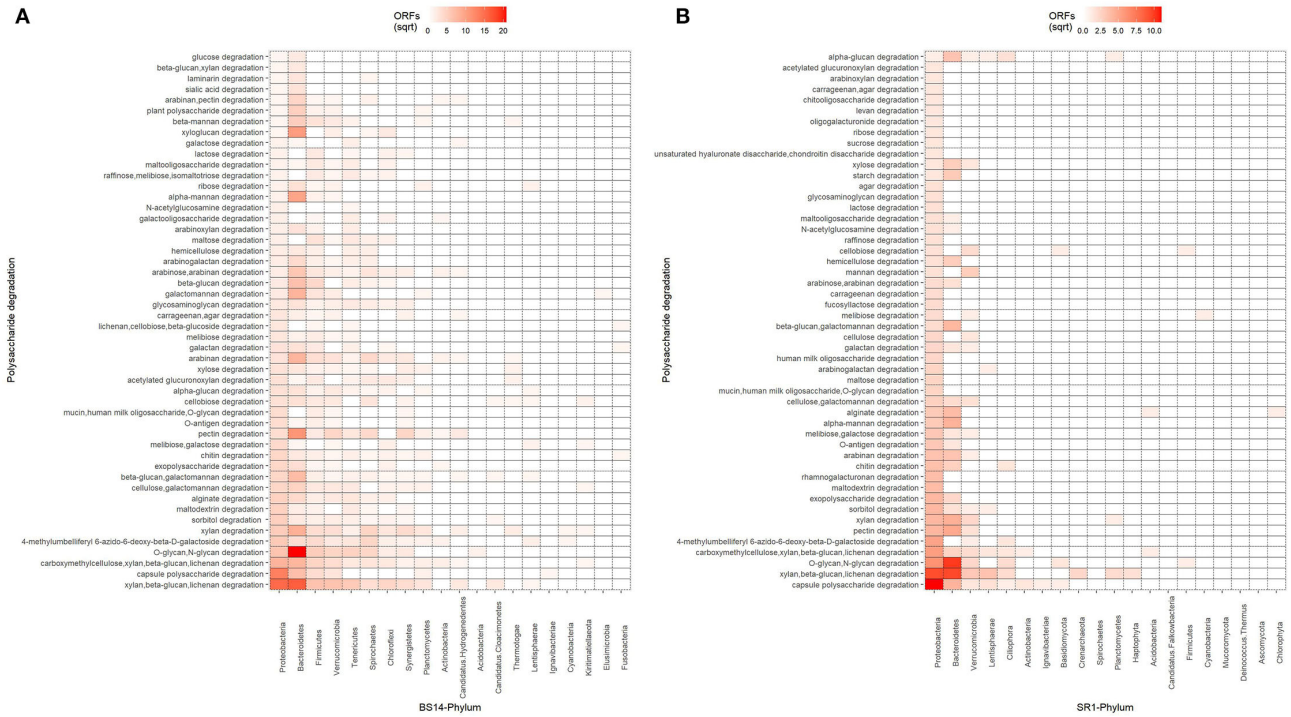


FIGURE 11 | Heatmap showing the distribution of polysaccharide metabolism of **(A)** BS14 and **(B)** SR1 at the phylum level.

the MAGs obtained from BS14 had a high abundance of the Entner-Doudoroff pathway than SR1. All MAGs were compared for the presence of enzymes participating in rTCA pathway and observed their predominance in MAGs obtained from BS14 (**Figure 12A**). Among the bins of BS14, BS14_bin12 belonging to Desulfobacterota had a high abundance of rTCA enzymes.

The nitrogen and sulfur metabolism presence was also observed in the MAGs of SR1 and BS14, which indicated the diverse biogeochemical cycles prevailing in the niche. Nitrogen and sulfur metabolism plays a crucial role in chemolithotropism by acting as electron donors and mostly prevailing in harsh conditions (Osburn et al., 2014). The enzymes responsible for nitrogen metabolism were equally distributed among the MAGs of BS14 and SR1 (**Figure 12B**), whereas for sulfur metabolism, the BS14_bin12 of phylum Desulfobacterota has a high abundance of thiosulfate reductase (EC: 1.8.5.5). The BS14_bin17 and SR1_bin13 belonging to the phylum Proteobacteria have a high abundance of sulfur-oxidizing enzymes (**Figure 12C**). The MAGs of BS14 have a high abundance of enzymes responsible for methanogenesis as compared with SR1 (**Figure 12D**). The MAGs were also compared for the presence of the enzymes involved in hydrocarbon biodegradation. The MAGs of BS14 were found more enriched with hydrocarbon degradation than SR1 (**Supplementary Figures S8A,B**). The protocatechuate and benzene modules were found predominant in both metagenomes. For SR1, metagenome SR1_bin2 and SR1_bin8 have the maximum number of genes responsible for benzene degradation, whereas for BS14, BS14_bin7, BS14_bin17, and BS14_bin22 have the highest number of genes responsible for hydrocarbon degradation.

Comparing Electron Transport Chain (ETC) complexes, an almost similar distribution pattern in the MAGs of both SR1 and BS14 was observed, but the MAGs obtained from BS14 also have a high abundance of cytochrome c oxidase, cbb3 type among them (**Supplementary Figures S7A,B**). At the same time, the presence of arsenate reductase was dominant in the MAGs of BS14 compared with SR1. The photosynthesis module was not observed in any MAG, whereas the module of methanogenesis and methanotrophy, short-chain fatty acid (SCFA), and alcohol conversions were observed in both BS14 and SR1.

The annotation of MAGs of both SR1 and BS14 indicated the microbiome's diverse functional metabolic potential that balances the community's geochemical and energy flux.

CONCLUSION

The CETP, Baddi, receives wastewater from hundreds of small and large-scale industries, thus providing an artificial enriched medium to encourage the metabolism of its native microbial community. This study involved high-throughput metagenomic sequencing of activated sludge collected from CETP and river water from their drainage point to decipher the structure

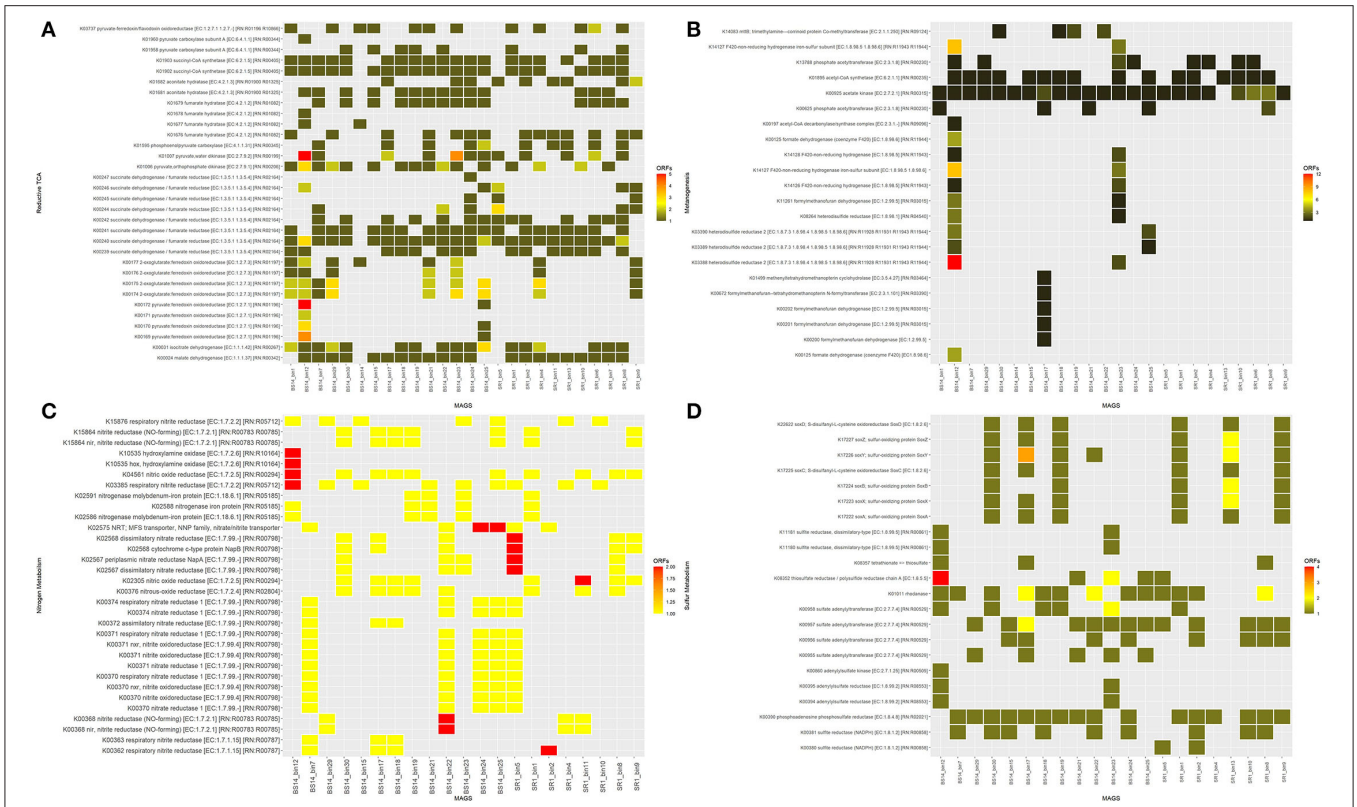


FIGURE 12 | Functional annotation of metagenome-assembled genomes (MAGs) (BS14 and SR1) for the presence of crucial enzymes involved in **(A)** reductive TCA; **(B)** nitrogen metabolism; **(C)** sulfur metabolism; and **(D)** methanogenesis.

of microbial communities and their metabolic potential. The AMR profile and their origin were also determined to study the shared resistome of activated sludge and river water. Our study concluded the prevalence of the bacterial kingdom in activated sludge with the dominance of proteobacteria. On the contrary, the eukarya were dominant in the SR1, with Ciliophora the most abundant phyla. The shared resistome profile was observed in comparing the AMR profile of BS14 and SR1. AMR against cephalosporin, penam, fluoroquinolones, macrolides, and tetracycline was highest in BS14 and SR1. *Azoarcus* and *Desulfococcus* possessed the highest antimicrobial resistance genes against cephalosporins and penams in BS14, whereas, in the case of SR1, *Pseudomonas*, *Acinetobacter*, and *Shewanella* contained the highest AMR genes for beta-lactam antibiotics. To further explore the spread of disease-causing bacteria and the presence of probable toxins in the SR1 which may originate from BS14, the VFDB analysis revealed that the majority of virulent factors originated from *Desulfuromonas* (6.81%), *Thauera* (6.38%), *Azoarcus* (5.35%), *Acinetobacter* (4.48%), *Geoalkalibacter* (3.25%), *Desulfobulbus* (1.86%) in BS14, and *Pseudomonas* (30.91%) contributed to more virulent factors followed by *Shewanella* (9.71%), *Acinetobacter* (9.28%), *Polynucleobacter* (7.36), *Flavobacteriia* (3.88%), and *Fluviibacter* (3.37%) in SR1. The toxins, i.e., Hemolysin A&B and cyclolysin secretion protein (Cya), were found in both the BS14 and SR1 samples. However, it can be assumed that cyclolysin may also be present in both samples. Colibactin (clbI; a genotoxic metabolite) was found in SR1 but its resistance mechanism was found in BS14. The CAZy annotation showed that the carbohydrate-active enzymes belonging to class GH and GT were most prevalent in both BS14 and SR1. The top carbohydrate-active enzymes obtained from the KEGG database showed the abundance of lipid metabolism, citric acid intermediary enzymes, glycogen synthesis, and enzymes of other central metabolic pathways. On analysis of BS14 and SR1 with PULDB, it was observed that the major community belonged to Proteobacteria and Bacteroidetes for the metabolism of capsular polysaccharides, xylan, lichenan, beta-glucan, etc., mainly in identifying the capability to utilize the 2nd generation feedstock. This study also reported various unique KEGG orthology terms related to different functional categories, such as the biodegradation of xenobiotics and aromatics. The COG and PFAM annotation of the BS14 and SR1 samples depicted the abundance of lipid metabolism, signal transduction mechanisms, membrane integrity, carbohydrate metabolism, and other essential cell regulatory processes. During the functional annotation of MAGs, the pathways related to rTCA and chemolithotropism were observed, which proved the diversity in energy flux and biogeochemical cycles operated in the microbial community. More research is clearly needed to better

understand the fate of these genetic components throughout WWTPs to avoid their release from WWTP effluents and reuse of treated water.

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

GV, HS, and SP designed the experiments, analyzed the data, and wrote the manuscript. AP supervised the project and reviewed the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by the Council of Scientific and Industrial Research (CSIR), India BSC-0402 (CSIR), OLP-804 (CSIR), and the Department of Biotechnology, Government of India (BT/PR7368/INF/22/177/2012).

ACKNOWLEDGMENTS

GV acknowledges the University Grant Commissions (UGC) and HS acknowledges CSIR-IMTECH for fellowships.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.933373/full#supplementary-material

REFERENCES

Alcock, B. P., Raphenya, A. R., Lau, T. T. Y., Tsang, K. K., Bouchard, M., Edalatmand, A., et al. (2020). CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Res.* 48, D517–D525. doi: 10.1093/nar/gkz935

Ali, M., Almohana, A. I., Alali, A. F., Kamal, M. A., Khursheed, A., Khursheed, A., et al. (2021). Common effluent treatment plants monitoring and process augmentation options to conform non-potable reuse. *Front. Environ. Sci.* 9, 1–16. doi: 10.3389/fenvs.2021.741343

Alneberg, J., Bjarnason, B. S., De Bruijn, I., Schirmer, M., Quick, J., Ijaz, U. Z., et al. (2014). Binning metagenomic contigs by coverage and composition. *Nat. Methods* 11, 1144–1146. doi: 10.1038/nmeth.3103

Andrews, S. (2010). *FastQC: A Quality Control Tool for High Throughput Sequence Data*. Available online at: https://www.Bioinformatics.Babraham.Ac.Uk/Projects/Fastqc/

Balskus, E. P. (2015). Colibactin: understanding an elusive gut bacterial genotoxin. *Nat. Prod. Rep.* 32, 1534–1540. doi: 10.1039/C5NP00091B

Ben Khedher, M., Ghedira, K., Rolain, J-. M., Ruimy, R., and Croce, O. (2022). Application and challenge of 3rd generation sequencing for clinical bacterial studies. *Int. J. Mol. Sci.* 23, 1395. doi: 10.3390/ijms23031395

Bersanelli, M., Mosca, E., Remondini, D., Giampieri, E., Sala, C., Castellani, G., et al. (2016). Methods for the integration of multi-omics data: mathematical aspects. *BMC Bioinf.* 17, 167–177. doi: 10.1186/s12859-015-0857-9

Bock, C., Farlik, M., and Sheffield, N. C. (2016). Multi-omics of single cells: strategies and applications. *Trends Biotechnol.* 34, 605–608. doi: 10.1016/j.tibtech.2016.04.004

Bodor, A., Bounedjoum, N., Vincze, G. E., Erdeiné Kis, Á., Laczi, K., Bende, G., et al. (2020). Challenges of unculturable bacteria: environmental perspectives. *Rev. Environ. Sci. Biotechnol.* 19, 1–22. doi: 10.1007/s11157-020-09522-4

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Buchfink, B., Xie, C., and Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* 12, 59–60. doi: 10.1038/nmeth.3176

Bushnell, B. (2014). BBMap: A Fast, Accurate, Splice-Aware Aligner.

Chaumeil, P. A., Mussig, A. J., Hugenholtz, P., and Parks, D. H. (2020). GTDB-Tk: a toolkit to classify genomes with the genome taxonomy database. *Bioinformatics* 36, 1925–1927. doi: 10.1093/bioinformatics/btz848

Chu, B. T. T., Petrovich, M. L., Chaudhary, A., Wright, D., Murphy, B., Wells, G., et al. (2018). Metagenomics reveals the impact of wastewater treatment plants on the dispersal of microorganisms and genes in aquatic sediments. *Appl. Environ. Microbiol.* 84. doi: 10.1128/AEM.02168-17

Cluster, B. I. (2018). *Baddi Industrial Cluster*.

Costessi, A., Van Den Bogert, B., May, A., Ver Loren Van Themaat, E., Roubos, J. A., Kolkman, M. A. B., et al. (2018). Novel sequencing technologies to support industrial biotechnology. *FEMS Microbiol. Lett.* 365, 1–13. doi: 10.1093/femsle/fny103

Drula, E., Garron, M. L., Dogan, S., Lombard, V., Henrissat, B., Terrapon, N., et al. (2022). The carbohydrate-active enzyme database: functions and literature. *Nucleic Acids Res.* 50, D571–D577. doi: 10.1093/nar/gkab1045

Griggs, D., Stafford-Smith, M., Gaffney, O., Rockström, J., Öhman, M. C., Shyamsundar, P., et al. (2013). Policy: Sustainable development goals for people and planet. *Nature* 495, 305–307. doi: 10.1038/495305a

Handelsman, J. (2004). Metagenomics: application of genomics to uncultured microorganisms. *Microbiol. Mol. Biol. Rev.* 68, 669–85. doi: 10.1128/MMBR.68.4.669-685.2004

Hendriksen, R. S., Munk, P., Njage, P., van Bunnik, B., McNally, L., Lukjancenko, O., et al. (2019). Global monitoring of antimicrobial resistance based on metagenomics analyses of urban sewage. *Nat. Commun.* 10. doi: 10.1038/s41467-019-08853-3

Hubeny, J., Harnisz, M., Korzeniewska, E., Buta, M., Zieliński, W., Rolbiecki, D., et al. (2021). Industrialization as a source of heavy metals and antibiotics which can enhance the antibiotic resistance in wastewater, sewage sludge and river water. *PLoS ONE* 16, 1–24. doi: 10.1371/journal.pone.0252691

Hyatt, D., and Chen, G-. L., LoCascio, P. F., Land, M. L., Larimer, F. W., and Hauser, L. J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinform.* 11, 119. doi: 10.1186/1471-2105-11-119

Kanehisa, M., Sato, Y., and Morishima, K. (2016). BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J. Mol. Biol.* 428, 726–731. doi: 10.1016/j.jmb.2015.11.006

Kang, D. D., Li, F., Kirton, E., Thomas, A., Egan, R., An, H., et al. (2019). MetaBAT 2: An adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ.* 2019, 1–13. doi: 10.7717/peerj.7359

Kraemer, S. A., Ramachandran, A., and Perron, G. G. (2019). Antibiotic pollution in the environment: From microbial ecology to public policy. *Microorganisms* 7, 1–24. doi: 10.3390/microorganisms7060180

Krassowski, M., Das, V., Sahu, S. K., and Misra, B. B. (2020). State of the Field in Multi-Omics Research: From Computational Needs to Data Mining and Sharing. *Front. Genet.* 11, 1–17. doi: 10.3389/fgene.2020.610798

Li, H. (2017). *GitHub Repos*. Available online at: https://github.com/lh3/bioawk

Madsen, E. L. (2011). Microorganisms and their roles in fundamental biogeochemical cycles. *Curr. Opin. Biotechnol.* 22, 456–464. doi: 10.1016/j.copbio.2011.01.008

Marvasi, M., Pangallo, D., Cavalieri, D., and Poyatos-Jiménez, F. (2021). Editorial: multi-omics revolution in microbial cultural heritage conservation. *Front. Microbiol.* 12, 1–2. doi: 10.3389/fmicb.2021.720509

Matviichuk, O., Mondamert, L., Geffroy, C., Gaschet, M., Dagot, C., Labanowski, J., et al. (2022). River biofilms microbiome and resistome responses to wastewater treatment plant effluents containing antibiotics. *Front. Microbiol.* 13, 1–14. doi: 10.3389/fmicb.2022.795206

McMichael, A. J. (2000). The urban environment and health in a world of increasing globalization: Issues for developing countries. *Bull. World Health Organ.* 78, 1117–1126.

Menzel, P., Ng, K. L., and Krogh, A. (2016). Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nat. Commun.* 7, 11257. doi: 10.1038/ncomms11257

Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E. L. L., et al. (2020). Pfam: The protein families database in 2021. *Nucleic Acids Res.* 49, D412–D419. doi: 10.1093/nar/gkaa913

Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E. L. L., et al. (2021). Pfam: The protein families database in 2021. *Nucleic Acids Res.* 49, D412–D419. doi: 10.1093/NAR/gkaa913

Mukherjee, M., Laird, E., Gentry, T. J., Brooks, J. P., and Karthikeyan, R. (2021). Increased antimicrobial and multidrug resistance downstream of wastewater treatment plants in an urban watershed. *Front. Microbiol.* 12. doi: 10.3389/fmicb.2021.657353

Narayanan, C. M., and Narayan, V. (2019). Biological wastewater treatment and bioreactor design: a review. *Sustain. Environ. Res.* 1, 1–17. doi: 10.1186/s42834-019-0036-1

Nguyen, A. Q., Vu, H. P., Nguyen, L. N., Wang, Q., Djordjevic, S. P., Donner, E., et al. (2021). Monitoring antibiotic resistance genes in wastewater treatment: current strategies and future challenges. *Sci. Total Environ.* 783, 0–57. doi: 10.1016/j.scitotenv.2021.146964

Nurk, S., Meleshko, D., Korobeynikov, A., and Pevzner, P. A. (2017). MetaSPAdes: A new versatile metagenomic assembler. *Genome Res.* 27, 824–834. doi: 10.1101/gr.213959.116

Osburn, M. R., LaRowe, D. E., Momper, L. M., and Amend, J. P. (2014). Chemolithotrophy in the continental deep subsurface: Sanford underground research facility (SURF), USA. *Front. Microbiol.* 5, 1–14. doi: 10.3389/fmicb.2014.00610

Padalkar, A. V., and Kumar, R. (2018). Common effluent treatment plant (CETP): Reliability analysis and performance evaluation. *Water Sci. Eng.* 11, 205–213. doi: 10.1016/j.wse.2018.10.002

Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., and Tyson, G. W. (2015). CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* 25, 1043–1055. doi: 10.1101/gr.186072.114

Reddy, B., Pandey, J., and Dubey, S. K. (2019). Assessment of environmental gene tags linked with carbohydrate metabolism and chemolithotrophy associated microbial community in River Ganga. *Gene* 704, 31–41. doi: 10.1016/j.gene.2019.04.004

Satterthwaite, D., McGranahan, G., and Tacoli, C. (2010). Urbanization and its implications for food and farming. *Philos. Trans. R. Soc. B Biol. Sci.* 365, 2809–2820. doi: 10.1098/rstb.2010.0136

Sayers, E. W., Bolton, E. E., Brister, J. R., Canese, K., Chan, J., Comeau, D. C., et al. (2022). Database resources of the national center for biotechnology information. *Nucleic Acids Res.* 50, D20–D26. doi: 10.1093/nar/gkab1112

Shaffer, J. P., Nothias, L-. F., Thompson, L. R., Sanders, J. G., Salido, R. A., Couvillion, S. P., et al. (2021). Multi-omics profiling of Earth's biomes reveals that microbial and metabolite composition are shaped by the environment. *bioRxiv*. doi: 10.1101/2021.06.04.446988

Shaffer, M., Borton, M. A., McGivern, B. B., Zayed, A. A., La Rosa, S. L., Solden, L. M., et al. (2020). DRAM for distilling microbial metabolism to automate the curation of microbiome function. *Nucleic Acids Res.* 48, 8883–8900. doi: 10.1093/nar/gkaa621

Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell Syst. Tech. J.* 27, 379–423. doi: 10.1002/j.1538-7305.1948.tb01338.x

Sieber, C. M. K., Probst, A. J., Sharrar, A., Thomas, B. C., Hess, M., Tringe, S. G., et al. (2018). Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. *Nat. Microbiol.* 3, 836–843. doi: 10.1038/s41564-018-0171-1

Simpson, E. H. (1949). Measurement of Diversity. *Nature 163*, 688–688. doi: 10.1038/163688a0

Tamames, J., and Puente-Sánchez, F. (2019). SqueezeMeta, a highly portable, fully automatic metagenomic analysis pipeline. *Front. Microbiol.* 10, 1–10. doi: 10.3389/fmicb.2018.03349

Tatusov, R. L., Galperin, M. Y., Natale, D. A., and Koonin, E. V. (2000). The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res.* 28, 33–36. doi: 10.1093/nar/28.1.33

Terrapon, N., Lombard, V., Drula, É., Lapébie, P., Al-Masaudi, S., Gilbert, H. J., et al. (2018). PULDB: the expanded database of Polysaccharide Utilization Loci. *Nucleic Acids Res.* 46, D677–D683. doi: 10.1093/nar/gkx1022

Tripathi, P., Shine, E. E., Healy, A. R., Kim, C. S., Herzon, S. B., Bruner, S. D., et al. (2017). ClbS Is a Cyclopropane Hydrolase That Confers Colibactin Resistance. *J. Am. Chem. Soc.* 139, 17719. doi: 10.1021/jacs.7b09971

Unamba, C. I. N., Nag, A., and Sharma, R. K. (2015). Next generation sequencing technologies: The doorway to the unexplored genomics of non-model plants. *Front. Plant Sci.* 6, 1074. doi: 10.3389/fpls.2015.01074

Valderrama, J. A., Durante-Rodríguez, G., Blázquez, B., García, J. L., Carmona, M., Díaz, E., et al. (2012). Bacterial degradation of benzoate: cross-regulation between aerobic and anaerobic pathways. *J. Biol. Chem.* 287, 10494–10508. doi: 10.1074/jbc.M111.309005

Vilanova, C., and Porcar, M. (2016). Are multi-omics enough? *Nat. Microbiol.* 1, 1–2. doi: 10.1038/nmicrobiol.2016.101

Waldrop, M. M. (2021). Microbes for better sewage treatment. *Proc. Natl. Acad. Sci. U. S. A.* 118, 1–5. doi: 10.1073/pnas.2112863118

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. New York, NY: Springer. Available online at: https://ggplot2.tidyverse.org

Wu, Y. W., Simmons, B. A., and Singer, S. W. (2016). MaxBin 2.0: An automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* 32, 605–607. doi: 10.1093/bioinformatics/btv638

Zieliński, W., Korzeniewska, E., Harnisz, M., Drzymała, J., Felis, E., Bajkacz, S., et al. (2021). Wastewater treatment plants as a reservoir of integrase and antibiotic resistance genes—An epidemiological threat to workers and environment. *Environ. Int.* 156, 106641. doi: 10.1016/j.envint.2021.106641

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