frontiers | Research Topics

Omics techniques in deciphering environmental, industrial and therapeutic applications of microbes

Edited by
Helianthous Verma, Utkarsh Sood, Janmeajy Pandey and Sanjay Kumar Singh Patel

Published in
Frontiers in Microbiology

ISSN 1664-8714
ISBN 978-2-8325-4108-1
DOI 10.3389/978-2-8325-4108-1

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

Omics techniques in deciphering environmental, industrial and therapeutic applications of microbes

Topic editors

Helianthous Verma – University of Delhi, India

Utkarsh Sood – University of Delhi, India

Janmeajy Pandey – Central University of Rajasthan, India

Sanjay Kumar Singh Patel – Konkuk University, Republic of Korea

Citation

Verma, H., Sood, U., Pandey, J., Patel, S. K. S., eds. (2023). *Omics techniques in deciphering environmental, industrial and therapeutic applications of microbes*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-4108-1

Table of contents

frontiers | Frontiers in Microbiology

TYPE Editorial
PUBLISHED 28 November 2023
DOI 10.3389/fmicb.2023.1327368

OPEN ACCESS

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

*CORRESPONDENCE
Helianthous Verma
✉ helianthousverma@ramjas.du.ac.in

RECEIVED 24 October 2023
ACCEPTED 14 November 2023
PUBLISHED 28 November 2023

CITATION
Sood U, Pandey J, Patel SKS and Verma H (2023) Editorial: Omics techniques in deciphering environmental, industrial and therapeutic applications of microbes. *Front. Microbiol.* 14:1327368. doi: 10.3389/fmicb.2023.1327368

Editorial: Omics techniques in deciphering environmental, industrial and therapeutic applications of microbes

Utkarsh Sood[1], Janmejay Pandey[2], Sanjay Kumar Singh Patel[3] and Helianthous Verma[4]*

[1]Department of Zoology, Kirori Mal College, University of Delhi, New Delhi, India, [2]Department of Biotechnology, Central University of Rajasthan, Ajmer, India, [3]Department of Chemical Engineering, Konkuk University, Seoul, Republic of Korea, [4]Departmet of Zoology, Ramjas College, University of Delhi, New Delhi, India

KEYWORDS

Omics, genomics, metagenomics, metaproteomics, biotechnologic application, therapeutic, environmental microbe

Editorial on the Research Topic
Omics techniques in deciphering environmental, industrial and therapeutic applications of microbes

Microbes are essential for life to exist on Earth. These tiny organisms play an important role in the sustainability of the environment and human health. They produce a large number of biotechnologically important products and also certain metabolites that can affect humans in both positive and negative ways. Omics techniques such as genomics, transcriptomics, proteomics and metabolomics, have revolutionized our understanding of microbes and their role in environment, industrial and therapeutic applications. These techniques let investigators analyze the microbe's genetic, protein, and metabolic profiles and provide valuable insights into their responses to environmental factors, biotechnological potential and their functional interactions.

The Research Topic aims to highlight the application of omic techniques in deciphering the environmental, industrial and therapeutic application of microbes. Nine articles have been published on this topic, which provide insights into the diverse roles that microbes play in different settings.

Sharma et al. reported a novel hyperactive (142 μg/mL/min) and thermostable keratinase from *Bacillus pacificus* RSA27 for the valorization of chicken feather waste to peptides rich in essential amino acids. *In silico* analysis on the protein sequence of the keratinase validated the high-affinity calcium-binding site (Asp128, Leu162, Asn164, Ile166, and Val168) and a catalytic triad of Asp119, His151, and Ser308 of keratinase belongs to serine protease. The biotechnologically important product was scaled to 5L fermeter, and it was to achieve ~94% hydrolysis with a total of 154 μmol/mL amino acids production without any cytotoxic influence proving its broad biotechnological applications potential.

Bilinski et al. used next-generation sequencing microbes culturing methods (aerobic and anaerobic conditions) and flow cytometry (viability measurements), for the assessments of fecal microbiota transplantation as fresh vs. frozen stool without any cryoprotectants. They

reported that freezing stool samples enormously impacted cultivable bacterial community structure, evidenced by a significant drop in *Actinobacteria* and *Bacilli* that was also confirmed using the amplicon sequencing.

Genomics plays a pivotal role in uncovering the biotechnological potential of microbes by revealing their genetic makeup and functional capabilities, aiding in the identification of valuable genes for applications in medicine, agriculture, industry, and environmental remediation. It allows for the targeted manipulation and optimization of microbial strains for enhanced production of biofuels, biopharmaceuticals, enzymes, and other bioproducts. Yadav et al. isolated and characterized a human intestinal bacterium *Cytobacillus oceanisediminis* NB2 and subjected it to whole genome sequencing. The genomic analysis of this strain identified the presence of gene clusters for diverse bio-catalytic activity, stress response, and antimicrobial activity, as well as indicated the absence of pathogenic gene islands. The strain being native to the human gut also displayed probiotic potential as its genetic repertoire coded for functional features like anti-amylase, anti-lipase, glutenase, prolyl endopeptidase, lactase, bile salt hydrolase, cholesterol oxidase and anti-pathogenic activity.

Gattoni et al. isolated a novel low-abundant rare biosphere marine bacterium *Leeuwenhoekiella parthenopeia* sp. nov. Mr9^T and sequenced its complete genome. The genome sequence validated the strain Mr9^T was a novel species of the phylum Bacteroidota. The genome was also mapped to the Mediterranean Sea metagenomes revealing an abundance of ∼0.003% of the bacterial population in the metagenomes. The genomic information of the strain revealed the presence of five biosynthetic gene clusters with biotechnological applications. The total lipid content (lipidome) was found to inhibit tumor cell viability.

Ngamcharungchit et al. isolated a new marine actinomycete, strain MCN248 from coastal sediment in Thailand. It shared genetic similarities with *Nonomuraea harbinensis* and *Nonomuraea ferruginea* but exhibited low DNA-DNA hybridization relatedness. Genomic analysis revealed the potential for biosynthesis of various secondary metabolites, including anticancer compounds. The crude extract from this strain inhibited colorectal cancer cells.

Metagenomics, which focuses on analyzing the ecological sample's genetic pool, has been significantly used to investigate microbial communities' diversity and functional capacity. It has uncovered the microbial communities' vast genomic variety and their roles in the ecosystem, such as nutrient cycling and pollutants degradation. Nimonkar et al. reported that oligotrophs and copiotrophic microorganisms at a hypersaline and hyper-alkaline Lonar Lake (Maharashtra, India). Further, they showed that compositional changes in the culture media led to a significant alteration in the selection of organisms within the same sample. Isolated oligotrophs with valuable enzyme production potential can be a valuable resource for cost-effective industrial enzyme production. Additionally, these oligotrophs can serve as a tool for OMICS studies to gain insights into how microbes adapt and survive in environments with very limited nutritional resources, and this knowledge can be applied to similar situations.

The study by Wing-Tung Lee et al. investigated microbial communities in the built environment, aiming to improve the assessment of pathogenic risks beyond traditional bacterial culture methods. The research utilizes 16S rRNA gene analysis with nanopore sequencing, comparing various taxonomic classifiers (ARGpore2, Emu, Kraken2/Bracken, and NanoCLUST) in terms of their performance. The results indicate that NanoCLUST is a preferred option for microbial profiling, displaying high concordance with dominant species and similar profiles to MegaBLAST. For identifying culturable species, Emu demonstrates the highest accuracy (81.2%) and F1 score (29%). This research contributes valuable insights for future microbial community studies, particularly those employing nanopore 16S rRNA sequencing in complex environments.

Zhang et al. reported the genomic diversity and characteristics of Monascus species, which are significant in the food industry but can also produce citrinin mycotoxin. Through genomic analyses and comparative genomics, they identified two major clades within the *Monascus purpureus* clade and the *M. pilosus-M. ruber* clade, with *M. pilosus* and *M. ruber* being closely related. The research highlights differences in gene content related to environmental adaptation and reveals gene clusters for pigment synthesis and citrinin production. Notably, the citrinin gene cluster is found only in *M. purpureus*, while the monacolin K gene cluster is present in *M. pilosus* and *M. ruber*, with variations in sequence conservation. This study enhances our understanding of Monascus species in terms of classification, metabolism, and safety, providing valuable insights for further research in the field.

Metaproteomics, an essential field in microbiome research, characterizes microbial community proteomes. The study by Ascandari et al. employed bibliometric analysis to assess the global landscape of metaproteomic research and Africa's contribution. The number of metaproteomic publications has risen significantly, with notable contributions from the USA, Germany, China, and Canada. Frontiers in Microbiology is a key publishing platform. While Africa's contribution is limited (2.2% of total publications), South Africa stands out. Importantly, more than half of Africa's publications have a high Field-Weighted Citation Impact (FWCI), indicating their significance. To enhance Africa's involvement in this field, investments, collaborations, and mentorship programs are vital for future progress.

Articles published in the Research Topic, hence, are of wide "Omics" disciplines which holds valuable information on microbes and their ability to produce biotechnological, therapeutical and industrially important compounds. Also, microbial adaptations to the environment with minimal nutrient requirement for cost effective enzyme production and better methods such as NanoCLUST for assessment beyond the traditional method has been reported in the Research Topic. Thus, these studies enhanced the knowledge of bacterial compounds and the OMICs methods which can further be exploted to obtain useful information in exploring microbial world.

Author contributions

HV: Writing – original draft, Writing – review & editing. US: Writing – original draft. JP: Writing – review & editing. SP: Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

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: 25 April 2022
doi: 10.3389/fmicb.2022.882902

Chicken Feather Waste Valorization Into Nutritive Protein Hydrolysate: Role of Novel Thermostable Keratinase From *Bacillus pacificus* RSA27

Chhavi Sharma[1], *Svetlana Timorshina*[2], *Alexander Osmolovskiy*[2], *Jyoti Misri*[3] *and Rajni Singh*[1]*

[1] *Amity Institute of Microbial Technology, Amity University, Noida, India,* [2] *Department of Microbiology, Faculty of Biology, Lomonosov Moscow State University, Moscow, Russia,* [3] *Division of Animal Science, Indian Council of Agricultural Research, New Delhi, India*

OPEN ACCESS

Edited by:
Sanjay Kumar Singh Patel, Konkuk University, South Korea

Reviewed by:
Sunil Kumar Suman, Indian Institute of Petroleum (CSIR), India
Mohammad Saghir Khan, Aligarh Muslim University, India

***Correspondence:**
Rajni Singh
rsingh3@amity.edu

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

Received: *24 February 2022*
Accepted: *14 March 2022*
Published: *25 April 2022*

Citation:
Sharma C, Timorshina S, Osmolovskiy A, Misri J and Singh R (2022) Chicken Feather Waste Valorization Into Nutritive Protein Hydrolysate: Role of Novel Thermostable Keratinase From Bacillus pacificus RSA27. Front. Microbiol. 13:882902. doi: 10.3389/fmicb.2022.882902

Microbial keratinases exhibit a momentous role in converting keratin biowastes into exceedingly valuable protein supplements. This study reports a novel, highly stable keratinase from *Bacillus pacificus* RSA27 for the production of pure peptides rich in essential amino acids from chicken feathers. Purified keratinase showed a specific activity of 38.73 U/mg, 2.58-fold purification, and molecular weight of 36 kDa. Kinetic studies using a chicken feather as substrate report K_m and V_{max} values of 5.69 mg/ml and 142.40 μg/ml/min, respectively, suggesting significant enzyme-substrate affinity/biocatalysis. Identification and *in silico* structural-functional analysis of keratinase discovered the presence of distinct amino acid residues and their positions. Besides, keratinase possesses a high-affinity calcium-binding site (Asp^{128}, Leu^{162}, Asn^{164}, Ile^{166}, and Val^{168}) and a catalytic triad of Asp^{119}, His^{151}, and Ser^{308}, known attributes of serine protease (subtilisin family). Furthermore, a scale-up to 5 L fermenter revealed complete feather hydrolysis (94.5%) within 24 h with high activity (789 U/ml) and total amino acid of 153.97 μmol/ml. Finally, cytotoxicity evaluation of protein hydrolysate resulted in negligible cytotoxic effects (1.02%) on the mammalian hepatoblastoma cell line, signifying its potential biotechnological applications.

Keywords: keratinolytic protease, purification, characterization, structural modeling, scale-up, amino acid production

INTRODUCTION

Feather waste from poultry agro-industries constitutes a major environmental contaminant and accounts for several million tons of highly accumulating recalcitrant waste worldwide (Li, 2019; Mazotto et al., 2022). The disposal of such animal-derived wastes in incinerators or landfills increases both monetary and environmental costs (Wu et al., 2017). Chicken feather shows immense rigidity against physical, chemical, biological agents, and resists proteolysis by proteases like trypsin, pepsin, and papain, which is ultimately a significant obstacle in managing these costs (Qiu et al., 2022). Keratin, a rigid protein with high cysteine and disulfide bonds in β-keratin is the

main component of a chicken feather that contributes to its extreme stability/resistance to degradation (McKittrick et al., 2012; Gupta and Singh, 2014; Nnolim et al., 2020a). Various high energy-consuming, non-eco-friendly thermal and chemical methods are extensively used and explored to decompose and reuse keratin (Li, 2021). But such methods ensue in the loss of some thermosensitive amino acids like tryptophan, lysine, and methionine and add-on to non-nutritive amino acids like lanthionine and lysinoalanine (Lee et al., 2016; Li, 2021).

Keratinases are extracellular-specific proteases capable of degrading keratin (Nigam, 2013; Nnolim et al., 2020a). Hydrolysis of keratinous waste (chicken feather) using microbial keratinase has gained renewed interest as it could be used in the non-polluting processes for environmental cleanup through recycling. Therefore, microbial keratinases have gained significant importance in chicken feather degradation and yield feather-feed rich in amino acids (serine, proline, and cysteine) for value addition as an animal protein supplement (Latha et al., 2016; Suman et al., 2018; Li, 2019; Qiu et al., 2022). Research suggests other promising applications of microbial keratinolytic proteases as efficient fertilizers (Singh et al., 2019), along with industrial applicability in peptide synthesis (Verma et al., 2017), therapeutic use (Gupta et al., 2017), biodegradable inexpensive thermoplastics (Jin et al., 2011), detergent, cosmetic, and leather industry (Brandelli, 2008). Thus, microbial decomposition of keratinous waste is a low-expense pivotal approach with no loss of essential amino acids and energy (Gupta and Ramnani, 2006; Brandelli, 2008).

Diverse groups of microorganisms are reported to produce keratinases (Kalia and Purohit, 2008), such as numerous *Bacillus* sp. (Gupta and Singh, 2013; Gupta et al., 2015; Dong et al., 2017; Hamiche et al., 2019; Nnolim et al., 2020b; Almahasheer et al., 2022; Devi et al., 2022; Mazotto et al., 2022), *Aspergillus oryzae* (Farag and Hassan, 2004), *Paecilomyces marquandii* and *Doratomyces microsporum* (Gradisar et al., 2005), *Streptomyces* sp. (Tatineni et al., 2008), *Stenotrophomonas maltophilia* (Cao et al., 2009), *Myrothecium verrucaria* (Moreira-Gasparin et al., 2009), *Thermoactinomyces* sp. RM4 (Verma et al., 2016), and *Fusarium* sp. (Calin et al., 2017). The limitations of existing keratinolytic enzymes constitute their low efficiency on complex feather substrates and low stability (Vidmar and Vodovnik, 2018) and suggest an intensive hunt for novel safe substitutes. Also, expanding the enzyme production process from a laboratory-scale unit to a commercial is challenging as it shows intricacy in evaluating factors such as media components, substrates, and physiological conditions during cultivation. Till date, few scale-up studies have been reported to accomplish almost complete chicken feather degradation (Peng et al., 2019).

Therefore, our research intends to suggest a novel keratinase by *Bacillus pacificus* RSA27. The enzyme showed distinct activity with high stability within a wide pH and temperature range along with metal ions/inhibitors/surfactants. Our study further reports the successful purification, characterization, and *in silico* structural-functional analysis of keratinase. As per our knowledge, this is one of the few studies which present keratinase as a potential alternative for biotransformation of chicken feathers into hydrolysate enriched with high-protein content. The scale-up of enzyme production was successfully studied in 5 L fermenter with complete chicken feather degradation within 24 h of cultivation. Henceforth, the research reveals a highly potent/stable keratinase ensuing high-protein content and might serve as an effectual alternative for chicken feather waste management.

MATERIALS AND METHODS

Ethical Clearance

Chicken feathers were collected from a small unit of poultry production (Noida, Uttar Pradesh, India) and no live animal or human subjects were sacrificed. We have not used any live material, which could raise any ethical issue, and therefore, the study did not require an ethics review process.

Chemicals and Reagents

All the chemicals and reagents used in the study were purchased from Sigma Aldrich, St. Louis, MO, USA.

Inoculum Source, Culture Conditions, and Production of Keratinase

Keratinase-producing strains were isolated from soil samples (12–15 cm deep) collected from a poultry-feather dumping site at Ghazipur (New Delhi, India). Soil samples collected in sterile screw-capped vials were serially diluted and 100 µl of each dilution was spread on a sterilized nutrient agar medium. The plates were kept at 37°C for 24 h. The colony obtained on plates was screened for keratinase production by cultivating it in a sterile media containing peptone (1%), glucose (1%), KH_2PO_4 (0.9%), and K_2HPO_4 (0.3%) enriched with a chicken feather (0.5%) as substrate and microbial isolates (1%). Feathers were rinsed with tap water followed by deionized water to remove dust particles and unwanted skin, which were further air-dried at normal room temperature before their addition to the culture medium and autoclaving. The keratinolytic potential of strains was evaluated by incubating the medium containing cultures at temperature 37°C and 180 rpm for 20 days. The strains were finally screened on the basis of their feather degrading efficiency with respect to time and the highly potent strain was used for further study.

Degradation Rate and Keratinolytic Activity of the Potent Screened Strain

The rate of degradation and keratinolytic activity of the selected isolate was further examined. Unhydrolyzed chicken feathers were harvested at fixed time intervals of 12, 24, 36, and 48 h of incubation and filtered through Whatman filter paper 1. These were then thoroughly washed with deionized water to remove culture cells and other soluble materials. The feathers were further oven-dried overnight at 60°C and the degradation rate of the screened strain was determined [chicken feather degradation (g/L) = dry weight of feathers before degradation—dry weight of feathers after degradation] using analytical laboratory weighing balance. Keratinolytic activity was determined using a chicken feather as substrate. The cell-free extract (after centrifugation at 10,000 rpm for 10 min) of keratinase production medium was

used for the assay. The reaction mixture comprised keratinase (100 μl) and glycine-NaOH buffer (4.9 ml, 0.5 M, pH 9) with chicken feather (0.5%) as substrate was incubated for 1 h at 60°C. Trichloroacetic acid (4 ml of 5%) was employed to stop the reaction. The reaction mixture was filtered and the filtrate was spectrophotometrically measured at 280 nm. One unit (U/ml) of keratinase production was defined as an increase in 0.01/h absorbance at 280 nm under the assay conditions as described above.

Scanning Electron Microscopy

Feather hydrolysis achieved by selected inoculum was pictured through SEM. The digested feathers were collected at fixed time intervals (0, 24, and 48 h) and washed thoroughly with sterile distilled water. The samples were further dried and coated with a thin gold-palladium layer using a gold sputter coater. SEM analysis was executed using Leo 435VP Microscope at an accelerating voltage of 7–21 kV (All India Institute of Medical Sciences, New Delhi, India).

Identification of Microbial Strain

The potent selected isolate inducing chicken feather degradation was then identified for its morphological, biochemical, and molecular characteristics through Bergey's brochure for identification of bacteriology and 16S rDNA sequencing. Genomic DNA extraction of the strain was performed using a modified phenol-chloroform extraction technique. The ∼1.5 kb, 16S-rDNA fragment was amplified using high-fidelity PCR polymerase in DNA Thermal Cycler (Mastercycler pro, Eppendorf). PCR amplification reaction mixture (100 μl) comprised 1 μl of template DNA, 400 ng of 16S forward primer (degenerated 5′ AGAGTTTGATCMTGGCTCAG 3′), 400 ng of 16S reverse primer (degenerated 5′ TACGGYTACCTTGTTACGACTT 3′), 4 μl of dNTPs (2.5 mM), 10 μl of Taq DNA Polymerase Assay Buffer (10×), 1 μl of Taq DNA Polymerase Enzyme (3 U/μl), and water. PCR amplification, agarose gel electrophoresis, purification of amplified DNA, and sequencing were performed as per the standardized protocol of Sharma et al. (2020). The obtained sequence was then subjected to a nucleotide blast at NCBI (https://blast.ncbi.nlm.nih.gov) and identified to species level. Evolutionary analyses were conducted through multiple sequence alignment by Clustal Omega and the maximum likelihood method. Bioinformatic tool: Molecular evolutionary genetic analysis software (MEGAX64.exe) was used to create the phylogenetic tree.

Purification of Keratinase

The keratinase from *B. pacificus* RSA27 was purified through precipitation and chromatographic techniques to procure its purest form. The fermented medium was centrifuged at 6,000 rpm for 30 min at 4°C to remove cell debris. The protein hydrolysate (supernatant) was then treated with chilled ethanol (95%) and incubated at −20°C overnight. The mixture was centrifuged at 6,000 rpm for 15 min to procure the precipitated proteins and further purified using gel filtration chromatography. The proteins were dissolved in phosphate buffer (50 mM, pH 7.5) and loaded on Sephadex G-75 column (50 × 15 mm), which was pre-equilibrated with Tris-HCL buffer (50 mM, pH 7.5). Protein fractions were eluted at a flow rate of 1.0 ml/min with Tris-HCL buffer (50 mM, pH 7.5) and assayed for their activity and protein content (Bradford assay). Active fractions were pooled and further purified through ion-exchange chromatography using the Q Sepharose column (65 × 10 mm). The column was pre-equilibrated with glycine-NaOH buffer (10 mM, pH 9), and pooled active fractions were subjected to the column for purification. Fractions were eluted with a linear gradient of NaCl (0.6 M) at a flow rate of 1.0 ml/min. The fractions with keratinolytic activity were pooled, freeze-dried, and examined for its activity/protein concentration and molecular weight (SDS-PAGE).

Characterization of Keratinase

Effect of pH and Temperature on Activity and Stability of Keratinase

The optimal pH of keratinase activity was determined using different buffers of pH values from 4 to 12. Buffers *viz.* citrate phosphate for pH 3.0–6.0, Tris-HCL for pH 7.0–8.0, and glycine-NaOH for pH 9.0–12.0 were used. The enzyme was incubated for 1 h with 0.5% chicken feather as substrate and thereafter assayed for its activity. To evaluate the optimal temperature of enzyme activity, keratinase was incubated for 1 h with 0.5% chicken feather as substrate at different temperatures ranging from 20 to 100°C and thereafter assayed for its activity. Furthermore, we have also evaluated the stability of keratinase at different pH values of 3–12 and temperatures ranging from 20 to 80°C. The enzyme was preincubated at the aforementioned pH and temperatures for 2 h without substrate and then assayed for its activity after every 30 min of time interval (30–120 min).

Keratinase Activity in the Presence of Metal Ions, Inhibitors, and Surfactants

The keratinolytic activity of the enzyme in the presence of potential metal ions, inhibitors, and surfactants was evaluated. Metal ions such as Na^+, K^+, Zn^{2+}, Co^{2+}, Ca^{2+}, Fe^{2+}, Mn^{2+}, Cu^{2+}, Mg^{2+}, and Hg^{2+} and inhibitors such as ethylene-diamine tetra acetic acid (EDTA), β-mercaptoethanol, phenylmethylsulfonyl fluoride (PMSF), urea, indole-3-acetic acid (IAA), N-bromosuccinimide, 1,10-phenanthroline, and dithiothreitol (DTT) of 2 and 5 mM concentration were used. Besides, the effects of various (1 %) surfactants *viz.* saponin, sodium cholate, sodium dodecyl sulfate (SDS), Tween 80, and Triton X 100 on enzyme activity were also studied. The keratinase was preincubated for 1 h with the aforementioned chemical reagents and then assayed for its activity at optimal pH of 9 and temperature of 60°C. The sample without any chemical reagent was treated as a control.

Enzyme Kinetics

GraphPad Prism 9.0 Ink was used to evaluate kinetic parameters (K_m and V_{max}) of keratinase using non-linear regression, Michaelis-Menten curve with various concentrations of a chicken feather as substrate (0.5–5.0 mg/ml). The experiments were conducted under optimum temperature and pH conditions.

Ker Gene Sequencing and Keratinase Prediction

The identification of gene (*ker*) responsible for the production of keratinase was identified through sequencing. Forward primer 5′-AAAAGGAGAGGGTAAAGAGT-3′ and reverse primer 5′-AGCAGGTATGGAGGAGCCTG-3′ were used for PCR amplification. Amplification reaction, gel electrophoresis, and sequencing were performed as per the standardized protocol of Gupta et al. (2017). Bioinformatics tool ExPASY (http://web.expasy.org/translate) was used to translate the obtained *ker* gene sequence into the amino acid sequence to further get an insight into the potent keratinase responsible for feather degradation. The deduced sequence was analyzed using PSI-BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins), aligned using CLUSTAL W (http://www.ebi.ac.uk/clustalw) for the identification of conserved residues and sequence logos were generated using WebLogo 3.7.4 program (http://weblogo.threeplusone.com/) for a clear alignment of identified conserved domains.

Structural Modeling and *in vitro* Validation

The three-dimensional (3D) structure of keratinase was predicted using homology modeling tools—Swiss-model (https://swissmodel.expasy.org/) and Phyre2 (http://www.sbg.bio.ic.ac.uk/phyre2). Modeling server Swiss-model predicted one calcium-binding site (Ca1) of keratinase, which was visualized using Ligplot+ version 2.2 (https://www.ebi.ac.uk/thornton-srv/software/LigPlus/). *In vitro* validation evaluating the effect of Ca ions on the thermostability of keratinase was further studied by incubating keratinase for 2 h with 5 mM of $CaCl_2$ at different temperatures (20–80°C). Keratinolytic activity was performed as per our standard protocol. A reaction mixture without $CaCl_2$ served as a control.

Scale-Up in a Laboratory 5-L Batch Fermenter

The *in vitro* feather degradation was scaled up in a 5 L continuous stirred tank bioreactor (Biostat-B-Lite, Laboratory fermenter system, Sartorius, Germany). A total of 3 L of optimized media was sterilized and inoculated with 2% of overnight grown culture and 0.5% chicken feather waste. The fermentation was carried out for 24 h at 37°C at 180 rpm agitation, uncontrolled pH, and sterile air with a constant rate of 2 vvm. Samples were withdrawn at fixed intervals (2 h) and analyzed for keratinase production and change in pH with respect to time. The fermentation parameters such as temperature, pH, and airflow were continuously monitored throughout the process.

Amino Acid Profiling

The amino acid profiling of *B. pacificus* RSA27 keratinase was accomplished using cell-free extract of digested feathers in triplicates at the time interval of 0, 12, and 24 h using high-performance chromatography (Agilent 1100 HPLC, Santa Clara, United States) and was performed as per the standardized protocol of Gupta and Singh (2014).

In vitro Cytotoxicity Evaluation of Protein Hydrolysate

Cell-free supernatant 0.09–50% (v/v) of digested feathers was analyzed for cytotoxicity using mammalian hepatoblastoma cell line HepG2 (8 × 103 cells/well). Doxorubicin hydrochloride (0.10–5.00 μM) and untreated cells served as positive and negative controls, respectively. The cells were then incubated at 37°C for 72 h, and cytotoxicity was assessed using MTT [3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide] assay. MTT (20 μl of 5 mg/ml) was added to each well and incubated for 3 h at 37°C. The supernatant was aspirated and dimethyl sulfoxide (150 μl) was added to each well to dissolve formazan crystals. Microplate reader (BioTek Winooski, VT, USA) was used to measure absorbance at 540 nm, and cytotoxicity (%) was determined.

RESULTS AND DISCUSSION

Chicken Feather Degradation Rate of Keratinase

In total, 44 strains out of the 200 isolates exhibited keratinolytic ability. Out of which feather degradation with 24 strains was observed in 15–20 days, 19 strains in 5–10 days, and 1 strain in 2 days. The highest feather degrading capacity was observed with isolate RSA27, which initiated the fastest feather degradation within 24 h and completely degraded feather within 48 h of incubation at 180 rpm and temperature 37°C (**Figure 1A**). The breakdown of a chicken feather was visualized using SEM after 0, 24, and 48 h of incubation. Incubation for 24 h resulted in degradation of feather barbs, whereas calamus degradation was observed after 48 h of incubation, leading to a complete breakdown of chicken feathers (**Figure 1B**). A similar result was reported by Gupta and Singh while estimating chicken feather hydrolyzing ability of two *Bacillus subtilis* strains E163 and E165 after 24, 48, and 72 h of cultivation (Gupta and Singh, 2014). In another study, keratinase from *B. subtilis* 8 resulted in a complete breakdown of the chicken feather-shaft structure within 48 h when analyzed and identified through SEM (He et al., 2018). A significant decrease in the weight of chicken feathers hydrolyzed by RSA27 was observed with the increase in incubation time (0–48 h). The observed residual weight (g) was 4.17, 3.39, 1.95, and 0.1 after incubation of 12, 24, 36, and 48 h, respectively. Complete feather degradation was observed after 48 h of incubation (**Figure 1C**). The keratinolytic activity of RSA27 strain after 48 h was found to be 125.10 U/ml.

Identification of Keratinolytic Strain

The isolate RSA27 morphologically exhibited non-translucent, white, and circular (2–3 mm in diameter) colonies. It was observed as non-motile, rod-shaped, and gram-positive *Bacillus* strain. Biochemical characterization imparted its positive catalase, gelatinase, and oxidase activities. Results were positive for citrate utilization, Voges-Proskauer test, and arginine dihydrolase while negative for lysine decarboxylase, β-galactosidase, H_2S, indole, and acid production from arabinose, amygdalin, mannitol, sorbitol, sucrose, glucose, rhamnose, inositol, and melibiose. Such phenotypic and biochemical

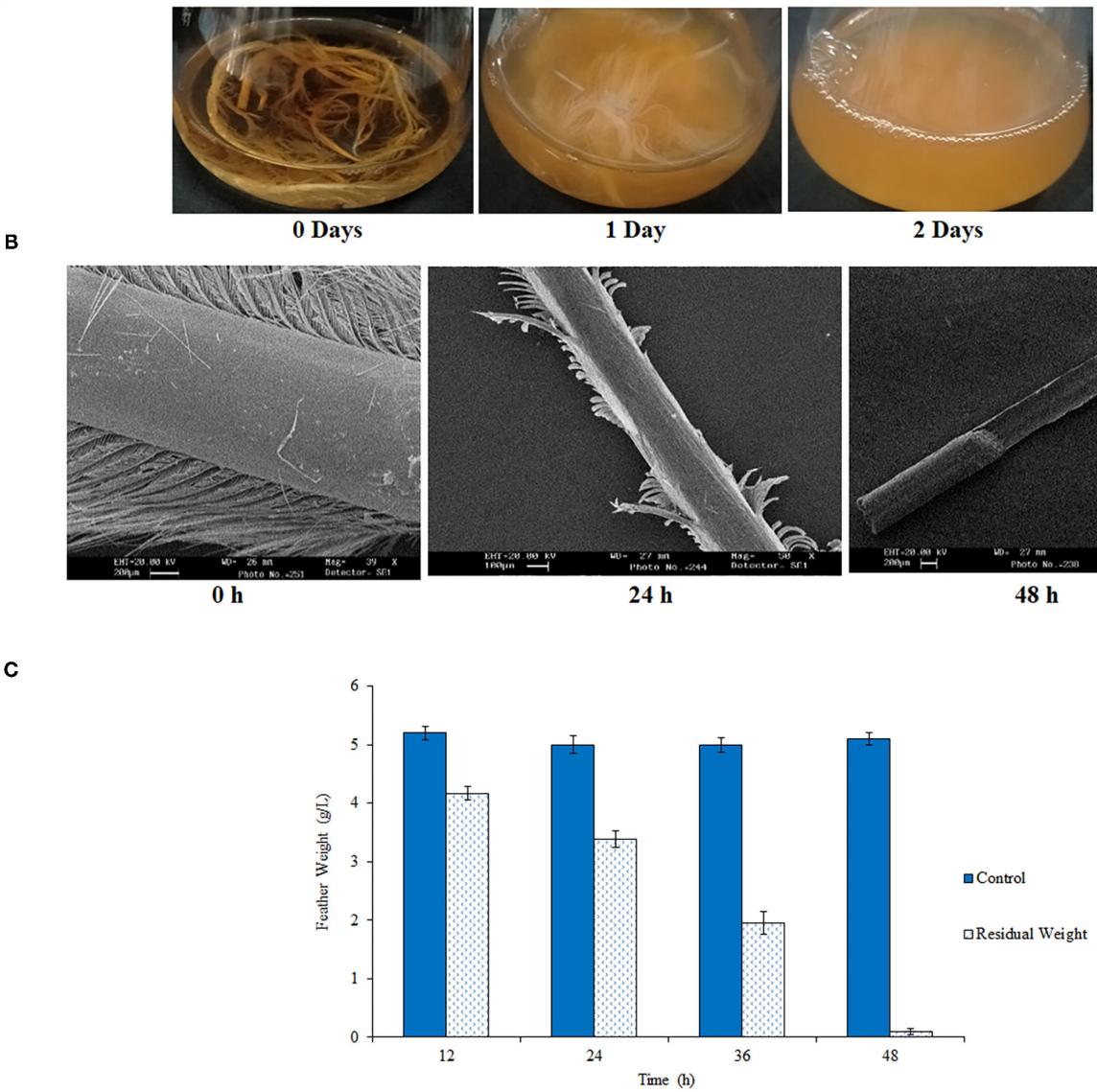


FIGURE 1 | Chicken feather degradation by RSA27. **(A)** RSA27 efficacy with respect to time at 180 rpm and 37°C. 0 days: control, 1 day: RSA27 initiated feather degradation after 24 h, 2 days: complete degradation of feathers after 48 h. **(B)** SEM of feather degradation at 0 h (intact feather: control), 24 h (degradation of feather barbs), and 48 h (calamus degradation leading to complete degradation of the feather). **(C)** A residual dry weight of feathers after 12, 24, 36, and 48 h.

specifications were similar to strain *B. pacificus* EB422T of *Bacillus cereus* group (Liu et al., 2017). So, the strain was further confirmed and identified as *B. pacificus* RSA27 (NCBI Accession number MT180833.1) by 16S rDNA sequencing. Phylogenetic analysis by maximum likelihood methodology conveyed details about the distinct position and evolutionary association of *B. pacificus* RSA27 with other *Bacillus* strains (**Supplementary Figure 1**).

Purification

Purification of keratinase through chilled ethanol precipitation and Sephadex G-75 column resulted in 75.02% yield with 1.25-fold purification and 54.48% yield with 2.05-fold purification, respectively. Furthermore, the enzyme was 2.58-fold purified with a yield of 54.48% relative to crude keratinase by Q Sepharose (**Table 1**). A specific activity of 38.73 U/ml was assessed for the ultimate keratinase. The molecular weight of ∼36 kDa was analyzed by SDS-PAGE (**Figure 2**). Literature reports numerous keratinolytic enzymes of apparently similar molecular weight of 35 kDa from *Bacillus licheniformis* (Suntornsuk et al., 2005), 42 kDa from *Microbacterium* sp. (Thys and Brandelli, 2006), and 44 kDa from *Streptomyces* sp. (Tatineni et al., 2008).

Characterization of Keratinase

Activity and Stability of Keratinase Under Varied pH and Temperature

We have further quantified the keratinase activity and stability to evaluate enzymatic hydrolysis within wide pH and temperature

TABLE 1 | Purification parameters of keratinase from *Bacillus pacificus* RSA27.

Purification steps	Total activity (U/mL)	Total protein (mg/mL)	Specific activity (U/mg)	Yield (%)	Fold purification
Crude enzyme	125.10 ± 5.21	8.34 ± 0.52	15.00 ± 1.09	100	1
Ethanol precipitation	93.85 ± 3.76	4.99 ± 0.31	18.81 ± 1.21	75.02	1.25
Sephadex G-75	68.15 ± 3.11	2.21 ± 0.17	30.84 ± 1.17	54.48	2.05
Q Sepharose	64.29 ± 2.86	1.66 ± 0.09	38.73 ± 1.24	51.39	2.58

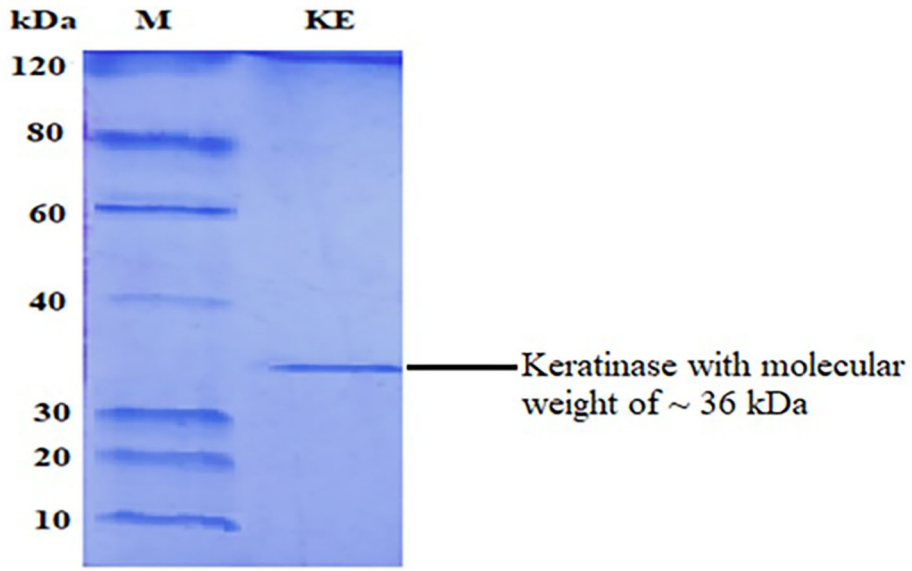


FIGURE 2 | SDS-PAGE of purified keratinase. M, protein marker; KE, keratinase.

range. Keratinase exhibited 100% enzymatic activity with a chicken feather as substrate (0.5%) at pH 9 (**Figure 3A**). The keratinase retained its full stability (100%) at pH 9 even after 2 h of incubation, while a slight reduction in activity was observed from pH 3 to 8 and pH 10 to 12 (**Figure 3C**). Most of the reported proteases are alkaline in nature and thus beneficial for various biotechnological applications (Rahman et al., 2005; Verma et al., 2016; Pandey et al., 2019; Sharma et al., 2020). Our results are in resemblance with some other previous studies as well, which again suggests that alkaline pH alters cysteine residues to lanthionine and aids in keratinolytic activity with high industrial applicability (Rissen and Antranikian, 2001; Selvam et al., 2013). Also, *B. pacificus* RSA27 originated keratinase is stable even under acidic conditions (pH 3–6) with no major loss in activity, suggesting its utilization as an animal feed supplement, which can retain its functionality even after passing through the mammalian digestive system (Wu et al., 2017). Research supports the vision of the use of biowastes as feed, which makes the process economical (Patel et al., 2021).

The temperature profile of keratinase demonstrates optimum activity at 60°C (100%) with wide stability within 20–80°C (**Figure 3B**). The enzyme showed 100% stability at 20°C and a slight reduction in activity of 98.84, 97.97, 96.69, 90.98, 81.92, and 80.59% with an increase in temperature from 30 to 80°C, correspondingly (**Figure 3D**). Our results were in contrast with some earlier reports where after 2 h of incubation, keratinase of *Thermoactinomycetes* sp. showed 100% stability at 70°C (Verma et al., 2016). Another keratinase rMtaker was also found stable at 70°C after 2 h of incubation (Wu et al., 2017). Also, as mentioned, the maximum activity of our keratinase was observed at 60°C (with substrate) and full stability at 20°C (without substrate). Hence, it becomes necessary to mention here that the presence of protein substrates (here chicken feather) tend to increase the thermostability and activity of heated enzymes. A similar phenomenon was reported and discussed in other studies as well (Chang and Mahoney, 1995; Wu et al., 2017).

Keratinase Efficacy in the Presence of Metal Ions, Inhibitors, and Surfactants

Bacillus pacificus RSA27 derived keratinase was treated with different potential metal ions/inhibitors (2 and 5 mM) and surfactants (1%), and the activity was measured using a chicken feather as substrate. Keratinase activity was significantly enhanced in the presence of monovalent *viz.* Na^+ and K^+ and divalent metal ions *viz.* Mn^{2+}, Fe^{2+}, Mg^{2+}, Cu^{2+}, and Ca^{2+}, at both the assayed concentrations. An activity increase of 32.84, 36.85, 42.54, 29.58, 18.59, 45.99, and 47.01% was observed in the presence of 5 mM concentration of Na^+, K^+, Mn^{2+}, Fe^{2+}, Mg^{2+}, Cu^{2+}, and Ca^{2+}, respectively. Such an outcome suggests the requirement of these metal ions for optimal activity of the

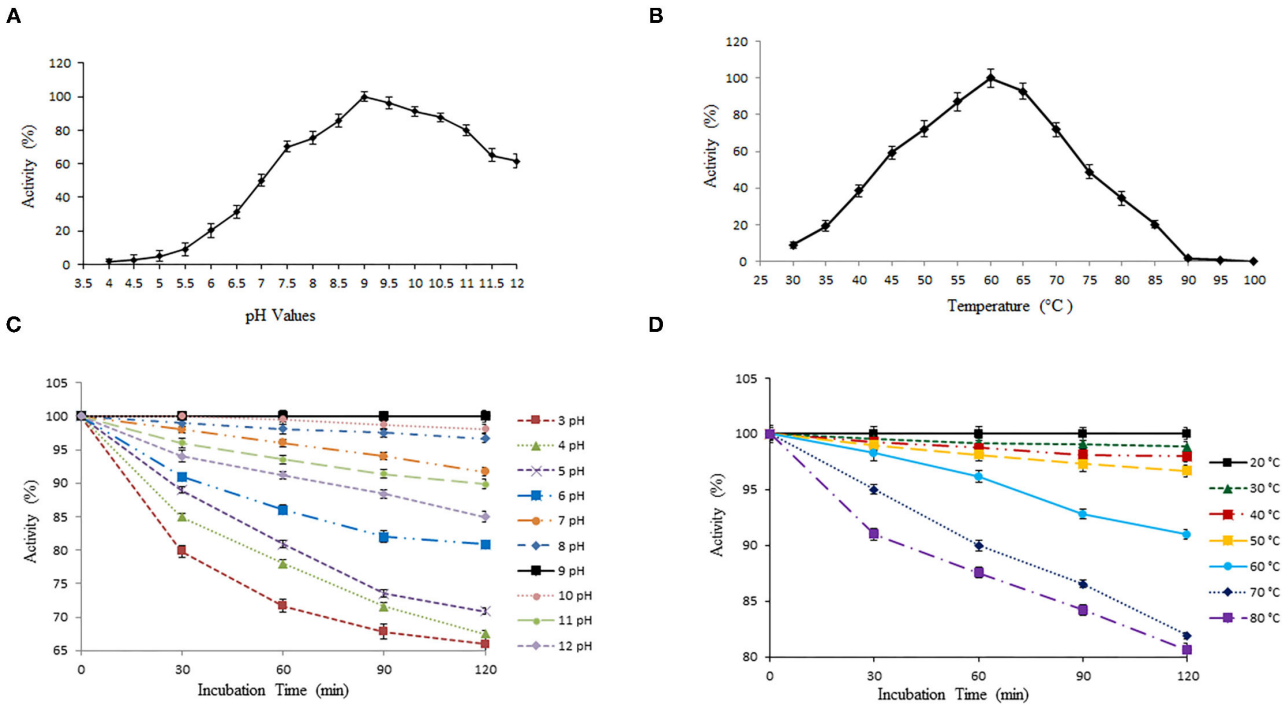


FIGURE 3 | Effect of temperature and pH on activity and stability of keratinase. **(A,B)** The optimal activity of keratinase was observed at pH 9 and 60°C. **(C,D)** Keratinase exhibits stability within wide pH (3–12) and temperature (20–80°C) range.

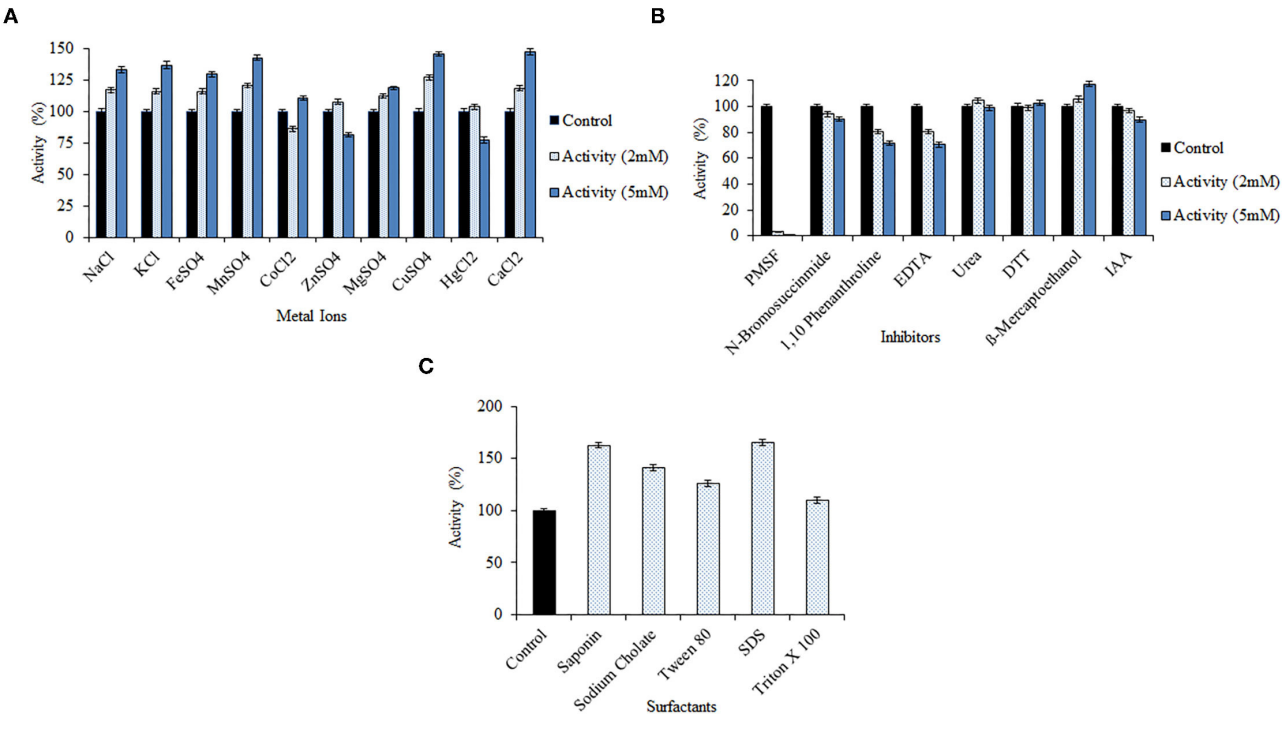


FIGURE 4 | Effect of various metal ions/inhibitors/surfactants on the keratinolytic activity of the enzyme from *Bacillus pacificus* RSA27. **(A,B)** Effect of metal ions and inhibitors (2 and 5 mM). **(C)** Effect of surfactants (1% solution with water).

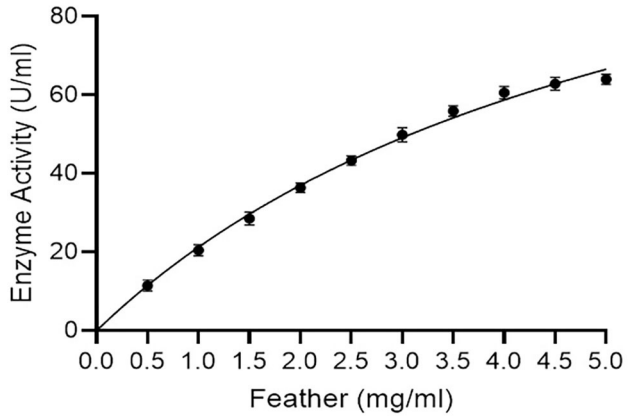


FIGURE 5 | Enzyme kinetics (Michaelis-Menton plot) for keratinase using a chicken feather as substrate ($R^2 = 0.9902$).

enzyme (Patel et al., 2019, 2022; Kondaveeti et al., 2022). Besides, slight inhibition with 86.24% of keratinase activity was observed at low concentration (2 mM) while a slight increase of 10.32% in activity was testified at high concentration (5 mM) of metal ion Co^{2+}. On the contrary, an increase of 7.51 and 3.58% in enzyme activity was observed at low concentration of Zn^{2+} and Hg^{2+}, respectively. However, when used in high concentration, Zn^{2+} and Hg^{2+} resulted in the slight inhibition of keratinase activity (**Figure 4A**). Our results are in contrast to some previous research reports where no significant effect on keratinase activity was observed with monovalent ions (Na^+, Li^+, and K^+; Jaouadi et al., 2013). However, results similar to our study were reported by Verma et al. (2016) wherein some remarkable activity of keratinase from *Thermoactinomycetes* sp. RM4 was observed in the presence of monovalent ions Na^+ and K^+. Another study by Yong et al. (2020) detailed the stimulation of keratinase activity with Mg^{2+} and Ca^{2+} and complete inhibition by Cu^{2+}. Our study also reports high keratinase activity at both concentrations of Ca^{2+}, which is normally considered a standard attribute of serine proteases (Yong et al., 2020). Furthermore, inhibition inactivity of keratinase from *Bacillus subtilis* K-5 was observed with metal ions Na^+, K^+, Zn^{2+}, Mn^{2+}, Hg^{2+}, Cu^{2+}, and Ca^{2+} (Singh et al., 2014).

The keratinase was found to be serine protease in nature, as the presence of both high and low concentrations of PMSF resulted in complete loss of keratinase activity. Keratinase activity of 3.07% and 0.09% was observed with 2 and 5 mM of serine protease inhibitor PMSF, respectively. Also, with the increase in the concentration of inhibitors *viz.* N-bromosuccinimide, 1,10-phenanthroline, EDTA, IAA, and urea, the activity of keratinase gradually decreased. No significant decrease in enzyme activity was observed with DTT and mercaptoethanol at their respective concentrations (**Figure 4B**). BsKER71 keratinase by *Bacillus subtilis* S1-4 was also strongly inhibited by PMSF and reported as serine protease (Nnolim et al., 2020b). Similar observations were stated from keratinases of *Brevibacillus brevis* US575 (Jaouadi et al., 2013) and *Aspergillus parasiticus* (Anitha and Palanivelu, 2013). Also, studies have shown that metalloprotease inhibitors (EDTA and 1,10-phenanthroline) might slightly inhibit the catalytic activity of some serine proteases, which recommends the requirement of transition metals for their hydrolysis (Nnolim et al., 2020b).

However, a significant increase in keratinolytic activity was noticed with surfactants like saponin (1.63-fold), sodium cholate (1.41-fold), Tween 80 (1.26-fold), SDS (1.65-fold), and Triton X 100 (1.09-fold) (**Figure 4C**). Our observations are similar to keratinase from *Bacillus polyfermenticus* B4 wherein an increase in relative activity (%) of the enzyme was observed with Triton X 100 (109 ± 5), Tween 80 (100 ± 1), and SDS (279 ± 2; Dong et al., 2017). Another surfactant stable keratinase was obtained from *Brevibacillus parabrevis* CGMCC 10798, wherein 11 and 30% increment in enzyme's activity was observed with Tween 40 and Triton X 100, respectively (Zhang et al., 2016). Remarkable keratinolytic activity was observed with saponin (1.91-fold) and sodium cholate (1.87-fold) in other scientific reports as well (Rajput et al., 2010).

Enzyme Kinetics

K_m and V_{max} values of keratinase using different concentrations of a chicken feather as substrate (0.5–5.0 mg/ml) were 5.69 mg/ml and 142.40 μg/ml/min, respectively, with R^2 of 0.9902 (**Figure 5**). K_m and V_{max} values of keratinase SLSP-k using casein (substrate) were 0.64 mM and 420 μmol/ml/min, respectively (Gurunathan et al., 2021). Another research on alkaline protease reports K_m and V_{max} values of 7.0 mg/ml and 54.30 μmol/ml/min, respectively (Mushtaq et al., 2012).

Ker Gene Sequencing and Prediction of Keratinase Sequence

The *ker* gene sequence study revealed the presence of 1,197 bp nucleotides encoding for 399 amino acids (**Supplementary Figure 2**). The longest reading frame (362 amino acids) from position 26 to 387 was chosen for further *in silico* analysis and is mentioned underneath.

MAFSNMSAQAAGKSSTEKKYIVGFKQTMSAMSSAKKKD
VISEKGGKVQKQFKYVNAAAATLDEKAVKELMIPADLGKK
GADRPIIKGAQSVPYGISQIKAPALHSQGYTGSNVKVAVIDS
GIDSSHPDLNVRGGASFVPSETNPYQDGSSHGTHVAGTIAA
LNNSIGVLGVAPSASLYAVKVLDSTGSGQYSWIINGIEWAIS
NNMDVDASKENATFGHINSAQANDPSVSSGIVVAAAAGN
EGSSGSTSTVGYPAKYPSTIAVGAVNSSNQRASFSSAGSELD
VMAPGVSIQSTLPGGTYGAYNGTSMATPHVAGAAALILSK
HPTWTNAQVRDRLESTATYLGSSFYYGKGLINVQAAAQ

Prosite search revealed that keratinase belongs to the subtilase family (serine protease) showing three distinct active sites ^{115}VAVIDSGIdssH^{126} (Asp active site), ^{151}HGThVAGtIAA^{161} (His active site), and ^{306}GTSmAtPhVAG^{316} (Ser active site). Further, PSI-BLAST depicted that the top 10 homologous enzymes were used for conserved domain analysis of keratinase. Subtilisin AprE enzymes from *Bacillus* (WP_041335567.1 and WP_017695350.1) with a maximum score of 631, a total score of 631, query cover of 100%, *e*-value of 0, the identity of 90.06%, and length of 381 were detected with maximum sequence homology with keratinase. Additionally, moderate sequence

FIGURE 6 | Weblogo for conserved domain analysis of keratinase. Blue, green, and black colored are hydrophilic, neutral, and hydrophobic residues, respectively. The relative frequency of each residue is symbolized by height of symbol within stack and overall height of stack demarcates degree of conservation.

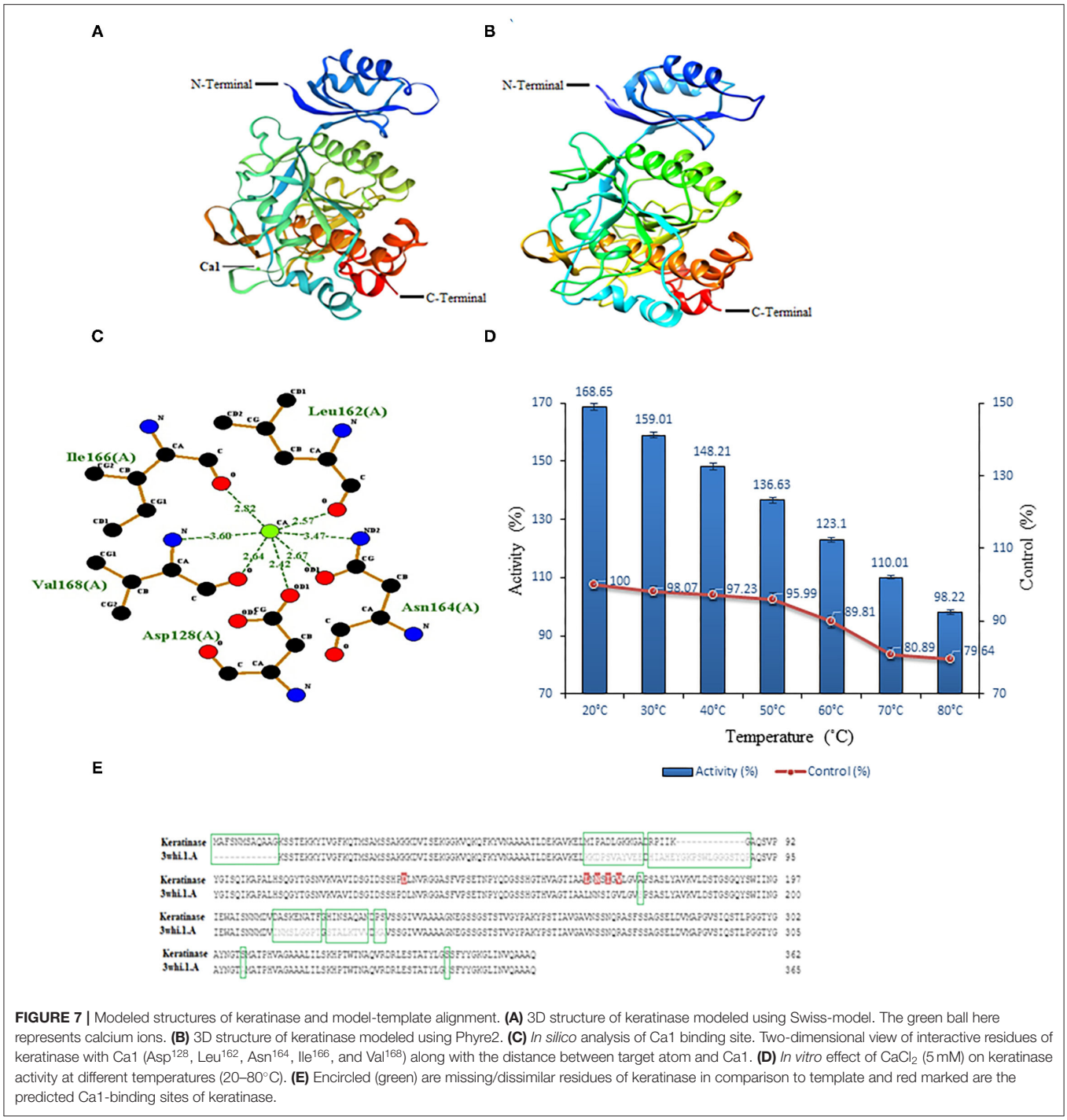


FIGURE 7 | Modeled structures of keratinase and model-template alignment. **(A)** 3D structure of keratinase modeled using Swiss-model. The green ball here represents calcium ions. **(B)** 3D structure of keratinase modeled using Phyre2. **(C)** *In silico* analysis of Ca1 binding site. Two-dimensional view of interactive residues of keratinase with Ca1 (Asp^{128}, Leu^{162}, Asn^{164}, Ile^{166}, and Val^{168}) along with the distance between target atom and Ca1. **(D)** *In vitro* effect of $CaCl_2$ (5 mM) on keratinase activity at different temperatures (20–80°C). **(E)** Encircled (green) are missing/dissimilar residues of keratinase in comparison to template and red marked are the predicted Ca1-binding sites of keratinase.

similarity was observed with keratinase from *Bacillus subtilis* (AIY62812.1) with a maximum score of 630, a total score of 630, query cover of 100%, *e*-value of 0, the identity of 90.06%, and length of 362. A clear alignment of Clustal Omega identified that conserved domains were generated using Weblogo. Results indicate that six large stretches of highly conserved domains in homologous regions were observed from positions (Gln^{28}-Lys^{65}, Gln^{67}-Leu^{88}, Ala^{107}-Pro^{146}, Leu^{148}-Asp^{226}, Val^{249}-Ser^{297}, and Gly^{299}-Ala^{336}) and low keratinase conserved amino acids at Met^{89}-Ala^{99}, Arg^{100}-Gly^{106}, Asp^{228}-Phe^{236}, His^{238}-Asn^{245}, Pro^{247}, and Ser^{248} (**Figure 6**; **Supplementary Figure 3**). Alignment of a novel keratinase KerUS from *Brevibacillus brevis* US575 with homologous enzymes revealed distinct keratinase residues in the mature region at Glu^{13}, Lys^{51}, Ala^{65}, Ser^{87}, Asp^{104}, Pro^{129}, Asn^{130}, Phe^{159}, Thr^{162}, Asp^{172}, Gly^{175}, Ser^{182}, Asn^{186}, Ala^{219}, Val^{233}, Tyr^{238}, Thr^{242}, and Asn^{243} (Jaouadi et al.,

TABLE 2 | Chicken feather degradation in 5 L fermenter depicting change in keratinase activity and pH in reaction system with due course of time.

Time (h)	Keratinase activity (U/mL)	pH
0	0	7.20 ± 0.05
2	21.57 ± 1.01	7.35 ± 0.08
4	47.20 ± 2.33	7.55 ± 0.10
6	100.21 ± 3.91	7.69 ± 0.10
8	189.36 ± 5.95	7.90 ± 0.15
10	248.51 ± 7.15	8.00 ± 0.17
12	356.98 ± 8.41	8.25 ± 0.18
14	441.47 ± 8.97	8.41 ± 0.20
16	509.92 ± 8.94	8.70 ± 0.25
18	627.23 ± 9.76	8.85 ± 0.22
20	789.00 ± 10.43	9.10 ± 0.35
22	780.87 ± 10.02	9.18 ± 0.39
24	768.04 ± 10.18	9.20 ± 0.41

TABLE 3 | The concentration of amino acids detected in feather hydrolysate after 12 and 24 h of degradation.

Amino acid	Concentration (μmol/mL)		
	0 h	12 h	24 h
Aspartic acid	0.29 ± 0.06	0.41 ± 0.05	0.71 ± 0.07
Glutamic acid	2.94 ± 0.17	3.73 ± 0.22	3.13 ± 0.25
Serine	4.48 ± 0.51	5.52 ± 0.81	5.46 ± 0.21
Histidine	0.51 ± 0.05	0.57 ± 0.05	0.62 ± 0.04
Glycine	14.97 ± 0.84	12.75 ± 0.80	12.51 ± 0.97
Threonine	3.62 ± 0.25	3.81 ± 0.19	3.95 ± 0.62
Alanine	16.27 ± 0.27	14.92 ± 0.75	14.06 ± 0.47
Arginine	1.32 ± 0.09	1.48 ± 0.11	1.92 ± 0.10
Tyrosine	5.29 ± 0.21	9.34 ± 0.44	26.87 ± 0.47
Valine	2.16 ± 0.13	7.13 ± 0.16	14.55 ± 0.81
Methionine	12.25 ± 0.87	15.49 ± 0.70	28.45 ± 0.69
Phenylalanine	1.15 ± 0.07	1.61 ± 0.08	3.78 ± 0.37
Isoleucine	3.41 ± 0.14	5.20 ± 0.41	9.56 ± 0.25
Leucine	4.99 ± 0.60	8.35 ± 0.73	14.10 ± 0.41
Lysine	6.88 ± 0.81	7.96 ± 0.18	9.31 ± 0.25
Cysteine	2.41 ± 0.09	3.10 ± 0.21	4.99 ± 0.20
Proline	ND	ND	ND
Tryptophan	ND	ND	ND
Glutamide	ND	ND	ND
Asparagine	ND	ND	ND

2013). Another study of a novel thermostable keratinase from *Deinococcus geothermalis* reports numerous conserved residues along with four conserved stretches from Asp^{163}-Ile^{166}, Gly^{195}-Thr^{203}, Asn^{252}-Gly^{257}, and Gly^{346}-Met^{349} (Tang et al., 2021). The outcomes thus suggest that keratinase exhibits no absolute similarity with already reported *Bacillus* sp. protein and is a novel keratinolytic enzyme with significant feather hydrolysis potential.

Structural Modeling and *in vitro* Validation

Swiss-model modeled keratinase structure with GMQE (0.86) and QMEAN (−2.01) and predicted 88.86% identity with "3whi.1.A" (subtilisin E, Crystal structure of unautoprocessed form of IS1-inserted pro-subtilisin E; **Figure 7A**). Besides, another modeling server Phyre2 predicted keratinase structure with 100% confidence, 96% coverage, and 89% identity with template "c3whiA" (**Figure 7B**). The modeled keratinase was found to exhibit Ca1 (calcium-binding site) and amino acid residues Asp^{128}, Leu^{162}, Asn^{164}, Ile^{166}, and Val^{168} were present amid 4 Å for ligand contacts (**Figure 7C**). Subtilisin enzymes from *Bacillus* sp. have two high-affinity Ca1 and low-affinity Ca2-binding sites, which have (specifically Ca1) aids in protection against autolysis and thermal stability (Sharma et al., 2021a,b). Our *in silico* study supports the *in vitro* characterization results wherein an increase in activity was observed in the presence of $CaCl_2$ (2/5 mM). The effect of Ca ions on thermostability was further analyzed to draw conclusive results. A significant increase of 68.65, 60.94, 50.98, 40.64, 33.29, 29.12, and 18.58% was examined in the presence of $CaCl_2$ (5 mM) at temperatures 20, 30, 40, 50, 60, 70, and 80°C, respectively, thus validating our *in silico* outcome (**Figure 7D**). Enhanced thermostability of proteases M179 and AprE176 from *Bacillus subtilis* HK176 was accessed with Ca ions wherein M179 retained 36% of its activity after 5 h at 45°C while AprE176 retained 11% activity (Jeong et al., 2015). RFEA1 from *B. cereus* RSA1 showed 30.87% of increase in its activity in the presence of $CaCl_2$ and an increased thermostability of 10.58, 21.12, 26.71, 30.17, 36.98, 41.25, and 45.65% at temperatures 80, 70, 60, 50, 40, 30, and 20°C, respectively (Sharma et al., 2021a). In another study, *Bacillus subtilis* DC27-produced enzyme was reported with enhanced activity (122.02 ± 5.71%) in the presence of 5 mM Ca ions (Hu et al., 2019).

Model-template alignment revealed that stretch of residues from Met^{1} to Gly^{12} is lacking in the template with respect to keratinase and from Tyr^{77} to Gln^{89} is lacking in keratinase with respect to the template. Besides, a large number of dissimilar residues (Met^{70}-Ala^{80}, Arg^{82}-Lys^{86}, Gly^{87}, Ala^{172}, Asp^{209}-Phe^{217}, His^{219}-Asn^{216}, Pro^{228}, Ser^{229}, Ser^{308}, and Ser^{346}) were reported in keratinase in comparison to a template. Additionally, huge dissimilarity in positions of residues was observed amid keratinase and its template (**Figure 7E**). Research suggests that despite high identity, homologous enzymes might exhibit distinct residues and lack some stretches of amino acids (Sharma et al., 2021a).

Scale-Up and Amino Acid Profiling

Scale-up to a 5 L fermenter was performed to improve efficiency and outstretch the broad industrial conditions of chicken feather hydrolyzing novel *B. pacificus* RSA27 isolate. Predictably, all chicken feathers degraded within 24 h of inoculation with a degradation rate of 94.5% and left with the remaining 0.825 g undecomposed feather in the culture, which is extremely difficult to decompose. Keratinase production started within 2 h of culture and retained its maximum activity (789.00 U/ml), which was 6.31-fold increment than the original activity (125.01 U/ml). A gradual increase in pH (9.20) was observed with respect to time (**Table 2**). Besides, the scale-up cultivation in a 5 L bioreactor has led to an increase in efficiency of feather hydrolysis by lowering degradation

TABLE 4 | Cytotoxicity evaluation of protein hydrolysate (0.09−50%) using MTT assay in HepG2 cells.

Compound	Concentration	Cytotoxicity (%)
Positive control (doxorubicin hydrochloride)	0.10 μM	49.10 ± 2.19
	1.00 μM	75.62 ± 2.67
	5.00 μM	81.52 ± 3.72
Protein hydrolysate (cell-free supernatant)	0.09%	1.02 ± 0.08
	0.25%	2.51 ± 0.06
	1.00%	4.60 ± 0.14
	1.75%	10.20 ± 0.53
	2.50%	18.29 ± 0.83
	5.00%	21.29 ± 0.99
	10.00%	35.21 ± 1.59
	25.00%	59.21 ± 2.65
	50.00%	61.29 ± 2.36

time to 24 h, which is half the degradation time in 100-ml Erlenmeyer flasks. Therefore, such large-scale optimization using novel *B. pacificus* RSA27 exhibits significant potential for industrial poultry feather waste degradation. In recent research, 81.8% rate of chicken feather degradation was observed in a 3-L fermenter and an increase in chicken feather degrading efficiency was accomplished by reducing the degradation time to half the actual time (Peng et al., 2019). Bioreactor scale-up of enzyme production has also been performed and reported in many other previous reports, wherein Ni et al. (2011) stated five-fold increase in keratinase production in 5 L bioreactor using optimized medium. Similarly, Fang et al. (2013) has observed enhanced production of keratinase in fed-batch conditions using keratin waste (117.7%).

The hydrolyzed chicken feather consists of protein β-keratin (91%), which was analyzed through high-performance liquid chromatography at the time intervals of 12 and 24 h of the degradation process. The concentration of hydrolyzed sample (12 and 24 h) is depicted in **Table 3**, wherein the total composition of amino acids was observed 153.97 μmol/ml after 24 h of hydrolysis. An increase in the concentration of aspartic acid (2.45-fold), tyrosine (5.08-fold), valine (6.74-fold), methionine (2.32-fold), phenylalanine (3.29-fold), isoleucine (2.80-fold), leucine (2.82-fold), lysine (1.35-fold), and cysteine (2.07-fold) was observed over the period of time taken for complete degradation (24 h). The obtained amino acid concentrations are greater than those stated previously (Jeong et al., 2010; Fang et al., 2013; Ramakrishna et al., 2017; Peng et al., 2019; **Supplementary Table 1**). Tyrosine, leucine, valine, and phenylalanine are essential amino acids that cannot be synthesized in the body. Also, sulfur-containing methionine and cysteine along with tyrosine, phenylalanine, and threonine are important for feather and hair keratin synthesis, whereas amino acid arginine contributes to cats' urea cycle. Hydrolyzed chicken feathers thus have a prodigious potential to produce amino acids, which can be further utilized as animal protein supplements (Peng et al., 2019).

Cytotoxicity of Protein Hydrolysate on Mammalian Cell Lines

Metabolism of nutrients and digestives (amino acids, glucose, and fatty acids) is performed by liver tissue, which encompasses essential enzymes for such requisite functions (Sadi et al., 2015). Also, hepatocellular carcinoma (liver cancer) is a major global concern and therefore HepG2 cells were used for *in vitro* cytotoxicity evaluation of hydrolyzed feathers. Furthermore, cytotoxicity evaluation of an enzyme is extremely important for its commercialization/applicability, and limited attempts are scientifically reported in this regard. Our cytotoxicity data revealed that in *in vitro* conditions 0.09% of protein hydrolysate has very less/negligible cytotoxicity (1.02%), whereas an increase in toxicity was observed at higher concentrations (**Table 4**). As a hydrolyzed feather at 0.09% concentration exhibited negligible cytotoxic effect as compared to the positive control, it might be explored as an efficient fertilizer and animal feed supplement at low concentrations along with several industrial applications. However, *in vivo* animal modeled examination is essential to assess any allergic or immunomodulatory reaction of the hydrolyzed feather.

CONCLUSION

Agro-environmental organic waste valorization to beneficial produce is highly essential for sustainable development and has extensively nurtured the bio-economy worldwide. Henceforth, keratinase by *B. pacificus* RSA27 is a novel and efficient eco-friendly alternative for bio-converting chicken feathers into nutritive protein hydrolysate. Scanning electron microscopy has evidently displayed the keratinase potential of chicken feather hydrolysis pattern over time. The protease showed significant keratinolytic activity even at extreme temperature (20–80°C) and pH range (3–12). Here, observations suggest that the keratinase will retain its activity even in the mammalian digestive tract as there is no major loss in its activity even at extreme acidic conditions of pH 3–6. Also, the protease showed significant activity with divalent/monovalent metal ions and surfactants, imperative for optimum production/utility. The keratinase was observed as a serine protease with complete activity inhibition with PMSF (0.09%). K_m and V_{max} values of 5.69 mg/ml and 142.40 μg/ml/min indicate significant enzyme-substrate affinity/biocatalysis. Deduced sequence/structural analysis confirms its novelty with distinct residues, active sites, and high-affinity Ca1-binding site (at diverse positions) when compared with homologous enzymes. Scale-up revealed complete feather hydrolysis with significant activity (789 U/ml) and protein content (153.97 μmol/ml), indicating its substantial use in numerous biotechnological applications. Conclusively, effectual feather degradation by novel keratinase supports the prospective of recalcitrant keratinous waste valorization into valuable produce.

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

RS contributed to the conception and design of the study. CS conducted the experiments and wrote the manuscript. ST assisted CS in performing experiments and manuscript writing. RS, AO, and JM supervised the study. All authors contributed to manuscript revision, read, and approved the submitted version.

FUNDING

The research has been financially supported by the Indian Council of Agricultural Research (ICAR) under F. No. AS/22/5/2018-ASR-IV.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.882902/full#supplementary-material

REFERENCES

Almahasheer, A. A., Mahmoud, A., El-Komy, H., Alqosaibi, A. I., Aktar, S., AbdulAzeez, S., et al. (2022). Novel feather degrading keratinases from *Bacillus cereus* group: biochemical, genetic and bioinformatics analysis. *Microorganisms* 10:93. doi: 10.3390/microorganisms10010093

Anitha, T. S., and Palanivelu, P. (2013). Purification and characterization of an extracellular keratinolytic protease from a new isolate of *Aspergillus parasiticus*. *Protein Expr. Purif.* 88, 214–220. doi: 10.1016/j.pep.2013.01.007

Brandelli, A. (2008). Bacterial keratinases: useful enzymes for bioprocessing agroindustrial wastes and beyond. *Food Bioproc. Tech.* 1, 105–116. doi: 10.1007/s11947-007-0025-y

Calin, M., Constantinescu, A. D., Alexandrescu, E., Raut, I., Badea, D. M., Arsene, M. L., et al. (2017). Degradation of keratin substrates by keratinolytic fungi. *Electron. J. Biotechn.* 28, 101–112. doi: 10.1016/j.ejbt.2017.05.007

Cao, Z. J., Zhang, Q., Wei, D. K., Chen, L., Wang, J., Zhang, X. Q., et al. (2009). Characterization of a novel Stenotrophomonas isolate with high keratinase activity and purification of the enzyme. *J. Ind. Microbiol. Biot.* 36, 181–188. doi: 10.1007/s10295-008-0469-8

Chang, B. S., and Mahoney, R. R. (1995). Enzyme thermostabilization by bovine serum albumin and other proteins: evidence for hydrophobic interactions. *Biotechnol. Appl. Biochem.* 22, 203–214.

Devi, S., Chauhan, A., Bishist, R., Sankhyan, N., Rana, K., and Sharma, N. (2022). Production, partial purification and efficacy of keratinase from Bacillus halotolerans L2EN1 isolated from the poultry farm of Himachal Pradesh as a potential laundry additive. *Biocatal. Biotransform.* 40, 1–21. doi: 10.1080/10242422.2022.2029851

Dong, Y. Z., Chang, W. S., and Chen, P. T. (2017). Characterization and overexpression of a novel keratinase from Bacillus polyfermenticus B4 in recombinant *Bacillus subtilis*. *Bioresour. Bioprocess.* 4:47. doi: 10.1186/s40643-017-0177-1

Fang, Z., Zhang, J., Liu, B., Du, G., and Chen, J. (2013). Biodegradation of wool waste and keratinase production in scale-up fermenter with different strategies by *Stenotrophomonas maltophilia* BBE11-1. *Bioresour. Technol.* 140, 286–291. doi: 10.1016/j.biortech.2013.04.091

Farag, A., and Hassan, M. (2004). Purification, characterization and immobilization of a keratinase from *Aspergillus oryzae*. *Enzyme Microb. Technol.* 34, 85–93. doi: 10.1016/j.enzmictec.2003.09.002

Gradisar, H., Friedrich, J., Krizaj, I., and Jerala, R. (2005). Similarities and specificities of fungal keratinolytic proteases: comparison of keratinases of *Paecilomyces marquandii* and Doratomyces microsporus to some known proteases. *Appl. Environ. Microbiol.* 71, 3420–3426. doi: 10.1128/AEM.71.7.3420-3426.2005

Gupta, R., and Ramnani, P. (2006). Microbial keratinases and their prospective applications: an overview. *Appl. Microbiol. Biotechnol.* 70, 21–33. doi: 10.1007/s00253-005-0239-8

Gupta, S., Nigam, A., and Singh, R. (2015). Purification and characterization of a *Bacillus subtilis* keratinase and its prospective application in feed industry. *Acta Biol. Szeged.* 59, 197–204. Available online at: https://bipublication.com/files/IJABR-V4I1-2013-24.pdf

Gupta, S., and Singh, R. (2013). Statistical modeling and optimization of keratinase production from newly isolated *Bacillus subtilis* RSE163. *Int. J. Adv. Biotechnol. Res.* 4, 167–174. Available online at: https://www2.sci.u-szeged.hu/ABS/2015/Acta%20HPe/59197.pdf

Gupta, S., and Singh, R. (2014). Hydrolyzing proficiency of keratinases in feather degradation. *Indian J. Microbiol.* 54, 466–470. doi: 10.1007/s12088-014-0477-5

Gupta, S., Tewatia, P., Misri, J., and Singh, R. (2017). Molecular modeling of cloned *Bacillus subtilis* keratinase and its insinuation in Psoriasis treatment using docking studies. *Indian J. Microbiol.* 57, 485–491. doi: 10.1007/s12088-017-0677-x

Gurunathan, R., Huang, B., Ponnusamy, V. K., Hwang, J. S., and Dahms, H. U. (2021). Novel recombinant keratin degrading subtilisin like serine alkaline protease from *Bacillus cereus* isolated from marine hydrothermal vent crabs. *Sci. Rep.* 11:12007. doi: 10.1038/s41598-021-90375-4

Hamiche, S., Mechri, S., Khelouia, L., Hattab, M., Badis, A., and Jaouadi, B. (2019). Purification and biochemical characterization of two keratinases from *Bacillus amyloliquefaciens* S13 isolated from marine brown alga *Zonaria tourneforti* with potential keratin-biodegradation and hide-unhairing activities. *Int. J. Biol. Macromol.* 122, 758–769. doi: 10.1016/j.ijbiomac.2018.10.174

He, Z., Sun, R., Tang, Z., Bu, T., Wu, Q., Li, C., et al. (2018). Biodegradation of feather waste keratin by the keratin-degrading strain *Bacillus subtilis* 8. *J. Microbiol. Biotechnol.* 28, 314–322. doi: 10.4014/jmb.1708.08077

Hu, Y., Yu, D., Wang, Z., Hou, J., Tyagi, R., Liang, Y., et al. (2019). Purification and characterization of a novel, highly potent fibrinolytic enzyme from *Bacillus subtilis* DC27 screened from Douchi, a traditional Chinese fermented soybean food. *Sci. Rep.* 9:9235. doi: 10.1038/s41598-019-45686-y

Jaouadi, N. Z., Rekik, H., Badis, A., Trabelsi, S., Belhoul, M., Yahiaoui, A. B., et al. (2013). Biochemical and molecular characterization of a serine keratinase from *Brevibacillus brevis* US575 with promising keratin-biodegradation and hide-dehairing activities. *PLoS ONE.* 8:e76722. doi: 10.1371/journal.pone.0076722

Jeong, J. H., Park, K. H., Oh, D. J., Hwang, D. Y., Kim, H. S., Lee, C. Y., et al. (2010). Keratinolytic enzyme-mediated biodegradation of recalcitrant feather by a newly isolated *Xanthomonas* sp. P5. *Polym. Degrad. Stab.* 95, 1969–1977. doi: 10.1016/j.polymdegradstab.2010.07.020

Jeong, S. J., Heo, K., Park, J. Y., Lee, K. W., Park, J. Y., Joo, S. H., et al. (2015). Characterization of AprE176, a fibrinolytic enzyme from *Bacillus subtilis* HK176. *J. Microbiol. Biotechnol.* 25, 89–97. doi: 10.4014/jmb.1409.09087

Jin, E., Reddy, N., Zhu, Z., and Yang, Y. (2011). Graft polymerization of native chicken feathers for thermoplastic applications. *J. Agric. Food Chem.* 59, 1729–1738. doi: 10.1021/jf1039519

Kalia, V. C., and Purohit, H. J. (2008). Microbial diversity and genomics in aid of bioenergy. *J. Ind. Microbiol. Biotechnol.* 35, 403–419. doi: 10.1007/s10295-007-0300-y

Kondaveeti, S., Park, G. D., Shanmugam, R., Pagolu, R., Patel, S. K., Bisht, A., et al. (2022). Investigating the role of metals loaded on nitrogen-doped carbon-nanotube electrodes in electroenzymatic alcohol dehydrogenation. *Appl. Catal. B.* 307:121195. doi: 10.1016/j.apcatb.2022.121195

Latha, P. P., Singh, R. K., Kukrety, A., Saxena, R. C., Bhatt, M., and Jain, S. L. (2016). Poultry chicken feather derived biodegradable multifunctional additives for lubricating formulations. *ACS Sustain. Chem. Eng.* 4, 999–1005. doi: 10.1021/acssuschemeng.5b01071

Lee, Y. S., Phang, L. Y., Ahmad, S. A., and Ooi, P. T. (2016). Microwave-alkali treatment of chicken feathers for protein hydrolysate production. *Waste Biomass Valori*. 7, 1147–1157. doi: 10.1007/s12649-016-9483-7

Li, Q. (2019). Progress in microbial degradation of feather waste. *Front. Microbiol*. 10:2717. doi: 10.3389/fmicb.2019.02717

Li, Q. (2021). Structure, application, and biochemistry of microbial keratinases. *Front. Microbiol*. 12:674345. doi: 10.3389/fmicb.2021.674345

Liu, Y., Du, J., Lai, Q., Zeng, R., Ye, D., Xu, J., et al. (2017). Proposal of nine novel species of the *Bacillus cereus* group. *Int. J. Syst. Evol. Microbiol*. 67, 2499–2508. doi: 10.1099/ijsem.0.001821

Mazotto, A. M., Cedrola, S. M. L., de Souza, E. P., Couri, S., and Vermelho, A. B. (2022). Enhanced keratinase production by *Bacillus subtilis* amr using experimental optimization tools to obtain feather protein lysate for industrial applications. *3 Biotech*. 12:90. doi: 10.1007/s13205-022-03153-y

McKittrick, J., Chen, P., Bodde, S. G., Yang, W., Novitskaya, E. E., and Meyers, M. A. (2012). The structure, functions, and mechanical properties of keratin. *JOM* 64, 449–468. doi: 10.1007/s11837-012-0302-8

Moreira-Gasparin, F. G., De-Souza, C. G., Costa, A. M., Alexandrino, A. M., Bracht, C. K., Boer, C. G., et al. (2009). Purification and characterization of an efficient poultry feather degrading protease from *Myrothecium verrucaria*. *Biodegradation* 20, 727–736. doi: 10.1007/s10532-009-9260-4

Mushtaq, Z., Irfan, M., Nadeem, M., Naz, M., and Syed, Q. (2012). Kinetic study of extracellular detergent stable alkaline protease from *Rhizopus oryzae*. *Braz. Arch. Biol. Technol*. 58, 175–184. doi: 10.1590/S1516-8913201400071

Ni, H., Chen, Q., Chen, F., Fu, M., Dong, Y., and Cai, H. (2011). Improved keratinase production for feather degradation by *Bacillus licheniformis* ZJUEL31410 in submerged cultivation. *Afr. J. Biotechnol*. 10, 7236–7244. doi: 10.5897/AJB11.168

Nigam, P. S. (2013). Microbial enzymes with special characteristics for biotechnological applications. *Biomolecules* 3, 597–611. doi: 10.3390/biom3030597

Nnolim, N. E., Mpaka, L., Okoh, A. I., and Nwodo, U. U. (2020b). Biochemical and molecular characterization of a thermostable alkaline metallo-keratinase from *Bacillus* sp. Nnolim-K1. *Microorganisms* 8:1304. doi: 10.3390/microorganisms8091304

Nnolim, N. E., Udenigwe, C. C., Okoh, A. I., and Nwodo, U. U. (2020a). Microbial keratinase: next generation green catalyst and prospective applications. *Front. Microbiol*. 11:3280. doi: 10.3389/fmicb.2020.580164

Pandey, D., Patel, S. K., Singh, R., Kumar, P., Thakur, V., and Chand, D. (2019). Solvent-tolerant acyltransferase from Bacillus sp. APB-6: purification and characterization. *Indian J. Microbiol*. 59, 500–507. doi: 10.1007/s12088-019-00836-8

Patel, S. K., Das, D., Kim, S. C., Cho, B. K., Kalia, V. C., and Lee, J. K. (2021). Integrating strategies for sustainable conversion of waste biomass into dark-fermentative hydrogen and value-added products. *Renew. Sustain. Energy Rev*. 150:111491. doi: 10.1016/j.rser.2021.111491

Patel, S. K., Gupta, R. K., Kumar, V., Mardina, P., Lestari, R., Kalia, V. C., et al. (2019). Influence of metal ions on the immobilization of β-glucosidase through protein-inorganic hybrids. *Indian J. Microbiol*. 59, 370–374. doi: 10.1007/s12088-019-00796-z

Patel, S. K. S., Kalia, V. C., Kim, S. Y., Lee, J. K., and Kim, I. W. (2022). Immobilization of laccase through inorganic-protein hybrids using various metal ions. *Indian J. Microbiol*. doi: 10.1007/s12088-022-01000-5

Peng, Z., Mao, X., Zhang, J., Du, G., and Chen, J. (2019). Effective biodegradation of chicken feather waste by co-cultivation of keratinase producing strains. *Microb. Cell Fact*. 18:84. doi: 10.1186/s12934-019-1134-9

Qiu, J., Barrett, K., Wilkens, C., and Meyer, A. S. (2022). Bioinformatics based discovery of new keratinases in protease family M36. *New Biotechnol*. 68, 19–27. doi: 10.1016/j.nbt.2022.01.004

Rahman, R. N. Z. R., Geok, L. P., Basri, M., and Salleh, A. B. (2005). Physical factors affecting the production of organic solvent-tolerant protease by Pseudomonas aeruginosa strain K. *Bioresour. Technol*. 96, 429–436. doi: 10.1016/j.biortech.2004.06.012

Rajput, R., Sharma, R., and Gupta, R. (2010). Biochemical characterization of a thiol-activated, oxidation stable keratinase from *Bacillus pumilus* KS12. *Enzyme Res.* 2010:132148. doi: 10.4061/2010/132148

Ramakrishna, M. R., Sathi, K. R., Ranjita, Y. C., Bee, H., and Reddy, G. (2017). Effective feather degradation and keratinase production by *Bacillus pumilus* GRK for its application as bio-detergent additive. *Bioresour. Technol*. 243, 254–263. doi: 10.1016/j.biortech.2017.06.067

Rissen, S., and Antranikian, G. (2001). Isolation of Thermoanaerobacter *Keratinophilus* sp. nov., a novel thermophilic, anaerobic bacterium with keratinolytic activity. *Extremophiles* 5, 399–408. doi: 10.1007/s007920100209

Sadi, G., Emsen, B., Kaya, A., Kocabaş, A., Çinar, S., and Kartal, D. (2015). Cytotoxicity of some edible mushrooms extracts over liver hepatocellular carcinoma cells in conjunction with their antioxidant and antibacterial properties. *Pharmacogn. Mag*. 1, 6–18. doi: 10.4103/0973-1296.157665

Selvam, K., Vishnupriya, B., and Yamuna, M. (2013). Isolation and description of keratinase producing marine actinobacteria from South Indian Coastal Region. *Afr. J. Biotechnol*. 12, 19–26. doi: 10.5897/AJB12.2428

Sharma, C., Nigam, A., and Singh, R. (2021a). Computational-approach understanding the structure-function prophecy of fibrinolytic protease RFEA1 from *Bacillus cereus* RSA1. *PeerJ*. 9:e11570. doi: 10.7717/peerj.11570

Sharma, C., Osmolovskiy, A., and Singh, R. (2021b). Microbial fibrinolytic enzymes as anti-thrombotics: production, characterisation and prodigious biopharmaceutical applications. *Pharmaceutics* 13:1880. doi: 10.3390/pharmaceutics13111880

Sharma, C., Salem, G. E. M., Sharma, N., Gautam, P., and Singh, R. (2020). Thrombolytic potential of novel thiol-dependent fibrinolytic protease from *Bacillus cereus* RSA1. *Biomolecules* 10:3. doi: 10.3390/biom10010003

Singh, R. S., Singh, T., and Pandey, A. (2019). "Microbial enzymes," in *Advances in Enzyme Technology*, eds R. S. Singh, R. R. Singhania, A. Pandey, C. Larroche (Oxford: Elsevier), 1–40. doi: 10.1016/B978-0-444-64114-4.00001-7

Singh, S., Gupta, P., Sharma, V., Koul, S., Kour, K., and Bajaj, B. K. (2014). Multifarious potential applications of keratinase of *Bacillus subtilis* K-5. *Biocatal. Biotransfor*. 32, 333–342. doi: 10.3109/10242422.2014.978306

Suman, S. K., Patnam, P. L., Ghosh, S., and Jain, S. L. (2018). Chicken feather derived novel support material for immobilization of laccase and its application in oxidation of veratryl alcohol. *ACS Sustain. Chem. Eng*. 7, 3464–3474. doi: 10.1021/acssuschemeng.8b05679

Suntornsuk, W., Tongjun, J., Onnim, P., Oyama, H., Ratanakanokchai, K., Kusamran, T., et al. (2005). Purification and characterisation of keratinase from a thermotolerant feather-degrading bacterium. *World J. Microbiol. Biotechnol*. 21:1111. doi: 10.1007/s11274-005-0078-x

Tang, Y., Guo, L., Zhao, M., Gui, Y., Han, J., Lu, W., et al. (2021). A novel thermostable keratinase from *Deinococcus geothermalis* with potential application in feather degradation. *Appl. Sci*. 11:3136. doi: 10.3390/app11073136

Tatineni, R., Doddapaneni, K. K., Potumarthi, R. C., Vellanki, R. N., Kandathil, M. T., Kolli, N., et al. (2008). Purification and characterization of an alkaline keratinase from *Streptomyces* sp. *Bioresour. Technol*. 99, 1596–1602. doi: 10.1016/j.biortech.2007.04.019

Thys, R. C., and Brandelli, A. (2006). Purification and properties of a keratinolytic metalloprotease from *Microbacterium* sp. *J. Appl. Microbiol*. 101, 1259–1268. doi: 10.1111/j.1365-2672.2006.03050.x

Verma, A., Singh, H., Anwar, M. S., Kumar, S., Ansari, M. W., and Agrawal, S. (2016). Production of thermostable organic solvent tolerant keratinolytic protease from *Thermoactinomyces* sp. RM4: IAA production and plant growth promotion. *Front. Microbiol*. 7:1189. doi: 10.3389/fmicb.2016.01189

Verma, A., Singh, H., Anwar, S., Chattopadhyay, A., Tiwari, K. K., Kaur, S., et al. (2017). Microbial keratinases: industrial enzymes with waste management potential. *Crit. Rev. Biotechnol*. 37, 476–491. doi: 10.1080/07388551.2016.1185388

Vidmar, B., and Vodovnik, M. (2018). Microbial keratinases: enzymes with promising biotechnological applications. *Food Technol. Biotechnol*. 56, 312–328. doi: 10.17113/ftb.56.03.18.5658

Wu, W. L., Chen, M. Y., Tu, I. F., Lin, Y. C., Kumar, N. E., Chen, M. Y., et al. (2017). The discovery of novel heat stable keratinases from Meiothermus taiwanensis WR-220 and other extremophiles. *Sci. Rep*. 7:4658. doi: 10.1038/s41598-017-04723-4

Yong, B., Fei, X., Shao, H., Xu, P., Hu, Y., Ni, W., et al. (2020). Recombinant expression and biochemical characterization of a novel keratinase BsKER71 from feather degrading bacterium *Bacillus subtilis* S1-4. *AMB Expr*. 10:9. doi: 10.1186/s13568-019-0939-6

Zhang, R. X., Gong, J. S., Su, C., Zhang, D. D., Tian, H., Dou, W. F., et al. (2016). Biochemical characterization of a novel surfactant-stable serine keratinase with no collagenase activity from *Brevibacillus parabrevis* CGMCC 10798. *Int. J. Biol. Macromol*. 93, 843–851. doi: 10.1016/j.ijbiomac.2016.09.063

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: 01 July 2022
doi: 10.3389/fmicb.2022.872735

Fresh Versus Frozen Stool for Fecal Microbiota Transplantation—Assessment by Multimethod Approach Combining Culturing, Flow Cytometry, and Next-Generation Sequencing

Jaroslaw Bilinski[1†], Mikolaj Dziurzynski[2*†], Pawel Grzesiowski[3], Edyta Podsiadly[4], Anna Stelmaszczyk-Emmel[5], Tomasz Dzieciatkowski[6], Karol Lis[1], Martyna Tyszka[1], Krzysztof Ozieranski[7], Łukasz Dziewit[2] and Grzegorz W. Basak[1]*

[1] Department of Hematology, Transplantation and Internal Medicine, Medical University of Warsaw, Warsaw, Poland, [2] Department of Environmental Microbiology and Biotechnology, Faculty of Biology, Institute of Microbiology, University of Warsaw, Warsaw, Poland, [3] Foundation for the Infection Prevention Institute, Warsaw, Poland, [4] Department of Pharmaceutical Microbiology, Medical University of Warsaw, Warsaw, Poland, [5] Department of Laboratory Diagnostics and Clinical Immunology of Developmental Age, Medical University of Warsaw, Warsaw, Poland, [6] Department of Medical Microbiology, Medical University of Warsaw, Warsaw, Poland, [7] First Department of Cardiology, Medical University of Warsaw, Warsaw, Poland

The objective of this work was to compare the quality of FMT preparations made from fresh feces with those made from feces frozen at –30°C without any pre-processing or cryopreservation additives. The research hypothesis was that such preservation protocol (frozen whole stool, then thawed and processed) is equipotent to classical fresh FMT preparation. For that, three complementary methods were applied, including: (i) culturing in aerobic and anaerobic conditions, (ii) measuring viability by flow cytometry, and (iii) next-generation sequencing. Flow cytometry with cell staining showed that the applied freezing protocol causes significant changes in all of the observed bacterial fractions. Alive cell counts dropped four times, from around 70% to 15%, while the other two fractions, dead and unknown cell counts quadrupled and doubled, with the unknown fraction becoming the dominant one, with an average contribution of 57.47% per sample. It will be very interesting to uncover what this unknown fraction is (e.g., bacterial spores), as this may change our conclusions (if these are spores, the viability could be even higher after freezing). Freezing had a huge impact on the structure of cultivable bacterial communities. The biggest drop after freezing in the number of cultivable species was observed for Actinobacteria and Bacilli. In most cases, selected biodiversity indices were slightly lower for frozen samples. PCoA visualization built using weighted UniFrac index showed no donor-wise clusters, but a clear split between fresh

OPEN ACCESS

Edited by:
Utkarsh Sood,
The Energy and Resources Institute (TERI), India

Reviewed by:
Roshan Kumar,
Magadh University, India
Peter Kilbride,
Cytiva, United Kingdom
Om Prakash,
National Centre for Cell Science, India

***Correspondence:**
Jaroslaw Bilinski
jaroslaw.bilinski@gmail.com
Mikolaj Dziurzynski
mikolaj.dziurzynski@biol.uw.edu.pl

† 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: *09 February 2022*
Accepted: *19 May 2022*
Published: *01 July 2022*

Citation:
Bilinski J, Dziurzynski M, Grzesiowski P, Podsiadly E, Stelmaszczyk-Emmel A, Dzieciatkowski T, Lis K, Tyszka M, Ozieranski K, Dziewit Ł and Basak GW (2022) Fresh Versus Frozen Stool for Fecal Microbiota Transplantation—Assessment by Multimethod Approach Combining Culturing, Flow Cytometry, and Next-Generation Sequencing. Front. Microbiol. 13:872735. doi: 10.3389/fmicb.2022.872735

and frozen samples. This split can be in part attributed to the changes in the relative abundance of Bacteroidales and Clostridiales orders. Our results clearly show that whole stool freezing without any cryoprotectants has a great impact on the cultivability and biodiversity of the bacterial community, and possibly also on the viability of bacterial cells.

Keywords: fecal microbiota transplantation, conservation, gut microbiota, culturing, next-generation sequencing, flow cytometry, viability

INTRODUCTION

Fecal microbiota transplantation (FMT) is a very effective treatment method in *Clostridioides difficile* infections (CDI) and other exploratory indications (Rokkas et al., 2019; DuPont et al., 2020; Ramai et al., 2021). To enable wide access to therapy for patients, various steps are taken to make FMT products commonly available. The most frequently used procedure is freezing the samples with the addition of cryoprotective agents, mostly glycerol.

Most studies applying frozen FMT have reported an overall CDI cure rate between 81 and 100% (Ramai et al., 2021). Two retrospective analyses (Hamilton et al., 2012; Satokari et al., 2015) and three randomized clinical trials (Youngster et al., 2014; Cammarota et al., 2015; Lee et al., 2016) directly compared fresh and frozen fecal preparations. All five studies reported no differences between fresh and frozen FMTs, despite the range in the storage time of frozen fecal matter (from 1 week to 6 months). Thus, current evidence indicates similar efficacy of frozen and fresh fecal preparations.

It is, however, postulated that glycerol can skew intestine microbiota composition and other natural methods should be applied (De Weirdt et al., 2010). Freezing the feces alone, without adding cryoprotectants, is the most common practice when collecting patient samples, but to the authors' knowledge, it is also practiced in stool storage for the production of fecal microbiota preparations for transfer into the gastrointestinal tract of the recipient now, not only historically (Gustafsson et al., 1999) but also with very good results not differing from others (Grzesiowski et al., 2015). Different cryoprotectants have been described and tested, for example, freeze-drying FMT capsules (Staley et al., 2017; Burz et al., 2019). However, sole freezing in −20°C of the whole stool sample before preparation has not been described to date, although descriptions of such freezing may suggest that collecting and freezing all feces can be very effective (Vandeputte et al., 2017).

The objective of this work was to compare the quality of FMT preparations made from fresh feces with those made from frozen at −30°C without cryopreserving and processing before thawing. The research hypothesis was that such preservation protocol (freezed whole stool, then thawed and processed) is equipotent to classical fresh FMT preparation. For testing this hypothesis three complementary methods were applied, including: (i) culturing in aerobic and anaerobic conditions, (ii) measuring viability by flow cytometric method, and (iiii) next-generation sequencing (Bilinski et al., 2020). We postulated that the stool has protective properties for bacteria itself, and its processing before freezing is not obligatory, since anaerobic conditions inside the feces and low hydrated matter content may be a natural cryo- and viability-preservative.

MATERIALS AND METHODS

Stool Donors, Stool Donations, and Processing Methodology

Ten consecutive stools donated by each of the three donors were used for the experiments (30 stools in total divided into two equal parts to prepare 60 fecal microbiota suspensions, 30 from fresh, and 30 from frozen stool). The donors were randomly selected (A and B) or intentionally chosen (C). In brief, one of them (donor C; male, 28 years old, healthy, with normal BMI) was a regular donor of feces to produce a preparation for the fecal microbiota transplantation (chosen from a stool donor bank) and the other two were randomly selected males (donor A—male, 16 years old with food allergy, recurrent aphthous stomatitis, and normal BMI; and donor B – male, 55 years old, with inhaled allergy, medical worker, with BMI 27). A medical questionnaire with basic data was received from each person. Each feces sample after donation and transportation to the laboratory was divided into two equal parts. Half of each stool was processed immediately and the second half was stored frozen, without any processing, at −30°C for a median of 15 days, and thereafter thawed and processed in the same way as the fresh stool sample. All samples were prepared in the same time frame and the same way in aerobic conditions by homogenizing, diluting in 0.9% NaCl, and sieving through sterile gauze or sieves to obtain a clear, homogeneous fluid being a suspension of feces. This is the regular way of producing feces for the use as FMT (Bilinski et al., 2017). The material both from fresh and frozen stool prepared in this way was then divided into three parts—one for assessment by the flow cytometry in the LIVE/DEAD method (Molecular Probes, Oregon, United States), one for performing classical culturing, and the third for immediate isolation of DNA for the V3-V4 16S rDNA variable regions sequencing (60 samples in total). **Figure 1** shows the research protocol we used in this work.

Flow Cytometry

Bacterial viability in fecal microbiota samples from fresh and frozen stool was measured by flow cytometry using LIVE/DEAD Bac*Light*™ Bacterial Viability and Counting Kit (L34856, Molecular Probes) according to manufacturer instructions (Molecular Probes, Oregon, United States). In brief, 977 μL

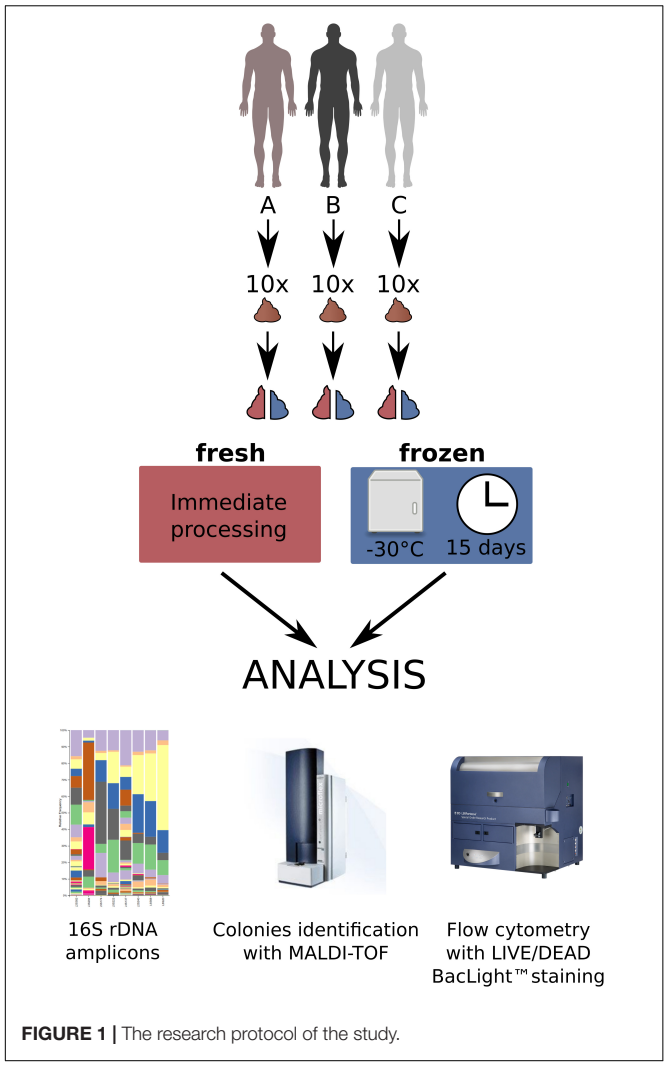


FIGURE 1 | The research protocol of the study.

of 0.9% NaCl, 1.5 μL of SYTO9, 1.5 μL of propidium iodide (PI), and 10 μL of the diluted sample were added to the flow cytometry analysis tube. Samples were 10-fold diluted in 0.9% NaCl. The tube was incubated for 15 min in the dark at room temperature and 10 μL of the microsphere suspension (beads) was added to the stained sample. The total volume of the sample in the flow cytometry analysis tube was 1000 μL. The samples were analyzed on an LSR Fortessa flow cytometer (Becton Dickinson, New Jersey, United States) with FACS Diva v8 software (Becton Dickinson). The gating strategy was as described in our first work (Bilinski et al., 2020). Three main cell populations were observed—alive, dead, and unknown (probably alive, probably dead) with a special not-alive-not-dead group of cells ($SYTO9^{-}PI^{-}$). The number of bacteria per mL in each analyzed gate was counted according to the following formula taken from the manufacturer materials:

$$\frac{\left(\left(\#of\ events \in gatedbacteriaregion\right) x\ \left(dillutionfactors\right)\right)}{\left[\left(\#of\ events \in beadregion\right) x10^{-6}\right]} = bacteria/ml$$

Cultivation of Stool Microbiota

Samples of fecal microbiota suspension derived from each fresh and frozen stool were plated on six different agar media and incubated under conditions as follows: (i) CNA medium (colistin nalidixic acid agar; Oxoid, Basingstoke, United Kingdom)—for cultivation of Gram-positive aerobes, an enriched agar medium containing sheep's blood, colistin, and nalidixic acid (to inhibit the growth of Gram-negative bacteria), and culture conditions: aerobic with 5% CO_2, 37°C, 48 h; (ii) MacConkey medium (bioMérieux, Marcy l'Etoile, France)—for the isolation of Gram-negative rods, contains bile salts and crystal violet (to inhibit the growth of Gram-positive bacteria), and culture conditions: aerobic, 37°C, 48 h; (iii) Bile and esculin (CC) medium (Oxoid)—for the isolation and identification of bacteria belonging to the genus *Enterococcus*, which grow well in the presence of bile and have the ability to break down esculin, and culture conditions: aerobic, 37°C, 48 h; (iv) Schaedler Anaerobe KV Selective Agar with freeze-dried horse blood and the addition of kanamycin and vancomycin (bioMérieux)—a highly nutritious medium for the selective growth and isolation of anaerobic bacteria, especially of the genus *Bacteroides* and *Prevotella*, culture conditions: anaerobic, 37°C, 4 days; (v) Schaedler Anaerobe KV Selective Agar with freeze-dried horse blood (bioMérieux)—a highly nutritious medium for the isolation of absolute and relative anaerobes, culture conditions: anaerobic, 37°C, 4 days; (vi) Sabouraud agar with gentamicin and chloramphenicol (Oxoid)—selective medium for cultivation of mold and yeast, high glucose concentration, and presence of antibiotics (chloramphenicol and gentamicin) and acidic pH inhibits bacterial growth; the presence of antibiotics is another selection factor, and culture conditions: aerobic, 37°C, 10 days. The anaerobic incubations were carried out in anaerobic jars and atmosphere generators (Oxoid).

After the initial sample processing, colonies were selected (at least one colony per morphology) for identification using a Microflex LT mass and the MBT Compass IVD Biotyper software (Bruker Daltonics, Bremen, Germany). The colonies were deposited on a MALDI-TOF (Bruker Daltonics) target microflex and extracted with 5% formic acid, air-dried, and then overlaid with a 1-μL matrix solution of α-cyano-4-hydroxycinnamic acid in 50% acetonitrile and 2.5% trifluoroacetic acid. Two spots were examined for each colony. The Biotyper software was used to compare the protein profile of the cultured bacteria from a database of Bruker consisting of 2750 protein profiles. A score >1.9 was considered a reliable identification at the species level and a score >1.7 indicated the identification at the genus level. Strains of bacteria with scores lower than 1.7 were considered as unidentified.

To enumerate the number of colony-forming units (CFU) in the stool samples, 0.2 g of stool was diluted in 1 mL of phosphate-buffered saline (PBS), and 1 to 5 μL of diluted sample was spread on each media. Bacterial counts were recorded as CFU per gram of feces for each isolated species.

DNA Sequencing

Total bacterial DNA was extracted using the Qiagen DNeasy Power Soil kit (Qiagen, Hilden, Germany) according to

the manufacturer's instructions and stored at –20°C. Using isolated DNA as a matrix, PCR reactions were performed in triplicate (to reduce PCR bias) using Bakt_341F 5′-CCTACGGGNGGCWGCAG-3′ and Bakt_805R 5′-GACTACHVGGGTATCTAATCC-3′ primer pair amplifying the variables V3 and V4 regions of the 16S rRNA genes (Herlemann et al., 2011; Klindworth et al., 2013). The electrophoretic analysis was performed for each of three replicates for the qualitative and quantitative evaluation of the PCR products. Then, products of three independent PCR reactions for each sample were mixed and used for the DNA sequencing as one amplicon to minimize the error due to the selectivity of the PCR reactions. The amplified PCR products were sequenced using the Illumina MiSeq instrument (Illumina, San Diego, CA, United States) in paired-end mode using a v3 chemistry kit (Illumina) at BIOBANK LAB (Chair and Department of Molecular Biophysics, University of Łódź, Poland).

Bioinformatic Analysis

Sequencing data were processed with Qiime2 (version 2020.10) package (Bolyen et al., 2019). The reads were imported into Qiime2 and run through the dada2 plugin to obtain Amplicon Sequence Variants (ASV) (Callahan et al., 2016). Taxonomy was assigned for each of the ASVs using a pre-trained Naive Bayes classifier, based on Silva 132 99% database (Quast et al., 2013), which was trimmed to include only the V3 and V4 regions of 16S rRNA gene, bound by Bakt_341F and Bakt_805R primers sequences. Alfa and beta diversity metrics were generated using the following Qiime2 plugins: phylogeny (including mafft aligner and FastTree tool), diversity, and emperor (Price et al., 2010; Katoh and Standley, 2013).

Statistical Analysis

Statistical differences between fresh and frozen samples were analyzed using Wilcoxon signed-rank test. To identify genera that differ in abundance between samples from different donors, ANCOM analysis was used (Mandal et al., 2015). All additional statistics and visualizations were generated using Python programming language with SciPy, matplotlib, pandas, numpy, and seaborn libraries (Hunter, 2007; Virtanen et al., 2020). Except when stated otherwise, p-values of less than 0.05 were considered statistically significant.

Ethics

All subjects gave their informed consent for inclusion before they participated in the study. The investigations were carried out following the rules of the Declaration of Helsinki. According to local Bioethics Committee rules, for the non-intervention studies, no approval was needed to conduct this study.

RESULTS

General Characteristics of Fresh Microbiota Samples

Results obtained for fresh samples have already been analyzed and described in our previous work (Bilinski et al., 2020). In brief, each method we applied (flow cytometry, classical culturing, and next-generation sequencing) has been shown to contribute to stool microbiota characterization differently, using a different perspective.

The flow cytometry analysis allowed for the evaluation of the total number of cells in each sample as well as live versus dead cell fractions analysis. In the performed analysis, a large fraction of cells not stained with one of the reagents, denoted Unknown, with an interesting subgroup, called $SYTO9^{-}PI^{-}$ fraction, composed of cells not stained by either of the reagents was revealed. Performed analysis showed that there were no statistically significant differences between cell numbers in fecal microbiota suspensions prepared from each donor stool and there were no significant differences in viability of cells for each donor. Noticeable domination of alive cells was observed.

The classical culturing to show whether this technique can reveal culturable bacterial indicators for "good" versus "bad" stool donors was applied. In total, 104 species representing 36 genera were identified. Classical culturing of fresh microbiota shows that donor C, being a regular stool donor, is characterized by the largest number of cultivable species (64) obtained from his stool versus other donors (48 species for donor A and 56 for donor B). Donor C's stool had the largest number of unique species (29). The cultivable core microbiota, detected in the sample from all donors, was composed of only 16 species. In the next step, we evaluated the presence of identified species over time (throughout 10 sampling days) and we have shown that the plethora of bacterial species occurred only on individual days suggesting that single sampling can deliver non-representative and possibly false results. *Escherichia coli* was the only species detected in all samples.

An amplicon-based approach (i.e., metabarcoding combined with high-throughput taxonomical identification of bacteria) showed 97.75% of ASVs classified down to the genus level. The bacterial ASVs represented 18 classes, with Bacteroidia and Clostridia relative abundance constituting an average of 49.9% and 40.0%, respectively. At a genus level, the most dominant taxa were *Bacteroides* and *Faecalibacterium*, with relative abundance in each sample no less than 35% and 11%, respectively.

Alpha-diversity analysis showed that the Shannon index was similar for donors A and B, with its mean values equal to 10.11 and 10.02, respectively, while it was slightly, but significantly higher for the donor C–10.39 ($p = 0.0191$ for donor A vs. C and $p = 0.0005$ for donor B vs. C according to Kruskal–Wallis test, H value = 12.18).

ANCOM analysis, testing for taxa differences between donors, showed that when donors A and C were compared, the first could be characterized by *Anaeroplasmatales* and *Gastranaerophilales* orders, while the latter was characterized by an increase in abundance of *Acidaminococcus* and *Paraprevotella* genera. ANCOM analysis on donor B versus donor C pair showed an increase of *Anaeroplasma* and *Holdermanella* genera with Muribaculaceae and Puniceicoccaceae families defining donor B, and *Lachnospiraceae* and *Dialister* relative abundances significantly increased in donor C samples.

Pearson correlation coefficients between the double negative group of cells ($SYTO9^{-}PI^{-}$) and genera-level taxonomy data

showed that the relative abundance of *Anaeroplasma* is positively correlated with the double negative group per sample percentage ($\rho = 0.6312$), followed by *Sanguibacteroides* ($\rho = 0.4592$).

Beta-diversity analysis on the Bray–Curtis Dissimilarity index showed that the overall internal similarity of time-resolved samples from donor C was much higher than for other donors. Clustering of the bacterial composition of feces in donor C indicated the most stable composition of intestinal microbiota over time.

General Characteristics of Frozen Microbiota Samples

Flow Cytometry

Frozen samples reported on average of $1.89*10^{10}$ cells per ml of prepared suspension, with donor C showing the highest average number of cells, $2.5*10^{10}$ cells/ml, followed by donor B, $2.09*10^{10}$ cells/ml, and donor A with $1.08*10^{10}$ cells/ml (**Figure 2**). Statistical testing with the Kruskal–Wallis test followed by the pairwise Dunn test showed a statistically significant difference between cell counts obtained from donor A and donor C samples ($\rho = 0.0069$). No other significant result was obtained when comparing the remaining pairs that were A versus B ($\rho = 0.06$) and B versus C ($\rho = 0.36$).

Among the three major cell groups, the Unknown group was usually the dominant one, with a mean sample contribution of 47.16%, 64.08%, and 61.16% for donors A, B, and C, respectively (**Figure 2**). Contrarily, alive and dead cells constituted comparably small percentages of cells across all samples, with a mean of 15.08% for alive and 27.45% for dead groups. Interestingly, $SYTO9^{-}PI^{-}$ (cells not stained with either dye) cells, a subgroup of the Unknown, nearly always constituted a regular percentage of the Unknown group, without major differences between donors. Furthermore, we noted that donor C samples could be characterized by their regular cell group distribution in every sample (**Figure 2**).

Classical Culturing

A three-way Venn diagram was prepared to show bacterial species detected in frozen samples from each donor (**Figure 3**). In total, 69 species representing four phyla (*Actinobacteria*, *Bacteroidetes*, *Firmicutes,* and *Proteobacteria*) were found. Culturing experiments detected 44, 31, and 33 bacterial species in samples from donors A, B, and C, respectively. Core microbiota, common to all donors, was made of 12 different species. Frozen samples from donor A contained the highest number of unique bacterial species (22), while the unique number of species in donors B and C were 6 and 14, respectively. Samples from donor C reported the highest species persistence, which is the number of species detected in all samples from a given donor. There were four species detected in all samples from donor C (*Bacteroides vulgatus*, *Escherichia coli*, *Lactococcus garvieae,* and *Weissella confusa*) followed by donor B with three such strains *(Bacteroides ovatus*, *Escherichia coli,* and *Lactococcus garvieae*) and donor A with only *Escherichia coli* detected in all samples.

Isolated strains were classified as either aerobic or anaerobic based on their assigned taxonomy. The highest number of unique aerobic and anaerobic species was recovered for donor A, 27 aerobic and 17 anaerobic, followed by donor C, 20 aerobic and 13 anaerobic, and donor B with 16 aerobic and 15 anaerobic strains. Statistical analysis showed significant differences in the number of recovered aerobes between donors B and C ($\rho = 0.037$).

Next-Generation Sequencing

A total of 5,974,017 reads were obtained from Illumina MiSeq sequencing, with reads per sample ranging from 107,041 to 264,007. Quality control and merging of paired-ended reads using the *dada2* software package, resulted in the retention of, on average, 88,513.23 paired reads (SD = 16,406.09) per sample (**Supplementary Table 1**). Both the Nonpareil 3 and alpha rarefaction analysis (Qiime2 *diversity* plugin) showed sequencing depth close to 100%. Overall, 11,131 different ASVs were discovered, with feature frequency ranging from 1 to 10,539 merged sequences.

Taxonomy for each ASV was assigned using the Naive Bayes Classifier, trained on the Silva 132 database, and showed 93.38% of ASVs classified down to the genus level. Overall classification showed that 99.49% of all ASVs were bacterial, 0.34% archeal, and less than 0.17% were unclassified. The bacterial ASVs represent 18 classes, with Bacteroidia and Clostridia relative abundance constituting on average of 24.32% and 62.38%, respectively. Frozen samples were dominated by *Faecalibacterium* (4.56%–31.42%), *Bacteroides* (0.98%–26.98%), *Agathobacter* (0.18%–25.74%), and *Ruminococcus* 2 (0.55%–24.78%) genera (**Figure 4**). Differential analysis of frozen samples composition between donors highlighted several statistically different genera. ANCOM analysis showed that *Dialister* and *Coprococcus 2* genera were reported more frequently in samples from donors A and C than B. *Lachnospiraceae* NK4A136 group was more abundant in samples from donors A and B. Samples from donors B and C, in contrast to samples from donor A, reported an increased relative abundance of *Paraprevotella* genus. *Holdemanella*, *Methanobrevibacter,* and *Ruminococcaceae* UCG-004 were unique to samples from donor B, while *Peptococcus* was regularly detected only in samples from donor C.

Alpha-diversity indices analysis showed relatively lower Shannon Diversity and Pielou's Evenness indices in samples from donor C when compared with other frozen samples; however, no statistically significant result was obtained. Comparative analysis of Faith's Phylogenetic Diversity (PD) between donors showed that frozen samples from donor B had the highest average Faith's PD, with medians equal to 17.00, 24.79, and 19.12 for donors A, B, and C, respectively (**Figure 5**).

Beta-diversity analysis using classical indices, such as Bray–Curtis dissimilarity index, Jaccard index, and unweighted Unifrac, showed clear clusters, one per donor. However, the clusters were much less visible on PCoAs plotted using the weighted Unifrac distance metric, which also takes relative abundances into account (**Supplementary Figure 1**). This shows that each donor has his unique microbiota species, but they are not the dominant ones, as incorporating relative abundances blends all samples into one big cluster. This cluster is further stretched along one axis, which could mean that there is an

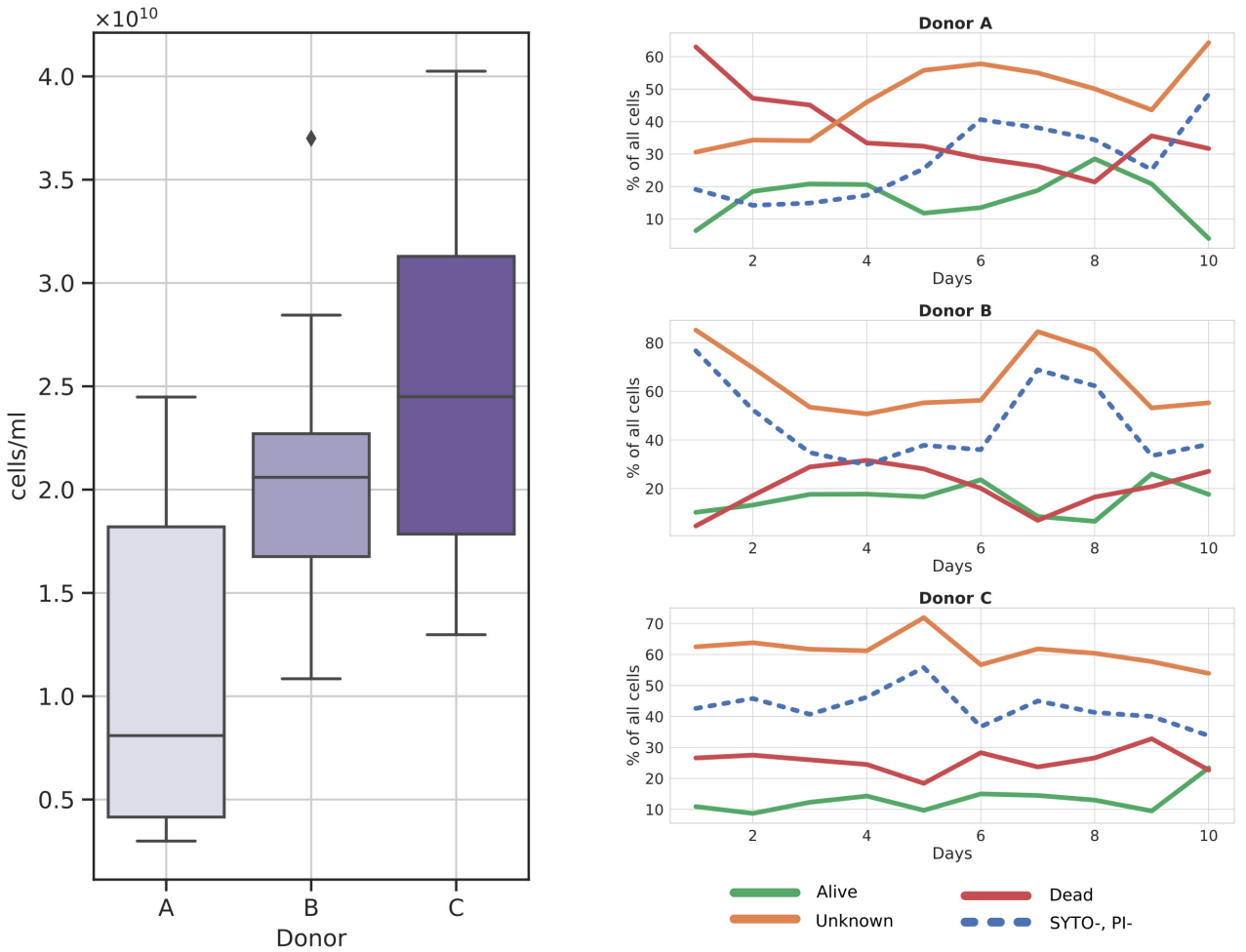


FIGURE 2 | Cytometry cell count charts in frozen samples. **(Left panel)** Total cell counts per donor over all samples; **(Right panel)** relative changes in cell fractions throughout the sampling days. SYTO-, PI- fraction is additionally dashed as it is a subgroup of the Unknown fraction.

important gradient in the relative abundance of one or more ASVs in all samples.

Comparative Analysis of Fresh and Frozen Fecal Microbiota Suspensions

Flow Cytometry

Cytometry reported an average of $2.27*10^{10}$ cells across all samples (median = $2.21*10^{10}$). When comparing fresh with frozen samples, only one pair, donor B samples, showed slightly but significantly higher cell counts in frozen suspensions when compared with the fresh ones (mean fresh cell count: $1.63*10^{11}$, mean frozen cell count: $2.08*10^{11}$) (**Figure 6**).

Stool suspension cell counts were divided into three main groups: alive, dead, and unknown (not stained with either one of the reagents). Alive or dead groups were named so, if there was no doubt that they were clustered as alive or dead. In addition, we decided to highlight one more group denoted as $SYTO9^{-}PI^{-}$. It is a subgroup of "unknown," not stained by both reagents considered as "double negative" (we think this may be a bacterial spores fraction or bacterial cells fraction with a particularly thick cell wall). A clear change in cell counts in each of those groups was detected as a result of whole stool freezing (**Figure 6**). Freezing and thawing the whole stool to prepare fecal microbiota suspensions resulted in a nearly 4-fold reduction of alive cell counts in samples from all donors. On average, 68.88% (SD = 10.85%) of cells in each fresh sample were categorized as alive, while alive cells constituted only 15.08% (SD = 6.08%) of all cells in frozen samples. The percentage of dead cells increased on average four times (fresh: 6.51%, frozen: 27.45%), while the unknown group grew by a factor of two (fresh: 24.60%, frozen: 57.47%). The $SYTO9^{-}PI^{-}$ subgroup reported 2.5-fold increase (fresh: 15.16%, frozen: 39.20%). All changes between fresh and frozen groups were statistically significant.

Samples from donor C appeared to react more consistently to freezing than samples from the other two donors. **Figure 6** shows that cell count distributions of frozen samples from donors A and B show much more variability than those from donor C. This could be explained by the more stable composition of donors'

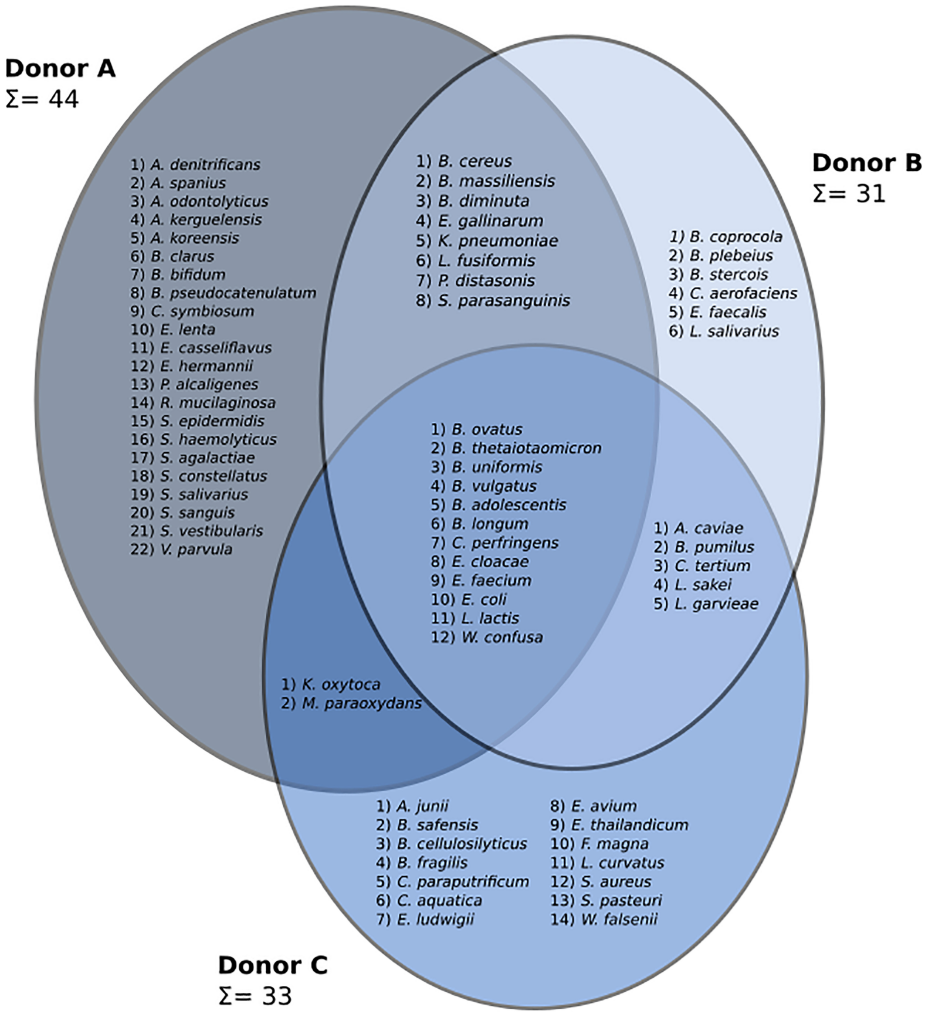


FIGURE 3 | A three-set Venn diagram showing species discovered in frozen samples from each patient. Identified genera are as follows: *Achromobacter* (*A. denitrificans*, *A. spanius*), *Acinetobacter* (*A. junii*), *Actinomyces* (*A. odontolyticus*), *Aeromonas* (*A. caviae*), *Arthrobacter* (*A. kerguelensis*, *A. koreensis*), *Bacillus* (*B. cereus*, *B. pumilus*, *B. safensis*), *Bacteroides* (*B. cellulosilyticus*, *B. clarus*, *B. coprocola*, *B. fragilis*, *B. massiliensis*, *B. ovatus*, *B. plebeius*, *B. stercois*, *B. thetaiotaomicron*, *B. uniformis*, *B. vulgatus*), *Bifidobacterium* (*B. adolescentis*, *B. bifidum*, *B. longum*, *B. pseudocatenulatum*), *Brevundimonas* (*B. diminuta*), *Clostridium* (*C. paraputrificum*, *C. perfringens*, *C. symbiosum*, *C. tertium*), *Collinsella* (*C. aerofaciens*), *Comamonas* (*C. aquatica*), *Eggerthella* (*E. lenta*), *Enterobacter* (*E. cloacae*, *E. ludwigii*), *Enterococcus* (*E. avium*, *E. casseliflavus*, *E. faecalis*, *E. faecium*, *E. gallinarum*, *E. thailandicum*), *Escherichia* (*E. coli*, *E. hermannii*), *Finegoldia* (*F. magna*), *Klebsiella* (*K. oxytoca*, *K. pneumoniae*), *Lactobacillus* (*L. curvatus*, *L. sakei*, *L. salivarius*), *Lactococcus* (*L. garvieae*, *L. lactis*), *Lysinibacillus* (*L. fusiformis*), *Microbacterium* (*M. paraoxydans*), *Parabacteroides* (*P. distasonis*), *Pseudomonas* (*P. alcaligenes*), *Rothia* (*R. mucilaginosa*), *Staphylococcus* (*S. aureus*, *S. epidermidis*, *S. haemolyticus*, *S. pasteuri*), *Streptococcus* (*S. agalactiae*, *S. constellatus*, *S. parasanguinis*, *S. salivarius*, *S. sanguinis*, *S. vestibularis*), *Veillonella* (*V. parvula*), *Wautersiella* (*W. falsenii*), and *Weissella* (*W. confusa*).

C microbiota over the sampling time, as was proved previously (Bilinski et al., 2020).

Classical Culturing

Freezing had a huge impact on the structure of cultivable bacterial communities. Overall, four different bacterial phyla were detected in samples from each donor. Variability between donors and their fresh and frozen samples is shown in **Figure 7**. When inspected on the class taxonomic level, the biggest drop in the number of cultivable species was observed for Actinobacteria and Bacilli. Donor C, a regular stool donor, reported the highest number of classes in fresh samples—12 classes. Five of those 12 classes were undetectable in suspensions from frozen samples, which were Betaproteobacteria, Erysipelotrichia, Flavobacteriia, Negativicutes, and Sphingobacteriia. Inspection of donor C's unique species, detected in both fresh and frozen samples yielded only three species: *Bacteroides cellulosilyticus*, *Enterococcus avium*, and *Enterococcus thailandicus*. Interestingly, suspensions prepared from donor A frozen samples showed two bacterial classes undetected in his fresh samples, which were Alphaproteobacteria and Betaproteobacteria.

Experiments on fresh samples detected 103 bacterial species, while after freezing, the same samples delivered only 69 species; 47 of those 69 species were also detected in the fresh

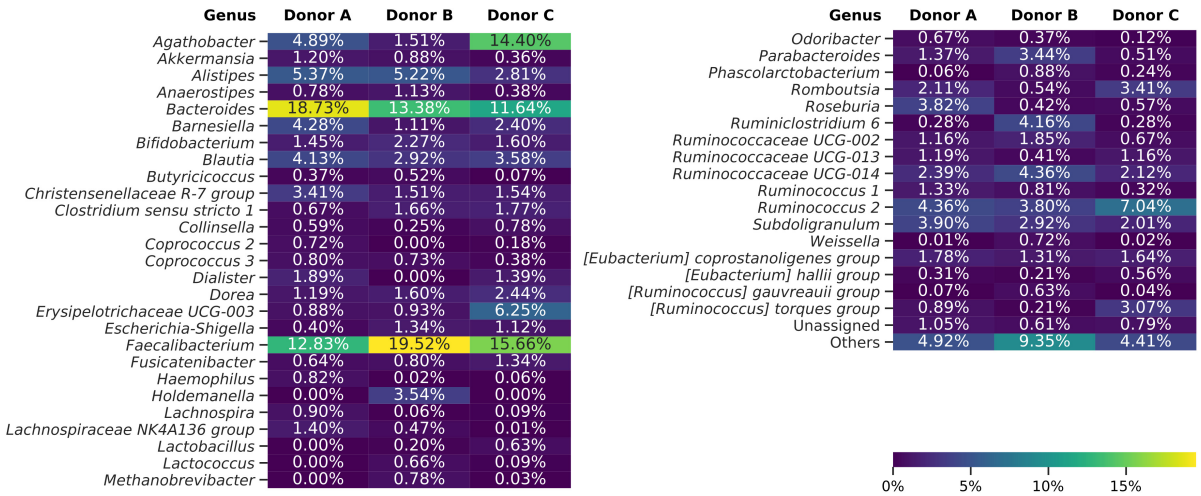


FIGURE 4 | Heat map showing bacterial genera detected using amplicon sequencing (V3–V4 region of 16S rDNA). Averaged data for each donor are presented. The "Others" group summarizes genera with individual abundances lower than 0.5% in each sample. Sequences unassigned at the genus level were grouped and named "Unassigned."

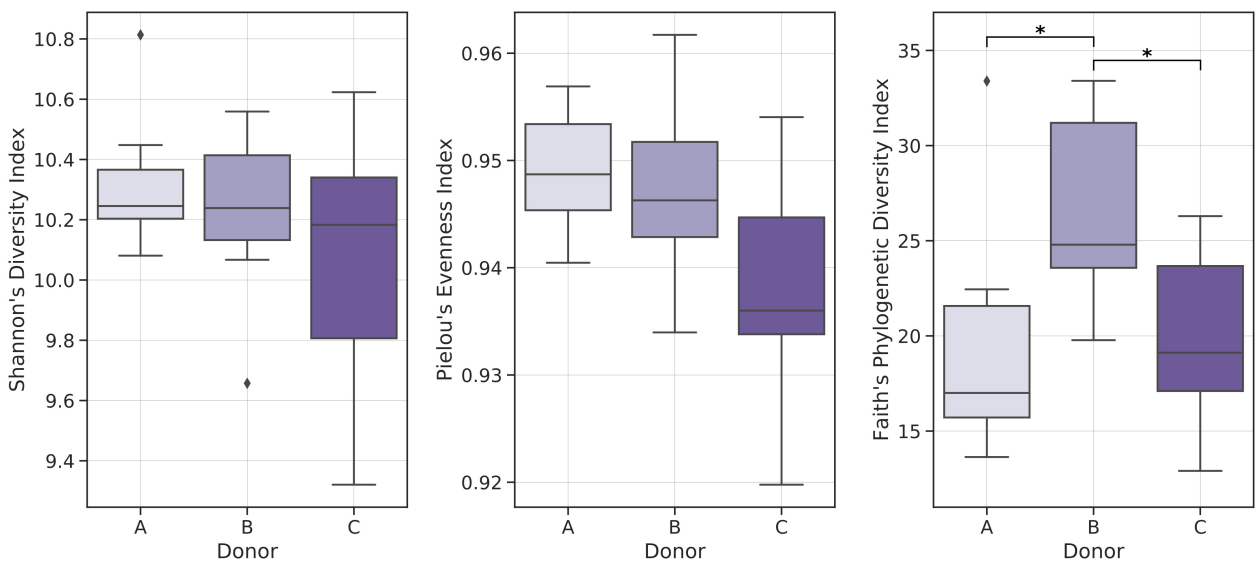


FIGURE 5 | Boxplots showing distributions of selected biodiversity indices. Kruskal–Wallis test was used to detect statistically significant differences. *$p < 0.01$.

samples, while 22 were unique to frozen samples. Donor C was characterized by the biggest drop in the number of detectable cultivable species, a drop from 63 to 33, followed by donor B, 53 to 31, and donor A, 48 to 44 (**Figure 7**). However, we observed very high variability on lower taxonomic levels such as genus and family levels. In a few cases, there were species-level shifts between fresh and frozen samples, for example, in donor A, *Clostridium tertium* present in fresh samples was not detectable in frozen samples, while frozen samples reported previously undetected *Clostridium perfringens*. In addition, it is important to notice that only 31 of 103 detected bacterial species were found in suspensions prepared from the same stool before and after freezing. This further supports our conclusion from the previous work, that although the MALDI-TOF method is very precise, it requires repetitive sampling overtime when exploring diverse bacterial communities. Interestingly, frozen samples have shown a higher number of "highly persistent" species, that is species detected in all samples from the same donor. In the fresh samples, only *Escherichia coli* has been detected in all samples, while in the frozen samples that was valid for *Bacteroides ovatus*, *Bacteroides vulgatus*, *Escherichia coli*, *Lactococcus garvieae,* and *Weissella confusa*.

Next-Generation Sequencing

A total of 60 amplicon libraries, 30 prepared from fresh and 30 from frozen suspensions, was included in this analysis. As all

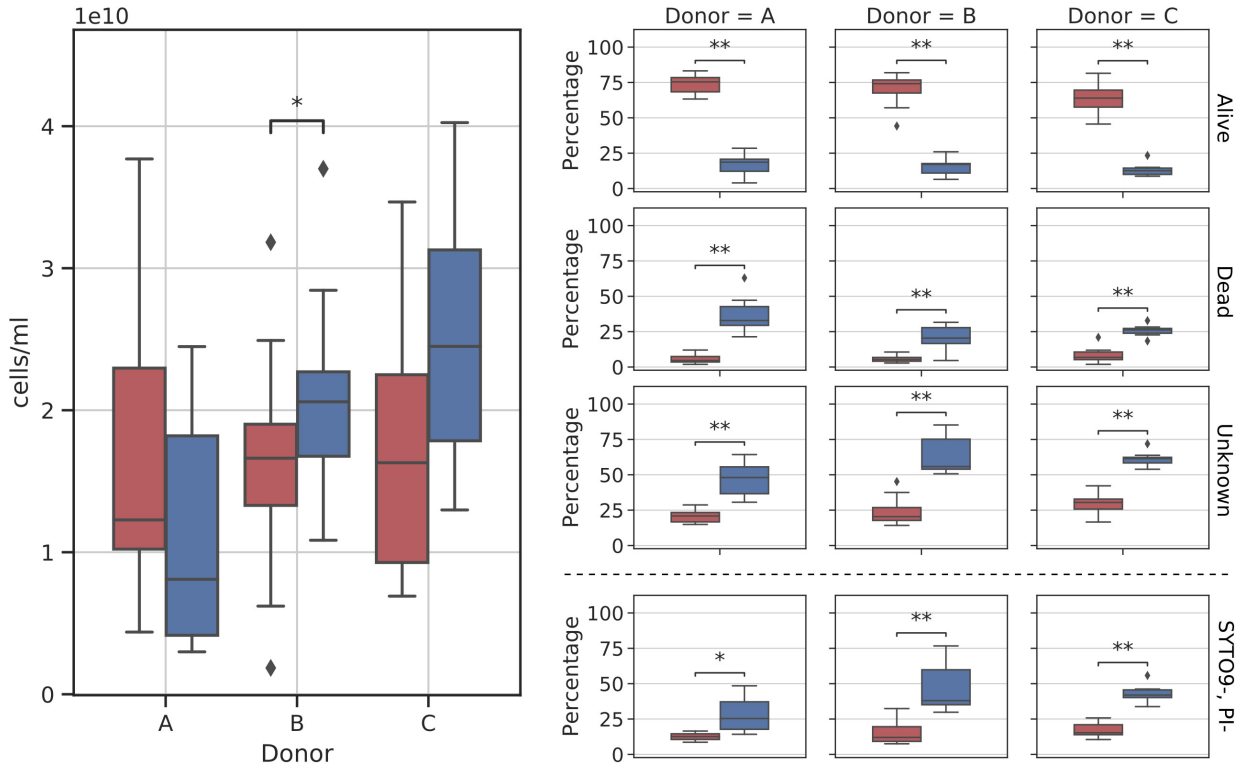


FIGURE 6 | Diagram comparing flow cytometry cell counts and group fractions obtained for fresh and frozen suspensions. The **(left panel)** shows the total donor cell counts comparison. The **(right panel)** shows changes in each of the inspected flow cytometry groups (alive, dead, unknown, and SYTO9$^-$, PI$^-$ subgroups). Red—fresh samples; blue—frozen samples; * – $p < 0.05$; ** – $p < 0.01$.

libraries were prepared simultaneously, there were no significant differences between the number of obtained sequencing reads between fresh and frozen samples. For the needs of comparative analysis, 49,706 sequences were sampled from each sample. Alpha-rarefaction analysis with observed ASV counts and Good's coverage showed that obtained subsamples accurately represented their original counterparts.

Alpha-diversity analysis was conducted using the Shannon Diversity index, Pielou's Evenness index, and Faith's Phylogenetic Diversity index. In most cases, selected biodiversity indices were slightly lower for frozen samples, but no statistically significant differences were detected when comparing fresh and frozen samples from the same donor. There were also no significant differences in comparisons of selected biodiversity indices between all fresh and all frozen samples.

Beta-diversity analyses using PCoA visualizations on Bray–Curtis (quantitative and qualitative) index and Jaccard (only qualitative) index showed clear, donor-wise clusters, with fresh and frozen samples clustering by the donor. This shows that although freezing introduced some substantial differences in community composition, those differences did not overcome intrinsic inter-donor characteristics. Interestingly, PCoA visualization built using a weighted UniFrac (phylogenetic and qualitative) index showed no donor-wise clusters, but a clear split between fresh and frozen samples (**Figure 8**). This split can be in part attributed to the changes in the relative abundance of Bacteroidales and Clostridiales orders. While Bacteroidales are more abundant in fresh samples, Clostridiales appear to be more abundant in some of donor A and C frozen samples, as donor B frozen samples are placed mainly in the lower right part of the diagram, and thus being uninfluenced or even negatively influenced by this trend.

Sequences obtained from fresh and frozen samples were classified using the same methodology, with Silva 132 as a reference database. **Figure 9** shows mean and per donor differences in relative abundances between fresh and frozen amplicons at the genus taxonomic level. *Bacteroides* genus was the main source of differences between investigated groups. We observed that *Bacteroides* were more abundant in fresh than frozen samples by an average of 21.51 percentage points. Members of Clostridiales class, which was shown to be characteristic of some of the frozen samples, were much more dispersed, with more than 50 genera contributing to this trend. The biggest contributors to this trend were *Faecalibacterium* (more abundant in frozen samples by 4.89 percentage points), *Ruminococcus 2* group (more abundant in frozen samples by 4.09 percentage points), and *Agathobacter* (more abundant in frozen samples by 3.16 percentage points). However, this trend was not uniform for all donors, as in the case of donor B, members of *Agathobacter* were slightly, by 0.76 percentage points, more abundant in fresh compared to frozen samples.

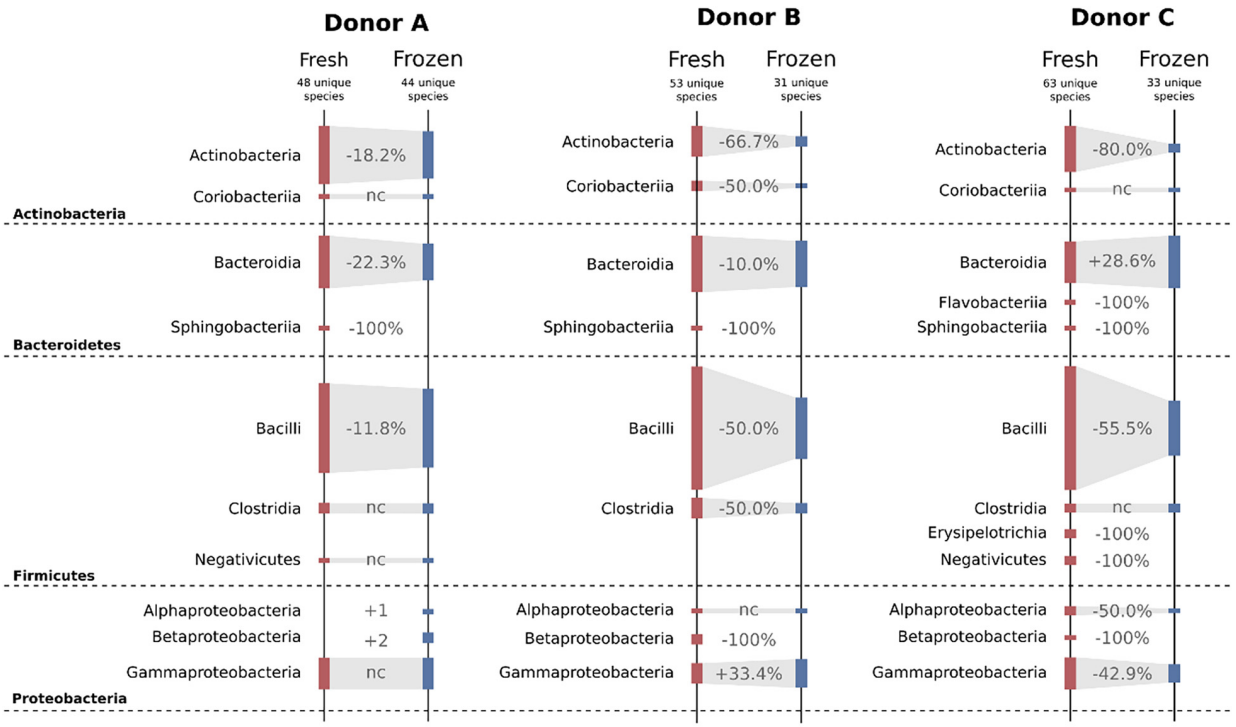


FIGURE 7 | Diagram showing changes in cultivable bacteria between suspensions prepared from fresh and frozen stools. Dashed horizontal lines separate bacterial phyla. Values between bars show changes in the number of detected species. nc – no changes.

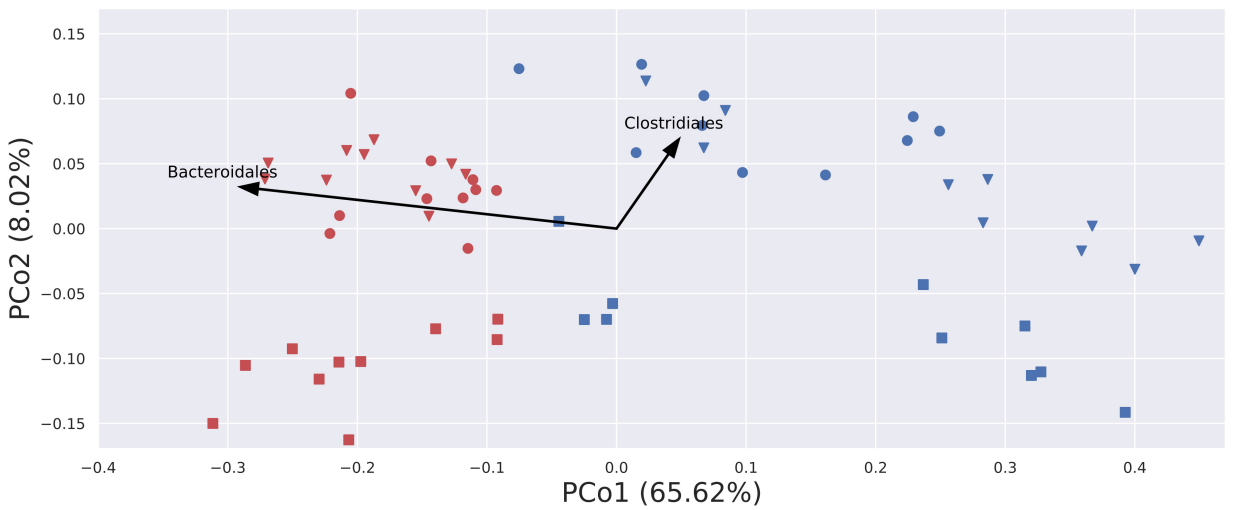


FIGURE 8 | Visualization of PCoA analysis on weighted UniFrac index. Samples are stretched along the x-axis, with a clear division for fresh (red) and blue (frozen) samples. Arrows describe a general trend in the relative abundance of given bacterial order. Only two arrows were plotted, as arrows for other bacterial orders would not be visible due to their short length. Blue – amplicons prepared from frozen stool; red – amplicons prepared from fresh stool; circle – donor A; square – donor B; triangle – donor C.

To statistically test observed differences, we employed ANCOM analysis. ANCOM analysis highlighted eight bacterial classes to differ between fresh and frozen samples. Members of Bacteroidia, Clostridia, Deltaproteobacteria, and Lentisphaeria were more abundant in fresh samples, while members of Actinobacteria, Bacilli, Coriobacteriia, and Erysipelotrichia were more abundant in frozen samples. This result is exceptionally interesting when interpreted with results obtained from a culture-based approach. Culturing showed that the number of bacterial species from Bacilli and Actinobacteria classes was considerably

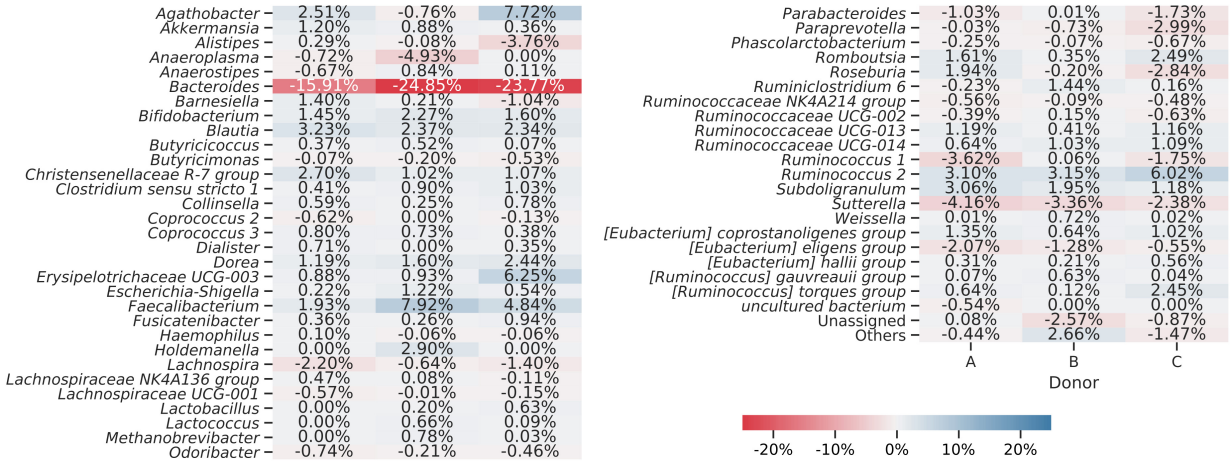


FIGURE 9 | Heat map showing differences in bacterial genera detected from amplicons prepared from the fresh and frozen stool. Averaged relative abundances from fresh samples were subtracted from results obtained for frozen samples, so negative values show that a certain genus was more abundant in fresh samples and positive values show that a genus was more abundant in frozen samples. The "Others" group summarizes genera with individual abundances lower than 0.5% in every sample. Sequences unassigned at the genus level were grouped and named "Unassigned."

lower in frozen than fresh samples, while NGS analysis showed that their relative abundances were higher in frozen samples. Furthermore, culturing showed similar numbers of species of Bacteroidia class in fresh and frozen samples, while ANCOM analysis showed that Bacteroidia are significantly more abundant in fresh samples. This could mean that in the first case, whole stool freezing, in general, lowers species diversity in Bacilli and Actinobacteria classes. In the second case, we observed a rapid drop in the relative abundance of Bacteroidia, but their species diversity stayed at the same level.

Interestingly, ANCOM analysis on genus level highlighted only two genera—*Sutterella* and *[Eubacterium] eligens* group (member of Lachnospiraceae family), both significantly more abundant in fresh than frozen samples. Although these two taxa constituted no more than 8% of a given sample, due to freezing their relative abundances dropped by the order of 100, and in some samples, they became undetectable.

ANCOM comparisons between fresh and frozen samples from the same donor also showed *Suterella* to be significantly more abundant in fresh samples in each pair. In addition, donors A and B showed two more and donor C one more statistically significant difference at the genus level. In donor A, an "uncultured" genus of Eggerthellaceae family was more abundant in frozen samples. Moreover, ANCOM analysis of donor A also highlighted *Lachnospiraceae ND3007* group to be more abundant in fresh than frozen samples. In donor B, the analysis showed *Anaeroplasma* and an "uncultured" member of *Puniceicoccaceae* family to be more abundant in fresh compared to frozen samples, while comparisons on donor C amplicons highlighted *Lachnoclostridium* genus to be more abundant in fresh samples.

Correlation analysis detected several possible links between NGS and flow cytometry results, for suspensions prepared from both fresh and frozen stools (**Supplementary Table 2**). Our results show that the biggest number of correlations can be observed for donor B samples and the SYTO9$^-$, PI$^-$ flow cytometry subfraction. *Anaeroplasma* genus was positively correlated (correlation = 0.9377) with SYTO9$^-$, PI$^-$ subfraction, and *Holdemanella* was negatively correlated (correlation = –0.6982) with the same subfraction in donor B fresh samples.

DISCUSSION

Recent reports show that FMT interventions using frozen material deliver similar results to interventions with fresh material only (Burz et al., 2019). In the course of this work, we decided to further elucidate the impact of whole stool freezing on the FMT-ready suspensions microbial community. Our results clearly show that whole stool freezing without any cryopreserving buffer significantly alters stool microbiota. This change can be detected using either of the methods we employed, that is, flow cytometry with cell staining, classical culturing, and metabarcoding.

Flow cytometry with cell staining showed that freezing causes significant changes in all of the observed fractions. Alive cell counts dropped four times, from around 70% to 15%, while the other two fractions, dead and unknown, cell counts quadrupled and doubled, with the Unknown fraction becoming the dominant one, with an average contribution of 57.47% per sample. These changes were observed for all three donors and we did not detect any significant differences between them. However, it is important to notice that if we were to apply flow cytometry without cell staining (discriminating dead and alive cells), we would observe no differences between fresh and frozen samples. This observation additionally highlights the fact that here NGS results may be significantly biased as being based mainly on DNA from the dead cells, and therefore the differences in actual alive, active community composition may be much bigger if cell sorting for alive and dead fraction was used before DNA

isolation. Nevertheless, FMT with frozen material still appears to deliver satisfactory results, as the whole stool freezing approach is routinely used in one of the polish stool banks with an overall CDI cure rate approx. 90% (Grzesiowski et al., 2015). We also consider it very important to note that the lack of characterization of cells from the Unknown group, and especially from the group of "double negative" cells, does not allow us to clearly state whether the viability of bacterial cells has actually decreased, whether it has not changed or even increased during freezing. This has not been the subject of this experiment. However, the next step must be to characterize this population after FACS cell sorting and then plating to see if these cells start to proliferate and what species they are. It may turn out that these "double negative cells" are simply bacterial spores that will start to proliferate when they are returned to favorable incubation conditions, which would mean that, after freezing, the viability of the preparations could even increase. It may also be a shift toward cells with a thick cell wall, which may be suggested by the fact that the Unknown fraction correlates, for example, with bacteria of the *Anaeroplasma* genus, known as bacteria with a thick cell wall.

Classical culturing showed a significant drop in the number of detected bacterial species as a result of freezing. Among four detected bacterial phyla, Bacteroidetes and Bacilli (of Firmicutes) showed the highest decrease in the number of observed cultivable bacterial species (**Figure 7**). As in the case of flow cytometry, the changes were similar between all donors; however, donor C (the regular stool donor) showed the highest drop in the number of unique cultivable species, with 63 species detected in fresh and only 31 in frozen samples. Applied sampling methodology, 10 stools from each donor over 10 days, also allowed us to notice a very high variability in the number of detected bacterial species. That led to the conclusion that time-resolved sampling is crucial to obtain replicable results if only culturing is employed for stool quality assessment. It is also very interesting when combining culture data with flow cytometry and NGS. The biggest drop of culturable cells in donor C, with the information of mostly sporulating bacteria as microbiota components in this donor and a significant increase in the "unknown" fraction in flow cytometry, may lead us to the conclusion that freezing may induce massive spore-forming in those sporulating bacteria. However, this needs further investigation.

Metabarcoding was the method that delivered the most comprehensive results; however, it must be noted that we did not perform any cell fractioning before DNA isolation, and as a result, all DNA available in the samples, both from alive and dead cells, was jointly sequenced. Inspection of basic alpha-diversity indices showed no significant differences between fresh and frozen samples from most donors. We observed a decrease in the relative abundance of bacteria from Bacteroidales order, while the *Faecalibacterium* genus relative abundances increased in FMT suspensions after freezing. PCoA visualization of weighted UniFrac distance metric showed clear separation of fresh and frozen samples, with Bacteroidales being more abundant in fresh than frozen samples (**Figure 8**). ANCOM analysis of NGS results highlighted several taxa with statistically significant differences between investigated groups. Members of Bacteroidia, Clostridia, Deltaproteobacteria, and Lentisphaeria were more abundant in fresh samples, while members of Actinobacteria, Bacilli, Coriobacteriia, and Erysipelotrichia were more abundant in frozen samples. Surprisingly, ANCOM analysis on genus level highlighted only two genera—*Sutterella* and *[Eubacterium] eligens* group (member of Lachnospiraceae family), both significantly more abundant in fresh than frozen samples.

We also investigated the possibility that whole stool freezing without any pre-processing could promote the preservation of anaerobic bacteria. Culturing experiments showed no statistically significant differences in this matter. We were not able to validate this hypothesis on NGS data, as there are no databases that would allow us to properly classify obtained ASVs as either aerobic or anaerobic.

In addition, we hypothesized that the obtained results would allow us to highlight some of the key characteristics of "a good donor." We based the correlations with donor C samples (regular stool donor). Interestingly, freezing had the biggest impact on donor C samples. Classical cultivation experiments showed the biggest decrease in the number of cultivable bacterial species in this case, as 5 out of 12 bacterial classes identified in fresh samples were undetectable after freezing. This decrease in diversity is also visible upon inspection of Shannon's Diversity and Faith's Phylogenetic Diversity indices. While donor C's fresh samples showed the highest mean values in these indices, biodiversity in suspensions prepared from the frozen stool samples of donor C was lower than for other donors. Cultivation experiments showed three bacterial species unique only to donor C, which were *Bacteroides cellulosilyticus*, *Enterococcus avium,* and *Enterococcus thailandicus*. All three are common human gut microbiota, with the first one characterized by its versatile metabolic potential, especially in carbohydrate hydrolysis, and its high activity in heparinase I (Ali-Ahmad et al., 2017). *Enterococcus avium*, while more often found in birds than human microflora, is considered a human pathogen due to its wide array of virulence factors (Yu et al., 2019). Not much is known about *Enterococcus thailandicus*, but its presence only in donor C is exceptionally interesting, as a recent report by Li et al. (2021) described *E. thailandicus* d5b strain with very potent antimicrobial activity against *C. difficile* (Li et al., 2021). These final observations raise several novel questions that will be used as a foundation for our further research.

DATA AVAILABILITY STATEMENT

The original contributions presented in this study are publicly available. This data can be found here: http://www.ebi.ac.uk/, PRJEB36368.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and

institutional requirements. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

JB, MD, ŁD, and GB designed the study, performed all analyses, and wrote the article. JB, MD, PG, EP, AS-E, TD, KL, MT, and KO performed the experiments and commented on the previous version of the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was funded by "Incubator of Innovation + 1.0" program financed by the Polish Ministry of Science and Higher Education and financed under EU funds (SG OP).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.872735/full#supplementary-material

REFERENCES

Ali-Ahmad, A., Garron, M.-L., Zamboni, V., Lenfant, N., Nurizzo, D., Henrissat, B., et al. (2017). Structural insights into a family 39 glycoside hydrolase from the gut symbiont *Bacteroides* cellulosilyticus WH2. *J. Struct. Biol.* 197, 227–235. doi: 10.1016/j.jsb.2016.11.004

Bilinski, J., Dziurzynski, M., Grzesiowski, P., Podsiadly, E., Stelmaszczyk-Emmel, A., Dzieciatkowski, T., et al. (2020). Multimodal approach to assessment of Fecal Microbiota donors based on three complementary methods. *J. Clin. Med.* 9:2036. doi: 10.3390/jcm9072036

Bilinski, J., Grzesiowski, P., Sorensen, N., Madry, K., Muszynski, J., Robak, K., et al. (2017). Fecal Microbiota transplantation in patients with blood disorders inhibits gut colonization with antibiotic-resistant bacteria: results of a prospective. Single-Center Study. *Clin. Infect. Dis.* 65, 364–370. doi: 10.1093/cid/cix252

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.

Burz, S. D., Abraham, A.-L., Fonseca, F., David, O., Chapron, A., Béguet-Crespel, F., et al. (2019). A guide for *ex vivo* handling and storage of stool samples intended for fecal microbiota transplantation. *Sci. Rep.* 9:8897. doi: 10.1038/s41598-019-45173-4

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

Cammarota, G., Masucci, L., Ianiro, G., Bibbò, S., Dinoi, G., Costamagna, G., et al. (2015). Randomised clinical trial: faecal microbiota transplantation by colonoscopy vs. vancomycin for the treatment of recurrent *Clostridium difficile* infection. *Aliment. Pharmacol. Ther.* 41, 835–843. doi: 10.1111/apt.13144

De Weirdt, R., Possemiers, S., Vermeulen, G., Moerdijk-Poortvliet, T. C. W., Boschker, H. T. S., Verstraete, W., et al. (2010). Human faecal microbiota display variable patterns of glycerol metabolism. *FEMS Microbiol. Ecol.* 74, 601–611. doi: 10.1111/j.1574-6941.2010.00974.x

DuPont, H. L., Jiang, Z.-D., DuPont, A. W., and Utay, N. S. (2020). Abnormal intestinal microbiome in medical disorders and potential reversibility by Fecal Microbiota Transplantation. *Dig. Dis. Sci.* 65, 741–756. doi: 10.1007/s10620-020-06102-y

Grzesiowski, P., Hermann, A., Dubaniewicz, A., Kasprzyk, J., Pawlik, D., and Zak-Pulawska, Z. (2015). Effectiveness of FMT in recurrent *Clostridium difficile* infection. *Antimicrob. Resist. Infect. Control* 4:P27.

Gustafsson, A., Berstad, A., Lund-Tønnesen, S., Midtvedt, T., and Norin, E. (1999). The effect of faecal enema on five microflora-associated characteristics in patients with antibiotic-associated diarrhoea. *Scand. J. Gastroenterol.* 34, 580–586. doi: 10.1080/003655299750026038

Hamilton, M. J., Weingarden, A. R., Sadowsky, M. J., and Khoruts, A. (2012). Standardized frozen preparation for transplantation of fecal microbiota for recurrent *Clostridium difficile* infection. *Am. J. Gastroenterol.* 107, 761–767. doi: 10.1038/ajg.2011.482

Herlemann, D. P., Labrenz, M., Jurgens, K., Bertilsson, S., Waniek, J. J., and Andersson, A. F. (2011). Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea. *ISME J.* 5, 1571–1579. doi: 10.1038/ismej.2011.41

Hunter, J. D. (2007). Matplotlib: a 2D graphics environment. *Comput. Sci. Eng.* 9, 90–95.

Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010

Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M., et al. (2013). Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res.* 41:e1. doi: 10.1093/nar/gks808

Lee, C. H., Steiner, T., Petrof, E. O., Smieja, M., Roscoe, D., Nematallah, A., et al. (2016). Frozen vs Fresh Fecal Microbiota transplantation and clinical resolution of Diarrhea in patients with recurrent *Clostridium difficile* infection: a Randomized Clinical Trial. *JAMA* 315, 142–149. doi: 10.1001/jama.2015.18098

Li, T., Lyu, L., Zhang, Y., Dong, K., Li, Q., Guo, X., et al. (2021). A newly isolated E. thailandicus strain d5B with exclusively antimicrobial activity against C. difficile might be a novel therapy for controlling CDI. *Genomics* 113, 475–483. doi: 10.1016/j.ygeno.2020.09.032

Mandal, S., Van Treuren, W., White, R. A., Eggesbo, M., Knight, R., and Peddada, S. D. (2015). Analysis of composition of microbiomes: a novel method for studying microbial composition. *Microb. Ecol. Health Dis.* 26:27663. doi: 10.3402/mehd.v26.27663

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

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., et al. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* 41, D590–D596. doi: 10.1093/nar/gks1219

Ramai, D., Zakhia, K., Fields, P. J., Ofosu, A., Patel, G., Shahnazarian, V., et al. (2021). Fecal Microbiota Transplantation (FMT) with colonoscopy is superior to enema and nasogastric tube while comparable to capsule for the treatment of recurrent *Clostridioides difficile* infection: a systematic review and meta-analysis. *Dig. Dis. Sci.* 66, 369–380. doi: 10.1007/s10620-020-06185-7

Rokkas, T., Gisbert, J. P., Gasbarrini, A., Hold, G. L., Tilg, H., Malfertheiner, P., et al. (2019). A network meta-analysis of randomized controlled trials exploring the role of fecal microbiota transplantation in recurrent *Clostridium difficile* infection. *United Eur. Gastroenterol. J.* 7, 1051–1063. doi: 10.1177/2050640619854587

Satokari, R., Mattila, E., Kainulainen, V., and Arkkila, P. E. T. (2015). Simple faecal preparation and efficacy of frozen inoculum in faecal microbiota transplantation for recurrent *Clostridium difficile* infection–an observational cohort study. *Aliment. Pharmacol. Ther.* 41, 46–53. doi: 10.1111/apt.13009

Staley, C., Hamilton, M. J., Vaughn, B. P., Graiziger, C. T., Newman, K. M., Kabage, A. J., et al. (2017). Successful Resolution of Recurrent *Clostridium difficile* Infection using Freeze-Dried, Encapsulated Fecal Microbiota; Pragmatic Cohort Study. *Am. J. Gastroenterol.* 112, 940–947. doi: 10.1038/ajg.2017.6

Vandeputte, D., Tito, R. Y., Vanleeuwen, R., Falony, G., and Raes, J. (2017). Practical considerations for large-scale gut microbiome studies. *FEMS Microbiol. Rev.* 41, S154–S167. doi: 10.1093/femsre/fux027

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., et al. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* 17, 261–272.

Youngster, I., Sauk, J., Pindar, C., Wilson, R. G., Kaplan, J. L., Smith, M. B., et al. (2014). Fecal microbiota transplant for relapsing *Clostridium difficile* infection using a frozen inoculum from unrelated donors: a randomized, open-label, controlled pilot study. *Clin. Infect. Dis.* 58, 1515–1522. doi: 10.1093/cid/ciu135

Yu, T., Li, L., Zhao, Q., Wang, P., and Zuo, X. (2019). Complete genome sequence of bile-isolated *Enterococcus avium* strain 352. *Gut Pathog.* 11:16. doi: 10.1186/s13099-019-0294-9

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: 13 July 2022
doi: 10.3389/fmicb.2022.932795

Isolation and Characterization of Human Intestinal Bacteria *Cytobacillus oceanisediminis* NB2 for Probiotic Potential

Monika Yadav[1], *Tarun Kumar*[1], *Akshay Kanakan*[2], *Ranjeet Maurya*[2,3], *Rajesh Pandey*[2,3] *and Nar Singh Chauhan*[1]*

[1] Department of Biochemistry, Maharshi Dayanand University, Rohtak, India, [2] Integrative GENomics of Host-PathogEn (INGEN-HOPE) Laboratory, Council of Scientific and Industrial Research-Institute of Genomics and Integrative Biology (CSIR-IGIB), New Delhi, India, [3] Academy of Scientific and Innovative Research (AcSIR), Ghaziabad, India

Systemic characterization of the human gut microbiota highlighted its vast therapeutic potential. Despite having enormous potential, the non-availability of their culture representatives created a bottleneck to understand the concept of microbiome-based therapeutics. The present study is aimed to isolate and evaluate the probiotic potential of a human gut isolate. Physiochemical, morphological, and phylogenetic characterization of a human gut isolate identifies it as a rod-shaped gram-negative microbe taxonomically affiliated with the *Cytobacillus* genus, having an optimal growth at 37°C in a partially alkaline environment (pH 8.0). This human gut isolate showed continuous growth in the presence of salts (up to 7% NaCl and 10% KCl), antibiotics, metals and metalloids [silver nitrate (up to 2 mM); lead acetate (up to 2 mM); sodium arsenate (up to 10 mM); potassium dichromate (up to 2 mM)], gastric and intestinal conditions, diverse temperature (25–50°C), and pH (5–9) conditions making it fit to survive in the highly variable gut environment. Genomic characterization identified the presence of gene clusters for diverse bio-catalytic activity, stress response, and antimicrobial activity, as well as it indicated the absence of pathogenic gene islands. A combination of functional features like anti-amylase, anti-lipase, glutenase, prolyl endopeptidase, lactase, bile salt hydrolase, cholesterol oxidase, and anti-pathogenic activity is indicative of its probiotic potential in various disorders. This was further substantiated by the CaCo-2 cell line assay confirming its cellular adherence and biosafety. Conclusively, human gut isolate possessed significant probiotic potential that can be used to promote animal and human health.

Keywords: probiotic, human gut microbe, microbiome therapeutics, microbial characterization, microbial isolation

OPEN ACCESS

Edited by:
Sanjay Kumar Singh Patel,
Konkuk University, South Korea

Reviewed by:
Asiya Nazir,
Abu Dhabi University,
United Arab Emirates
Prabhanshu Tripathi,
Indian Institute of Toxicology Research (CSIR), India
Abhay Bajaj,
National Environmental Engineering Research Institute (CSIR), India

***Correspondence:**
Nar Singh Chauhan
nschauhan@mdurohtak.ac.in

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: *13 July 2022*

Citation:
Yadav M, Kumar T, Kanakan A, Maurya R, Pandey R and Chauhan NS (2022) Isolation and Characterization of Human Intestinal Bacteria Cytobacillus oceanisediminis NB2 for Probiotic Potential. Front. Microbiol. 13:932795. doi: 10.3389/fmicb.2022.932795

INTRODUCTION

Human gut microbes play an important role in the maintenance of human health through active participation in host metabolism, immunity, gut homeostasis, and pathogen eradication (Yadav et al., 2018; Yadav and Chauhan, 2022). Gut microbes are being characterized for their therapeutic potential to treat human disorders (Thaiss and Elinav, 2017). *Christensenella* sp. is shown to reduce depression and anxiety-like behavior (Verma et al., 2020). *Akkermansia*

muciniphila augments relief to the host from metabolic disorders (Kalia et al., 2022a), as well as protects against atherosclerosis by reducing gut permeability and preventing inflammation (Li et al., 2016). *Lactobacillus johnsonii* protects the host against the onset of cancer (Marcial et al., 2017). *Bifidobacterium longum* reduces the severity of Crohn's disease (Yao et al., 2021) and repairs the mucus layer integrity impaired due to a high-fat diet (Schroeder et al., 2018). *Oxalibacterium formigenes* prevent kidney stones by ensuring oxalic acid breakdown (Jalanka-Tuovinen et al., 2011). *Bacteroides* sp. protects against adiposity (Walker and Parkhill, 2013). *Lactobacillus johnsonii, Akkermansia muciniphila, Bifidobacterium longum, Bacteroides* sp., *Roseburia intestinalis, Faecalibacterium prausnitzii,* and *Bacillus* sp. are characterized for certain health-promoting benefits such as anti-cancer, anti-diabetic, anti-obesity, anti-pathogenic, as well as cholesterol-removing properties (Aswathy et al., 2008). These studies highlighted the scope of harnessing the potential of gut microbes as probiotic strains in disease therapeutics (Yadav and Chauhan, 2022). The impact of probiotics on human as well as animal health has promoted their use as food additives even on a commercial scale (Cuello-Garcia et al., 2015; Varankovich et al., 2015). These strains are also being used as food additives to improve the health of poultry animals for disease prevention and increased meat production (Kalia et al., 2022b). Several strains of *Lactobacillus*, *Bifidobacterium,* and *Bacillus* have also been used as potential probiotics (Lee et al., 2019). *B. cereus, B. clausii, B. coagulans, B. licheniformis, B. polyfermenticus, B. pumilus,* and *B. subtilis* are well-characterized commercial probiotic strains (Lee et al., 2019). Despite the enormous therapeutic potential, the majority of human gut microbes could not be exploited for their probiotic potential, which is attributed to the lack of their cultured representative. Scientific explorations are required to culture human gut microbes to harness their probiotic potential. The current study was designed to culture a human gut bacterium and characterize it for its probiotic potential. The probiotic potential of this microbe can be efficiently used to improve animal as well as human health.

METHODS

Sample Collection and Ethical Statement

Bacterial isolate strain NB2 was cultured from a fecal sample collected from a healthy individual (age 28 years, female, BP 120/80, blood sugar 100–120 mg/dl, BMI 26.4, with no symptoms of any illness). A total of 100 mg of fecal sample was homogenized and serially diluted (10^{-1}-10^{-5}) in phosphate buffer saline (pH 7.4). About 100 μl of each serial dilution was plated on a nutrient agar medium plate. The culture plates were incubated at 37°C till the appearance of microbial colonies. Sub-culturing of microbial colonies was performed in LB (Luria Bertini) agar medium at 37°C. The study was conducted after receiving ethical clearance from the Human ethical committee at M. D. University, Rohtak Haryana, India. Strict human ethical guidelines were followed, and written consent was sought from the individual included in this study.

Molecular and Phenotypic Characterization

Gram staining of bacterial isolate strain NB2 was performed using a commercially available gram-staining kit (Himedia, K001-1KT). The morphology was observed using a compound microscope. The microbial growth pattern was analyzed after continuously culturing the bacterial isolate strain NB2 (0.01 at OD_{600nm}) in LB broth for 24 h at 37°C. Taxonomic affiliation of human gut isolate NB2 was performed using 16S rRNA gene analysis (Kumar Mondal et al., 2017). The substrate preference of bacterial isolate strain NB2 was checked using the Hi-Carbo kit (Himedia, KB009A-1KT, KB009B1-1KT, KB009C-1KT) at 37°C for 24 h following the manufacturer's instructions.

Genome Characterization of Bacterial Isolate Strain NB2

The genomic DNA of bacterial isolate strain NB2 was sequenced on Illumina MiSeq using Nextera XT DNA Library Prep kit following the manufacturers' protocol (https://sapac.illumina.com/content/dam/illumina-marketing/documents/products/datasheets/datasheet_nextera_xt_dna_sample_prep.pdf). Raw reads were quality checked using FASTQC v0.11.9 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc) and fastQ Validator v0.1.1 (https://github.com/statgen/fastQValidator). Removal of contaminated reads was performed to get the error corrected reads. The SPAdes v3.15.1 assembler was used for the *de novo* assembly which uses an automatic k-mer optimization approach and is thereby a good tool for bacterial genome assembly. It uses Bayes Hammer to perform read error correction on each data set and Mismatch Corrector, a post-processing tool, to reduce the number of mismatches in assembly using the BWA-0.7.17 tool. Further, the BUSCO v5.0.0 assessment tool was used with the latest bacterial orthologous catalog (bacteria_odb10) for analyzing the completeness of a set of predicted genes in bacterial genome assemblies (https://busco.ezlab.org/). The sequenced genome was compared with the reference genomes of the *Bacillus* species (**Supplementary Table S1**) to assess the evolutionary and phylogenetic relationships of the sequenced bacterial isolate strain NB2 with sequenced *Bacillus* genomes. The phylogenomic relationship of the bacterial isolate strain NB2 was assessed with other *Bacillus* genomes using M1CR0B1AL1Z3R webserver (https://microbializer.tau.ac.il/). The average nucleotide identity and tetra-correlation values were calculated using J-species software (http://jspecies.ribohost.com/jspeciesws/). The assembled genome was annotated with PROKKA-v1.12 annotation pipeline (Seemann, 2014). The pathogenic islands within the sequenced genome were detected with the Island Viewer 4 (https://www.pathogenomics.sfu.ca/islandviewer/resources/) following default parameters (https://www.pathogenomics.sfu.ca/islandviewer/about/). The antibiotic resistance genes were identified using the comprehensive antibiotic resistance database (CARD) (https://card.mcmaster.ca/) and ResFinder-4.1 server (https://cge.food.dtu.dk/services/ResFinder/). The dbCANmeta server (https://bcb.unl.edu/dbCAN2/blast.php) was used to identify CAZymesin, the sequenced genome.

Biosafety Assessments of the Bacterial Isolate Strain NB2

Hemolytic activity of the bacterial isolate strain NB2 bacterial culture was assessed using the blood agar plate (5% v/v) (Barik et al., 2021). Cellular toxicity of bacterial isolate strain NB2 was assessed against Caco-2 cell lines (Dowdell et al., 2020).

Stress Resistance Physiology

Growth of the bacterial isolate strain NB2 was assessed in gastric (pH 2.0; pepsin for 2 h) and intestinal (pH 8.0; trypsin for 6 h) conditions (AlKalbani et al., 2019). NB2 growth was also observed in the presence of the bile salts (Nami et al., 2019). The growth pattern of human gut isolate NB2 was observed in presence of salts (NaCl and KCl) and metal/metalloids [silver nitrate (0.1–2 mM), cadmium chloride (0.1–2 mM), lead acetate (0.1–2 mM), potassium dichromate (0.1–2 mM), and sodium arsenate (0–50 mM); **Supplementary Table S2**]. The growth pattern of the bacterial isolate strain NB2 was continuously assessed with an interval of 2 h after growing active microbial culture [0.05 OD (600 nm)] for 24 h in LB broth supplemented with a respective stressor. Resistance of the bacterial isolate strain NB2 was observed against lysozyme activity (Samedi and Linton Charles, 2019). Antibiotic susceptibility of the bacterial isolate strain NB2 was observed against antibiotic discs of amikacin, Amoxicillin, Bacitracin, Cephalothin, Erythromycin, Novobiocin, Oxytetracycline, Vancomycin, Ceflnaxone, Ceftazidime, Cefotaxime, Lincomycin, Netilin, and Ofloxacin (Himedia, OD034R-1PK, and OD003R-1PK) using disc diffusion assay after recording the zone of the growth inhibition (mm) on the LB agar medium after incubation for 24 h at 37°C.

Auto-Aggregation and Cell Surface Hydrophobicity

The auto-aggregation tendency and cell surface hydrophobicity of the bacterial isolate strain NB2 were also observed (Collado et al., 2008; Dowarah et al., 2018).

Health-Promoting Features of Bacterial Isolate Strain NB2

The anti-pathogenic property of the bacterial isolate strain NB2 was screened against the pathogens *Staphylococcus aureus* (MTCC No. 96), *E. coli* (MTCC No. 443), and *Salmonella typhi* (MTCC No. 98) with a disc diffusion assay (Kumar et al., 2021). Co-aggregation tendency of the bacterial isolate strain NB2 with pathogens [*Staphylococcus aureus* (MTCC No. 96), *E. coli* (MTCC No. 443), and *Salmonella typhi* (MTCC No. 98)] was also assessed (Valeriano et al., 2014). Human gut bacterial isolate strain NB2 was assessed for anti-amylase activity (Sekhon-Loodu and Rupasinghe, 2019), anti-lipase activity (Jaradat et al., 2020), cholesterol removal activity (Shobharani and Halami, 2016), and bile salt hydrolysis activity (Shobharani and Halami, 2016). Bacterial isolate strain NB2 was screened for the activity of glutenase (Shobharani and Halami, 2016), prolylendopeptidase (Kumar et al., 2018), lactase (Leksmono et al., 2018), laccase (Mandic et al., 2019), peroxidase (https://www.sigmaaldrich.com/IN/en/technical-documents/protocol/protein-biology/enzyme-activity-assays/enzymatic-assay-of-peroxidase), and phosphatase (Ndubuisil et al., 2002).

RESULTS

Characterization of Bacterial Isolate Strain NB2

Microscopic observation of human gut isolate NB2 indicated it is a rod-shaped gram-negative bacteria. Growth pattern analysis of gut isolate NB2 indicated that this microbe attains log phase after 2 h (**Figure 1**), and a doubling time for the isolated gut microbe was observed to be 53.3 min in aerobic growth conditions. Additionally, bacterial isolate strain NB2 also showed growth (0.413 OD at 600 nm) after incubating the culture for 24 h at 37°C in anaerobic growth conditions. Good growth in aerobic conditions in comparison to anaerobic conditions indicates its growth preference in aerobic conditions. It also indicates its facultative nature. The 16S rRNA gene of the gut isolate NB2 shared 99 and 96% nucleotide similarity with *Cytobacillus oceanisediminis* 2691 (CP015506.1) and *Bacillus firmus* (AY833571.2), respectively (**Supplementary Table S3**) indicating its taxonomic affiliation with *Cytobacillus oceanisediminis*. Phylogenetic analysis of gut isolate 16S rRNA gene also indicated a similar observation (**Figure 2**). Combining the 16S rRNA gene homology and polygenetic analysis, bacterial isolate strain NB2 was labeled as *Cytobacillus* sp. NB2 till further taxonomic characterization. Substrate utilization assay of the human gut isolate NB2 indicated its potential to utilize xylose, maltose, raffinose, trehalose, melibiose, sucrose, L-arabinose, mannose, esculin, and citrate out of the given 35 substrates (**Table 1**; **Supplementary Table S4**). Substrate-utilization preference of gut isolate NB2 was found similar to the *Cytobacillus oceanisediminis* as compared to *Bacillus firmus* (**Supplementary Table S4**). Even the antibiotic susceptibility assay of gut isolate NB2 (**Supplementary Table S5**) indicates its higher similarity with *Cytobacillus oceanisediminis* than *Bacillus firmus*. Substrate utilization assay, antibiotic susceptibility along with 16S rRNA gene homology, and phylogenetic analysis indicate that the gut isolate NB2 is a strain of *Cytobacillus oceanisediminis*, hence labeled as *Cytobacillus oceanisediminis* NB2.

Genomic Characterization of *Cytobacillus oceanisediminis* NB2

Genome sequence assembly of *Cytobacillus oceanisediminis* NB2 resulted in 203 contigs amounting to 5,235,740 base pairs with 41.41% GC content (**Supplementary Table S6**). BUSCO v5.0.0 assessment tool was used with the latest bacterial ortholog catalog (bacteria_odb10) for analyzing the completeness of the set of predicted genes in the bacterial genome assembly (**Supplementary Table S6**). A quantitative assessment of the completeness in terms of the expected gene content of a genome assembly or annotated gene set (https://busco.ezlab.org/) was done. The BUSCO assessment resulted in 100% genome assembly and 124 complete, 123

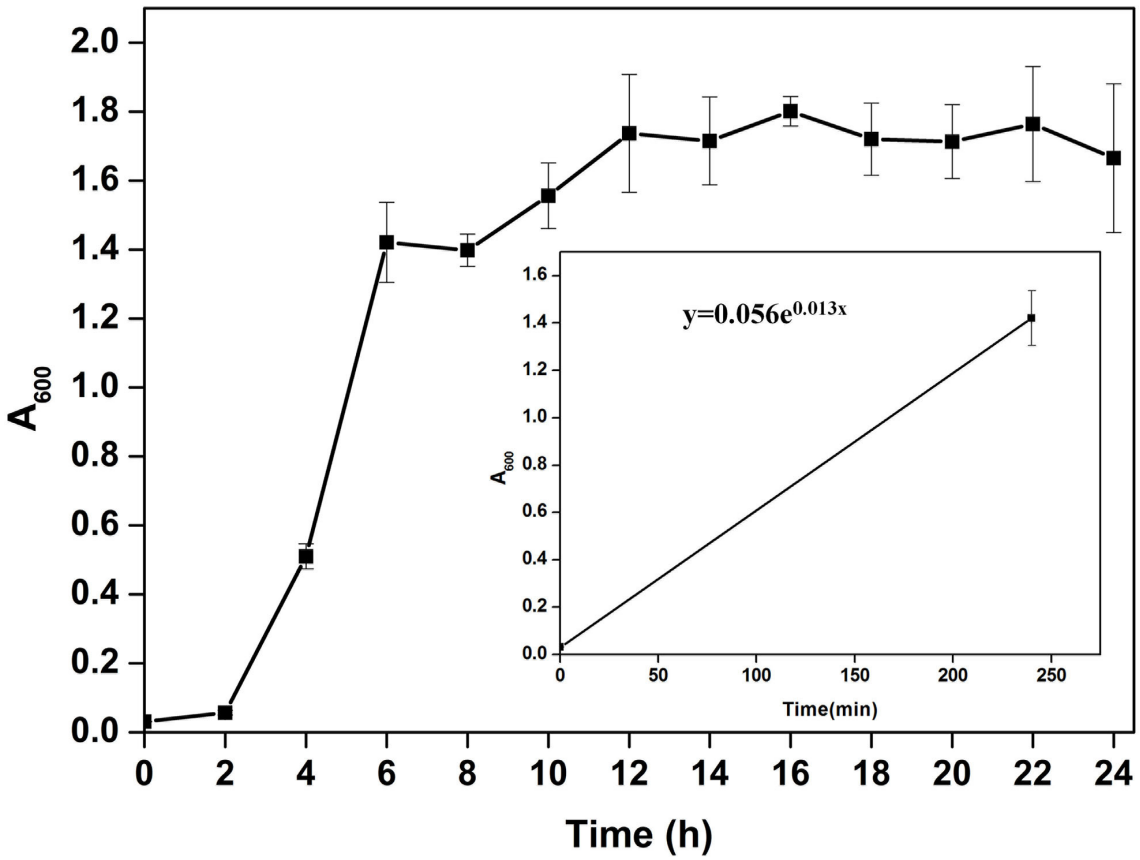


FIGURE 1 | Growth pattern analysis of the bacterial isolate strain NB2 in Luria-Bertani broth for 24 h at 37°C with constant shaking at 200 rpm. Each point in the graph is the mean value of readings observed in triplicate experiments.

single copies, 1 duplicated copy, 0 fragmented, and 0 missing conserved proteins within the bacterial genome. Genome annotation has identified 5,195 coding genes and 128 RNAs in the genome of *Cytobacillus oceanisediminis* NB2 (**Figure 3**; **Supplementary Table S7**). Genome characterization identified the presence of 254 genome-encoded protein features associated with metal/metalloid toxicity resistance, 147 features associated with antibiotic resistance, 414 protein features associated with oxidative stress tolerance, and 108 features associated with heat tolerance (**Supplementary Table S8**). Additionally, CAZymes annotation with HMMER resulted in a total of 35 CAZymes clusters (**Supplementary Table S9**). No pathogenic islands/genes were determined within the genome of the microbial isolate. Additionally, the virulence genes were manually searched within the genome of the microbial isolate resulting in the absence of many genes related to the pathogenic behavior of the isolated microbe.

Genome Comparison of *Cytobacillus oceanisediminis* NB2

The sequenced genome of *Cytobacillus oceanisediminis* NB2 was compared with the genomes of the *Bacillus* species isolated from the various sources (**Supplementary Table S1**) to elucidate genome-level similarities and uniqueness. The comparison was made for genome size, coding sequences, tRNA, and rRNA (**Supplementary Table S1**). Additionally, the average ANI value among all *Bacillus* species was ∼66–97%, which is toward the lower end of the 62–100% spectrum of interspecies variation within a genus (Kim et al., 2014), suggesting substantial genomic diversity. This observation was reaffirmed by tetra correlation among member species, highlighted by a wide distribution of *z*-scores (**Supplementary Table S10**). The isolated gut microbe shared high ANI (>98.0%) with *Bacillus oceansedimins* while ANIs with *Bacillus mediterraneensis, Bacillus subtilis, Bacillus clausii, Bacillus* sp. *bd59s,* and *Bacillus megaterium* were found to be 69.5867.16, 66.00, 67.25, and 68.05%, respectively (**Table 2**). A z-score value of 0.99806 during tetra-correlation scoring corroborates its similarity with *Bacillus oceansedimins*. Other *Bacillus* strains also shared good similarities with the gut isolate (*z*-score ∼0.90–0.99; **Supplementary Table S9**). Genome-based phylogenetic analysis of *Cytobacillus oceanisediminis* NB2 also indicated its similarity with *Cytobacillus oceanisediminis* (**Figure 4**). Genome characterization indicated the presence of a total of 19736 COGs (**Supplementary Figure S1**). These results indicate genome plasticity, which could be attributed due to niche-specific genome evolution (Woodcock et al., 2017).

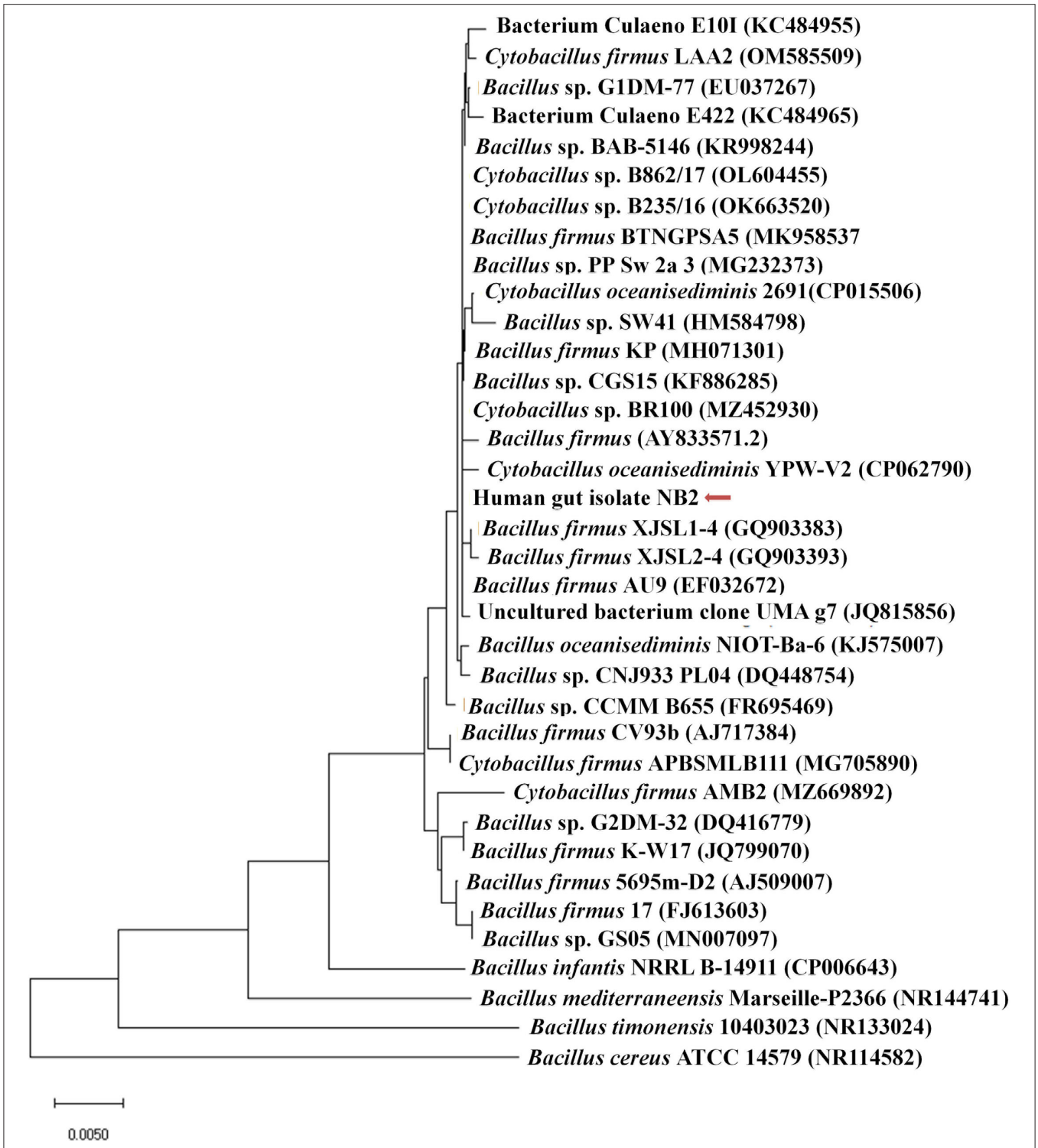


FIGURE 2 | Phylogenetic affiliation of bacterial isolate strain NB2 with the other *Bacillus* species. Phylogenetic tree was constructed with the neighbor-joining method of phylogenetics using the 16S rRNA gene sequences of bacterial isolate strain NB2 and NCBI database homologs using MEGAX software. Numbers at the node represent bootstrap values in percent for the node (based on 500 bootstrap sampling). Out-group was represented by *Bacillus cereus* ATCC 14579 SSU rRNA gene sequence.

TABLE 1 | Physiological, morphological, and biochemical characterization of *Cytobacillus oceanisediminis* NB2.

Property	Term
Gram stain	Negative
Cell shape	Rod
Temperature range	20–40°C
Optimum temperature	37°C
pH range	5–9
Optimum pH	8.0
Habitat	Human gut
Salinity/metal/metalloid resistance	Upto 7% NaCl and 10% KCl; Silver nitrate (upto 2 mM); Lead acetate (upto 2 mM); Sodium arsenate (upto 10 mM); Potassium dichromate (upto 2 mM)
Substrate utilization preference	Xylose, Maltose, Raffinose, Trehalose, Melibiose, Sucrose, L-arabinose, Mannose, Esculin, and Citrate
Oxygen requirement	Aerobic/Anaerobic
Biotic relationship	Host-associated
Pathogenicity	Non-pathogenic

Hemolytic and Cytotoxicity Activity

Evaluation of bacterial toxicity is essential before considering a bacterial strain a probiotic, as it should not be harmful to the host cells. Toxicity and hemolytic activity assessment of *Cytobacillus oceanisediminis* NB2 indicated no hemolytic activity in Blood agar plate assay after 24 h of incubation at 37°C. Even cytotoxicity analysis of *Cytobacillus oceanisediminis* NB2 demonstrated that Caco-2 cells showed 91.82 ± 5.04% and 89.28 ± 7.85% viability after 24 h of exposure with the cell-free supernatant and cell lysate of *Cytobacillus oceanisediminis* NB2, respectively.

Stress-Response Physiology

Cytobacillus oceanisediminis NB2 showed continued growth within the pH range of 5.0–9.0, with an optimum growth at pH 8.0 (**Figure 5A**). *Cytobacillus oceanisediminis* NB2 showed continued growth within the temperature range of 25–50°C, while an optimum growth was observed at 30–35°C (**Figure 5B**). *Cytobacillus oceanisediminis* NB2 also indicated continued growth in the LB medium supplemented up to 7.0% NaCl (w/v) and 10.0% KCl (w/v). Similarly, *Cytobacillus oceanisediminis* NB2 showed growth in the presence of various metals (silver, lead,

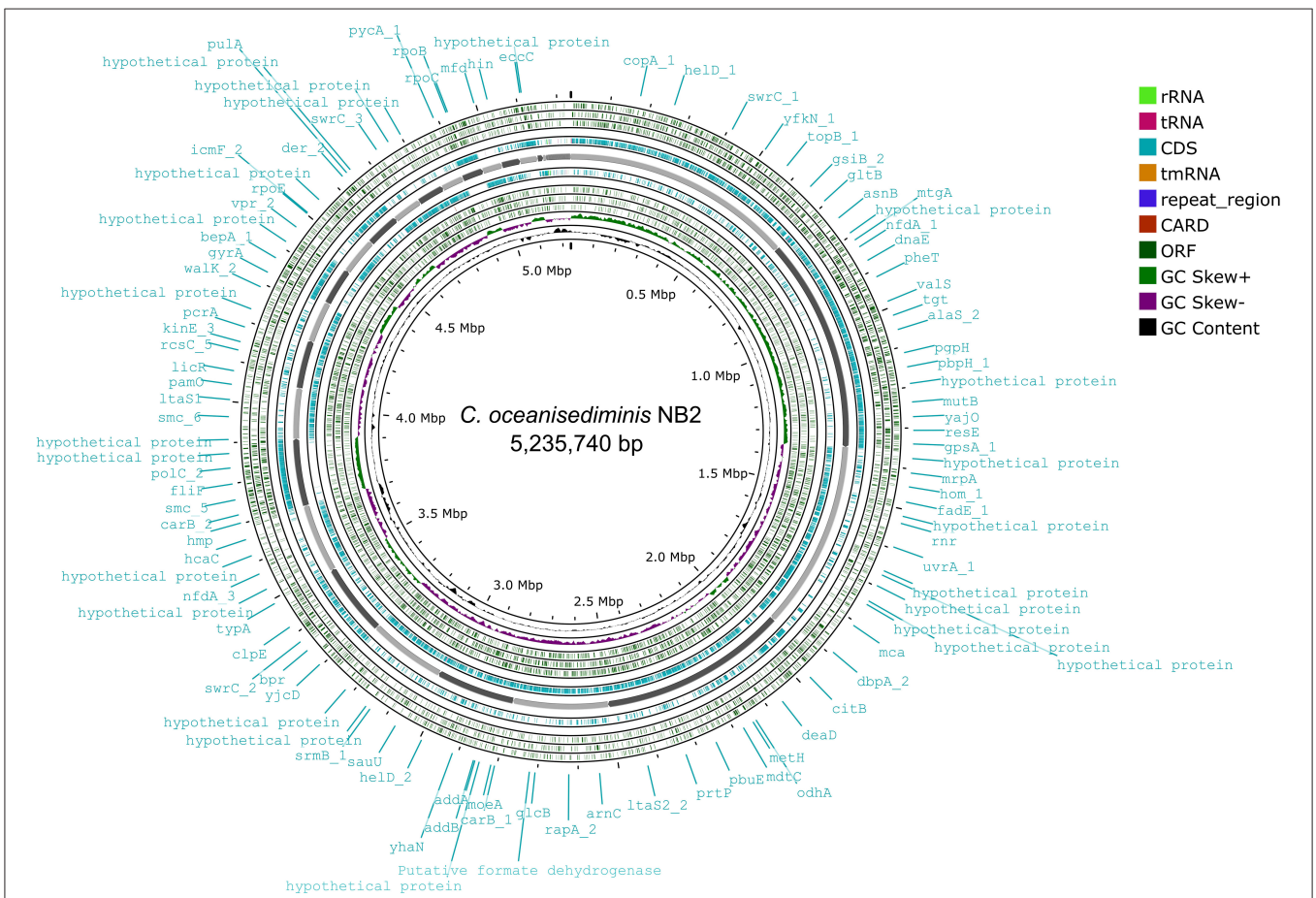


FIGURE 3 | Genome map of the *Cytobacillus oceanisediminis* NB2. The circular genome map was drawn using Proksee online tool (https://proksee.ca/) that uses a complete genome sequence and annotates it using PROKKA. It also identifies resistance gene using CARD identifier.

TABLE 2 | Average nucleotide identity (ANI) of *Cytobacillus oceanisediminis* NB2 with reference to other *Bacillus* species.

	***Cytobacillus oceanisediminis* NB2 (current study)**	***Bacillus clausii* strain ENTPr**	***Bacillus coagulans* strain HM-08**	***Bacillus infantis* NRRLB-14911**	***Bacillus subtilis* strain B-1**	***Bacillus subtilis* strain UD1022**	***Bacillus subtilis* subsp. Spizizenii str. W23**	***Bacillus subtilis* subsp. subtilis str.168**	***Cytobacillus firmus***	***Bacillus velezensis* strain BIMB-454D**	***Cytobacillus oceanisediminis* 2691**	***Cytobacillus oceanisediminis* strain YPW-V2**
Cytobacillus oceanisediminis NB2 (current study)	*	66.12 [17.61]	67.75 [20.05]	70.98 [39.56]	67.52 [21.36]	67.83 [23.30]	67.96 [23.27]	67.83 [23.37]	88.39 [63.37]	67.58 [21.54]	98.00 [89.18]	97.89 [90.28]
Bacillus clausii strain ENTPr	66.00 [20.90]	*	66.01 [17.64]	65.68 [19.23]	66.12 [19.35]	65.89 [21.37]	66.19 [21.03]	65.97 [21.36]	65.91 [19.63]	66.06 [19.91]	65.91 [20.96]	65.98 [21.17]
Bacillus coagulans strain HM-08	67.85 [28.25]	66.36 [20.27]	*	68.08 [28.83]	67.71 [26.92]	67.34 [27.28]	67.35 [26.65]	67.33 [27.29]	68.25 [27.98]	67.78 [27.16]	67.87 [28.37]	67.77 [28.83]
Bacillus infantis NRRLB-14911	71.15 [41.79]	65.81 [17.42]	68.06 [21.95]	*	67.78 [23.34]	67.74 [24.21]	67.74 [24.24]	67.72 [24.17]	71.41 [39.28]	67.78 [23.57]	71.18 [41.90]	71.16 [42.21]
Bacillus subtilis strain B-1	67.48 [27.18]	65.95 [21.06]	67.33 [24.47]	67.63 [27.57]	*	76.21 [72.75]	76.30 [72.81]	76.21 [73.32]	67.74 [26.31]	97.34 [93.01]	67.61 [27.32]	67.54 [27.45]
Bacillus subtilis strain UD1022	68.05 [29.49]	66.33 [23.29]	67.50 [25.34]	68.13 [28.66]	76.28 [72.23]	*	92.47 [89.08]	98.12 [94.29]	68.27 [28.03]	76.43 [72.95]	68.11 [29.75]	68.06 [29.95]
Bacillus subtilis subsp. *spizizenii* str. W23	68.07 [29.24]	66.24 [23.32]	67.41 [24.65]	67.71 [29.34]	76.43 [71.64]	92.60 [88.82]	*	92.50 [88.18]	68.17 [27.79]	76.55 [72.26]	68.06 [29.43]	68.04 [29.63]
Bacillus subtilis subsp. *subtilis* str.168	68.06 [28.04]	66.27 [22.16]	67.49 [24.26]	67.98 [27.50]	76.34 [69.33]	97.93 [90.56]	92.23 [85.12]	*	68.31 [26.49]	76.47 [71.31]	68.07 [28.40]	68.10 [28.48]
Cytobacillus firmus	88.56 [69.24]	66.41 [19.07]	68.52 [23.11]	71.31 [41.19]	68.11 [23.49]	68.29 [24.69]	68.27 [24.67]	68.21 [24.90]	*	68.07 [23.69]	88.77 [70.40]	88.71 [71.22]
Bacillus velezensis strain BIMB-454D	67.94 [25.78]	66.27 [20.79]	67.90 [23.63]	67.94 [26.89]	97.31 [87.52]	76.34 [69.92]	76.44 [69.43]	76.33 [71.12]	68.05 [25.22]	*	67.98 [26.02]	67.88 [26.24]
Cytobacillus oceanisediminis 2691	97.97 [85.73]	66.18 [17.40]	67.91 [19.84]	71.12 [38.50]	67.94 [20.89]	68.19 [22.72]	68.18 [22.59]	68.18 [22.74]	88.60 [61.99]	67.95 [21.02]	*	98.72 [87.37]
Cytobacillus oceanisediminis strain YPW-V2	98.02 [88.62]	66.18 [18.63]	67.87 [20.89]	71.19 [40.03]	67.82 [21.83]	68.03 [23.48]	68.11 [23.54]	68.04 [23.55]	88.65 [64.30]	67.82 [22.09]	98.81 [89.27]	*

**Indicates the ANI between the same genomes.*

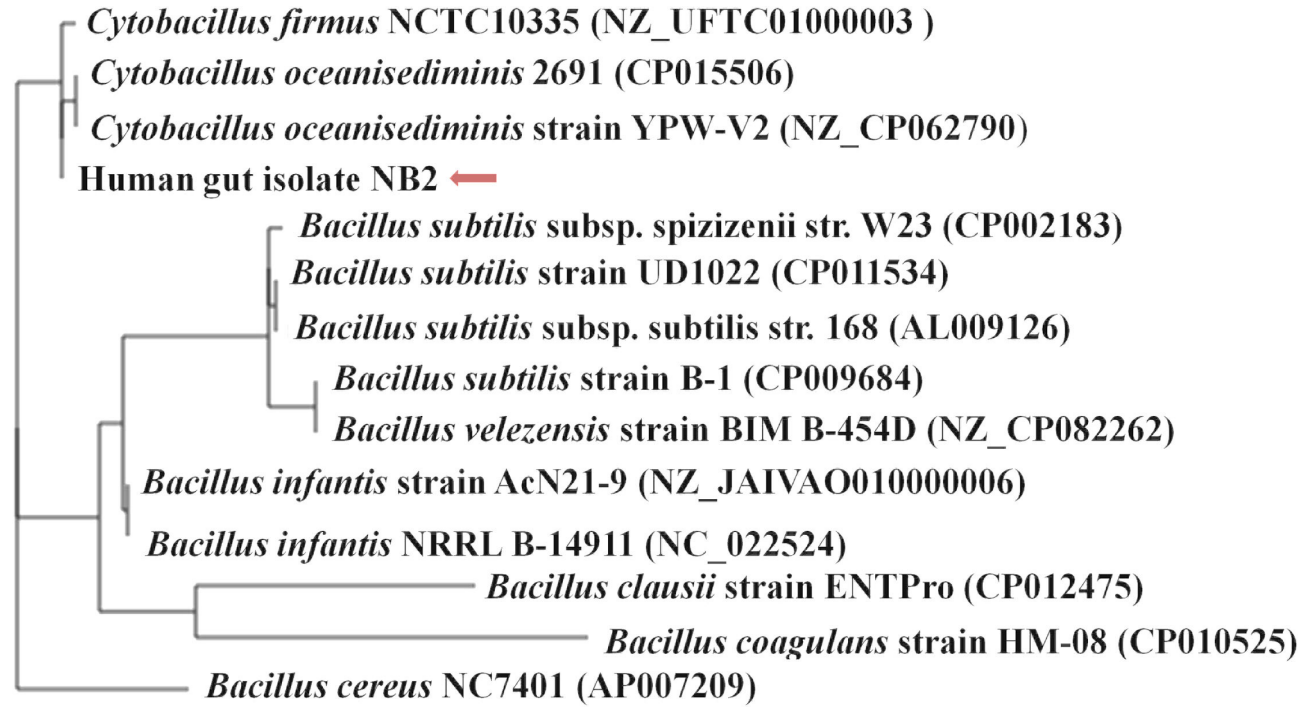


FIGURE 4 | Phylogenetic assessment of the genome of *Cytobacillus oceanisediminis* NB2 with the genomes of other *Bacillus* species. The phylogenetic relationship of the gut bacterial isolate strain NB2 was assessed with other *Bacillus* genomes using M1CR0B1AL1Z3R webserver (https://microbializer.tau.ac.il/). The online tool extracts ORFs, detects OGs, extracts OG sequences, infers a core proteome, and reconstructs the species' phylogeny. The tree was drawn by taking Maximal e-value cutoff: 0.01; identity minimal percent cutoff: 80.0%; minimal percentage for core: 100.0% with no bootsrapping.

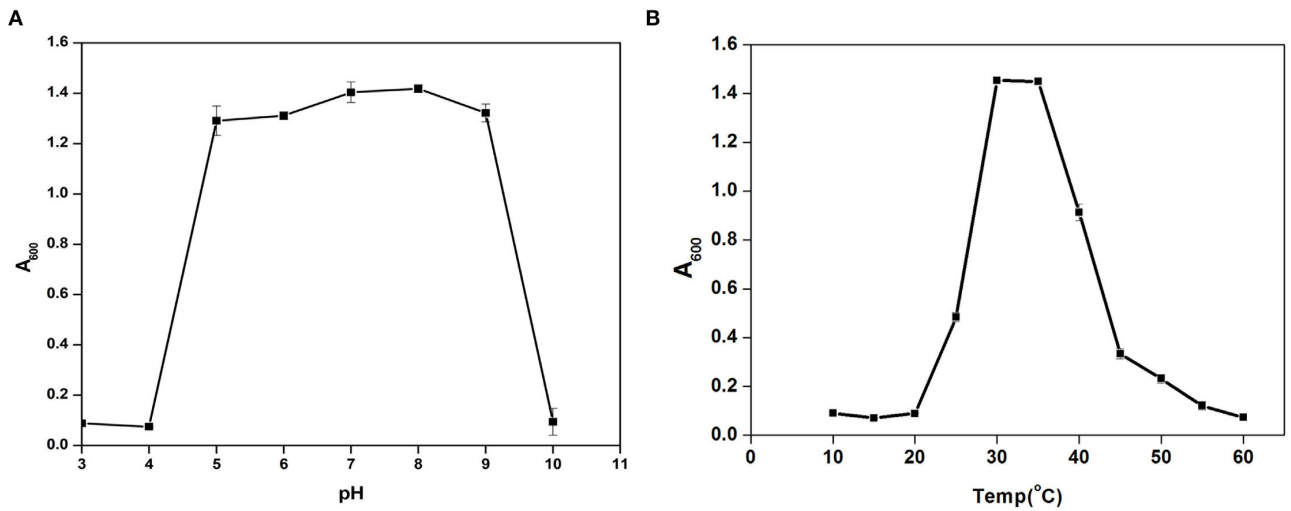


FIGURE 5 | Growth pattern assessment of the *Cytobacillus oceanisediminis* NB2in Luria-Bertani broth with diverse pH (3–10 with an interval of 1.0 pH) **(A)** and temperature (10–60°C with an interval of 5°) **(B)** conditions for 24 h with constant shaking at 200 rpm. Each point in the graph is the mean value of readings observed in triplicate experiments.

cadmium, and potassium) and metalloid arsenic. *Cytobacillus oceanisediminis* NB2 also exhibited resistance to Cephalothin (30 μg), Ceflnaxone (30 μg), Ceftazidime (30 μg), and Ofloxacin (2 μg), moderate susceptibility to Amoxicillin (10 μg), Bacitracin (10 μg), and Lincomycin (2 μg), and high susceptibility to Amikacin (10 and 30 μg), Erythromycin (15 μg), Novobiocin (30 μg), Oxytetracycline (30 μg), Vancomycin (30 μg), Cefotaxime (30 μg), and Netilin (30 μg). *Cytobacillus oceanisediminis* NB2

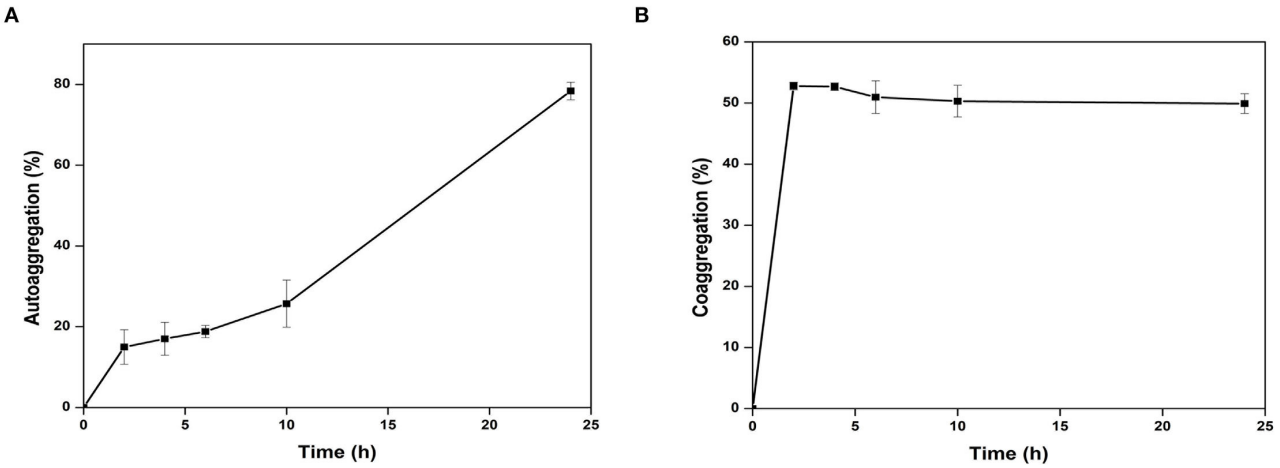


FIGURE 6 | Autoaggregation **(A)** and Co-AGGREGATION **(B)** of *Cytobacillus oceanisediminis* NB2 with pathogen mixture of *E. coli, Staphylococcus aureus,* and *Salmonella typhimurium*.

did not show any bile salt hydrolysis activity. Even, a 62.90 ± 0.5% growth suppression of *Cytobacillus oceanisediminis* NB2 was observed in the bile-enriched medium. *Cytobacillus oceanisediminis* NB2 did not show any growth suppression in gastric conditions, while a 48.6% growth suppression was observed in intestinal conditions. A high concentration of lysozyme (100 mg/L) suppressed the *Cytobacillus oceanisediminis* NB2 growth, while no effect was observed at lower lysozyme concentration (1 mg/L).

Auto-Aggregation and Cell Surface Hydrophobicity

The adherence properties of microbial cells are due to their aggregation abilities. The gut isolates showed adherence to the epithelial cells and mucosa due to their auto-aggregation activity (Krausova et al., 2019). *Cytobacillus oceanisediminis* NB2 cells showed low adherence to toluene (0.97% ± 0.87), which confirms the hydrophilic nature of the isolate and indicated its electron-donating nature. *Cytobacillus oceanisediminis* NB2 showed 78.43 ± 0.97% auto-aggregation after 24 h, with auto-aggregation of 15.03 ± 2.04%, 17.05 ± 2.10%, 18.82 ± 1.57%, and 25.81 ± 2.58% after 2, 4, 6, and 10 h, respectively (**Figure 6A**). *Cytobacillus oceanisediminis* NB2 also showed 15.89 ± 2.45 % adherence to the Caco-2 cells. Adherence to intestinal cells is an essential feature for successful establishment and colonization (Yadav and Chauhan, 2022). The cellular adherence property of *Cytobacillus oceanisediminis* NB2 indicated the possibility of its successful establishment in the gut environment.

Health-Promoting Properties of *Cytobacillus oceanisediminis* NB2

The probiotics can modulate pathogenic abundance by co-aggregating with them (Yadav and Chauhan, 2022). *Cytobacillus oceanisediminis* NB2 was found to co-aggregate with the pathogenic strains. A time-dependent co-aggregation was observed for Staphylococcus *aureus* (MTCC No. 96), *E. coli* (MTCC No. 443), and *Salmonella typhi* (MTCC No. 98) (**Figure 6B**). Disc diffusion assay showed an anti-pathogenic activity for *Cytobacillus oceanisediminis* NB2 against three pathogenic strains of *Staphylococcus aureus* (MTCC No. 96), *E. coli* (MTCC No. 443), and *Salmonella typhi* (MTCC No. 98). A zone of 10.8 ± 1.0 mm, 12.8 ± 1.0 mm, and 13.5 ± 0.5 mm growth inhibition was observed, respectively, for Staphylococcus *aureus* (MTCC No. 96), *E. coli* (MTCC No. 443), and *Salmonella typhi* (MTCC No. 98) that indicated *Cytobacillus oceanisediminis* NB2-induced growth inhibition.

Anti-glycemic and anti-lipogenic effects are considered therapeutic targets to overcome diabetic mellitus (Type-II), obesity, and cardiovascular pathological conditions (Salehi et al., 2020). Thus, α-amylase inhibition seems to be the prime therapeutic target. *Cytobacillus oceanisediminis* NB2 has shown 9.82 ± 0.55% inhibition in the amylase activity. The isolated microbial culture shows 14.79 ± 1.44% inhibition of the lipase activity. Similarly, the *Cytobacillus oceanisediminis* NB2 showed cholesterol-oxidizing activity. *Cytobacillus oceanisediminis* NB2 also showed a significant prolylendopeptidase activity (0.318 units/mg microbial pellet). The presence of this enzyme activity could be helpful in removing the gluten antigen to overcome gluten-induced celiac diseases. *Cytobacillus oceanisediminis* NB2 also showed a significant lactase activity (38.796 units/mg of the bacterial pellet), which could help overcome lactose indigestibility issues for lactose-intolerant individuals. *Cytobacillus oceanisediminis* NB2 was found to possess alkaline phosphatase (9.54 ± 0.04 units/mg bacterial pellet) and acid phosphatase activity (and 190.8 ± 0.16 units/mg bacterial pellet), respectively. Phosphatase activity could play an important role in cell proliferation and differentiation. This microbe was also found to have peroxidase activity (2.4804 ± 0.02 units/mg bacterial pellet) of the enzyme, which could help to overcome oxidative stress. Laccase is a multi-copper oxidase

that was characterized to play a vital role in host health (Janusz et al., 2020). *Cytobacillus oceanisediminis* NB2 was also found to have laccase enzyme activity (0.004452 units/mg bacterial pellet).

DISCUSSION

Diet has an important impact on a healthy life (Lindefeldt et al., 2019). Supplementation of diet with probiotic strains can further augment human health (Wang et al., 2019). The application of any bacteria as a probiotic strain requires extensive characterization for safety and applicability (Yadav and Chauhan, 2022). Recently, probiotic bacteria are being extensively explored and characterized for their potential therapeutics for various human disorders (Yadav and Chauhan, 2022). Probiotic strains were identified from various sources like dairy products (Haghshenas et al., 2017; Karami et al., 2017), fermented drinks (Angelescu et al., 2019; Setta et al., 2020), plants (Rahman et al., 2018; Samedi and Charles, 2019), soil (Siraj et al., 2017), and animals (Abdou et al., 2018; Li et al., 2021). Human gut microbiota are being extensively characterized for their health-promoting benefits; however the full potential for their usage as microbiome therapeutics has humongous possibilities (Yadav and Chauhan, 2022). These characterizations are primarily performed as a consortium; however, individual-specific microbial diversity has not been characterized to date for assessing their suitability as probiotic strain, especially in different geographical regions of the world as well as within large countries like India. The lack of pure culture for a majority of gut microbes is the major bottleneck toward their functional usage (Lagier et al., 2015). Efforts are being made to culture human gut microbes in laboratory conditions to characterize them for their probiotic potential (Tang et al., 2020). Thus, the current study was planned to explore the human gut microbiota to isolate the human gut bacterium for probiotic applications.

In the current study, a bacterial culture was isolated from the human feces. Biochemical, physiological, and taxonomic characterization identifies it as a species of *Bacillus*. Members of the *Bacillus* were characterized by diverse habitats, including human feces, and showed a wide range of biotechnological potential (Singh et al., 2009; Zhang et al., 2010; Kumar et al., 2013, 2014; Patel et al., 2014; Boucherba et al., 2017). Various *Bacillus* species have already been characterized for probiotic potential (Elshaghabee et al., 2017). *B. cereus, B. clausii, B. coagulans, B. licheniformis, B. polyfermenticus, B. pumilus*, and *B. subtilis* are commercially used probiotics. Although various strains of *Bacillus* were studied in various organisms such as mice and pigs, studies regarding the *Bacillus* strains as a probiotic in the human body are still evolving (Hong et al., 2009). The bacterial isolate strain NB2 was assessed for its substrate utilization tendency where it has shown the positive esculin hydrolysis and citrate utilization that is in line with the substrate utilization characteristic of the other *Bacillus* strains (Beesley et al., 2010). The whole-genome analysis of the bacterial isolate strain NB2 indicated the presence of COGs associated with the general adaptive and metabolic mechanisms required for the microbial cell survival within the human body, thus suggesting a strong affiliation to survive and thrive within the human host (Yadav et al., 2020, 2021). The gut isolate was found to possess no pathogenic islands indicating its safety considerations for probiotic features (Li et al., 2018). The bacterial isolate strain NB2 contains a total of 35 CAZymes. The *Bacillus* strains have unique anti-cancer, anti-oxidant, anti-diabetic, as well as anti-obesity characteristics (Elshaghabee et al., 2017). Within the human body, the microbe may suffer various stressful conditions such as gastric environment, heat, temperature, and pH stress. Thus, a probiotic bacterium should possess significant features to resist all these stressful conditions. *Bacillus* strain is well-known to adapt and thrive within the host's body (Yadav et al., 2018). Likewise, *Cytobacillus oceanisediminis* NB2 was identified to thrive in high salt, variable pH, and temperature conditions indicating its suitability to survive in a highly variable gut ecosystem. Bile salts pose a major challenge to microbial survival (Bustos et al., 2018). The differential expression of bile salt resistance proteins may influence bile tolerance of the isolated microbe (Hamon et al., 2011).

Cytobacillus oceanisediminis NB2 showed growth in the presence of bile salt, despite slight growth suppression indicating its bile tolerance property. The strain/species-specific acid tolerance might have influenced the bacterial survival in the acidic gastric conditions (Nami et al., 2019) since certain microbial strains are adapted to thrive in acidic conditions (Guan and Liu, 2020). The *Cytobacillus oceanisediminis* NB2 did not show any growth suppression in the gastric pH in presence of pepsin while it showed partial growth suppression in the human intestinal conditions. These results indicate its survivability in diverse environments, making it suitable to apply in the human gastrointestinal ecosystem. *Bacillus* strains were known to produce toxins and may transfer antibiotic resistance; thus, a safety evaluation needs to be done (Kotowicz et al., 2019). *Cytobacillus oceanisediminis* NB2 toxicity was assessed against the Caco-2 cells. The 24-h bacterial exposure showed 89.3% cell viability, thus, indicating it is a safe and non-toxic microbe. Furthermore, the bacterial isolate strain NB2 did not show hemolysis. Auto-aggregation and co-aggregation properties are important for the anti-pathogenic potential of probiotics (Collado et al., 2008). Auto-aggregation enables the microbes to bind with each other and form the first line of defense against the pathogens (Trunk et al., 2018), while co-aggregation enables the microbial assessment for their binding capacity with the pathogens. *Cytobacillus oceanisediminis* NB2 was found to have different levels of aggregation and co-aggregation indicating differential environmental and internal factors (Vlková et al., 2008). Thus, this microbe can impart several benefits to maintain intestinal health by protecting it from pathogens. The probiotics improve the host's health without posing adverse effects on other microbial groups as well as the host due to antibiotics-led dysbiosis (Plaza-Diaz et al., 2019). The microbial factors can suppress the pathogen's survival and thus modulate the risk of infection. The gut isolate suppressed the growth of three pathogenic strains, i.e., *Salmonella typhi, E. coli,* and *Staphylococcus*. Thus, the anti-pathogenic activity indicates that microbes can be easily used to eradicate the overgrown pathogens within the host. The bacterial isolate

strain NB2 was detected to possess various enzymatic activities against various substrates. Different levels of activities were obtained, and positive enzymatic activities indicate that the microbe can be used to promote the metabolic capacity of the host. The antibiotic treatment is a major threat to the host's health as it may modulate the other microbial strains as well as develop microbial dysbiosis leading to infections and health risks. The probiotics must contain the genetic features that may provide resistance to antibiotics. The bacterial isolate strain NB2 has shown resistance against a diverse range of antibiotics like (Amikacin, Amoxicillin, Bacitracin, Cephalothin, Erythromycin, Novobiocin, Oxytetracycline, Vancomycin, Ceflnaxone, Ceftazidime, Cefotaxime, Lincomycin, Netilin, and Ofloxacin). The presence of resistance against a range of antibiotics would allow it to survive in presence of antibiotic compounds. It also extends suitability when ingested with antibiotic drugs. The adherence properties of cells are due to their aggregation abilities. The bacterial isolate strain NB2 showed adherence to the epithelial cells and mucosa due to its auto-aggregation activity (Krausova et al., 2019). In the present study, the adherence ability is linked to auto-aggregation. The bacterial isolate strain NB2 cells showed low adherence to toluene ($0.97 \pm 0.87\%$) confirming the hydrophilic nature of the isolate and indicating its electron-donating nature. The bacterial isolate strain NB2 showed various health-promoting features such as anti-amylase, anti-lipase, lactase, laccase, protease, prolyl endopeptidase, and cholesterol-removing activities. Dietary polyphenols induce hyperglycemic effects by binding with the glucose transporters and inhibiting the activity of the digestive enzymes. Carbohydrate utilization by α-amylase produces glucose that causes an increase in blood glucose. Thus, α-amylase inhibition is the prime target in the case of diabetes mellitus Type-II pathophysiological condition. In the present study, *Cytobacillus oceanisediminis* NB2 has shown 9.83% inhibition in the amylase activity. Phosphatases are required for cell proliferation and differentiation (Krausova et al., 2019). The presence of alkaline, as well as acid phosphatase activities within the gut isolate indicated its significant survival within the human body. The bacterial isolate strain NB2 also possessed significant peroxidase activity that enables it to survive under oxidative stress. Lactase is required to convert lactose to glucose. Thus, the lactase enzyme is required for the treatment of lactose intolerance. The significant lactase activity within the *Cytobacillus oceanisediminis* NB2 indicated that ingestion/introduction of this microbe can be beneficial for the treatment of lactose intolerance. The presence of laccase activity indicated the role of the microbe in the digestion and metabolism of various phenolic compounds. To treat obesity, the gastrointestinal absorption of fats should be first reduced (Apovian et al., 2015). The presence of anti-lipase activity within the *Cytobacillus oceanisediminis* NB2 strongly indicates its potential for the treatment of obesity. Hypercholesterolemia is a major concern in the modern lifestyle. The removal of cholesterol from the blood can treat this disorder. The *Cytobacillus oceanisediminis* NB2 possesses significant cholesterol-removing ability. Though these initial characterizations strongly indicate the probiotic potential of *Cytobacillus oceanisediminis* NB2, further *in vivo* investigations are needed to validate its efficacy. The health-promoting tendencies of the isolated gut microbe can thus be harnessed to treat various disorders such as diabetes, lactose intolerance, hypercholesterolemia, celiac disease, as well as obesity. Microbiome engineering can thus be a significant effort for human healthcare.

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/genbank/, SUB11205683.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Institutional Human Ethical Committee, Maharshi Dayanand University, Rohtak. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

NC designed the study and experiments. NC, MY, and RP wrote the manuscript. MY and TK carried out the experiments. MY, TK, AK, and RM did the characterization. NC, MY, AK, and RM analyzed the data. All authors edited the manuscript and approved the final draft of the manuscript.

ACKNOWLEDGMENTS

The authors acknowledge the CSIR-Institute of Genomics and Integrative biology, New Delhi, India for DNA sequencing facility.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.932795/full#supplementary-material

REFERENCES

Abdou, A. M., Hedia, R. H., Omara, S. T., Mahmoud, M. A. E.-F., Kandil, M. M., and Bakry, M. A. (2018). Interspecies comparison of probiotics isolated from different animals. *Vet. World* 11, 227–230. doi: 10.14202/vetworld.2018.227-230

AlKalbani, N. S., Turner, M. S., and Ayyash, M. M. (2019). Isolation, identification, and potential probiotic characterization of isolated lactic acid

bacteria and *in vitro* investigation of the cytotoxicity, antioxidant, and antidiabetic activities in fermented sausage. *Microb. Cell Factor.* 18, 188. doi: 10.1186/s12934-019-1239-1

Angelescu, I.-R., Zamfir, M., Stancu, M.-M., and Grosu-Tudor, S.-S. (2019). Identification and probiotic properties of *lactobacilli* isolated from two different fermented beverages. *Ann. Microbiol.* 69, 1557–1565. doi: 10.1007/s13213-019-01540-0

Apovian, C. M., Aronne, L. J., Bessesen, D. H., McDonnell, M. E., Murad, M. H., Pagotto, U., et al. (2015). Pharmacological management of obesity: an endocrine society clinical practice guideline. *J. Clin. Endocrinol. Metab.* 100, 342–362. doi: 10.1210/jc.2014-3415

Aswathy, R. G., Ismail, B., John, R. P., and Nampoothiri, K. M. (2008). Evaluation of the probiotic characteristics of newly isolated lactic acid bacteria. *Appl. Biochem. Biotechnol.* 151, 244–255. doi: 10.1007/s12010-008-8183-6

Barik, A., Patel, G. D., Sen, S. K., Rajhans, G., Nayak, C., and Raut, S. (2021). Probiotic characterization of indigenous Kocuria flava Y4 strain isolated from Dioscorea villosa leaves. *Probiotics Antimicrob. Proteins*. doi: 10.1007/s12602-021-09877-2

Beesley, C. A., Vanner, C. L., Helsel, L. O., Gee, J. E., and Hoffmaster, A. R. (2010). Identification and characterization of clinical *Bacillus* spp. isolates phenotypically similar to *Bacillus anthracis*. *FEMS Microbiol. Lett.* 313, 47–53. doi: 10.1111/j.1574-6968.2010.02120.x

Boucherba, N., Gagaoua, M., Bouanane-Darenfed, A., Bouiche, C., Bouacem, K., Kerbous, M. Y., et al. (2017). Biochemical properties of a new thermo- and solvent-stable xylanase recovered using three phase partitioning from the extract of Bacillus oceanisediminis strain SJ3. *Bioresour. Bioprocess.* 4, 29. doi: 10.1186/s40643-017-0161-9

Bustos, A. Y., Font de Valdez, G., Fadda, S., and Taranto, M. P. (2018). New insights into bacterial bile resistance mechanisms: the role of bile salt hydrolase and its impact on human health. *Food Res. Int.* 112, 250–262. doi: 10.1016/j.foodres.2018.06.035

Collado, M. C., Meriluoto, J., and Salminen, S. (2008). Adhesion and aggregation properties of probiotic and pathogen strains. *Eur. Food Res. Technol.* 226, 1065–1073. doi: 10.1007/s00217-007-0632-x

Cuello-Garcia, C. A., Brozek, J. L., Fiocchi, A., Pawankar, R., Yepes-Nuñez, J. J., Terracciano, L., et al. (2015). Probiotics for the prevention of allergy: a systematic review and meta-analysis of randomized controlled trials. *J. Allergy Clin. Immunol.* 136, 952–961. doi: 10.1016/j.jaci.2015.04.031

Dowarah, R., Verma, A. K., Agarwal, N., Singh, P., and Singh, B. R. (2018). Selection and characterization of probiotic lactic acid bacteria and its impact on growth, nutrient digestibility, health and antioxidant status in weaned piglets. *PLoS ONE* 13, e0192978. doi: 10.1371/journal.pone.0192978

Dowdell, P., Chankhamhaengdecha, S., Panbangred, W., Janvilisri, T., and Aroonnual, A. (2020). Probiotic activity of enterococcus faecium and *Lactococcus lactis* isolated from Thai fermented sausages and their protective effect against Clostridium difficile. *Probiotics Antimicrob. Proteins* 12, 641–648. doi: 10.1007/s12602-019-09536-7

Elshaghabee, F. M. F., Rokana, N., Gulhane, R. D., Sharma, C., and Panwar, H. (2017). Bacillus as potential probiotics: status, concerns, and future perspectives. *Front. Microbiol.* 8, 1490. doi: 10.3389/fmicb.2017.01490

Guan, N., and Liu, L. (2020). Microbial response to acid stress: mechanisms and applications. *Appl. Microbiol. Biotechnol.* 104, 51–65. doi: 10.1007/s00253-019-10226-1

Haghshenas, B., Nami, Y., Almasi, A., Abdullah, N., Radiah, D., Rosli, R., et al. (2017). Isolation and characterization of probiotics from dairies. *Iran. J. Microbiol.* 9, 234–243.

Hamon, E., Horvatovich, P., Izquierdo, E., Bringel, F., Marchioni, E., Aoudé-Werner, D., et al. (2011). Comparative proteomic analysis of *Lactobacillus plantarum* for the identification of key proteins in bile tolerance. *BMC Microbiol.* 11, 63. doi: 10.1186/1471-2180-11-63

Hong, H. A., Khaneja, R., Tam, N. M. K., Cazzato, A., Tan, S., Urdaci, M., et al. (2009). *Bacillus subtilis* isolated from the human gastrointestinal tract. *Res. Microbiol.* 160, 134–143. doi: 10.1016/j.resmic.2008.11.002

Jalanka-Tuovinen, J., Salonen, A., Nikkilä, J., Immonen, O., Kekkonen, R., Lahti, L., et al. (2011). Intestinal microbiota in healthy adults: temporal analysis reveals individual and common core and relation to intestinal symptoms. *PLoS ONE* 6, e23035. doi: 10.1371/journal.pone.0023035

Janusz, G., Pawlik, A., Swiderska-Burek, U., Polak, J., Sulej, J., Jarosz-Wilkołazka, A., et al. (2020). Laccase properties, physiological functions, and evolution. *Int. J. Mol. Sci.* 21, 966. doi: 10.3390/ijms21030966

Jaradat, N., Abualhasan, M. N., Qadi, M., Issa, L., Mousa, A., Allan, F., et al. (2020). Antiamylase, antilipase, antimicrobial, and cytotoxic activity of Nonea obtusifolia (Willd.) DC. from Palestine. *BioMed. Res. Int.* 2020, 1–8. doi: 10.1155/2020/8821319

Kalia, V. C., Gong, C., Shanmugam, R., et al. (2022a). The emerging biotherapeutic agent: *Akkermansia*. *Indian J. Microbiol.* 62, 1–10. doi: 10.1007/s12088-021-00993-9

Kalia, V. C., Shim, W. Y., Patel, S. K. S., Gong, C., and Lee, J.-K. (2022b). Recent developments in antimicrobial growth promoters in chicken health: opportunities and challenges, *Sci. Total Environ.* 834, 15530. doi: 10.1016/j.scitotenv.2022.155300

Karami, S., Roayaei, M., Hamzavi, H., Bahmani, M., Hassanzad-Azar, H., Leila, M., et al. (2017). Isolation and identification of probiotic Lactobacillus from local dairy and evaluating their antagonistic effect on pathogens. *Int. J. Pharm. Investig.* 7, 137–141. doi: 10.4103/jphi.JPHI_8_17

Kim, S., Kim, D., Cho, S. W., Kim, J., and Kim, J.-S. (2014). Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome Res.* 24, 1012–1019. doi: 10.1101/gr.171322.113

Kotowicz, N., Bhardwaj, R. K., Ferreira, W. T., Hong, H. A., Olender, A., Ramirez, J., et al. (2019). Safety and probiotic evaluation of two *Bacillus* strains producing antioxidant compounds. *Benef. Microbes* 10, 759–771. doi: 10.3920/BM2019.0040

Krausova, G., Hyrslova, I., and Hynstova, I. (2019). *In vitro* evaluation of adhesion capacity, hydrophobicity, and auto-aggregation of newly isolated potential probiotic strains. *Fermentation* 5, 100. doi: 10.3390/fermentation5040100

Kumar Mondal, A., Kumar, J., Pandey, R., Gupta, S., Kumar, M., Bansal, G., et al. (2017). Comparative genomics of host–symbiont and free-living oceanobacillus species. *Genome Biol. Evol.* 9, 1175–1182. doi: 10.1093/gbe/evx076

Kumar, J., Verma, M. K., Kumar, T., Gupta, S., Pandey, R., Yadav, M., et al. (2018). S9A serine protease engender antigenic gluten catabolic competence to the human gut microbe. *Indian J. Microbiol.* 58, 294–300. doi: 10.1007/s12088-018-0732-2

Kumar, N., Mittal, A., Yadav, M., Sharma, S., Kumar, T., Chakraborthy, R., et al. (2021). Photocatalytic TiO2/CdS/ZnS nanocomposite induces Bacillus subtilis cell death by disrupting its metabolism and membrane integrity. *Indian J. Microbiol*. 61, 487–496. doi: 10.1007/s12088-021-00973-z

Kumar, P., Patel, S. K. S., Lee, J.-K., and Kalia, V. C. (2013). Extending the limits of Bacillus for novel biotechnological applications. *Biotechnol. Adv.* 31, 1543–1561. doi: 10.1016/j.biotechadv.2013.08.007

Kumar, P., Singh, M., Mehariya, S., Patel, S. K. S., Lee, J.-K., and Kalia, V. C. (2014). Ecobiotechnological approach for exploiting the abilities of bacillus to produce co-polymer of polyhydroxyalkanoate. *Indian J. Microbiol.* 54, 151–157. doi: 10.1007/s12088-014-0457-9

Lagier, J.-C., Edouard, S., Pagnier, I., Mediannikov, O., Drancourt, M., and Raoult, D. (2015). Current and past strategies for bacterial culture in clinical microbiology. *Clin. Microbiol. Rev.* 28, 208–236. doi: 10.1128/CMR.00110-14

Lee, N.-K., Kim, W.-S., and Paik, H.-D. (2019). Bacillus strains as human probiotics: characterization, safety, microbiome, and probiotic carrier. *Food Sci. Biotechnol.* 28, 1297–1305. doi: 10.1007/s10068-019-00691-9

Leksmono, C. S., Manzoni, C., Tomkins, J. E., Lucchesi, W., Cottrell, G., and Lewis, P. A. (2018). Measuring lactase enzymatic activity in the teaching lab. *J. Vis. Exp.* 54377. doi: 10.3791/54377

Li, B., Zhan, M., Evivie, S. E., Jin, D., Zhao, L., Chowdhury, S., et al. (2018). Evaluating the safety of potential probiotic Enterococcus durans KLDS6.0930 using whole genome sequencing and oral toxicity study. *Front. Microbiol.* 9, 1943. doi: 10.3389/fmicb.2018.01943

Li, J., Lin, S., Vanhoutte, P. M., Woo, C. W., and Xu, A. (2016). *Akkermansia Muciniphila* protects against atherosclerosis by preventing metabolic endotoxemia-induced inflammation in $Apoe^{-/-}$ mice. *Circulation* 133, 2434–2446. doi: 10.1161/CIRCULATIONAHA.115.019645

Li, Y., Jia, D., Wang, J., Li, H., Yin, X., Liu, J., et al. (2021). Probiotics isolated from animals in Northwest China improve the intestinal performance of mice. *Front. Vet. Sci.* 8, 750895. doi: 10.3389/fvets.2021.750895

Lindefeldt, M., Eng, A., Darban, H., Bjerkner, A., Zetterström, C. K., Allander, T., et al. (2019). The ketogenic diet influences taxonomic and functional

composition of the gut microbiota in children with severe epilepsy. *NPJ Biofilms Microbiomes* 5, 5. doi: 10.1038/s41522-018-0073-2

Mandic, M., Djokic, L., Nikolaivits, E., Prodanovic, R., O'Connor, K., Jeremic, S., et al. (2019). Identification and characterization of new laccase biocatalysts from pseudomonas species suitable for degradation of synthetic textile dyes. *Catalysts* 9, 629. doi: 10.3390/catal9070629

Marcial, G. E., Ford, A. L., Haller, M. J., Gezan, S. A., Harrison, N. A., Cai, D., et al. (2017). *Lactobacillus johnsonii* N6.2 modulates the host immune responses: a double-blind, randomized trial in healthy adults. *Front. Immunol.* 8, 655. doi: 10.3389/fimmu.2017.00655

Nami, Y., Vaseghi Bakhshayesh, R., Mohammadzadeh Jalaly, H., Lotfi, H., Eslami, S., and Hejazi, M. A. (2019). Probiotic properties of enterococcus isolated from artisanal dairy products. *Front. Microbiol.* 10, 300. doi: 10.3389/fmicb.2019.00300

Ndubuisil, M. I., Kwok, B. H. B., Vervoort, J., Koh, B. D., Elofsson, M., and Crews, C. M. (2002). Characterization of a novel mammalian phosphatase having sequence similarity to *Schizosaccharomyces pombe* PHO2 and *Saccharomyces cerevisiae* PHO13. *Biochemistry* 41, 7841–7848. doi: 10.1021/bi0255064

Patel, S. K. S., Kumar, P., Mehariya, S., Purohit, H. J., Lee, J.-K., and Kalia, V. C. (2014). Enhancement in hydrogen production by co-cultures of Bacillus and Enterobacter. *Int. J. Hydrog. Energy* 39, 14663–14668. doi: 10.1016/j.ijhydene.2014.07.084

Plaza-Diaz, J., Ruiz-Ojeda, F. J., Gil-Campos, M., and Gil, A. (2019). Mechanisms of action of probiotics. *Adv. Nutr.* 10, S49–S66. doi: 10.1093/advances/nmy063

Rahman, M., Sabir, A. A., Mukta, J. A., Khan, M.d,., M. A., Mohi-Ud-Din, M., et al. (2018). Plant probiotic bacteria Bacillus and Paraburkholderia improve growth, yield and content of antioxidants in strawberry fruit. *Sci. Rep.* 8, 2504. doi: 10.1038/s41598-018-20235-1

Salehi, B., Machin, L., Monzote, L., Sharifi-Rad, J., Ezzat, S. M., Salem, M. A., et al. (2020). Therapeutic potential of quercetin: new insights and perspectives for human health. *ACS Omega* 5, 11849–11872. doi: 10.1021/acsomega.0c01818

Samedi, L., and Charles, A. L. (2019). Isolation and characterization of potential probiotic Lactobacilli from leaves of food plants for possible additives in pellet feeding. *Ann. Agric. Sci.* 64, 55–62. doi: 10.1016/j.aoas.2019.05.004

Samedi, L., and Linton Charles, A. (2019). Evaluation of technological and probiotic abilities of local lactic acid bacteria. *J. Appl. Environ. Microbiol.* 7, 9–19. doi: 10.12691/jaem-7-1-3

Schroeder, B. O., Birchenough, G. M. H., Ståhlman, M., Arike, L., Johansson, M. E. V., Hansson, G. C., et al. (2018). Bifidobacteria or fiber protects against diet-induced microbiota-mediated colonic mucus deterioration. *Cell Host Microbe* 23, 27–40.e7. doi: 10.1016/j.chom.2017.11.004

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30, 14. doi: 10.1093/bioinformatics/btu153

Sekhon-Loodu, S., and Rupasinghe, H. P. V. (2019). Evaluation of antioxidant, antidiabetic and antiobesity potential of selected traditional medicinal plants. *Front. Nutr.* 6, 53. doi: 10.3389/fnut.2019.00053

Setta, M. C., Matemu, A., and Mbega, E. R. (2020). Potential of probiotics from fermented cereal-based beverages in improving health of poor people in Africa. *J. Food Sci. Technol.* 57, 3935–3946. doi: 10.1007/s13197-020-04432-3

Shobharani, P., and Halami, P. M. (2016). *In vitro* evaluation of the cholesterol-reducing ability of a potential probiotic *Bacillus* spp. *Ann. Microbiol.* 66, 643–651. doi: 10.1007/s13213-015-1146-6

Singh, M., Patel, S. K., and Kalia, V. C. (2009). *Bacillus subtilis* as potential producer for polyhydroxyalkanoates. *Microb. Cell Factor.* 8, 38. doi: 10.1186/1475-2859-8-38

Siraj, N. M., Sood, K., and Yadav, R. N. S. (2017). Isolation and identification of potential probiotic bacteria from cattle farm soil in Dibrugarh District. *Adv. Microbiol.* 7, 265–279. doi: 10.4236/aim.2017.74022

Tang, Q., Jin, G., Wang, G., Liu, T., Liu, X., Wang, B., et al. (2020). Current sampling methods for gut microbiota: a call for more precise devices. *Front. Cell. Infect. Microbiol.* 10, 151. doi: 10.3389/fcimb.2020.00151

Thaiss, C. A., and Elinav, E. (2017). The remedy within: will the microbiome fulfill its therapeutic promise? *J. Mol. Med.* 95, 1021–1027. doi: 10.1007/s00109-017-1563-z

Trunk, T., Khalil, H. S., and Leo, J. C. (2018). Bacterial autoaggregation. *AIMS Microbiol.* 4, 140–164. doi: 10.3934/microbiol.2018.1.140

Valeriano, V. D., Parungao-Balolong, M. M., and Kang, D.-K. (2014). *In vitro* evaluation of the mucin-adhesion ability and probiotic potential of *Lactobacillus mucosae* LM1. *J. Appl. Microbiol.* 117, 485–497. doi: 10.1111/jam.12539

Varankovich, N. V., Nickerson, M. T., and Korber, D. R. (2015). Probiotic-based strategies for therapeutic and prophylactic use against multiple gastrointestinal diseases. *Front. Microbiol.* 6, 685. doi: 10.3389/fmicb.2015.00685

Verma, H., Phian, S., Lakra, P., Kaur, J., Subudhi, S., Lal, R., et al. (2020). Human gut microbiota and mental health: advancements and challenges in microbe-based therapeutic interventions. *Indian J. Microbiol.* 60, 405–419. doi: 10.1007/s12088-020-00898-z

Vlková, E., Rada, V., Smehilová, M., and Killer, J. (2008). Auto-aggregation and co-aggregation ability in bifidobacteria and clostridia. *Folia Microbiol.* 53, 263–269. doi: 10.1007/s12223-008-0040-z

Walker, A. W., and Parkhill, J. (2013). Fighting obesity with bacteria. *Science* 341, 1069–1070. doi: 10.1126/science.1243787

Wang, Z.-B., Xin, S.-S., Ding, L.-N., Ding, W.-Y., Hou, Y.-L., Liu, C.-Q., et al. (2019). The potential role of probiotics in controlling overweight/obesity and associated metabolic parameters in adults: a systematic review and meta-analysis. *Evid. Based Complement. Alternat. Med.* 2019, 1–14. doi: 10.1155/2019/3862971

Woodcock, D. J., Krusche, P., Strachan, N. J. C., Forbes, K. J., Cohan, F. M., Méric, G., et al. (2017). Genomic plasticity and rapid host switching can promote the evolution of generalism: a case study in the zoonotic pathogen Campylobacter. *Sci. Rep.* 7, 9650. doi: 10.1038/s41598-017-09483-9

Yadav, M., and Chauhan, N. S. (2022). Microbiome therapeutics: exploring the present scenario and challenges. *Gastroenterol. Rep.* 10, goab046. doi: 10.1093/gastro/goab046

Yadav, M., Lomash, A., Kapoor, S., Pandey, R., and Chauhan, N. S. (2021). Mapping of the benzoate metabolism by human gut microbiome indicates food-derived metagenome evolution. *Sci. Rep.* 11, 5561. doi: 10.1038/s41598-021-84964-6

Yadav, M., Pandey, R., and Chauhan, N. S. (2020). Catabolic machinery of the human gut microbes bestow resilience against vanillin antimicrobial nature. *Front. Microbiol.* 11, 588545. doi: 10.3389/fmicb.2020.588545

Yadav, M., Verma, M. K., and Chauhan, N. S. (2018). A review of metabolic potential of human gut microbiome in human nutrition. *Arch. Microbiol.* 200, 203–217. doi: 10.1007/s00203-017-1459-x

Yao, S., Zhao, Z., Wang, W., and Liu, X. (2021). Bifidobacterium longum: protection against inflammatory bowel disease. *J. Immunol. Res.* 2021, 1–11. doi: 10.1155/2021/8030297

Zhang, J., Wang, J., Fang, C., Song, F., Xin, Y., Qu, L., et al. (2010). *Bacillus oceanisediminis* sp. nov., isolated from marine sediment. *Int. J. Syst. Evol. Microbiol.* 60, 2924–2929. doi: 10.1099/ijs.0.019851-0

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

TYPE Original Research
PUBLISHED 04 August 2022
DOI 10.3389/fmicb.2022.939984

OPEN ACCESS

EDITED BY
Sanjay Kumar Singh Patel,
Konkuk University, South Korea

REVIEWED BY
Nar Singh Chauhan,
Maharshi Dayanand University, India
Mamtesh Singh,
University of Delhi, India
Pradipta Saha,
University of Burdwan, India

*CORRESPONDENCE
Om Prakash
prakas1974@gmail.com;
omprakash@nccs.res.in

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

RECEIVED 09 May 2022
ACCEPTED 11 July 2022
PUBLISHED 04 August 2022

CITATION
Nimonkar YS, Godambe T, Kulkarni A, Patel T, Paul D, Paul D, Rale V and Prakash O (2022) Oligotrophy vs. copiotrophy in an alkaline and saline habitat of Lonar Lake. *Front. Microbiol.* 13:939984. doi: 10.3389/fmicb.2022.939984

Oligotrophy vs. copiotrophy in an alkaline and saline habitat of Lonar Lake

Yogesh S. Nimonkar[1], Tejashree Godambe[1], Apurva Kulkarni[1], Tarachand Patel[1], Dhreej Paul[1], Debarati Paul[2], Vinay Rale[3] and Om Prakash[1]*

[1]National Centre for Microbial Resource, National Centre for Cell Science, Pune, India, [2]Amity Institute of Biotechnology, Amity University Uttar Pradesh (AUUP), Noida, India, [3]Symbiosis School of Biological Sciences (SSBS) Symbiosis International (Deemed University) & Symbiosis Centre for Research & Innovation (SCRI), Symbiosis International (Deemed University), Pune, India

We reported our comparative observations on oligotrophs vs. copiotrophs from a hyper-alkaline and hypersaline habitat, Lonar Lake, situated in the Buldhana district of Maharashtra, India. Cell numbers of oligotrophic and copiotrophic microbes from the sediment were enumerated by the three-tube most probable number (MPN) method using an array of nutrient-rich and oligotrophic (≈10–20 mg carbon L^{-1}) media offering simulated natural conditions of pH and salinity. A total of 50 strains from 15 different genera and 30 different species were isolated from the highest positive dilutions of MPN to identify the taxa of oligotrophs and copiotrophic microorganisms dominating in Lonar Lake. We did not get any true oligotrophs due to their adaptation to higher carbon levels during the isolation procedure. On the contrary, several true copiotrophs, which could not adapt and survive on a low-carbon medium, were isolated. It is also observed that changes in medium composition and nutrient level altered the selection of organisms from the same sample. Our data indicate that copiotrophic microorganisms dominate the eutrophic Lonar Lake, which is also supported by the past metagenomics studies from the same site. We also reported that quick depletion of carbon from oligotrophic medium worked as a limiting factor, inducing cell death after 2–3 generations and preventing the development of visible colonies on plates and sufficient optical density in liquid medium. Therefore, a long-term supply of low levels of carbon, followed by isolation on enriched media, can serve as a good strategy in isolation of novel taxa of microorganism, with industrial or environmental importance.

KEYWORDS

nutrition, growth-rate, microbial-ecology, industrial enzymes, nutrient cycling, lifestyle switch

Highlights

- The composition and concentration of culture media play an essential role in the selection of microorganisms.
- A continuous supply of low nutrient levels for the long term is imperative to enrich the oligotrophic microbes.

- Isolated oligotrophs with valuable enzyme production potential can be a valuable resource for cost-effective industrial enzyme production.
- Oligotrophs from this study can serve as a tool for OMICS studies to understand microbes' lifestyle and survival tactics under ultra-low nutritional support and can extend this experience to other comparable situations.

Introduction

Microbes have adapted to various growth and survival strategies to survive and reproduce. Ecologically challenging habitats with varying nutrient availability impose intense selective pressure on microbes that determine their lifestyle tactic (Lauro et al., 2009). The response of bacteria to carbon input in a particular ecological niche helps to understand the population dynamics of microbes and their interactions with other biotic and abiotic components (Watve et al., 2000; Shahbaz et al., 2020).

Depending on their ability to adapt to various substrate concentrations, especially carbon, microbes are classified as oligotrophs and copiotrophs (Koch, 1998; Saha and Chakrabarti, 2006; Ohhata et al., 2007). The bacteria able to grow on low nutrient concentrations but unable to survive in nutrient-rich conditions, especially high load of carbon, are known as "oligotrophic bacteria" (Hu et al., 1999; Koch, 2001; Ohhata et al., 2007). Oligotrophs can be designated as "obligates" when they strictly require low concentrations of carbon for growth and "facultative" when they grow in low as well as high levels of carbon (Ishida and Kadota, 1981). Slow growth and metabolic rates, higher substrate affinity, and generally low population density are the characteristics of oligotrophic organisms (Ho et al., 2017). These properties allow them to preponderate in natural habits with limited available/utilizable carbon (Amy and Morita, 1983; Morita, 1988; Fegatella and Cavicchioli, 2000; Ferrari et al., 2005; Temperton et al., 2011).

In contrast, copiotrophs predominate nutrient-rich environments such as sewage lagoons, carbon-rich soils, and decomposing wood. They are adapted to use carbon sources rapidly and are especially suited to habitats with high nutrient flux (Fry, 1990). Thus, the oligotrophs have a competitive advantage over copiotrophs and can flourish in nutrient-limited systems (Button, 1991). Most aquatic and terrestrial habitats are nutrient-depleted and oligotrophic.

Habitat condition, nutrient availability, biodiversity, and unique adaptability (Paul and Clark, 1996) are the key factors influencing the plethora of microbes having the necessary and varied ability for carbon utilization. Oligotrophic and copiotrophic strategies give an insight into the adaptations and strategies employed by microorganisms in response to different nutrient requirements. Thus, studying contrasting survival strategies of both oligotrophs and copiotrophs is essential to track the microbial evolution and for an in-depth understanding of the two modes of livelihood or survival (Koch, 2001). An improved understanding of trophic strategies is also essential to understand the biogeochemical cycles and microbial responses to changes in the bioavailability of resources (Antony et al., 2013). From a historical perspective, it is interesting to note that the development of nutrition strategies for isolation has undergone sea change from the methodologies of the "Delft School" (liberal nutrition) to the demise of the concept of autotrophy *sensu stricto* (autotrophs shown to use organic carbon) (Kelly, 1971).

Lonar Lake, a hypersaline and hyper-alkaline soda lake located in Buldhana district, Maharashtra, India, harbors a unique ecological niche. It is the only crater formed due to the high-velocity meteoritic impact on basaltic rock. Many studies have reported the biodiversity associated with Lonar Soda Lake (Joshi et al., 2007; Pathak and Deshmukh, 2012; Tambekar and Dhundale, 2012; Sharma et al., 2016). Microbes isolated from the soda lake region manifest complex microbial food webs interconnecting various biological cycles *via* redox coupling. The inhabitants have been reported to possess biotechnologically important enzymes and biomolecules (Hinder et al., 1999; Antony et al., 2010; Grant and Sorokin, 2011). However, studies on the trophic interactions between these microbes and their role in the ecosystem are not adequately documented. The microbial diversity of such ecosystems has the potential to throw light on adaptability to extreme conditions and possibly pave the way to hunt for novel and industrially valuable biomolecules. In this study, we have cultured, quantified, identified, and metabolically characterized the oligotrophs and copiotrophs isolated from the hypersaline Lonar Lake sediment samples. In addition, we also tried to investigate the strategy that is prevalently used by indigenous organisms to survive in a unique extreme habitat such as Lonar Lake.

Materials and methods

Site description and sampling

Sediment samples used for cultivation and enumeration studies were collected from the hyper-alkaline and hypersaline Lonar Lake located in Buldhana district, Maharashtra, India (19°59′N, 76°31′E), using the grab sampling method. The sampling site's geographic location and physiochemical features are described anywhere else (Joshi et al., 2008; Antony et al., 2010, 2013, 2014; Surakasi et al., 2010; Borul, 2012; Paul et al., 2016; Bagade et al., 2020). For microbiological investigation, sediment samples from different locations of Lonar Lake were aseptically collected in pre-sterilized Falcon tubes, stored immediately on ice, and subsequently transported to the laboratory within 24 h. For cultivation and quantification

purposes, samples were stored at 4°C and processed within 4 days.

Cultivation and enumeration of bacteria

Oligotrophic and copiotrophic bacterial populations from the sample were enumerated by using a three-tube most probable number (MPN) method (Böllmann and Martienssen, 2020). For this, samples were serially diluted using oligotrophic and nutrient-rich growth media. Media such as nutrient broth (NB), Luria broth (LB), tryticase soy broth (TSB), and Ravan and Reasoner's 2A agar (R2A agar) were used for copiotrophs. For oligotrophs NB, LB, and TSB (1,000-fold diluted); (R2A) (100-fold diluted); Ravan (10-fold diluted); and a synthetic medium (fortified with 20 mg carbon L^{-1}) were deployed for the cultivation and enumeration of oligotrophic organisms (Tada et al., 1995; Saha and Chakrabarti, 2006; HuiXia et al., 2007; Matsuoka and Yoshida, 2018). Compositions of Ravan and the synthetic groundwater media have been reported previously by Nagarkar et al. (2001) and Green et al. (2010), respectively. Diluted media were supplemented with filter-sterilized vitamins and trace elements (Widdel and Bak, 1992). To mimic the *in situ* pH and salinity conditions of Lonar Lake, medium pH and salinity were adjusted to 9 and 1% (w/v), respectively. Carbon concentration of diluted oligotrophic medium was tentatively estimated by dividing the actual nutrient load of rich medium with dilution factors.

MPN series were prepared in loosely capped tubes and incubated aerobically at 30°C for 72 h in case of copiotrophs and up to 1 month for oligotrophs to grow. Growth of the isolates was monitored visually by observing the change in turbidity compared to uninoculated autoclaved controls. MPN was determined from statistical tables published by the American Public Health Association (https://microbeonline.com/probable-number-mpn-test-principle-procedure-results/). For isolation of dominant bacterial groups cultivated in different media, 100 µl aliquots from the highest positive dilutions of MPN series were spread plated on respective media to be incubated at 30°C till the emergence of colonies. Culture purity was obtained by repeated subculturing on respective solid media. Isolated pure cultures were preserved at −80°C with 20% glycerol (Prakash et al., 2013).

DNA extraction, 16S rRNA gene sequencing, and sequence analysis

Total genomic DNA of all the purified strains was extracted using Purelink-Pro 96-well genomic DNA isolation kit (Invitrogen, USA), following the instructions of the manufacturer. PCR amplification of 16S rRNA gene of the isolated genomic DNA was done using bacteria-specific universal primers 27F (5'-AGAGTTTGATCMTGGCTCAG-3') and 1492R (5'-TACGGYTACCTTGTTACGACTT-3') (Marathe et al., 2013). For 100 µl PCR reaction, 100 ng genomic DNA, 10 µl 10× buffer, 5 µl $MgCl_2$ (25 mM), 2 µl dNTP mix (10 mM), 10 moles of each primer, and 2U of Taq polymerase (Promega, USA) were used. The PCR reaction was done with an initial denaturation condition at 94°C for 5 min; followed by 30 cycles of denaturation at 94°C for 60 s, annealing at 58°C for 45 s, elongation at 72°C for 90 s, and final elongation at 72°C for 7 min. Amplified products were purified by polyethylene glycol method. Thereafter, sequencing was performed using Sanger's method (Big-dye terminator chemistry) as discussed in Prakash and Lal (2013) and Prakash et al. (2014).

The obtained sequences were manually checked for authenticity, and contigs were generated with the SeqMan program of DNASTAR. Identity of isolated pure cultures was determined by similarity search of the obtained sequences using the EzTaxon database (http://www.ezbiocloud.net/eztaxon) (Kim et al., 2013) and BLAST (NCBI) (http://blast.ncbi.nlm.nih.gov/Blast.cgi) program.

Determination of physiological features of the strains

Physiological parameters, such as range and optima of temperature, salinity, and pH, were determined according to Prakash et al. (2015). Oligotrophic, facultative oligotrophic, and copiotrophic nature of the strains was determined by supplying different concentrations (20, 200, 2,000, 20,000 mg L^{-1}) of carbon source to the growth medium with optimum nutrient, pH, and salinity. Plates were incubated at 30°C for 8 days, and growth response of the strains was recorded. It is speculated that during dilution of the complex medium the concentration of key nutrient elements such as carbon, nitrogen, and phosphorus is depleted and may adversely affect growth. To test the effect of nitrogen and phosphorus on bacterial growth, 100 mg L^{-1} NH_4Cl and K_2HPO_4 were added to the diluted medium. Facultative oligotrophs cultivated in a diluted medium were streaked on the plate, and their growth response was evaluated compared to unsupplemented control (Smith and Prairie, 2004).

Screening for enzyme production

An array of enzymatic activities such as DNase, amylase, protease, urease, cellulase, phosphatase, and gelatinase of the isolated strains was tested using standard protocols described in Smibert and Krieg (1994) unless mentioned otherwise. For DNase activity, bacterial strains were incubated in a medium containing tryptose 20 g L^{-1}, toluidine blue 0.1 g L^{-1}, and DNA 2 g L^{-1}. After 48 h incubation, the plates were flooded with

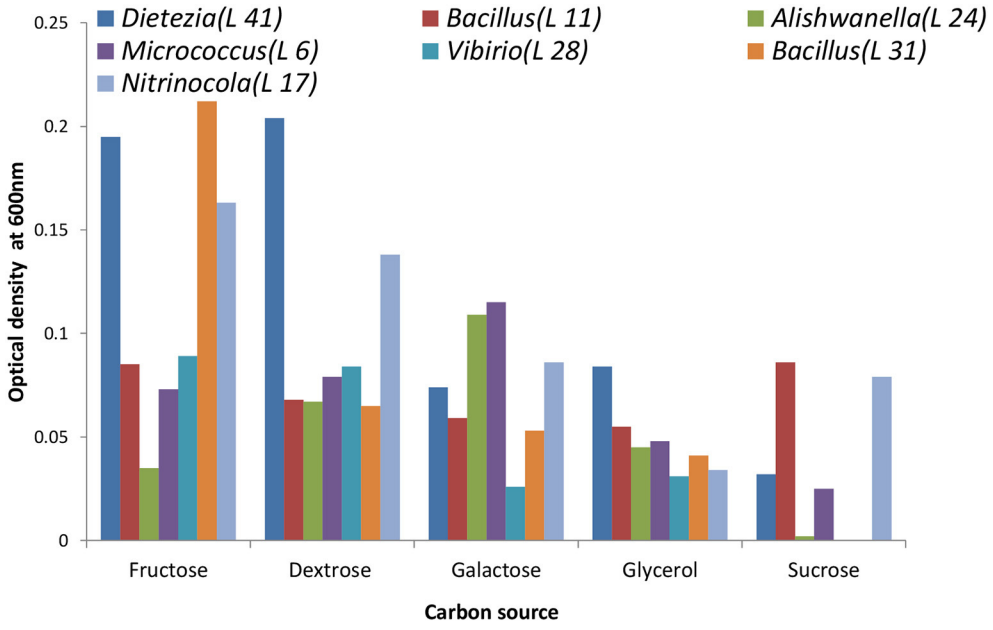


FIGURE 1
Utilization of different carbon sources by the isolates. Dextrose proved to be the most suitable one for stable growth.

1 N HCl solution. DNase-positive strains showed a clear zone around the colonies on plates. Amylolytic activity of the strains was tested using a starch agar medium. After 48 h incubation at 37°C, the plates were flooded with 0.6% KI solution, and a clear zone was observed around a positive strain due to the hydrolysis of starch. Proteolytic activities of the cultures were screened using 1% skim milk containing nutrient agar medium. A zone of clearance around the colonies appearing over the next 48 h was considered protease positive. Urease production was screened on *a urea* agar base (Oxoid) supplemented with 40% urea stock. The appearance of a zone of clearance encircling colonies after incubating for 48 h was considered positive. For cellulase production, carboxymethylcellulose agar was used. After 48 h of incubation, we flooded the plates with 0.03% Congo Red to look for a clear zone around the colonies. Screening of phosphatase activity was carried out on Pikovskaya agar medium, where positive strains were determined by visualization of a clear zone after 48 h incubation. Similarly, isolated strains were inoculated in NB with 12% gelatin for the gelatin assay and the organism's ability to liquefy gelatin after 48 h incubation was determined.

Carbon source optimization

It is recommended that for a comparative study of growth rate and generation time all strains should be cultivated on a common carbon source using the same type of synthetic medium and incubation condition. For that purpose, growth response of selected strains on different sugars (carbon source) was evaluated before the actual experiment. For this purpose, the chosen synthetic groundwater medium (Green et al., 2010) was supplemented with 50 mM of different sugars (dextrose, fructose, galactose, glycerol, and sucrose) as a carbon source, and an increase in optical density at 600 nm was detected using 96-well plate reader Bioscreen-C (Oy Growth Curves Ab Ltd., Finland). In brief, we prepared a synthetic groundwater medium, adjusted the pH to 8 using 1 N NaOH, and sterilized it by autoclaving. The medium was supplemented separately with 50 mM of filter-sterilized different sugars as a carbon source and inoculated with an equal volume of (50 μl) of culture inoculum (OD 0.5 at 600 nm) from all oligotrophic isolates in replicates of three. Plates were incubated in the incubator shaker at 30°C for 48 h at 150 rpm. Growth was monitored at different time points using a spectrophotometer at 600 nm. We then plotted clustered column chart using time and optical density data to determine the growth response of selected oligotrophs on different carbon sources (Figure 1).

Growth kinetics at different carbon levels

Among the tested carbohydrates, dextrose gives a satisfactory growth response to all the selected oligotrophs and was chosen for the growth kinetics experiment. The experiment was conducted in synthetic groundwater medium

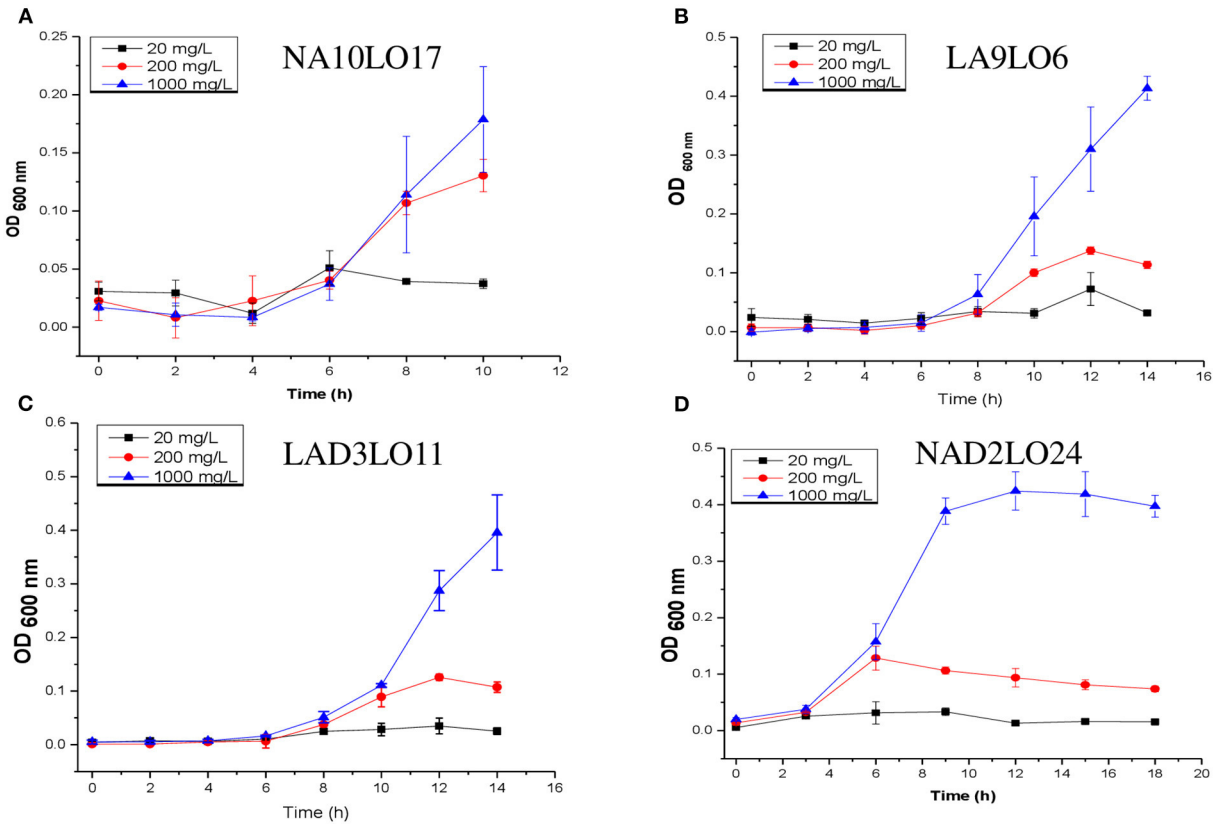


FIGURE 2
(A–D) Growth of **(A)** NA10LO17, **(B)** LA9LO6, **(C)** LAD3LO11, and **(D)** NAD2LO24 in different carbon concentrations (20, 200, and 1,000 mg L^{-1}).

using the different concentrations of dextrose as sources of carbon and energy and calculated the generation times from obtained optical density or CFU data (Figures 2A–D). In brief, synthetic groundwater medium was prepared and supplemented with different concentrations (20, 200, and 1,000 mg L^{-1}) of filter-sterilized dextrose. A volume of 20 ml culture medium was taken in a 100-ml capacity flask, inoculated with 1% inoculum (OD at 600 nm = 0.5), and incubated at 30°C at 150 rpm. Growth at different time intervals (0, 2, 4, 6, 8, …, 24 h) was monitored using a 96-well plate spectrophotometer (at 600 nm). Optical density was plotted against time to calculate the generation time and growth rate. The number of generations (n) was calculated using the formula $n = 3.3 \log b/B$ ("b" represents the optical density at the start of the exponential phase and "B" represents the optical density at the end of the exponential phase). Generation time (G) was calculated using G = t/n, where "t" represents a time period of exponential phase, and "n" represents the number of generations. The growth rate constant was calculated by K = 1/G, where "G" represents generation time. This relationship is valid only during the period when the population is in the exponential phase. Therefore, for different isolates, the "t" varied and the onset of the exponential phase also varied.

Statistics and sequence accession numbers

All the physiological data related to salinity, pH, and temperature were generated using a minimum of three biological replicates using a similar set of conditions. Experiments related to the screening of extracellular enzyme production were also conducted using multiple replicates, and the reported observations are based on the reproducibility of data. The experiments related to the screening of suitable carbohydrates for growth and growth kinetics were conducted in triplicates. The average and standard deviation of the data were calculated using Excel, the graph was generated in Excel, and the standard deviation in data was presented at each point of the graphs. The sequences retrieved from this study have been deposited in the GenBank database under accession no. ON562524 - ON562538, ON905511, ON905512, ON905513,

TABLE 1 Comparative study of bacterial load in oligotrophic and nutrient-rich mediums detected by MPN method and list of representative taxa cultivated from highest positive dilutions.

Media	Highest positive dilution	Cell number (MPN count)	Isolates
NB	10^{-9}, 10^{-10}, 10^{-11}	2.15×10^{10}	*Nitrincola*
LB	10^{-7}, 10^{-8}, 10^{-9}	$<24 \times 10^{8}$	*Pantoea, Micrococcus, Alishewanella*
TSB	10^{-9}, 10^{-10}, 10^{-11}	0.75×10^{10}	*Bacillus*
R2A	10^{-7}, 10^{-8}, 10^{-9}	4.62×10^{8}	*Alishewanella*
RAVAN	10^{-7}, 10^{-8}, 10^{-9}	1.5×10^{8}	*Vibrio*
Synthetic	10^{-6}, 10^{-7}, 10^{-8}	2.4×10^{7}	*Halomonas*
NB-D	10^{-3}, 10^{-4}, 10^{-5}	2.40×10^{4}	*Dietzia, Exiguobacterium*
LB-D	10^{-4}, 10^{-5}, 10^{-6}	0.93×10^{5}	*Bacillus, Rheinheimera*
TSB-D	10^{-4}, 10^{-5}, 10^{-6}	0.42×10^{5}	*Staphylococcus, Alishewanella*
R2A-D	10^{-3}, 10^{-4}, 10^{-5}	4.62×10^{4}	*Vibrio*
RAVAN-D	10^{-3}, 10^{-4}, 10^{-5}	2.4×10^{4}	Not cultivated

NB-D, LB-D, and TSB-D = 1,000-fold diluted; R2A = 100-fold diluted; Ravan-D = 10-fold diluted.

ON905514, ON905515, ON905516, ON905517, ON905518, ON905519, ON905520, ON905521, ON905522, ON905523, ON905524, ON905525, ON905526, ON905527, ON905528, and ON905529.

Results

Isolation, identification, and characterization

As expected, MPN counts of organisms cultivated in nutrient-rich media sets of media doubled are 2-fold higher than those grown in nutrient-deficient media. A total of 49 bacterial strains from the highest positive dilutions of MPN series made in different culture media were isolated in this study. Isolates obtained on different media used are shown in Table 1. Genera such as *Pantoea, Micrococcus, Alishewanella*, *Nitrincola,* and *Bacillus* were dominant from the highest dilutions (Table 1). 16S rRNA gene sequencing approaches showed that isolated strains belong to 30 different species from 15 genera, including *Bacillus, Exiguobacterium, Dietzia, Staphylococcus, Enterobacter, Vibrio, Penibacillus, Halomonas, Lysinibacillus, Pseudomonas, Pantoea, Alishewanella, Rheinheimera, Micrococcus,* and *Nitrincola*. Eight species of the genus *Bacillus* were detected among the isolated strains, while only 1–3 species of other genera were present. Tables 2, 3 show the strains' taxonomical affiliations and basic physiochemical traits. It was observed that all the isolated species of *Bacillus, Penibacillus,* and *Lysinibacillus* showed the potential to grow from low to high (20–2,000 mg L^{-1}) levels of carbon, while some species of the genus *Exiguobacterium, Staphylococcus, Vibrio, Halomonas,* and *Pseudomonas* were unable to grow at a carbon concentration of 20 mg L^{-1}. At 200 mg L^{-1} of carbon concentration, bacterial members *Pantoea*, *Alishewanella*, *Rheinheimera*, *Micrococcus,* and *Nitrincola* were unable to grow, indicating their true copiotrophic nature. Members of *Pantoea*, *Micrococcus*, *Nitrincola*, *Bacillus*, *Alishewanella*, and *Vibrio* showed the highest cell density (10^{-8} to 10^{-10}), and it indicated that they are mostly dominating in Loner Lake sediment. Organisms in oligotrophic media showed low cell density compared to organisms of carbon-rich media either due to longer generation time or suppression of growth due to excessive carbon shock. It was observed that *Bacillus* and *Alishewanella* grew on both lower and higher carbon concentration, but there was a stark difference in their growth pattern; *Alishewanella* quickly attained exponential phase, but this phase was of shorter duration (~2.5 h) compared to *Bacillus*, which showed a long lag phase but also a prolonged exponential phase (6 h) when supplemented with low levels of C (Figure 2). In contrast, *Vibrio*, *Halomonas*, and *Exiguobacterium* were isolated only from the oligotrophic medium. All 49 isolates were able to grow at a 3% salt concentration. Except *for Bacillus marisflavi*, *Penibacillus hunanensis*, *Paenibacillus massiliensis*, *Lysinibacillus macrolides*, *Pseudomonas stutzeri*, *Pantoea septica*, *Alishewanella tabrizica*, and *Pseudomonas stutzeri*, all other isolates showed growth at 10% NaCl concentration and indicated that they are well-adapted for the saline condition of the lake. In addition, *Halomonas ventosae*, *Halomonas kenyensis*, *Paenibacillus massiliensis*, and *Staphylococcus gallinarum* grew at 20% NaCl concentration, although it is yet to be proven whether they are true halophiles or halotolerant. Evaluation of growth at different pH values indicated that all the strains showed growth between pH 4 and 10. While *Exiguobacterium*, *Nitrincola*, *Pseudomonas,* and some species of *Bacillus* did not show any growth at pH 7 but only at pH 10 and are likely to be true alkalifiles. Only a few strains showed growth between pH 4 and 10. Enzyme production was screened to relate the role of oligotrophic and copiotrophic bacteria in that particular ecological habitat (Lonar Lake). Enzyme-producing ability varied amongst the isolates, and protease production was more prevalent, followed by the urease. This observation corroborated with the previous observation that *Bacillus* was the dominating genus found in Lonar Lake sediment (Table 4) and has been widely reported for its ability to produce alkaline protease (Pathak and Deshmukh, 2012).

We also studied the effect of additional sources of nitrogen and phosphorous on the growth of facultative oligotrophs. In some cases, bacterial growth was enhanced after adding extra nitrogen, while excess phosphorous seemed to limit or deter a few isolates, but it did not affect the growth of other isolates. This

TABLE 2 List of facultative oligotrophs and their physiological growth response on different levels of carbon, NaCl, and pH.

Isolate No.	Similarity with closest relative (%)	Carbon (≈mg L^{-1})			% NaCl Tolerance					pH tolerance		
		20	200	2,000	3	5	10	15	20	4	7	10
7	*Bacillus oceanisediminis* (99.8)	+	+	+	+	+	+	+	–	–	–	+
8	*Exiguobacterium aurantiacum* (100)	+	+	+	+	+	+	–	–	–	–	+
9	*Exiguobacterium himgiriensis* (95.8)	ND	ND	+	+	+	+	+	–	–	–	+
10	*Exiguobacterium aurantiacum* (100)	+	+	+	+	+	+	+	–	–	–	+
11	*Bacillus marisflavi* (100)	+	+	+	+	+	+	+	–	–	–	+
12	*Bacillus marisflavi* (99.4)	ND	+	+	+	–	–	–	–	–	+	+
22	*Exiguobacterium aurantiacum* (99.9)	+	+	+	+	+	+	–	–	–	–	+
23	*Bacillus clausii* (95.2)	+	+	+	+	+	+	–	–	–	+	+
24	*Dietzia cercidiphylli* (100)	+	+	+	+	+	+	–	–	–	+	+
25	*Staphyllococcus gallinarum* (100)	+	+	+	+	+	+	–	–	–	+	+
27	*S. gallinarum* (97.3)	ND	ND	+	+	+	+	+	+	+	–	+
30	*Bacillus altitudinis* (100)	+	+	+	+	+	+	–	–	–	+	+
32	*Enterobacter cancerogenus* (99.7)	+	+	+	+	+	–	–	–	+	+	+
33	*Vibrio metschnikovii* (98.2)	+	+	+	+	+	+	–	–	–	+	+
34	*Penibacillus hunanensis* (96.1)	+	+	+	+	+	–	–	–	–	+	+
35	*Paenibacillus massiliensis* (96.3)	ND	ND	+	+	+	–	+	+	–	+	+
36	*Paenibacillus massiliensis* (94.3)	ND	ND	+	+	+	–	–	–	–	+	+
37	*Bacillus cereus* (100)	+	+	+	+	+	+	+	–	–	+	+
38	*Bacillus safensis* (99.7)	+	+	+	+	+	+	+	–	–	+	+
39	*Bacillus thuringiensis* (96.6)	+	+	+	+	+	+	–	–	–	+	+
40	*Bacillus cereus* (100.0)	+	+	+	+	+	–	–	–	–	+	+
43	*Bacillus firmus* (99.4)	ND	ND	+	+	+	–	–	–	–	–	+
44	*Halomonas kenyensis* (98.8)	+	+	+	+	+	+	+	+	–	+	+
45	*Halomonas kenyensis* (98.8)	+	+	+	+	+	+	+	+	–	+	+
47A	*Lysinibacillus macrolides* (99.6)	+	+	+	+	+	–	–	–	–	+	+
47B	*Pseudomonas stutzeri* (99.6)	+	+	+	+	+	–	–	–	–	+	+

ND, no data.

is a known phenomenon that few organisms cease to grow in the presence of increased P- concentrations.

The screening of different sugars and glycerol as sources of carbon and energy showed that all the selected organisms grew satisfactorily on 50 mM dextrose. Out of seven, five showed almost similar ODs while the ODs of *Dietzia* and *Nitrincola* were higher than the other selected strains (Figure 1). Based on this observation, 50 mM dextrose was selected to study the growth kinetics. During our growth kinetics experiments, we were unable to detect any visible colonies on the agar plate supplemented with 20 mg L^{-1} carbon and could not determine their growth from the selected broth medium spectrophotometrically due to negligible optical density. The viability count was obtained using the CFU method on a high carbon medium to address this problem because all oligotrophs eventually adapt to higher carbon concentrations.

Furthermore, initially, we decided to take the CFU after every 2 h but did not get any colony due to immediate consumption of carbon from oligotrophic medium and entry of cells in death or viable but non-culturable (VBNC) stage. To address this, the sampling duration was curtailed to 30 min, which allowed determining the growth trend of organisms at low carbon concentration in an oligotrophic medium (Table 4). Our data also indicate that except *Nitrincola* the generation time of other strains shortened and the growth rate increased at higher carbon (200–1,000 mg L^{-1}) (Figure 2), but it was almost similar at 200 and 1,000 mg L^{-1} carbon.

Discussion

There is no commonly accepted definition for oligotrophic organisms specifying their exact carbon requirement (Poindexter, 1981; Schut et al., 1997; Hartke et al., 1998; Lee et al., 2021). According to Poindexter (1981), oligotrophs grow on a medium containing 1–15 mg of organic carbon L^{-1}. Upon subsequent culturing, they become cultivable in nutrient-rich medium (Poindexter, 1981; Van Gemerden and Kuenen,

TABLE 3 List of copiotrophic organism and their growth response on different levels of carbon, NaCl, and pH.

Isolate No.	Similarity with closest relative (%)	Carbon (≈mg L^{-1})			% NaCl					pH		
		20	200	2,000	3	5	10	15	20	4	7	10
1	*Pantoea septica* (99.78)	–	+	+	+	+	+	–	–	+	+	+
2	*Pantoea septica* (98.07)	–	+	+	+	+	–	–	–	–	+	+
3A	*Alishewanella tabrizica* (97.98)	–	+	+	+	+	–	–	–	–	+	+
3B	*Alishwanella tabrizica* (98.22)	–	+	+	+	+	–	–	–	–	+	+
4	*Rheinheimera longhuensis* (98.69)	–	+	+	+	+	+	–	–	+	+	+
6	*Micrococcus yunnanensis* (99.65)	–	+	+	+	+	+	–	–	–	+	+
13	*Rheinheimera longhuensis* (97.90)	–	+	+	+	+	+	–	–	–	+	+
14	*Micrococcus yunnanensis* (99.44)	–	+	+	+	+	+	+	–	–	+	+
17	*Nitrincola lacisaponensis* (99.71)	–	–	+	+	+	+	–	–	–	–	+
18	*Nitrincola lacisaponensis* (99.54)	–	–	+	+	+	+	–	–	–	–	+
19	*S. epidermidis* (100)	–	+	+	+	+	+	+	–	–	+	+
20	*S. epidermidis* (100)	–	+	+	+	+	+	+	–	+	+	+
21	*Exiguobacterium aurantiacum* (96.58)	–	–	+	+	+	+	+	–	–	–	+
28	*Alishewanella tabrizica* (98.18)	–	+	+	+	+	+	+	–	–	+	+
31	*Bacillus anthracis* (99.7)	ND	ND	+	+	+	+	–	–	+	+	+
41	*Vibrio metschnikovii* (98.64)	–	+	+	+	+	–	–	–	–	+	+
42	*Vibrio metschnikovii* (97.5)	–	+	+	+	+	–	–	–	–	+	+
46A	*Staphylococcus nepalensis* (98.57)	–	–	+	+	+	–	–	–	–	+	+
46B	*Pseudomonas stutzeri* (99.93)	–	–	+	+	–	–	–	–	–	–	+
46C	*Pseudomonas stutzeri* (99.93)	ND	ND	+	+	–	–	–	–	–	–	+
48	*Alishewanella solinquinati* (98.78)	–	–	+	+	+	–	–	–	–	+	+
49	*Halomonas ventosae* (97.79)	–	+	+	+	+	+	+	+	–	+	+
50	*Pantoea septica* (99.64)	–	+	+	+	+	+	–	–	–	+	+

1984; Schut et al., 1997; Hartke et al., 1998; Gao and Wu, 2018). Fry (1990) defined oligotrophy in marine environments where the carbon requirement was <10 mg L^{-1}. Hence, the carbon requirement of the oligotrophs may depend on the ecosystems, whether terrestrial or marine (Fry, 1990). In this study, the habitat, i.e., Lonar Lake, is a highly alkaline and saline soda lake. Because of its uniqueness, several studies related to geochemistry, cultivation, and characterization of microbial diversity from sediment and water samples of Lonar Lake have been carried out (Joshi et al., 2008; Antony et al., 2010, 2013, 2014; Surakasi et al., 2010; Paul et al., 2016; Bagade et al., 2020). However, there are no reports of oligotrophs/copiotrophic populations dominating the hypersaline and hyper-alkaline Lonar Lake ecosystem and their relation to geochemical features of the site, which may play cryptic roles in promoting them. In this study, we tried to elucidate whether k-strategist or r-strategist type of organisms dominate in hyper-alkaline and hypersaline habitats of Lonar Lake.

To quantify and differentiate the bacterial population growing at high and low levels of carbon in the sediment of Lonar Lake, MPN study was conducted. MPN study indicated that undiluted complex media like NB, LB, TSA, and R2A supported a maximum cell count of 10^{-8} to 10^{-10} per gram of wet weight of the sediment, while diluted complex media and other oligotrophic media used during this study supported only 10^{-4} to 10^{-7} cells per gram of wet weight. In addition, our result also suggests that the composition and concentration of media components play an essential role in selecting organisms from a habitat or while counting organisms using the MPN method. For instance, LB and NB components, though almost identical but owing to the double strength of LB *Pantoea, Micrococcus,* and *Alishewanella,* were favored while diluted NB promoted only *Nitrinicola*. Similar was the case with oligotrophic media like SGWM (promoted *Halomonas*), Ravan (promoted only *Vibrio*), and R2A (promoted *Alishewanella*) at higher dilution. Our observations indicated that though the genera *Nitrinicola* and *Bacillus* were dominant in the Lonar Lake they could not be manifested in TSB purely due to the media effect. Therefore, it may prove to be safer and wiser to include multiple media to get a fairer picture of cell count and dominance. Growth of certain organisms such as *Nitrinicola*, *Exiguobacterium*, *Staphylococcus*, *Pseudomonas,* and *Alishewanella* was deterred by carbon concentrations ≤200 mg L^{-1} and thus they may be considered true copiotrophs. The true oligotrophs that stopped

TABLE 4 Doubling time (G) and growth rate (k) of selected facultative oligotrophs on different carbon levels.

Strain ID.	Carbon (≈mg L^{-1})	Doubling Time G (h)	Growth Rate ($k = 1/G$ (h^{-1})
Nitrincola strain	20	0.96	1.04
NA10LO17	200	2.00	0.50
	1,000	1.45	0.69
Micrococcus strain LA9LO6	20	3.56	0.28
	200	1.33	0.75
	1,000	1.13	0.89
Bacillus strain LAD3LO11	20	3.17	0.32
	200	1.38	0.73
	1,000	1.86	0.54
Alishwanella strain	20	15.08	0.07
NAD2LO24	200	1.51	0.66
	1,000	2.67	0.37
Vibirio strain R2A8LO28	20	7.67	0.13
	200	6.14	0.16
	1,000	4.02	0.25
Bacillus strain TS9LO31	20	6.26	0.16
	200	1.78	0.56
	1,000	2.24	0.45
Dietezia strain RAV5LO41	20	1.88	0.53
	200	1.49	0.67
	1,000	2.22	0.45

growing ≥200 mg carbon L^{-1} were not found (Tables 2, 3). We also found that all the strains isolated using oligotrophic media could subsequently adapted to high carbon concentration and thus proved their ability to grow in complex carbon media. The lower cell count of these organisms in the MPN indicated their sustainability to high concentrations of carbon. However, due to the short generation time and faster growth rate of their competitor copiotrophs in the given milieu, these oligotrophs were kept at bay.

Based on our MPN results, we can predict that copiotrophy may dominate over oligotrophy in the Lonar Lake since the count of oligotrophic organisms (organisms in diluted medium) was lower than the number of copiotrophic (organisms growing in undiluted medium) organisms cultivated from the same sample. Physicochemical characterization of Lonar Lake in the past has revealed high Biological Oxygen Demand (BOD) and Chemical Oxygen Demand (COD) levels confirming high organic carbon facilitating eutrophic nutritional status. This is reflected in the MPN analyses favoring copiotrophs (Antony et al., 2010; Sharma et al., 2016).

The 20 isolates cultivated from oligotrophic media also show the ability to grow in the complex medium and can be classified as the second category of oligotrophs or facultative oligotrophs (Poindexter, 1981; Kuznetsov et al., 2003). Interestingly, we have reconfirmed earlier observations of Senechkin et al. (2010). These investigators deployed 200 different isolates, initially grown on oligotrophic media, to subsequently adapt and grow on highly nutritive media for 10 successive generations (accept two strains). Conversely, we also observed that six isolates obtained from undiluted rich media like MPN were unable to grow at carbon concentration lower than 200 mg L^{-1} and thus may prove to be true copiotrophs.

MPN data, along with cultivation, showed that *Bacillus* was predominant, followed by *Staphylococcus* and *Exiguobacterium*. As a matter of fact, the recovery of *Bacillus* from copiotrophic and oligotrophic media and its survivability in a wide spectrum of salinity and nutrients unequivocally establishes its ubiquity in various habitats (Lazar and Dumitru, 1973; Laiz et al., 2000, 2002a,b; Saiz-Jimenez and Laiz, 2000). A review of the past literature (Incerti et al., 1997; Joshi et al., 2008; Antony et al., 2010, 2013, 2014; Surakasi et al., 2010; Paul et al., 2016; Bagade et al., 2020) indicates that the closest relatives of our isolates such as *Nitrinicola*, *Alishwanella*, *Micrococcus*, *Halomonas,* and *Vibrio* were detected by culture-independent approach. This study demonstrated that they adapted to grow at higher alkalinity and salinity, and therefore they dominated in the Lonar Lake ecosystem.

Our results lend credence to the general notion that the cultivation of oligotrophs *sensu stricto* in low carbon regime is laborious and demands appropriate skills. Due to low cell biomass and optical density in the oligotrophic medium, it is difficult to visualize their growth with unaided eyes or spectrophotometrically, and they are left unattended by the usual isolation procedure. Furthermore, the data of culturing the strains on 20 mg L^{-1} carbon and the inability to form visible colonies after 2 h sampling indicated that despite the slow growth rate organisms utilized the available carbon in 2–3 generations and entered the death phase due to nutrient limitation, leading to the absence of visible colonies/detectable OD. Therefore, the strategy to supply low levels of nutrients continuously till visible OD was attained should be used for the successful cultivation of oligotrophs and also prevents contamination with copiotrophs, which only use high carbon-rich media. In addition, due to nutritional shock, true oligotrophs might not grow in a rich carbon medium when transferred directly from an oligotrophic environment.

Consequently, the pure oligotrophic strains could not be obtained as viable pure cultures. Therefore, it is recommended here that after successful cultivation or enrichment of cells in the liquid oligotrophic medium, they should be gradually exposed to high carbon concentrations (e.g., 50, 100, 200, 500, and 1,000 mg L^{-1}) to obtain visible colonies on solid media plate and for their successful isolation and purification. The data on growth rate suggested that low as well as high C levels both limited the growth rate of the organisms, and we can obtain sufficient biomass without overloading the medium with extra carbon. These observations might pave the way for

developing advanced protocols for the development of low-cost media and to add new link to the recent culturomics approaches. Media formulation similarities can be drawn with those from bioprocess engineering.

Life strategies and characteristic features of oligotrophic and copiotrophic organisms have been discussed in detail by Ho et al. (2017), which are very close to our findings that microbes cultivated using oligotrophic conditions get adapted to high nutrient conditions with gradual exposure to increasing concentrations of organic load in the medium (Ho et al., 2017). Using the computational model, Norris et al. (2021) demonstrated the occurrence of binding proteins is essential for oligotrophic life strategies (Newton and Shade, 2016; Norris et al., 2021). Weissman et al. (2021) estimated bacterial growth rates using genome sequences from cultured and uncultured representatives and concluded that oligotrophy and copiotrophy is a way of life strategies microorganisms adapt to sustain themselves in a particular habitat (Weissman et al., 2021). As rightly surmised, oligotrophic is not a taxonomic trait but a physiological adaptation and can be used as an attribute to capture the novel, not-yet-cultured slow-growing organisms (Bartelme et al., 2020). Oligotrophs can be a good source of industrial biomolecules, polysaccharides, and biofuels and can play a crucial role in waste management, but more in-depth studies are required to use them as a resource for these purposes (Porwal et al., 2008; Prakash et al., 2018; Lee et al., 2020).

Conclusion

The majority of organisms in Lonar Lake sediment are essentially copiotrophs though facultative oligotrophs are also present. The dominance of copiotrophs indicates the eutrophic nature of the ecosystem. A distinct paucity of information appears to exist on the trophic organismic interactions from the peculiar soda lake and their overall contribution to the maintenance of the ecosystem health. Naturally, these aspects deserve deeper investigation. We showed that members of the same genus behave differently toward varied carbon regimes and may perform either as oligotrophic or copiotrophic suiting to their adaptation strategies of the habitat.

One can appreciate this new nutrition "economy classes" of oligotrophs that can become strong contenders to the traditional workhorses to address specific issues like low-cost production of industrial biomolecules and need extra attention for their cultivation and preservation. Our study opens new insights into the incipient contributions of "unculturable" or rarely culturable microbes in balancing the carbon and other nutrient cycles by seemingly easy crossovers from oligotrophy to copiotrophy in tune with the increasing carbon levels due to human interventions in the lake's ecosystem.

Most of the transcriptomics and metaproteomics functionality data hinge around rich media used for cultivation. This has left a yawning knowledge gap on the actual functionality of oligotrophs in the natural environs of soil and water. We also propose that the oligotrophs handled by us can be used as model systems for functionality studies in the natural oligotrophic conditions of soil and water on OMICS platforms.

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: NCBI - ON562524 - ON562538 and ON905511 - ON905529.

Author contributions

Material preparation, data collection, and analysis were performed by YN, TG, AK, TP, and OP. The first draft of the manuscript was written by YN. Draft was edited by DhP, DeP, VR, and OP. Study conception and design was contributed by OP. All authors commented on previous versions of the manuscript and read and approved the final manuscript.

Funding

This study was supported by the Grant (BT/Coord.II/01/03/2016) provided by the Department of Biotechnology (DBT) Govt. of India.

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.939984/full#supplementary-material

References

Amy, P. S., and Morita, R. Y. (1983). Starvation-survival patterns of sixteen freshly isolated open-ocean bacteria. *Appl. Environ. Microbiol.* 45, 1109. doi: 10.1128/aem.45.3.1109-1115.1983

Antony, C. P., Kumaresan, D., Ferrando, L., Boden, R., Moussard, H., Scavino, A. F., et al. (2010). Active methylotrophs in the sediments of Lonar Lake, a saline and alkaline ecosystem formed by meteor impact. *ISME J.* 4, 1470–1480. doi: 10.1038/ismej.2010.70

Antony, C. P., Kumaresan, D., Hunger, S., Drake, H. L., Murrell, J. C., and Shouche, Y. S. (2013). Microbiology of Lonar Lake and other soda lakes. *ISME J.* 7, 468. doi: 10.1038/ismej.2012.137

Antony, C. P., Shimpi, G. G., Cockell, C. S., Patole, M. S., and Shouche, Y. S. (2014). Molecular characterization of prokaryotic communities associated with Lonar crater basalts. *Geomicrobiol. J.* 31, 519–528. doi: 10.1080/01490451.2013.849314

Bagade, A., Nandre, V., Paul, D., Patil, Y., Sharma, N., Giri, A., et al. (2020). Characterisation of hyper tolerant *Bacillus firmus* L-148 for arsenic oxidation. *Environ. Pollut.* 261, 114124. doi: 10.1016/j.envpol.2020.114124

Bartelme, R. P., Custer, J. M., Dupont, C. L., Espinoza, J. L., Torralba, M., Khalili, B., et al. (2020). Influence of substrate concentration on the culturability of heterotrophic soil microbes isolated by high-throughput dilution-to-extinction cultivation. *mSphere* 5, e00024-20. doi: 10.1128/mSphere.00024-20

Böllmann, J., and Martienssen, M. (2020). Comparison of different media for the detection of denitrifying and nitrate reducing bacteria in mesotrophic aquatic environments by the most probable number method. *J. Microbiol. Methods* 168, 105808. doi: 10.1016/j.mimet.2019.105808

Borul, S. B. (2012). Study of water quality of Lonar Lake. *J. Chem. Pharm. Res.* 4, 1716–1718. Available online at: https://www.jocpr.com/articles/study-of-water-quality-of-lonar-lake.pdf (Accessed June 28, 2022).

Button, D. K. (1991). Biochemical basis for whole-cell uptake kinetics: specific affinity, oligotrophic capacity, and the meaning of the michaelis constant. *Appl. Environ. Microbiol.* 57, 2033–2038. doi: 10.1128/aem.57.7.2033-2038.1991

Fegatella, F., and Cavicchioli, R. (2000). Physiological responses to starvation in the marine oligotrophic ultramicrobacterium *Sphingomonas* sp. strain RB2256. *Appl. Environ. Microbiol.* 66, 2037. doi: 10.1128/AEM.66.5.2037-2044.2000

Ferrari, B. C., Binnerup, S. J., and Gillings, M. (2005). Microcolony cultivation on a soil substrate membrane system selects for previously uncultured soil bacteria. *Appl. Environ. Microbiol.* 71, 8714. doi: 10.1128/AEM.71.12.8714-8720.2005

Fry, J. C. (1990). "Oligotrophs," in *Microbiology of Extreme Environments*, Ed C. A. Edwards (Oxford: Alden Press), 93–116.

Gao, Y., and Wu, M. (2018). Free-living bacterial communities are mostly dominated by oligotrophs. *bioRxiv*. 350348. doi: 10.1101/350348

Grant, W. D., and Sorokin, D. Y. (2011). "Distribution and diversity of soda lake alkaliphiles," in *Extremophiles Handbook*, ed K. Horikoshi (Tokyo, Japan: Springer), 27–54. doi: 10.1007/978-4-431-53898-1_3

Green, S. J., Prakash, O., Gihring, T. M., Akob, D. M., Jasrotia, P., Jardine, P. M., et al. (2010). Denitrifying bacteria isolated from terrestrial subsurface sediments exposed to mixed-waste contamination. *Appl. Environ. Microbiol.* 76, 3244–3254. doi: 10.1128/AEM.03069-09

Hartke, A., Giard, J. C., Laplace, J. M., and Auffray, Y. (1998). Survival of *Enterococcus faecalis* in an oligotrophic microcosm: changes in morphology, development of general stress resistance, and analysis of protein synthesis. *Appl. Environ. Microbiol.* 64, 4238. doi: 10.1128/AEM.64.11.4238-4245.1998

Hinder, B., Baur, I., Hanselmann, K., and Schanz, F. (1999). Microbial food web in an oligotrophic high mountain lake (Jöri Lake III, Switzerland). *J. Limnol.* 58, 162–168. doi: 10.4081/jlimnol.1999.162

Ho, A., Di Lonardo, D. P., and Bodelier, P. L. E. (2017). Revisiting life strategy concepts in environmental microbial ecology. *FEMS Microbiol. Ecol.* 93, fix006. doi: 10.1093/femsec/fix006

Hu, S. J., Van Bruggen, A. H. C., and Grünwald, N. J. (1999). Dynamics of bacterial populations in relation to carbon availability in a residue-amended soil. *Appl. Soil Ecol.* 13, 21–30. doi: 10.1016/S0929-1393(99)00015-3

HuiXia, P., ZhengMing, C., XueMei, Z., ShuYong, M., XiaoLing, Q., and Fang, W. (2007). A study on oligotrophic bacteria and its ecological characteristics in an arid desert area. *Sci. China Ser. D Earth Sci.* 50, 128–134. doi: 10.1007/s11430-007-5015-4

Incerti, C., B1anco-Varcla, M. T., Puertas, F., and Saiz-Jimenez, E. (1997). "Halotolerant and halophílic bacteria associated to effiorescences in Jerez cathcdraI," in *Origin, Mechanisms and efli!cts of Salts on Degradation of Monuments in Marine and Continental Environments*, Ed F. Zezza. Protection and Conscrvation of the European Cultural Heritage Research Report No. 4, 225–232.

Ishida, Y., and Kadota, H. (1981). Growth patterns and substrate requirements of naturally occurring obligate oligotrophs. *Microb. Ecol.* 123–130. doi: 10.1007/BF02032494

Joshi, A. A., Kanekar, P. P., Kelkar, A. S., Sarnaik, S. S., Shouche, Y., and Wani, A. (2007). Moderately halophilic, alkalitolerant *Halomonas campisalis* MCM B-365 from Lonar Lake, India. *J. Basic Microbiol.* 47, 213–221. doi: 10.1002/jobm.200610223

Joshi, A. A., Kanekar, P. P., Kelkar, A. S., Shouche, Y. S., Vani, A. A., Borgave, S. B., et al. (2008). Cultivable bacterial diversity of alkaline Lonar lake, India. *Microb. Ecol.* 55, 163–172. doi: 10.1007/s00248-007-9264-8

Kelly, D. P. (1971). Autotrophy: concepts of lithotrophic bacteria and their organic metabolism. *Ann. Rev. Microbiol.* 25, 177–210. doi: 10.1146/annurev.mi.25.100171.001141

Kim, M., Lee, K.-H., Yoon, S.-W., Kim, B.-S., Chun, J., and Yi, H. (2013). Analytical tools and databases for metagenomics in the next-generation sequencing era. *Genomics Inform.* 11, 102. doi: 10.5808/GI.2013.11.3.102

Koch, A. L. (1998). Microbial physiology and ecology of slow growth. *Microbiol. Mol. Biol. Rev.* 62, 248–248. doi: 10.1128/MMBR.62.1.248-248.1998

Koch, A. L. (2001). Oligotrophs versus copiotrophs. *Bioessays* 23, 657–661. doi: 10.1002/bies.1091

Kuznetsov, S. I., Dubinina, G. A., and Lapteva, N. A. (2003). Biology of oligotrophic bacteria. *Annu. Rev. Microbiol.* 33, 377–387. doi: 10.1146/annurev.mi.33.100179.002113

Laiz, L., Hermosin, B., Caballero, B., and Saiz-Jimenez, C. (2000). Bacteria isolated from the rocks supporting prehistoric paintings in two shelters from Sierra de Cazorla, Jaen, Spain. *Aerobiologia* 16, 119–124. doi: 10.1023/A:1007684904350

Laiz, L., Hermosin, B., Caballero, B., and Saiz-Jimenez, C. (2002a). "Facultatively oligotrophic bacteria in Roman mural paintings," in *Protection and Conservation of the Cultural Heritage of the Mediterranean Cities: Proceedings of the 5th International Symposium on the Conservation of Monuments in the Mediterranean Basin* (Sevilla), 173–178.

Laiz, L., Saiz-Jimenez, C., Cardell, C., and Rodriguez-Gordillo, J. (2002b). "Halotolerant bacteria in the efflorescences of a deteriorated church," in *Protection and Conservation of the Cultural Heritage of the Mediterranean Cities: Proceedings of the 5th International Symposium on the Conservation of Monuments in the Mediterranean Basin* (Sevilla, Spain), 183–189.

Lauro, F. M., McDougald, D., Thomas, T., Williams, T. J., Egan, S., Rice, S., et al. (2009). The genomic basis of trophic strategy in marine bacteria. *Proc. Natl. Acad. Sci.* 106, 15527–15533. doi: 10.1073/pnas.0903507106

Lazar, I., and Dumitru, L. (1973). Bacteria and their role in the deterioration of frescoes of the complex of monasteries from northern moldavia. *Rev. Roum. Biol.* 18, 3.

Lee, C., Song, H. S., Lee, S. H., Kim, J. Y., Rhee, J. K., and Roh, S. W. (2021). Genomic analysis of facultatively oligotrophic haloarchaea of the genera Halarchaeum, Halorubrum, and Halolamina, isolated from solar salt. *Arch. Microbiol.* 203, 261–268. doi: 10.1007/s00203-020-02027-2

Lee, J. K., Patel, S. K. S., Sung, B. H., and Kalia, V. C. (2020). Biomolecules from municipal and food industry wastes: an overview. *Bioresour. Technol.* 298, 122346. doi: 10.1016/j.biortech.2019.122346

Marathe, N. P., Regina, V. R., Walujkar, S. A., Charan, S. S., Moore, E. R. B., et al. (2013) A treatment plant receiving waste water from multiple bulk drug manufacturers is a reservoir for highly multi-drug resistant integron-bearing bacteria. *PLoS ONE* 8, e77310. doi: 10.1371/journal.pone.0077310

Matsuoka, T., and Yoshida, N. (2018). Establishment of an effective oligotrophic cultivation system for Rhodococcus erythropolis N9T-4. *Biosci. Biotechnol. Biochem.* 82, 1652–1655. doi: 10.1080/09168451.2018.1482196

Morita, R. Y. (1988). Bioavailability of energy and its relationship to growth and starvation survival in nature. *Can. J. Microbiol.* 34, 436–441. doi: 10.1139/m88-076

Nagarkar, P. P., Ravetkar, S. D., and Watve, M. G. (2001). Oligophilic bacteria as tools to monitor aseptic pharmaceutical production units. *Appl. Environ. Microbiol.* 67, 1371–1374. doi: 10.1128/AEM.67.3.1371-1374.2001

Newton, R. J., and Shade, A. (2016). Lifestyles of rarity: understanding heterotrophic strategies to inform the ecology of the microbial rare biosphere. *Aquat. Microb. Ecol.* 78, 51–63. doi: 10.3354/ame01801

Norris, N., Levine, N. M., Fernandez, V. I., and Stocker, R. (2021). Mechanistic model of nutrient uptake explains dichotomy between marine oligotrophic and copiotrophic bacteria. *PLoS Comput. Biol.* 17, e1009023. doi: 10.1371/journal.pcbi.1009023

Ohhata, N., Yoshida, N., Egami, H., Katsuragi, T., Tani, Y., and Takagi, H. (2007). An extremely oligotrophic bacterium, Rhodococcus erythropolis N9T-4, isolated from crude oil. *J. Bacteriol.* 189, 6824–6831. doi: 10.1128/JB.00872-07

Pathak, A. P., and Deshmukh, K. B. (2012). Alkaline protease production, extraction and characterization from alkaliphilic *Bacillus licheniformis* KBDL4: a Lonar soda lake isolate. *Indian J. Exp. Biol.* 50, 569–576. Available online at: https://pubmed.ncbi.nlm.nih.gov/23016494/ (Accessed June 28, 2022).

Paul, D., Kumbhare, S. V., Mhatre, S. S., Chowdhury, S. P., Shetty, S. A., Marathe, N. P., et al. (2016). Exploration of microbial diversity and community structure of Lonar Lake: the only hypersaline meteorite crater lake within basalt rock. *Front. Microbiol.* 6, 1553. doi: 10.3389/fmicb.2015.01553

Paul, E. A., and Clark, F. E. (1996). *Soil Microbiology and Biochemistry*, 2nd Edn. San Diego, CA; Orlando, FL: Academic Press.

Poindexter, J. S. (1981). Oligotrophy. *Adv. Microbial Ecol.* 5, 63–89. doi: 10.1007/978-1-4615-8306-6_2

Porwal, S., Kumar, T., Lal, S., Rani, A., Kumar, S., Cheema, S., et al. (2008). Hydrogen and polyhydroxybutyrate producing abilities of microbes from diverse habitats by dark fermentative process. *Bioresour. Technol.* 99, 5444–5451. doi: 10.1016/j.biortech.2007.11.011

Prakash, J., Sharma, R., Patel, S. K. S., Kim, I. W., and Kalia, V. C. (2018). Bio-hydrogen production by co-digestion of domestic wastewater and biodiesel industry effluent. *PLoS ONE* 13, e0199059. doi: 10.1371/journal.pone.0199059

Prakash, O., and Lal, R. (2013). Role of unstable phenanthrene-degrading *Pseudomonas species* in natural attenuation of phenanthrene-contaminated site. *Korean J. Microbiol. Biotechnol.* 41, 79–87. doi: 10.4014/kjmb.1207.07011

Prakash, O., Munot, H., Nimonkar, Y., Sharma, M., Kumbhare, S., and Shouche, Y. S. (2014). Description of *Pelistega indica* sp. nov., isolated from human gut. *Int. J. Syst. Evol. Microbiol.* 64, 1389–1394. doi: 10.1099/ijs.0.059782-0

Prakash, O., Nimonkar, Y., and Shouche, Y. S. (2013). Practice and prospects of microbial preservation. *FEMS Microbiol. Lett.* 339, 1–9. doi: 10.1111/1574-6968.12034

Prakash, O., Nimonkar, Y., Vaishampayan, A., Mishra, M., Kumbhare, S., Josef, N., et al. (2015). *Pantoea intestinalis* sp. nov., isolated from the human gut. *Int. J. Syst. Evol. Microbiol.* 65, 3352–3358. doi: 10.1099/ijsem.0.000419

Saha, P., and Chakrabarti, T. (2006). *Emticicia oligotrophica* gen. nov., sp. nov., a new member of the family "Flexibacteraceae", phylum Bacteroidetes. *Int. J. Syst. Evol. Microbiol.* 56, 991–995. doi: 10.1099/ijs.0.64086-0

Saiz-Jimenez, C., and Laiz, L. (2000). Occurrence of halotolerant/halophilic bacterial communities in deteriorated monuments. *Int. Biodeterior. Biodegradation* 46, 319–326. doi: 10.1016/S0964-8305(00)00104-9

Schut, F., Prins, R. A., and Gottschal, J. C. (1997). Oligotrophy and pelagic marine bacteria: facts and fiction. *Aquat. Microb. Ecol.* 12, 177–202. doi: 10.3354/ame012177

Senechkin, I. V., Speksnijder, A. G. C. L., Semenov, A. M., van Bruggen, A. H. C., and van Overbeek, L. S. (2010). Isolation and partial characterization of bacterial strains on low organic carbon medium from soils fertilized with different organic amendments. *Microb. Ecol.* 60, 829–839. doi: 10.1007/s00248-010-9670-1

Shahbaz, M., Kätterer, T., Thornton, B., and Börjesson, G. (2020). Dynamics of fungal and bacterial groups and their carbon sources during the growing season of maize in a long-term experiment. *Biol. Fertil. Soils* 56, 759–770. doi: 10.1007/s00374-020-01454-z

Sharma, R., Prakash, O., Sonawane, M. S., Nimonkar, Y., Golellu, P. B., and Sharma, R. (2016). Diversity and distribution of phenol oxidase producing fungi from soda lake and description of *Curvularia lonarensis* sp. nov. *Front. Microbiol.* 7, 1847. doi: 10.3389/fmicb.2016.01847

Smibert, R. M., and Krieg, N. R. (1994). "Phenotypic characterization," in *Methods for General and Molecular Bacteriology*, Eds P. Gerhardt, R. G. E. Murray, W. A. Wood, and N. R. Krieg (Washington, DC: American Society for Microbiology), 607–654.

Smith, E. M., and Prairie, Y. T. (2004). Bacterial metabolism and growth efficiency in lakes: the importance of phosphorus availability. *Limnol. Oceanogr.* 49, 137–147. doi: 10.4319/lo.2004.49.1.0137

Surakasi, V. P., Antony, C. P., Sharma, S., Patole, M. S., and Shouche, Y. S. (2010). Temporal bacterial diversity and detection of putative methanotrophs in surface mats of Lonar Crater Lake. *J. Basic Microbiol.* 50, 465–474. doi: 10.1002/jobm.201000001

Tada, Y., Ihmori, M., and Yamaguchi, J. (1995). Oligotrophic bacteria isolated from clinical materials. *J. Clin. Microbiol.* 33, 493. doi: 10.1128/jcm.33.2.493-494.1995

Tambekar, D. H., and Dhundale, V. R. (2012). Studies on the physiological and cultural diversity of bacilli characterized from Lonar Lake (Ms) India. *Biosci. Discov. Dep. Microbiol.* 3, 34–39. Available online at: http://www.biosciencediscovery.com (Accessed June 28, 2022).

Temperton, B., Gilbert, J. A., Quinn, J. P., and McGrath, J. W. (2011). Novel analysis of oceanic surface water metagenomes suggests importance of polyphosphate metabolism in oligotrophic environments. *PLoS ONE* 6, e16499. doi: 10.1371/journal.pone.0016499

Van Gemerden, H., and Kuenen, J. G. (1984). *Strategies for Growth and Evolution of Microorganisms in Oligotrophic Habitats*. Available online at: https://repository.tudelft.nl/islandora/object/uuid:ac0d1a1e-9462-4118-aeeb-fe5db9e7aa59 (accessed June 29, 2022).

Watve, M., Shejval, V., Sonawane, C., Rahalkar, M., Matapurkar, A., Shouche, Y., et al. (2000). The "K" selected oligophilic bacteria: a key to uncultured diversity? *Curr. Sci.* 78, 1535–1542. Available online at: https://www.researchgate.net/publication/259532213_The_ŠKŠ_selected_oligophilic_bacteria_A_key_to_uncultured_diversity (Accessed July 24, 2022).

Weissman, J. L., Hou, S., and Fuhrman, J. A. (2021). Estimating maximal microbial growth rates from cultures, metagenomes, and single cells via codon usage patterns. *Proc. Natl. Acad. Sci. U.S.A.* 118, e2016810118. doi: 10.1073/pnas.2016810118

Widdel, F., and Bak, F. (1992). "Gram-negative mesophilic sulfate-reducing bacteria," in *The Prokaryotes*, eds A. Balows, H. G. Trüper, M. Dworkin, W. Harder, and K.-H. Schleifer (New York: Springer), 3352–3378.

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 06 January 2023
DOI 10.3389/fmicb.2022.1090197

OPEN ACCESS

EDITED BY
Sanjay Kumar Singh Patel,
Konkuk University, South Korea

REVIEWED BY
Monika Yadav,
Maharshi Dayanand University, India
Helianthous Verma,
University of Delhi, India

*CORRESPONDENCE
Paulina Corral
✉ paulina.corral@unina.it

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

RECEIVED 05 November 2022
ACCEPTED 09 December 2022
PUBLISHED 06 January 2023

CITATION
Gattoni G, de la Haba RR, Martín J, Reyes F, Sánchez-Porro C, Feola A, Zuchegna C, Guerrero-Flores S, Varcamonti M, Ricca E, Selem-Mojica N, Ventosa A and Corral P (2023) Genomic study and lipidomic bioassay of *Leeuwenhoekiella parthenopeia*: A novel rare biosphere marine bacterium that inhibits tumor cell viability. *Front. Microbiol.* 13:1090197. doi: 10.3389/fmicb.2022.1090197

Genomic study and lipidomic bioassay of *Leeuwenhoekiella parthenopeia*: A novel rare biosphere marine bacterium that inhibits tumor cell viability

Giuliano Gattoni[1], Rafael R. de la Haba[2], Jesús Martín[3], Fernando Reyes[3], Cristina Sánchez-Porro[2], Antonia Feola[1], Candida Zuchegna[1], Shaday Guerrero-Flores[4], Mario Varcamonti[1], Ezio Ricca[1], Nelly Selem-Mojica[4], Antonio Ventosa[2] and Paulina Corral[1,2]*

[1]Department of Biology, University of Naples Federico II, Naples, Italy, [2]Department of Microbiology and Parasitology, Faculty of Pharmacy, University of Sevilla, Sevilla, Spain, [3]Fundación MEDINA, Granada, Spain, [4]Centro de Ciencias Matemáticas, Universidad Nacional Autónoma de México (UNAM), Morelia, Mexico

The fraction of low-abundance microbiota in the marine environment is a promising target for discovering new bioactive molecules with pharmaceutical applications. Phenomena in the ocean such as diel vertical migration (DVM) and seasonal dynamic events influence the pattern of diversity of marine bacteria, conditioning the probability of isolation of uncultured bacteria. In this study, we report a new marine bacterium belonging to the rare biosphere, *Leeuwenhoekiella parthenopeia* sp. nov. Mr9^T, which was isolated employing seasonal and diel sampling approaches. Its complete characterization, ecology, biosynthetic gene profiling of the whole genus *Leeuwenhoekiella*, and bioactivity of its extract on human cells are reported. The phylogenomic and microbial diversity studies demonstrated that this bacterium is a new and rare species, barely representing 0.0029% of the bacterial community in Mediterranean Sea metagenomes. The biosynthetic profiling of species of the genus *Leeuwenhoekiella* showed nine functionally related gene cluster families (GCF), none were associated with pathways responsible to produce known compounds or registered patents, therefore revealing its potential to synthesize novel bioactive compounds. *In vitro* screenings of *L. parthenopeia* Mr9^T showed that the total lipid content (lipidome) of the cell membrane reduces the prostatic and brain tumor cell viability with a lower effect on normal cells. The lipidome consisted of sulfobacin A, WB 3559A, WB 3559B, docosenamide, topostin B-567, and unknown compounds. Therefore, the bioactivity could be attributed to any of these individual compounds or due to their synergistic effect. Beyond the rarity and biosynthetic potential of this bacterium, the importance and novelty of this study is the employment of sampling strategies based on

ecological factors to reach the hidden microbiota, as well as the use of bacterial membrane constituents as potential novel therapeutics. Our findings open new perspectives on cultivation and the relationship between bacterial biological membrane components and their bioactivity in eukaryotic cells, encouraging similar studies in other members of the rare biosphere.

KEYWORDS

***Leeuwenhoekiella*, rare biosphere, diel vertical migration (DVM), biosynthetic profiling, membrane lipids, antiproliferative, tumor cells**

1. Introduction

The exploration of the uncultivated marine microbiota is a promising target in the search for alternative drugs to meet the urgent therapeutic needs where the current pharmacotherapy is ineffective (García-Davis et al., 2021; Zhou et al., 2021). Every year, new cases of cancer are registered worldwide, and many of them are resistant forms that do not respond to available treatments (Maeda and Khatami, 2018). This low therapeutic efficacy reduces life expectancy, as cancer is among the main leading causes of death (Siegel et al., 2021). In this scenario, the isolation of uncultured bacteria is a current trend, in addition to culture-independent studies, since there is a need to explore the true metabolic and intrinsic characteristics both in their habitat and *in vitro* (Pascoal et al., 2020; Sanz-Sáez et al., 2020). Most of the discovered bioactive compounds were derived from only 1% of the total microbial diversity known so far (Baker et al., 2021). An outstanding example is a salinosporamide A, an agent in phase III clinical trial for the treatment of glioblastoma produced by the rare marine actinobacteria of the genus *Salinispora* (Fenical et al., 2009; Roth et al., 2020; Boccellato et al., 2021). Despite this successful precedent, the huge biosynthetic potential remains in the great prokaryotic diversity not yet cultivated as predicted by microbiome studies (Paoli et al., 2022). The low rate of cultivation of new taxa is even more challenging when targeting the very low-abundant fraction (0.1 to 0.01% of relative abundance) known as the rare biosphere (Sogin et al., 2006; Pedrós-Alió, 2007). To overcome this limitation, prokaryotic diversity is being addressed by considering environmental variabilities that affect the microbial population distribution, such as biomass migration and seasonality (Shade et al., 2014; Cram et al., 2015; Mestre et al., 2020).

With the aim to cultivate novel and rare bacteria with anticancer potential, here, we employed sampling strategies considering seasonality and natural dynamic processes in the marine environment rather than culture techniques *per se*. The largest biomass dynamic event in the ocean is the diel vertical migration (DVM) where a part of the inhabitants, especially zooplankton, ascends from the depths to the surface waters during the night or light absence and moves downward during the day (Bosch and Rowland Taylor, 1973; Omand et al., 2021). In the same way, seasonality influences the pattern of microbial diversity as the water column in the ocean stratifies during the warm period and mixes during the cold months (Giovannoni and Vergin, 2012; Haro-Moreno et al., 2018; Mena et al., 2020). In this study, cold seasons and night were identified as the best sampling time since the water column is mixed and DVM brings planktonic particulate as drivers of microbial transfer from the depths during the night (Gilbert et al., 2012; Fuhrman et al., 2015; Kelly et al., 2019). In both cases, the microbial community is redistributed, increasing the possibility of reaching uncommon bacteria. Furthermore, it has been demonstrated that surface and deep-sea prokaryotic communities are strongly connected in the oceans through the transport of sinking particles (Mestre et al., 2018; Wenley et al., 2021). By employing this approach, a novel and rare marine bacterium, that we have designated as *Leeuwenhoekiella parthenopeia* strain Mr9^T, was isolated from reef seawater in the Gulf of Naples, Italy, from night sampling series during the late autumn. Herein, we fully characterized this strain and inferred its specialized metabolism through the genome by predicting unknown biosynthetic gene clusters (BGCs) of pharmacological relevance (Agrawal et al., 2017; Belknap et al., 2020; Scherlach and Hertweck, 2021). Similarly, the biosynthetic potential of all described species of the genus *Leeuwenhoekiella* was genomically assessed by determining the functional homologies based on gene cluster families (GCFs) (Navarro-Muñoz et al., 2020). Considering that most BGCs of the secondary metabolism are not expressed *in vitro* or are not released outside the cell (Letzel et al., 2017), such as in the case of *L. parthenopeia* Mr9^T, we investigated the structural constituents of the bacterial membrane, specifically the lipidome. The bacterial lipid composition variates to adapt in response to environmental changes maintaining the optimal functional properties of the membrane (Angelini et al., 2012; Chwastek et al., 2020). The lipid profile is used as a chemotaxonomic marker in the characterization of bacteria and archaea and can be different among species of the same genus (de la Haba et al., 2018), but there are no studies on bacterial membrane lipid composition and its effect on eukaryotic cells as such. Our study demonstrates that the

total lipid extract (TLE) from *L. parthenopeia* Mr9^T affects tumor cell viability of prostate adenocarcinoma DU-145, and glioblastoma U-87 MG cell lines with a lower effect on the normal cell line HaCaT. The genus *Leeuwenhoekiella*, the object of this study, includes chemoorganotrophic Gram-negative bacteria, producing non-diffusible yellow-pigmented colonies with gliding motility (Nedashkovskaya et al., 2005; Tahon et al., 2020). Currently, this genus includes only seven species (Parte et al., 2020) derived from marine or sea-related habitats, and no bioactive compounds that impair tumor cell viability produced by strains of these species have been reported to date.

2. Materials and methods

2.1. Sampling and isolation

Serial weekly nightly samplings were carried out in the reef of the Gulf of Naples, Italy 40°49′48.9″N 14°13′31.8″E, during the fall and winter of 2020–2021. The selection of diel and cold seasons was based on previous studies (Gilbert et al., 2012; Fuhrman et al., 2015; Kelly et al., 2019) that demonstrate that the water column is mixed during these periods and DVM brings planktonic particulate acting as drivers of microbial transfer from the depths during the night, thus suggesting fall, winter, and night as the best sampling time. The salinity, pH, conductivity, particulate (total dissolved solids, TDS), and temperature were measured *in situ* with a refractometer and a digital multiparameter pH/EC/TDS/TEMP, Hanna HI-9812-51.

Five liters of seawater were collected from shallow waters (∼1 m) each time and homogenized by magnetic stirring. Aliquots of 40 mL of each sample were vortexed for 10 min, and 1 mL was serially diluted 10-fold in 9 mL of sterile seawater to a 10^{-6} dilution. A total of 100 μL of each dilution was inoculated on four different solid minimal media designed to cultivate oligotrophic and heterotrophic bacteria, formulated with filter-sterilized natural seawater as the base component and varying the mineral composition. The plates were incubated on the laboratory bench to follow the light day cycle at room temperature (∼21°C) and aerobic conditions. The growth was checked daily, and the smallest colonies were picked and streaked on the same medium until obtaining axenic cultures.

The strain Mr9^T was isolated from the night sampling series collected in November 2020 at 1-m depth after 4 days of incubation in seawater oligotrophic medium (SWOM) with the following composition (g/L of natural seawater): casamino acids, 0.5; yeast extract, 1.0; tryptone, 1.0; and agar, 15.0. This medium was supplemented with glycerol, 0.1 M; mineral solution, 500 μL/L; and vitamin solution, 500 μL/L. All supplement solutions were previously sterilized by filtration and added after autoclaving.

The mineral solution compositions are as follows (mg/L): *Na-EDTA, 250.0; $ZnSO_4$ $7H_2O$, 1095.0; $FeSO_4$ 7 H_2O, 500; $MnSO_4$ H_2O, 154; $CuSO_4$ $5H_2O$, 39.2; $Co(NO_3)_2$ $6H_2O$, 24.8; $Na_2B_4O_7$ $10H_2O$, 17.7. Distilled water up to 1,000 mL. *Dissolve the EDTA and add a few drops of concentrated H_2SO_4 to retard the precipitation of the heavy metal ions.

Vitamin solution composition (mg/L): Biotin, 4.0; folic acid, 4.0; pyridoxine-HCl, 20.0; riboflavin, 10.0; thiamine-HCl $2H_2O$, 10.0; nicotinamide, 10.0; D-Ca-pantothenate, 10.0; vitamin B_{12}, 0.20. Distilled water up to 1,000 mL. Stored in dark at 4°C.

After primary isolation and purification in SWOM, no growth was observed in synthetic medium marine agar (MA) 2260 (Difco), but the strain Mr9^T was adapted to grow on MA medium by adding the same supplement solutions as those used for SWOM, where a poor growth was visible after 10 days of incubation. After successive cultivation, the strain Mr9^T was well adapted to routine growth in MA without supplemental solutions, establishing the optimal growing condition to 4 days of incubation at 25°C. For long-term conservation, strain Mr9^T was cultivated in marine broth (MB) for 72 h and stored at −80°C in cryovials with 40% (v/v) of glycerol.

2.2. Characterization of new taxa

2.2.1. Strain identification by amplification and sequencing of 16S rRNA gene

For the identification of strain Mr9^T, the 16S rRNA gene was amplified using universal primers: 27F (5′-AGAG TTTGATCCTGGCTCAG-3′) and 1492R (5′-GGTTACCTT GTTACGACTT-3′). The PCR amplification was performed using a single colony of a 48-h culture of Mr9^T and resuspending it in 40 μL of a master mix as follows: *Reaction buffer 5×, 10 μL; primer forward 27F (20 μM), 1 μL; primer reverse 1492R (20 μM), 1 μL; *Taq* DNA polymerase MyTaq (Bioline), 1 μL; and H_2O nuclease-free, 27 μL. *Reaction buffer 5× contains dNTPs, $MgCl_2$, and enhancers. The reaction was performed using a thermal cycler T100 (BioRad, CA, USA) with the following settings: Initial denaturation for 5′ at 95°C, 25 cycles of 30″ at 94°C, 30″ at 50°C, 90″ at 72°C, and final extension for 10′ at 72°C.

The amplification was verified by 0.8% agarose gel electrophoresis in TBE and stained with SYBR Safe DNA Gel Stain 10,000× in dimethyl sulfoxide (DMSO) (Invitrogen, MA, USA), and the size was checked using the 1 Kb Plus DNA Ladder (Invitrogen, MA, USA).

The PCR product was purified using a silica-based membrane GeneJET PCR purification kit (Invitrogen, MA, USA) following the manufacturer's protocol and sequenced by Stab Vida (Portugal) using the Sanger method. Sequence quality analysis and trimming were accomplished using the software 4Peaks V1.8 and contig assembly of the whole gene with ChromasPro [Research Resource Identifiers (RRIDs)]: (RRID: SCR_000229) v.2.1.10.

Sequence similarity search was performed against the EzBioCloud database (Yoon et al., 2017) with the 16S-based ID application, and with the NCBI BLAST (RRID: SCR_004870)

against NCBI nr/nt Nucleotide (RRID: SCR_004860) database restricted to type-sequenced taxa and then also including uncultured/environmental sample sequences.

2.2.2. Genome sequencing, assembly, annotation, and relatedness indexes

Genomic DNA was extracted using the DNeasy UltraClean kit (QIAGEN, Hilden, Germany) following the protocol for Gram-negative bacteria. DNA purity and integrity were checked by agarose gel electrophoresis (0.8%) and using NanoDrop 2000 Spectrophotometer (Thermo Scientific, MA, USA). The DNA was quantified employing the Qubit assay kit for double-stranded DNA broad range (dsDNA BR, Invitrogen) and measured using the Qubit 4 Fluorometer (Invitrogen, MA, USA).

The genome of strain Mr9^T was sequenced using Illumina NovaSeq 6000 platform (2 × 150 bp paired-end sequencing reads) at Novogene (Genomic sequencing center, Cambridge, UK). Quality control of raw reads was checked with FastQC (RRID: SCR_014583) v.0.11.9 and reads containing adapters and low-quality bases ($Q \leq 20$) were removed with Trimmomatic (RRID: SCR_011848) v.0.36 (Bolger et al., 2014). *De novo* assembly of the reads was performed using SPAdes (RRID: SCR_000131) v.3.15.4 (Bankevich et al., 2012) and SOAP3 (RRID: SCR_005502) (Liu et al., 2012). The quality of final contigs was assessed using CheckM (RRID: SCR_016646) v.1.0.5 (Parks et al., 2015) and QUAST (RRID: SCR_001228) v.5.1.0rc1 (Gurevich et al., 2013). The genome sequence was annotated using the NCBI Prokaryotic Genomes Automatic Annotation Pipeline (PGAP) (RRID: SCR_021329) (Tatusova et al., 2016).

To assess the taxonomic status of strain Mr9^T as a new taxon, the genome sequence similarities among closely related species were estimated by computing the overall genome relatedness index (OGRI) (Chun and Rainey, 2014). The average nucleotide identity (ANI) was calculated using the OrthoANIu tool (RRID: SCR_022562) (Liu et al., 2016). The digital DNA–DNA hybridization (dDDH) was calculated using the Genome-to-Genome Distance Calculator (GGDC) website, with formula 2 (Meier-Kolthoff et al., 2013). This genomic comparison was between strain Mr9^T and type species of validly published species names as well as two genomes of non-type cultivated species of the genus *Leeuwenhoekiella*. The results were evaluated based on the cutoff for species boundaries; ANI, 95–96% (Konstantinidis and Tiedje, 2005; Goris et al., 2007; Richter and Rosselló-Móra, 2009; Chun and Rainey, 2014), and dDDH, 70% (Stackebrandt and Goebel, 1994; Auch et al., 2010).

2.2.3. Phylogenetic and phylogenomic analysis

To infer the 16S rRNA gene-based phylogeny, the respective sequences of strain Mr9^T and its close relative species of the genus *Leeuwenhoekiella* were downloaded from DDBJ/ENA/GenBank databases. After sequence alignment, phylogenetic trees were constructed using ARB v.6.0.5 software package (Westram et al., 2011) with neighbor-joining (NJ) (Saitou and Nei, 1987), maximum-parsimony (MP) (Fitch, 1971), and maximum-likelihood (ML) (Felsenstein, 1981) algorithms. The distance matrix was corrected by applying the Jukes–Cantor model of DNA evolution (Jukes and Cantor, 1969). To infer ML phylogeny, the general time-reversible model (Tavaré and Miura, 1986) with gamma distribution and proportion of invariant sites to estimate rate heterogeneity over sites (GTR + Γ + I) were used. Branch support was assessed by 1,000 bootstrap pseudoreplicates (Felsenstein, 1985).

Since phylogenies based on the 16S rRNA gene sequence comparison have been demonstrated not to be totally reliable in determining evolutionary relationships (Corral et al., 2018; de la Haba et al., 2018; Infante-Domínguez et al., 2020), a more robust genome-based phylogeny was inferred by constructing a phylogenomic tree. For this purpose, 11 genome sequences of cultivated species of the genus *Leeuwenhoekiella* were retrieved from the NCBI GenBank database (**Supplementary Table 1**). Of those genomes, nine corresponded to type strains comprising seven *Leeuwenhoekiella* species (given the fact that two genomes from each of the type strains of *L. palythoae* and *Leeuwenhoekiella marinoflava* were available in the public databases, although with different culture collection designation key); the other two out of 11 genomes corresponded to cultivated non-type strains but closely related to *Leeuwenhoekiella* sp. strain Mr9^T.

Pan- and core-genome datasets were determined using an all-vs.-all BLAST comparisons among the coding sequences (CDS) and translated CDS features of the annotated genomes under study, as previously described (de la Haba et al., 2019). Then, single-copy core gene and amino acid sequences were individually aligned with muscle (Edgar, 2004) and concatenated into a super-nucleotide or super-protein alignment, which was further analyzed to generate the phylogenomic tree using the appropriate ML algorithm implemented in FastTreeMP v.2.1.8 (Price et al., 2010). Generalized time-reversible (GTR) model of nucleotide evolution (Tavaré and Miura, 1986) and Jones–Taylor–Thornton (JTT) model of amino acid evolution (Jones et al., 1992), both with a single rate for each site (CAT), were applied for nucleotide- and amino acid-based phylogenomic reconstructions, respectively. Tree branch support was inferred using the Shimodaira–Hasegawa test (Shimodaira and Hasegawa, 1999). For the display, annotation, and management of 16S rRNA gene-based and phylogenomic trees, the online tool iTOL (RRID: SCR_018174) v.6.5.2 (Letunic and Bork, 2021) was used.

2.2.4. Phenotypic and physiological characterization

The macroscopic features of *Leeuwenhoekiella* sp. strain Mr9^T were based on the colony morphology in solid media

SWOM and MA during incubation from the 4th to the 30th day. The colony appearance such as shape, pigmentation, and consistency was visually evaluated. Cell morphology, size, and motility were observed by phase-contrast microscopy (Olympus BX41, Tokio, Japan) from a 72-h liquid culture incubated at 25°C with agitation. The optimal growth conditions and tolerance to temperature, salinity (NaCl), and pH were determined in the following ranges: temperature, 4°C–55°C, intervals of 10°C; NaCl, 0–20% (w/v), intervals of 0.5%; and pH, 3.0–12.0, intervals of 1.0.

The basal medium for salinity and pH tests contained 0.1% (w/v) of yeast extract. For the salinity, the basal medium was prepared in distilled water, and for the pH test, the salinity was adjusted to 3.5% (w/v) of NaCl (sea salts, Sigma-Aldrich, MO, USA) and then adjusted with MES (pH 5.0–6.0), MOPS (pH 6.5–7.0), Tris (pH 7.5–8.5), or CHES (pH 9.0) at a final concentration of 50 mM.

Marine broth (MB) and seawater oligotrophic medium (SWOM) broth were used for temperature testing, both with original formulations to observe the difference in temperature among synthetic and natural media. For each test, 2 mL of the basal medium was inoculated with 100 μL of a 48-h liquid culture (0.5 McFarland scale = 1.5×10^8 CFU/mL). All tubes were incubated with agitation at 25°C, except those for the temperature tolerance test.

The capacity of strain Mr9^T to grow under anaerobic conditions was determined by inoculating MA and SWOM agar plates and incubation at 25°C for 72 h in an anaerobic jar with an anaerobic gas generator and anaerobic indicator (Oxoid). Oxidase activity was examined by using oxidase sticks (PanReac AppliChem, Darmstadt, Germany) containing tetramethyl-p-phenylenediamine. Catalase activity was assessed by adding a 3% (w/v) H_2O_2 solution to colonies on a solid medium.

The biochemical and physiological tests were evaluated following the methods for characterization tests described by Barrow and Feltham (1993). Hydrolysis of Tween 80, gelatin, and DNA, urease, indole production from tryptophan, methyl red and Voges–Proskauer tests, Simmons's citrate, production of H_2S, and anaerobic growth in the presence of nitrate, arginine, or DMSO were evaluated supplementing the medium with 3.5% (w/v) of NaCl. The physiological profiling was approached by testing the assimilation of a battery of carbohydrates, alcohols, organic acids, and amino acids as a sole source of carbon and energy. The basal medium contained 0.05% (w/v) of yeast extract and 3.5% (w/v) of NaCl supplemented with 1% (w/v) of the tested substrate.

For comparative purposes, the type strain of the closest species to *Leeuwenhoekiella* sp. strain Mr9^T according to the 16S rRNA gene sequence comparison, *Leeuwenhoekiella nanhaiensis* KCTC 42729^T, and the type strain of the type species of the genus, *L. marinoflava* Laboratory of Microbiology of Ghent University (LMG) 1345^T, were purchased from the Korean Collection for Type Cultures (KCTC) and from the Laboratory of Microbiology of Ghent University/Belgian Coordinated Collections of Microorganisms (BCCM), respectively. All phenotypic and physiological tests were determined in triplicate including the reference strains and under the same laboratory conditions. The results were reported after 48–72 h, based on the controls (without inoculum).

2.2.5. Chemotaxonomic analysis

The polar lipid analysis of *Leeuwenhoekiella* sp. strain Mr9^T was carried out by DSMZ services, Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures (Germany), from the deposited strain *L. parthenopeia* Mr9^T (DSM 112950). In brief, polar lipids were extracted using the modified Bligh and Dyer's (1959) method and separated by two-dimensional silica gel thin-layer chromatography. Total lipids were revealed by spraying molybdate-phosphoric acid and specific reagents to detect defined functional groups.

The composition of cellular fatty acids methyl esters was also determined following the protocol recommended by the MIDI Microbial Identification System (Sasser, 1990). This determination was carried out at the CECT, Spanish Type Culture Collection (Valencia, Spain). The biomass of strain Mr9^T for the fatty acid determination was obtained from a culture on MA after 72 h at 30°C of incubation. The cellular content of fatty acids was analyzed by gas chromatography with an Agilent 6850 gas chromatograph and identified according to the TSBA6 method using the Microbial Identification Sherlock software package (MIDI, 2012).

2.3. Relative abundance of *Leeuwenhoekiella* sp. strain Mr9^T in marine habitats

To explore the occurrence of the *Leeuwenhoekiella* sp. strain Mr9^T in marine microbial communities from different habitats, the genome of this strain was mapped against 41 metagenomes of the Mediterranean Sea. The mapping involved two metagenomic collections from Mediterranean Sea datasets (**Supplementary Table 2**) derived from the global survey Tara Oceans expedition (Pesant et al., 2015; Sunagawa et al., 2015) and the metagenomic profile study at different depths during seasonal stratification and mixing period along the water column (Haro-Moreno et al., 2017, 2018, 2021; López-Pérez et al., 2017).

The genome sequence of *Leeuwenhoekiella* sp. strain Mr9^T was recruited from the metagenomic datasets using BLASTn (RRID: SCR_001598) search, considering only the reads that matched the genome with a nucleotide similarity ≥99% over a minimum alignment length of ≥50 bp,

and discarding all recruitments with less than three reads per kilobase of genome per gigabase of metagenome (RPKG) (Haro-Moreno et al., 2018).

Relative abundance was calculated and plotted using the abundance scale from 0.0 to 0.01% for each of the five categories of metagenomic collections: winter; early autumn incubation samples; early autumn depth profile stratification; summer and early autumn; and Tara Oceans Expedition.

2.4. Comparative genomic analysis of secondary metabolite biosynthetic gene clusters

The potential of species of the genus *Leeuwenhoekiella* to produce novel compounds was predicted by exploring the secondary metabolite biosynthetic gene clusters (BGCs) encoded in their genomes using antiSMASH (RRID: SCR_022060) v.6.0 (Blin et al., 2021). All genome sequences derived from type strains of *Leeuwenhoekiella*, including *L. parthenopeia* strain Mr9^T, were analyzed. Cluster sequence similarities were obtained by comparison with experimentally characterized genes encoding the biosynthesis of known chemical molecules present in the Minimum Information about a Biosynthetic Gene cluster (MIBiG) database (Kautsar et al., 2019). BGC relationships among species were circularly displayed using Circos (RRID: SCR_011798) Table Viewer visualization software v.0.63-9 (Krzywinski et al., 2009). In addition to antiSMASH, the antimicrobial potential was evaluated by predicting gene clusters responsible for producing post-translationally modified peptides (RiPPs) and other bacteriocins using the BAGEL 4 web server (van Heel et al., 2018). To determine the BGC homologies with known and patented proteins, each sequence of core biosynthetic gene cluster was translated into amino acids and further searched against non-redundant proteins and patented protein sequences database (pataa) using BLASTP (RRID: SCR_001010). Assignation of proteins to families was carried out using the Pfam (RRID: SCR_004726) database (Mistry et al., 2021).

Biosynthetic gene clusters diversity and evolution across the genus *Leeuwenhoekiella* were assessed by calculating the Jaccard index (JI), adjacency index (AI), and domain sequence similarity (DSS) of each BGC by employing the software BiG-SCAPE (Biosynthetic Gene Similarity Clustering and Prospecting Engine) (RRID: SCR_022561) and CORASON (CORe Analysis of Syntenic Orthologs to prioritize Natural Product Biosynthetic Gene Clusters) (Navarro-Muñoz et al., 2020). Based on sequence similarity networks that encode the biosynthesis of highly similar or identical molecules, BGC sequences from each genome were linked to enzyme phylogenies to create gene cluster families (GCFs). This approach allowed us to obtain a global biosynthetic profile to determine the potential of all species of this genus to produce novel compounds.

2.5. Lipidomic bioassay of *Leeuwenhoekiella parthenopeia* Mr9^T on tumor cell viability

2.5.1. Total lipid extraction of bacterial cell membrane

As several antitumor compounds have been obtained from marine bacteria, we evaluate the *in vitro* activity of *L. parthenopeia* Mr9^T extracts on eukaryotic cells. Since the non-diffusible yellow pigment of strain Mr9^T is retained in the cells and low released to the medium, a poor yield of extract not displaying any activity was obtained from the free-cell supernatant. Therefore, the extraction of the total lipidic content of the cell membrane (lipidome) was performed following a modified Bligh and Dyer protocol (Bligh and Dyer, 1959). This protocol allows the extraction of total lipids constituents of cell matrix such as lipopolysaccharides, phospholipids, glycolipids, lipoproteins, carotenoid pigment content, and other medium-to-low polarity compounds included in the cells. As carotenoid pigments are liposoluble, these are extracted together as lipids. In detail, strain Mr9^T was cultivated in a mat over SWOM solid medium, and after 7 days, when the pigmentation was more intense, the cell biomass was collected and washed by resuspending it in sterile seawater or 3.5% (w/v) NaCl sterile solution and lately centrifuged.

The pellet was resuspended with 3.5% NaCl solution in equal proportion 1:1 (w/v), that is, to 1 g of pellet, 1 mL of 3.5% NaCl was added to obtain an aqueous cell suspension. A monophasic mixture of cell suspension, methanol, and chloroform was created in a ratio of 0.8:2:1 (v/v). This ratio ensures that all membrane components independent of their affinity (hydrophobic or hydrophilic) are solubilized in a monophase. After gently mixing by inversion for 1 h, the mixture was centrifuged, and the supernatant was separated from the colorless pellet. The extraction from the pellet was repeated and collected from both supernatants. The monophase of supernatants was disrupted by adding 150 μL of KCl 0.2 M. Bilayer phases were obtained after centrifugation, and the lower phase of chloroform was collected in a weighted empty glass vial and dried under a fume hood. A total of 200 μL of chloroform was added to the residual aqueous layer, mixed, and centrifuged again until residual pigments were removed from the suspension. The dried extracts were weighed and stored at −20°C. For screenings, a stock of the extract was prepared by dissolving it in DMSO at a final concentration of 100 mg/mL.

2.5.2. Cell line culture

The lipid extract from *L. parthenopeia* Mr9^T was tested against two human tumor cell lines: metastatic prostate adenocarcinoma DU-145 (ATCC HTB-81^{TM}) and glioblastoma U-87 MG (ATCC HTB-14^{TM}) of moderate

and high malignancy, respectively. As a normal cell line, immortalized human keratinocyte HaCaT (RRID: CVCL_0038) was used. All cell lines were cultured in Dulbecco's Modified Eagle Medium (DMEM) with high glucose, supplemented with inactivated 10% (v/v) of fetal bovine serum (FBS), 100 IU/mL of penicillin G, 100 mg/mL of streptomycin, and 2 mM L-glutamine, and incubated at 37°C in a 5% CO_2 humidified atmosphere (Zuchegna et al., 2020).

2.5.3. Antiproliferative assay and statistical analysis

The cell viability of tumor and normal cells after treatment with the lipid extract of *L. parthenopeia* Mr9^T was assessed by colorimetric MTT [3-(4,5-dimethylthiazolyl)-2,5-diphenyl tetrazolium bromide] assay. In detail, all cell lines were seeded with 3×10^3 cells/well in 96-well microtiter plates and incubated for 24 h at 37°C for cell adhesion. After this time, the medium was replaced with fresh medium with increasing concentration (3.5 to 1,000 μg/mL) of lipid extract of strain Mr9^T.

For dose–response curve, 9-point serial dilutions (1:2) were performed. The vehicle DMSO without lipid extract (zero drug: 0 μg/mL) was used as a control. Plates were incubated for 24, 48, and 72 h. After each test time, the medium was removed and 100 μL of a 0.5-mg/mL MTT solution was added to each well and incubated at 37°C for 3 h. Then, the MTT solution was gently removed and 100 μL of DMSO was added to dissolve the formazan crystals. The plate was gently mixed for about 2 min under light protection, and the optical density in each well was measured using a microplate spectrophotometer (TECAN Infinite® 200 PRO) at 570 nm. Three individual replicates of the experiment were performed, each time in triplicate.

Statistical analyses were performed using GraphPad Prism Software V9.4.1 (RRID: SCR_002798) v.9.4.1. Data are expressed as a percentage normalized with controls. Values are represented as the mean ± standard deviation (SD) of three independent experiments in triplicate (N 9). Statistical significance between groups was determined by two-way analysis of variance (ANOVA) (mixed model) followed by Dunnett's multiple comparisons test.

In addition, statistical significance between the different cell lines for each biological replicate (time tested) was determined by two-way ANOVA followed by Tukey's multiple comparisons test with a single pooled variance [p-value in detail: 0.002 (**), 0.0002 (***), <0.0001 (****)]. A p-value < 0.05 is considered statistically significant and a p-value < 0.0001 is considered statistically highly significant. The IC_{50} value of the total lipid extract (lipidome) was determined by non-linear regression/dose–response inhibition [curve fit (Inhibitor) vs. normalized response].

2.6. Dereplication: HPLC/UV/HRMS analysis

Analysis of the composition of the lipidic extract of *L. parthenopeia* Mr9^T was performed by liquid chromatography hyphenated to UV detection and high-resolution tandem mass spectrometry (HPLC-UV-HRMS) using a previously reported analytical method (Martín et al., 2014). Nuclear magnetic resonance (^{1}H NMR) spectroscopy and heteronuclear single quantum coherence (HSQC) analyses were performed in a mixture of CD_3OD–$CDCl_3$ at 24°C on a Bruker Avance III spectrometer equipped with a 1.7-mm micro-cryoprobe.

3. Results

3.1. Strain Mr9^T is a novel species of the phylum *Bacteroidota*

The use of sampling strategies that consider seasonal and diel factors was the premise of this study to reach the hidden microbiota. From night serial samplings carried out in late autumn 2020, a collection of isolates, including the strain Mr9^T, was obtained from reef seawater in the gulf of Naples, Italy.

The physicochemical parameters of seawater in the sampling point registered a temperature of 13.8°C, pH 8.17, TDS of 2,515 ppm, and salinity of 3.4%. The strain Mr9^T was first isolated on a natural seawater oligotrophic medium (SWOM) producing smooth and yellow colonies after 4 days of incubation at room temperature (**Supplementary Figure 1**). According to the 16S rRNA gene sequence comparison, the strain Mr9^T was affiliated with the genus *Leeuwenhoekiella*, most closely related to the species *L. nanhaiensis* G18^T, sharing 98.6% sequence similarity. The genomic similarities inferred by measurements of OGRIs established that strains Mr9^T and *L. nanhaiensis* G18^T have 88.0% of average nucleotide identity (OrthoANIu) and 33.9% of digital DNA–DNA hybridization (dDDH) (**Figure 1A**). Species delineation based on a 98.7% for the standard taxonomic marker 16D rRNA gene sequence (Chun et al., 2018), on a 95% ANI (Konstantinidis and Tiedje, 2005; Goris et al., 2007; Richter and Rosselló-Móra, 2009; Chun and Rainey, 2014), and 70% dDDH (Stackebrandt and Goebel, 1994; Auch et al., 2010), strain Mr9^T represents a new species of the genus *Leeuwenhoekiella*, since it exhibited percentages to known species below these established thersholds.

The phylogenomic tree (**Figure 1B**), based on the alignment of 2,103 translated single-copy core genes, clearly showed that *Leeuwenhoekiella* sp. Mr9^T forms a separate branch, clustering with *L. nanhaiensis* and it is not related to any other species affiliated with this genus. A similar result was observed in the phylogenetic tree inferred from the alignment of 16S rRNA gene sequences (**Supplementary Figure 2**). The genome-to-genome relatedness obtained from dDDH and OrthoANIu

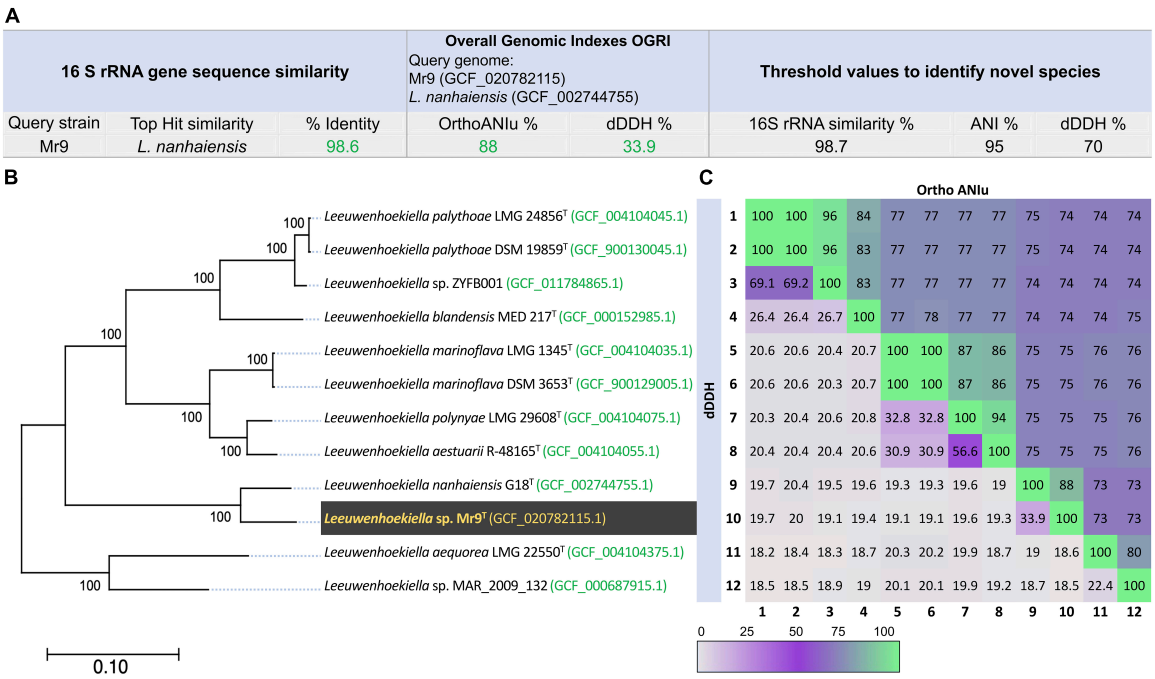


FIGURE 1
(A) 16S rRNA gene sequences similarity and overall genome relatedness index (OGRI) values among strain Mr9^T and *Leeuwenhoekiella nanhaiensis* G18^T. **(B)** Approximately maximum-likelihood phylogenomic tree based on the alignment of 2,103 translated single-copy core genes, showing the relationship among strain *Leeuwenhoekiella* sp. Mr9^T and the members of the genus *Leeuwenhoekiella*. The genome sequence accession numbers are shown in parentheses. Bootstrap percentages are shown above the branches. Bar, 0.10 substitutions per nucleotide position. **(C)** Heatmap plot of genome-to-genome analysis by OrthoANIu (upper triangle) and dDDH (lower triangle) showing the percentage of similarity among the genus under study. The colors are displayed according to the color key, with a 100% value when converging in the matrix with the same species.

analysis against the 11 genomes of species of the genus *Leeuwenhoekiella* (**Figure 1C**) showed that *Leeuwenhoekiella* sp. Mr9^T is not related to any described species of this genus, as previously confirmed by the limits for species delineation. Taxonomically, the genus *Leeuwenhoekiella* belongs to the family *Flavobacteriaceae* (Reichenbach, 1992), order *Flavobacteriales*, class *Flavobacteriia* (Bernardet, 2015), and phylum *Bacteroidota* (Oren and Garrity, 2021). Currently, this genus constituted only seven species (Parte et al., 2020), all derived from marine or sea-related habitats.

The differential phenotypic and physiological features of *Leeuwenhoekiella* sp. Mr9^T with respect to the closest related species *L. nanhaiensis* and the type species of the genus, *L. marinoflava*, are shown in **Table 1**. The polar lipid profile obtained by two-dimensional thin-layer chromatography (2D TLC) is shown in **Supplementary Figure 3**.

Based on the results obtained from the 16S rRNA gene sequences similarity, OGRIs, phylogenomic reconstruction, and physiological features, the strain Mr9^T constitutes unequivocally a new species of the genus *Leeuwenhoekiella*, for which the name *L. parthenopeia* sp. nov. is proposed. The type strain Mr9^T was deposited in three public culture collections: Spanish Type Culture Collection (CECT); Laboratory of Microbiology of Ghent University LMG/Belgian Coordinated Collections of Microorganisms (BCCM); and German Collection of Microorganisms and Cell Cultures GmbH (DSMZ). *L. parthenopeia* Mr9^T is available in these culture collections as CECT 30318^T, DSM 112950^T, and LMG 32428^T, respectively.

3.1.1. Description of *Leeuwenhoekiella parthenopeia* sp. nov.

Leeuwenhoekiella parthenopeia (par.the.no.pei'a. L. fem. adj. *parthenopeia*, of or belonging to Naples) cells are Gram-stain-negative rods with 0.6 × 1.0 μm; colonies are circular, entire, yellow-to-orange pigmented, with 2–4 mm in diameter in medium MA incubated at 25°C for 7 days; able to grow in oligotrophic medium with 3.0–8.0% (w/v) salts (optimal at 3.5%), in the pH range of 6.5–9.0 (optimal at pH 8.0) and from 10.0 to 37.0°C (optimal at 25°C); chemoorganotrophic and aerobic; and catalase positive and oxidase negative. Nitrate is reduced, but nitrite is not. Indole and H_2S are not produced. Voges–Proskauer and methyl red are negative. Aesculin is hydrolyzed, but casein, DNA, gelatin, Tween 80, and starch are not hydrolyzed. Acid is produced from L-arabinose, D-galactose, D-glucose, D-maltose, D-sucrose, and D-xylose but

TABLE 1 Differential features of strains: Mr9^T, *Leeuwenhoekiella marinoflava* LMG 1345^T, and *L. nanhaiensis* KCTC 42729^T.

Feature	Strain Mr9^T	*L. nanhaiensis* KCTC 42729^T	*L. marinoflava* LMG 1345^T
Cell size (width × length; μm)	0.6 × 1.0	0.4–07 × 1.4–4.1	1.0 × 1.5
Colony pigmentation	Yellow/orange	Orange	Yellow
Oxidase	–	–	+
Nitrate reduction	+	–	–
Hydrolysis of:			
Aesculin	+	+	–
Casein	–	–	+
DNA	–	+	–
Tween 80	–	+	+
Starch	–	+	–
Acid production from:			
L-Arabinose	+	+	–
D-Galactose	+	–	+
Glycerol	–	–	+
D-Glucose	+	+	–
D-Maltose	+	–	–
D-Sucrose	+	–	–
D-Xylose	+	+	–
DNA G + C (mol%, genome)	42.2	42.1	37.5

All data are from this study. +, positive reaction; –, negative reaction.

not from glycerol. Major fatty acids (>10%) are $C_{17:0}$ iso 3-OH and iso-$C_{15:0}$, followed by iso-$C_{15:1}$ G, iso-$C_{17:1}$ ω9c/$C_{16:0}$ 10-methyl, and $C_{16:1}$ ω6c/$C_{16:1}$ ω7c. The polar lipid profile consists of two aminolipids, one glycolipid, three lipids of undetermined composition, and a phospholipid with similar mobility to phosphatidylethanolamine (PE) as major lipid. The DNA G + C content is 42.2 mol% (genome).

The type strain is Mr9^T (=CECT 30318^T = DSM 112950^T = LMG 32428^T), isolated from seawater in the Gulf of Naples, Italy. The GenBank/EMBL/DDBJ accession numbers for the 16S rRNA and the complete genome of the type strain are MW785572 and JAJGMW000000000, respectively.

3.2. *Leeuwenhoekiella parthenopeia* belongs to the rare biosphere

The occurrence of *L. parthenopeia* Mr9^T in marine habitats was mapped against 41 Mediterranean Sea metagenomes categorized into five groups by depths and seasons.

As observed in the chart (**Figure 2**), *L. parthenopeia* showed a relatively higher abundance as follows: 0.0026% at 40-m depth during winter; 0.0066% at 60-m depth after 14 h of incubation of the sample from early autumn, October; 0.0047% at 15-m depth in a non-incubated sample from early autumn depth profile; and 0.0090% at 3,000-m depth across summer and early autumn. Finally, in the TARA Oceans expedition dataset, the higher abundance was 0.0065% at 5-m depth during December.

The results obtained from the metagenomic mapping across depths and seasons showed that *L. parthenopeia* is present in the microbial communities of 41 metagenomes investigated with a maximum detected relative abundance of 0.0090%. The average relative abundance of *L. parthenopeia* Mr9^T from all categories of metagenomic data from the Mediterranean Sea was as low as 0.0029%, revealing the scarce occurrence of this bacterium in the total microbial communities analyzed.

3.3. Biosynthetic potential of the genus *Leeuwenhoekiella*

3.3.1. Biosynthetic gene cluster analysis of *Leeuwenhoekiella parthenopeia* Mr9^T

The potential of *L. parthenopeia* Mr9^T to synthesize specialized compounds was inferred by predicting the secondary metabolite biosynthetic gene clusters (BGCs) encoded in its genome. Through antiSMASH analysis, the biosynthetic profile of *L. parthenopeia* Mr9^T consisted of five BGCs belonging to four typologies of secondary metabolite clusters: two terpenes; one type III polyketide synthase (T3-PKS); one NRPS; and one aryl polyene (APE). As shown in **Table 2**, just one BGC belonging to the terpene region exhibited a low homology of 28% with a carotenoid biosynthetic gene cluster from *Myxococcus xanthus*, while the rest of the BGC types do not show similarities with any experimentally characterized gene encoding the synthesis of known chemical molecules present in the MIBiG repository (Kautsar et al., 2019). Nevertheless, according to the protein family database Pfam, the amino acid sequence of core enzyme-coding genes of each BGC of *L. parthenopeia* Mr9^T was related with multifunctional enzymes as follows: terpene—squalene/phytoene synthase and lycopene cyclase protein; T3-PKS—chalcone and stilbene synthases; NRPS—AMP-binding enzyme/phosphopantetheine attachment site; APE—beta-ketoacyl synthase. Besides the BGCs predicted and their relatedness with protein families, the mining of gene clusters involved in the biosynthesis of bacteriocins, and RiPPs obtained with BAGEL 4 web server revealed two areas of interest in the genome with homologies higher than 97% with gene clusters related with the biosynthesis of Zoocin A (bacteriocin) and Sactipeptides (RiPP), both involved in the synthesis of natural complex products.

Furthermore, amino acid sequence homology with patented protein sequences (pataa) showed that the biosynthetic gene

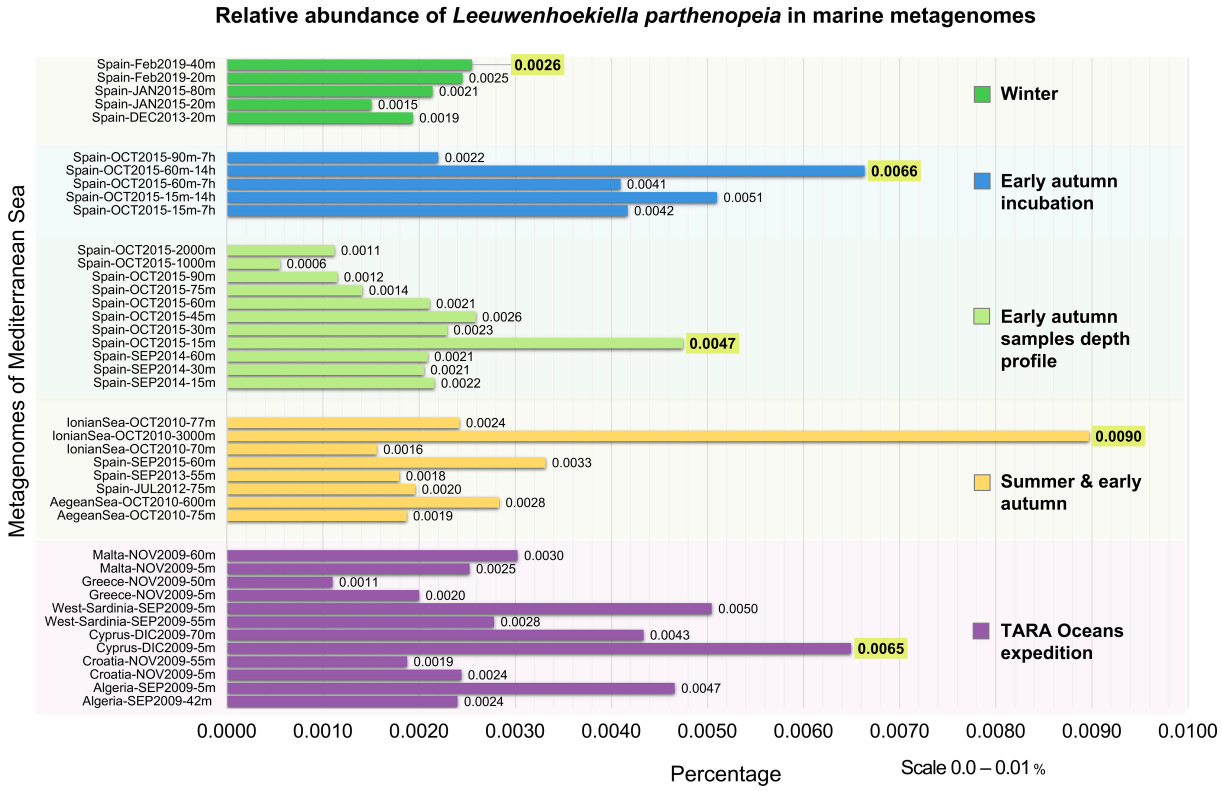


FIGURE 2
Occurrence of *Leeuwenhoekiella parthenopeia* Mr9^T in marine metagenomes from the Mediterranean Sea in the five different categories of metagenomic collections. Y-axis: winter, early autumn incubation, early autumn depth profile stratification, summer and early autumn, and Tara Oceans expedition. X-axis: relative abundance in percentage with the scale from 0.0 to 0.01%. The higher relative abundance for each data block is highlighted in yellow.

content of *L. parthenopeia* Mr9^T does not exhibit full similarity to any registered patent for various applications, showing a frequent match with patent US 8119385 that falls into the categories of terpene (carotenoids), T3-PKS, and bacteriocin. The identity percentages with five different patents ranged from 22.39 to 67.08%, thus revealing the high potential of *L. parthenopeia* Mr9^T for the synthesis of novel polypeptides for biotechnological application.

3.3.2. Diversity and distribution of BGCs across *Leeuwenhoekiella*

In the same way as the profile of *L. parthenopeia* Mr9^T, the specialized metabolite potential of all species of the genus *Leeuwenhoekiella*, including *L. parthenopeia* Mr9^T, was assessed by predicting the BGC content and diversity in each of the eight species through antiSMASH. A content of three to four BGCs was identified in each species, except *L. parthenopeia* Mr9^T, which harbors five BGCs as previously detailed (**Supplementary Figure 4**). A total of 30 BGCs (29 + 1 hybrid) were identified across the genus and, like the BGC profile of *L. parthenopeia* Mr9^T, only one of the two BGCs terpenes typologies exhibited 28% homology to a carotenoid gene cluster of *M. xanthus* (**Supplementary Table 3**). Moreover, this type of terpene is the unique gene cluster present in all species that shows homology with known BGCs; however, its similarity is very low and probably would not code for the synthesis of the same carotenoid as that of *M. xanthus*. The rest of the BGCs types showed no similarities to any experimentally characterized gene encoding the synthesis of known chemical molecules present in the MIBiG repository.

To determine how the gene cluster profile varies across *Leeuwenhoekiella*, a comparative genomic study was performed to show the correlation between species and categories of BGC. Through the chord diagram of **Figure 3**, it can be observed that three types of gene clusters are conserved in all eight species of the genus: two terpenes and one T3-PKS. It can also be evidenced by the different BGC contents between *L. parthenopeia* Mr9^T and the most closely related species *L. nanhaiensis* G18^T, where in the latter, the gene clusters NRPS and APE are absent. The exclusive characteristic of *L. parthenopeia* Mr9^T BGC pattern with respect to the other species of the genus is the presence of the NRPS gene cluster, which is absent in the rest of the species. Although an NRPS is also present in *L. polynyae*, it constitutes a hybrid

TABLE 2 Biosynthetic gene cluster types of *Leeuwenhoekiella parthenopeia* Mr9^T showing its relatedness with known molecules, protein families hits, and patented proteins based on sequence similarity.

BGC's content			
Prediction tool: AntiSMASH			
Type	Molecule related similarity (MIBiG)	Protein family identity (Pfam hit)	Patented protein sequences (pataa) similarity
Terpene	Carotenoid 28%	PF00494.19 Squalene/phytoene synthase	US 8119385 48.02% (Polynucleotides and polypeptides for pharmaceutical, agricultural, cosmetic, and nutraceutical context)
	Unknow	PF05834.12 Lycopene cyclase protein	US 7630836 34.69% (Polynucleotide array)
Aryl polyene	Unknow	PF00109.26 and PF02801.22 Beta-ketoacyl synthase, N and C terminal domain	US 6673910 34.95% (Aminoacidic sequences for diagnostics and therapeutics of *Moraxella catarrhalis*)
T3-PKS	Unknow	PF00195.19 and PF02797.15 Chalcone and stilbene synthases, N and C terminal domain	US 8119385 67.08%
NRPS-like	Unknow	PF00501.28 AMP-binding enzyme PF00550.25 Phosphopantetheine attachment site	US 7319142 22.39% (Insecticidal proteins for controlling insect infestation)
Bacteriocins and RiPPs content			
Prediction tool: BAGEL4			
Class Gene name	Function	Protein homology redundant	Patented protein sequences (pataa) similarity
(RiPP) Sactipeptides BmbF	GTP 3′,8-cyclase Peptoclostridium acidaminophilum	GTP 3′,8-cyclase MoaA 97.21%	US 11078247 43.34% (Insecticidal proteins for controlling insect infestation)
(Bacteriocin) Zoocin A	Zoocin A	Peptidase M23 97.36%	US 8119385 50.66%

The best score for amino acidic homology with patented protein sequences is marked in orange.

NRPS-like + T1-PKS cluster, since both BGCs overlap in the same region of the genome, therefore counting separately within its category.

3.3.3. BGC similarity networking and evolution of the genus *Leeuwenhoekiella*

To evaluate the diversity and evolutionary history of BGC repertoires across members of the genus *Leeuwenhoekiella*, a comparative interspecies analysis of conserved-coding genes was performed by means of sequence similarity networking and BGC phylogenetic reconstruction *via* BiG-SCAPE/CORASON (Navarro-Muñoz et al., 2020).

The similarity networking groups BGCs of each species into gene cluster families (GCF) according to the presence, copy number, organization, and sequence homology of the protein domains on antiSMASH-predicted regions. As shown in the similarity matrix calculation obtained through BiG-SCAPE (**Figure 4A**), the clustering of the total 30 BGC content of *Leeuwenhoekiella* gave rise to nine GCFs affiliated with four types of BGCs, the terpene category being the most represented. It is observed in the matrix that all eight species share the same core enzyme-coding genes that form two families: FAM_00027 and FAM_00026. These GCFs are affiliated with the BGC types terpene and T3-PKS, respectively. Conversely, three singletons were identified: FAM_00016 belonging to terpenes corresponding to *L. marinoflava*; FAM_00003 belonging to NRPS which was exclusive of *L. parthenopeia*, and FAM_00028 presents in *L. polynyae;* the latter corresponds to the chemical hybrid NRPS-like-T1-PKS, which explains the two times computing and its assignation to two categories, NRPS and PKS. These singletons result when the domains present in the BGC do not reach the thresholds that capture sequence similarity, synteny, or presence–absence of the Jaccard index against BGCs from other species to form a network, thus, a family of shared

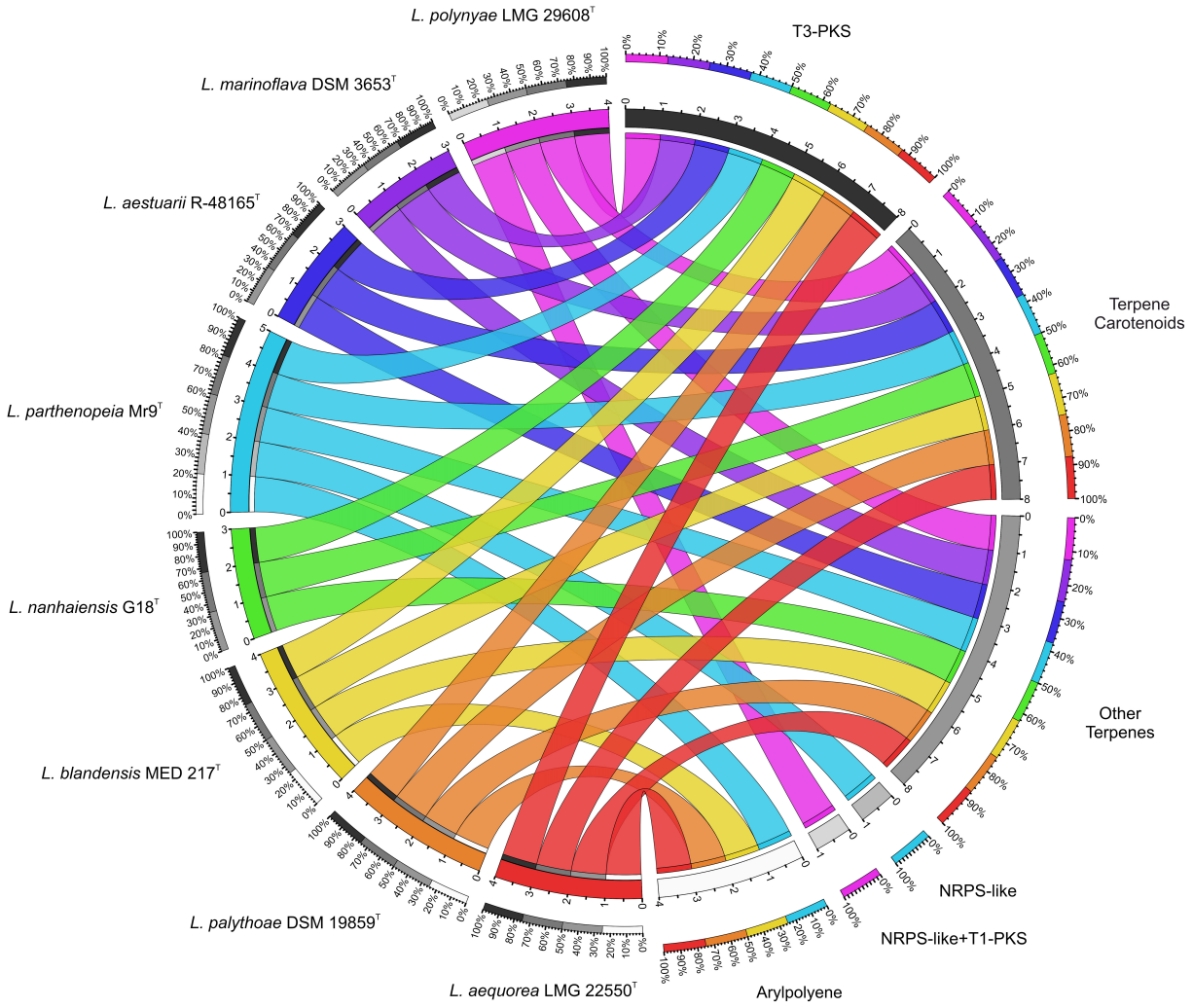


FIGURE 3
Circle plot showing the relationships among the eight species of *Leeuwenhoekiella* (left half of the circle) and the biosynthetic gene cluster (BGC)'s diversity by typologies (right half of the circle). Each genome is represented by different colors, which are segmented according to the number of BGC content in their genomes. The chords connect both sides indicating, according to their width, the number of BGCs belonging to each typology. The outer rings in the gray scale represent the percentage of presence of BGC and in the color scale the proportion of genomes per BGC.

gene groups could not be defined, remaining as an individual member. In all cases, whether as clusters or singletons, after the GCF's comparison with MIBiG reference BGCs, none of the nine GCFs match any pathways responsible to produce known compounds.

Evolutionary relationships between BGCs and within GCF were initially inferred by CORASON inside the GCF defined for BiG-SCAPE for each family. CORASON establishes phylogenies based on BGC's conserved core of common domains. The clades resulting from the phylogenetic reconstruction may be responsible for the biosynthesis of different types of chemistry due to the association of specific types of genes encoding additional enzymes. In correspondence with the families in the presence–absence matrix, the phylogenetic reconstructions of all BGCs from *Leeuwenhoekiella* (**Figure 4B**) generated six multi-locus approximately maximum-likelihood trees, and the three singletons (FAM_00016, FAM_00003, and FAM_00028) do not form clades. The featured conserved core domains and their related enzymes used to define the GCFs are specified for each tree. In addition, the genes that encode functionally uncharacterized proteins (domains of unknown function, DUF) were identified to see their cryptic potential of novel structures and functions.

To further investigate possible distant phylogenetic relationships among these families, CORASON was run outside the BiG-SCAPE-predicted BGC families. Instead of restricting the analysis to anti-SMASH-predicted regions, full *Leeuwenhoekiella* genomes were used as CORASON

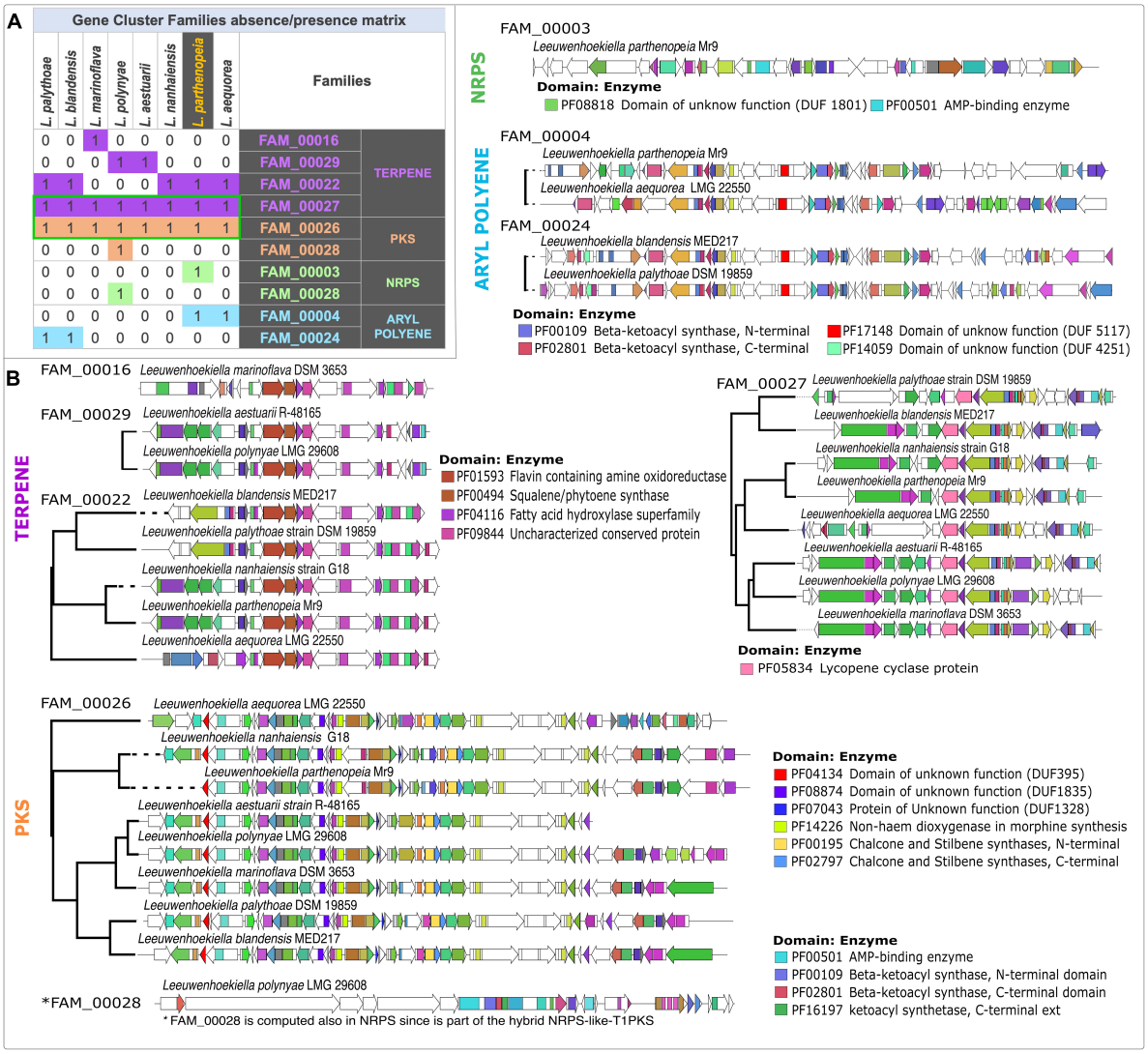


FIGURE 4
(A) Similarity network matrix based on the absence and presence of gene cluster families from each species of the genus *Leeuwenhoekiella*. Green boxes, gene cluster families (GCFs) are present in all members of *Leeuwenhoekiella*. **(B)** Multi-locus approximately maximum-likelihood trees based on the conserved core common domain.

inputs. As a result, families FAM_00016 (*L. parthenopeia*), FAM_00022 (*L. nanhaiensis, L. marinoflava, L. blandensis, L. palythoae, and L. aequorea*), and FAM_00029 (*L. polynyae and L. aestuarii*) were clustered in a single phylogenetic tree (Supplementary Figure 5).

With this observation, it is shown that a new superfamily that comprises a terpene BGC in each of the species is studied here. It is possible that due to the variation in gene content and sequence similarity, BiG-SCAPE split this superfamily into the three families shown in the presence–absence matrix, but these families shared a common evolutionary history as can be seen in Figure 4B. The same is true for aryl polyene families FAM_00004 and FAM_00024, and they share a common ancestor. In fact, though not all the genomes have an antiSMASH region predicted, the beta-ketoacyl synthase genomic vicinity is conserved in all eight genomes as can be seen in the Supplementary Figure 5 produced by CORASON.

According to the featured conserved core domains linked to enzymes, within the terpene category, the GCF: FAM_00016 (singleton), FAM_00029, and FAM_00022 encode the biosynthesis of highly similar or identical molecules to flavin, squalene, chalcones–stilbenes, and an unknown compound associated with an uncharacterized conserved protein. This category highlights the FAM_00027 which is composed of all species of *Leeuwenhoekiella* revealing the potential of members of this genus to synthesize carotenoid pigments as lycopene.

In the category PKS, the FAM_00026 is also clustered with all species of *Leeuwenhoekiella* and encodes the biosynthesis of

carotenoids, chalcone, and stilbenes. The singleton FAM_00028 (*L. polynyae*) also falls in this group, which is duplicated in the category NRPS, since *L. polynyae* contains the hybrid gene cluster NRPS-like-T1_PKS, computing in both groups. The featured conserved domains of this singleton are linked to the family protein AMP-binding enzyme involved in the synthesis of several bioactive compounds and beta-ketoacyl synthase involved in the synthesis of fatty acids. In both cases, given the high capacity to synthesize a wide spectrum of molecules, it is not possible to precise the most probable-related molecule. Regarding the NRPS category, the singleton FAM_00003 (*L. parthenopeia* Mr9^T) possesses the domain of the unknown function (DU1801) and a major facilitator superfamily domain; in both cases, the possible synthesis of a determined molecule is unknown.

Finally, in the category of aryl polyene (APE), two GCFs, FAM_00004 and FAM_00024, possess common conserved domains of unknown functions (DUF 5117 and DUF 4251) for which the eventually synthesized molecules remain unknown, and the beta-ketoacyl synthase is involved in the synthesis of fatty acids. The GCF analysis was performed, allowing us to detect the frequent presence of domains of unknown function (DUFs) in all GCFs by categories, strongly suggesting that this genus harbors the potential to synthesize novel compounds not described nor chemically tested yet.

3.4. Total lipid extract from *Leeuwenhoekiella parthenopeia* affects viability of tumor cells

The potential antiproliferative effect of total lipid extract (TLE) from *L. parthenopeia* Mr9 was evaluated using the colorimetric MTT assay, which measures cellular metabolic activity as an indicator of cell viability. Two human tumor cell lines: metastatic prostate adenocarcinoma DU-145 and glioblastoma U-87 MG of moderate and high malignancy, respectively, were tested. As a normal cell line, it was used to immortalize human keratinocyte HaCaT.

After the treatment of the cell lines with several doses of TLE from 3.5 to 1,000 μg/mL for 24, 48, and 72 h, it was possible to observe that the viability of tumor cells DU-145 and U-87 MG decreases gradually in time- and dose-dependent manner with lower effect on HaCaT (**Figure 5A**). In particular, the cell death of DU-145 and U-87 MG above 50% occurred with 120 μg/mL of extract after 24 h, while the cell death of HaCaT happens only with the highest concentration (1,000 μg/mL). After 48 h, DU-145 and U-87 MG are inhibited above 50% with 60 μg/mL, while HaCaT requires the highest concentration of the extract. In all cell lines, both tumor and normal, the greatest inhibitory effect was observed after 72 h of treatment, where the cell viability of DU-145 and U-87 MG decreased significantly, requiring a concentration of around 15 μg/mL to inhibit the 50% of the cells, while HaCaT requires more than 250 μg/mL. Between the two tumor cell lines, DU-145 and U-87 MG, it is observed that TLE is less effective on U-87 MG at doses lower than 60 μg/mL at 24 and 48 h of treatment, and successively at higher concentration, the difference of the inhibitory effect on both tumor cell line is reduced.

According to the cell viability results and the inhibitory concentration value IC_{50} (**Figure 5B**), the lipid extract of *L. parthenopeia* Mr9^T possesses inhibitory activity against metastatic prostate adenocarcinoma DU-145 and glioblastoma U-87 MG with greater effectiveness on DU-145. The minor effect on immortalized human keratinocyte HaCaT suggests a selective effect on tumor cells.

3.5. Dereplication: Total lipid composition

The chemical composition of the lipid extract from *L. parthenopeia* Mr9^T was analyzed by HPLC-UV-HRMS, and the molecular formulae of major components were determined. **Figure 6** shows the LC/UV trace obtained in the analysis. Searches of molecular formulae of the peaks in the figure against the Dictionary of Natural Products Database (DNP) tentatively identified some compounds and other molecules present in the extract with no matches in the database.

Following the above spectra, 20 peaks were identified, and their molecule formula and putative component are detailed in **Table 3**. From these peaks, nine molecular formulae do not match in the DNP, five corresponds to fermentation medium component (FMC), and six corresponds to known molecules as follows: sulfobacin-like component, sulfobacin A, WB 3559A, WB 3559B, docosenamide, and topostin B-567.

Searches of molecular formula of the peaks in the figure against the Dictionary of Natural Products Database (DNP) tentatively identified some compounds and other molecules present in the extract with no matches in the database. **Figure 7** represents the structures of the compounds tentatively identified in the lipidic extract. Not surprisingly, these compounds present in their structure units of $C_{17:0}$ iso 3-OH, $C_{15:0}$ iso, and $C_{15:1}$ iso fatty acids are identified as major fatty acids in Section "Description of *Leeuwenhoekiella parthenopeia* sp. nov." NMR analysis corroborated the presence of iso-fatty acids as the major components in the lipid extract.

4. Discussion

4.1. Isolation and ecology

The isolation of *L. parthenopeia* strain Mr9^T from a night seawater sample in autumn supports previous studies on the dynamic and seasonal effect on the microbial population. This

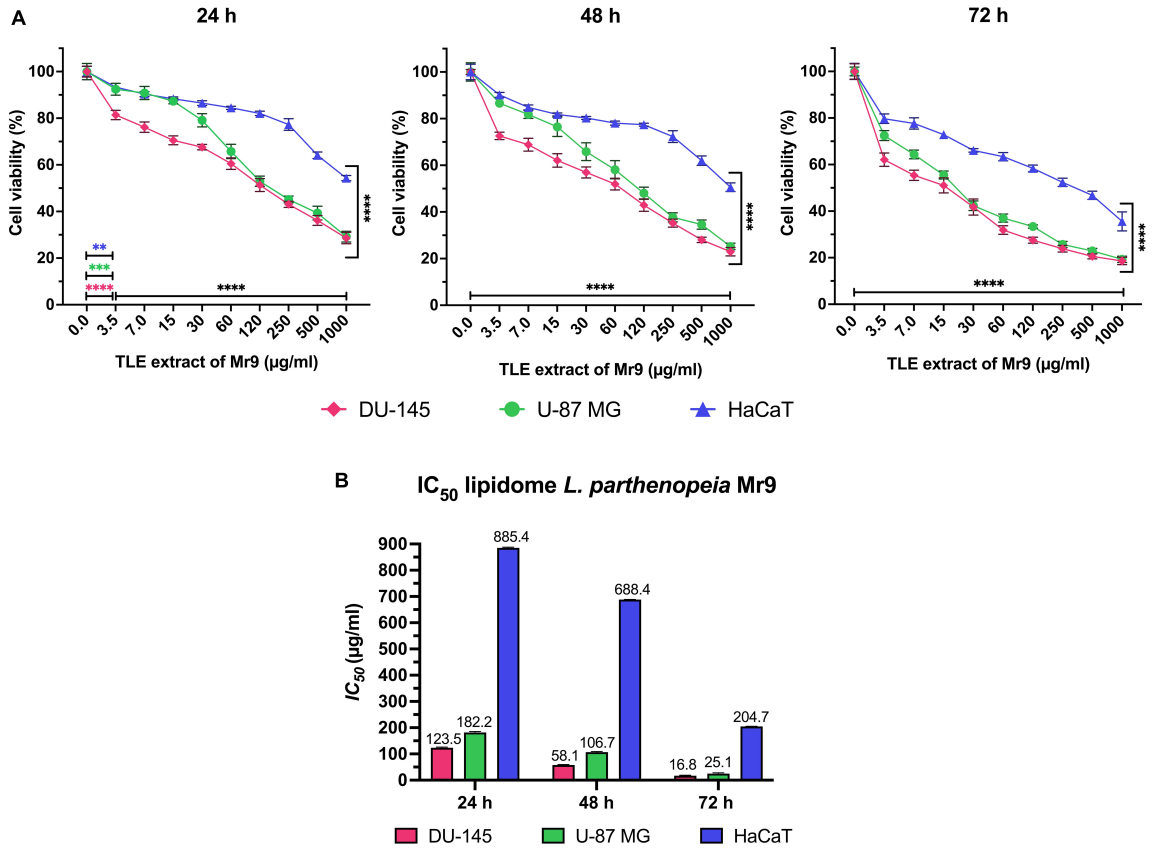


FIGURE 5
(A) Effect of total lipid extract from *Leeuwenhoekiella parthenopeia* Mr9^T cell extract in the cell viability of DU-145, U-87 MG, and HaCaT after 24, 48, and 72 h of stimulation with gradual doses. Untreated cells (0 µg/mL) were used as control and indicate cells treated with dimethyl sulfoxide (DMSO) without *L. parthenopeia* Mr9^T extract. Control corresponds to 100% of viability. The histograms represent the mean ± SD of three independent experiments performed in triplicate (n = 9), normalized to untreated control cells. Statistical analysis: p-value: 0.002 (**), 0.0002 (***), and <0.0001 (****) compared to the control. **(B)** IC_{50} inhibitory concentrations of the total lipid extract (TLE) required to inhibit DU-145, U-87 MG, and HaCaT after 24, 48, and 72 h of stimulation.

isolation could be associated with the natural phenomenon of the DVM and the mixing of the water itself during the cold months since these favor the ascent toward the surface of planktonic particles and solid matter, acting as vectors of the hidden microbiota. The particulate was ascertained by the visual observation and the measurement of TDS that showed high detection in night samples. Genomic and phenotypic features of this strain showed that it constitutes a new species of the genus *Leeuwenhoekiella*, for which we propose the new name *L. parthenopeia* sp. nov.

It is probable that *L. parthenopeia*, as a member of the *Flavobacteriia*, is trapped in small particles such as marine picoplankton, where this class is dominant and typically found at depth (Kirchman et al., 2003; Alonso et al., 2007; Gómez-Pereira et al., 2010; Mestre et al., 2017). In cold waters, a high abundance of some members of *Flavobacteriia* has also been reported (Ghiglione and Murray, 2012; Grzymski et al., 2012), and, in fact, two species of *Leeuwenhoekiella*, *L. aequorea* (Nedashkovskaya et al., 2005) and *L. polynyae* (Si et al., 2015), were isolated from Antarctic Sea water, which also explains that cold seasons would favor the presence of microbial population minorities as a rare biosphere. The lack of isolates of microorganisms belonging to this genus or any member of the phylum *Bacteroidota* in day samples, besides the reduction of particles, demonstrates that the cultivation of *L. parthenopeia* as well as other members within this phylum is not causality and indicates a close association with planktonic particulate, organic matter, algae or neustons colonizing their surface acting as ectosymbiont (Kirchman, 2002; Fernández-Gómez et al., 2013; Mestre et al., 2017; Baumas et al., 2021). Thus, the probability of isolation of similar bacteria may depend on the composition of phytoplankton and temporal dynamics (Rooney-Varga et al., 2005; Teeling et al., 2016; Camarena-Gómez et al., 2021).

Physiological characteristics showed that *L. parthenopeia* has a heterotrophic and versatile metabolism, assimilating various substrates as a carbon and energy source, and suggesting that this bacterium can thrive in a wide range of nutritional circumstances in its habitat. This versatility highly contributed

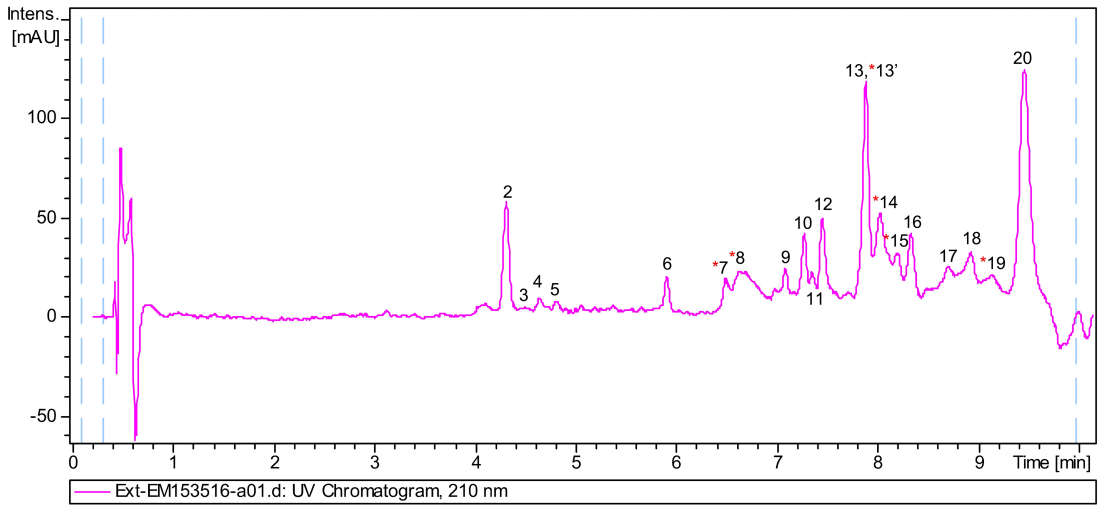


FIGURE 6
Liquid chromatography and high-resolution tandem mass spectrometry analysis of the lipidic extract from *Leeuwenhoekiella parthenopeia* Mr9^T showing 20 peaks in the chromatogram. Red asterisk, known components.

TABLE 3 Peaks with molecule formula and putative components.

Peak	Molecular formula	Putative dereplicated component
1, 2	$C_{22}H_{26}O_6$	Fermentation medium component
3	$C_{15}H_{31}NO_2$	No coincidences in the DNP
4	$C_{21}H_{33}N_3O_3$	Fermentation medium component
5	$C_{14}H_{18}O_5$	Fermentation medium component
6	$C_{21}H_{37}N$	Fermentation medium component
7	$C_{33}H_{67}NO_6S$	Sulfobacin like component
8	$C_{34}H_{69}NO_6S$	Sulfobacin A
9	$C_{33}H_{50}O_8P_2$	Fermentation medium component
10	$C_{35}H_{68}O_6S$	No coincidences in the DNP
11	$C_{36}H_{70}N_2O_6$	No coincidences in the DNP
12	$C_{37}H_{72}N_2O_6$	No coincidences in the DNP
13	$C_{33}H_{50}O_7P_2$	No coincidences in the DNP
13′	$C_{36}H_{66}N_2O_7$	WB 3559A
14	$C_{37}H_{68}N_2O_7$	WB 3559B
15	$C_{22}H_{43}NO$	Docosenamide
16	$C_{33}H_{60}O_3S_2$	No coincidences in the DNP
17	$C_{35}H_{62}O_3$	No coincidences in the DNP
18	$C_{53}H_{58}O_7P_2$	No coincidences in the DNP
19	$C_{34}H_{65}NO_5$	Topostin B-567
20	$C_{33}H_{52}O_6P_2$	No coincidences in the DNP

to its fast laboratory domestication, exhibiting the ability to grow in synthetic marine media without requiring natural seawater in culture media as the main component of primary cultures. Its role in nature might fulfill its function as a heterotroph, that is, degrading organic matter (Alonso et al., 2007; Gómez-Pereira et al., 2010; Xue et al., 2020).

The almost undetectable genomic recruitment patterns of the new species *L. parthenopeia* when mapping the 41 metagenomes from the Mediterranean Sea revealed that this species is poorly represented with a mean relative abundance as low as 0.0029%, considering that members of the class *Flavobacteriia* (phylum *Bacteroidota*) are prominent in marine environments across different seasons and depth, accounting from 4.7 to 13.9% (Cram et al., 2015) and from 3 to 20% in planktonic particles (Mestre et al., 2017; Baumas et al., 2021). This infimum percentage reveals the scarce occurrence of this novel bacterium in the total microbial communities analyzed. Based on the arbitrary threshold of 0.1 or 0.01% of relative abundance for defining rare biosphere (Pedrós-Alió, 2012), *L. parthenopeia*, similar to *L. blandensis* (Pedrós-Alió, 2013), is part of the low-abundant fraction of microbial population and therefore belongs to the rare biosphere.

According to all the categories of metagenomes analyzed, we could not observe a pattern of distribution that allows us to establish the depth or the specific season at which *L. parthenopeia* is relatively more abundant. However, we can affirm that despite its scarcity, *L. parthenopeia* is widely distributed in the Mediterranean Sea. A remarkable fact is the highest abundance of *L. parthenopeia* found in the Ionian Sea metagenome, at a depth of 3,000 m, which might suggest that *L. parthenopeia* could be more abundant at abyssal depths. In fact, its closest relative, the species *L. nanhaiensis*, was isolated at a depth of 2,000 m (Liu et al., 2016), while a culture-independent study about the depth and seasonal distribution of *Bacteroidota*, showed that the members of the genus *Leeuwenhoekiella*

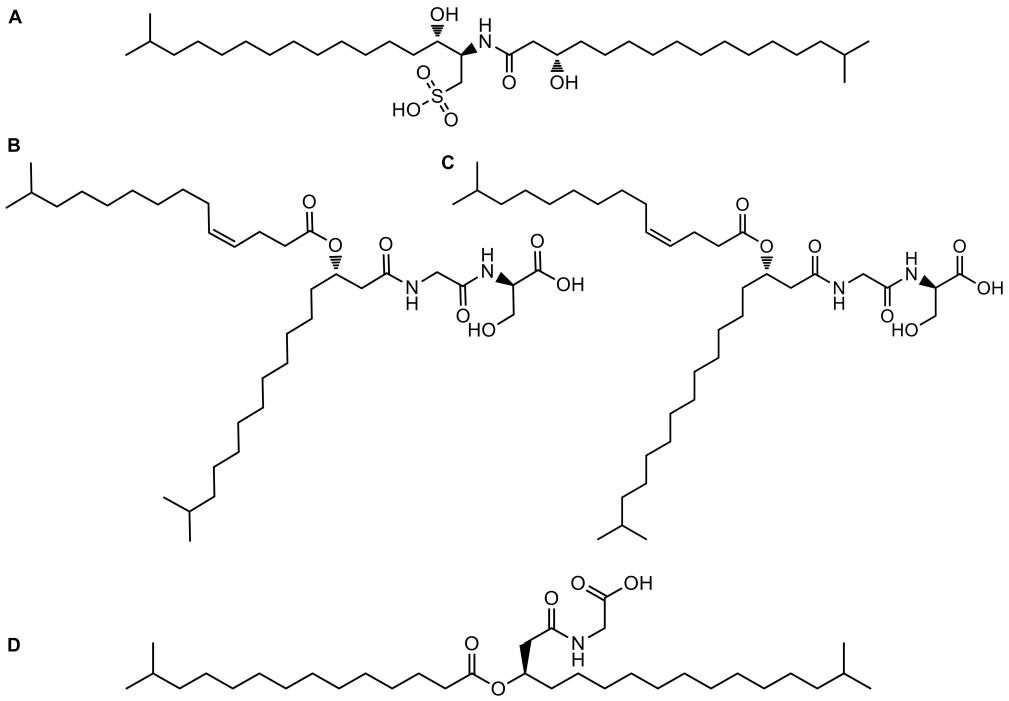


FIGURE 7
Structures of compounds tentatively identified in the liquid chromatography hyphenated to UV detection and high-resolution tandem mass spectrometry (HPLC-UV-HRMS) analysis of the lipidic extract of *Leeuwenhoekiella parthenopeia* Mr9^T: **(A)** sulfobacin A; **(B)** WB 3559A; **(C)** WB 3559B; **(D)** topostin B-567.

were detected in deep marine waters (Díez-Vives et al., 2019), therefore, supporting that vertical migration besides the mixing waters are determinant factors to reach the rare microbiota. Our study shows that the sample pre-incubation increases the abundance of *L. parthenopeia*, as observed in the metagenomes from incubated samples collected in October, where the presence of *L. parthenopeia* is doubled, especially after 14 h with respect to the other metagenomes studied. This result indicates that the pre-incubation of the sample before the culturing increases the possibility of isolation of poorly represented species. However, this procedure would contribute equally to the overgrowth of representatives of the most abundant phyla of the marine microbial community, such as *Pseudomonadota* (formerly *Proteobacteria*) and *Cyanobacteria* (Fernández-Gómez et al., 2013).

4.2. Biosynthetic potential

The biosynthetic gene profiling of *L. parthenopeia* Mr9^T and all species of the genus *Leeuwenhoekiella* showed that members of this genus have in common three BGCs: two terpenes and one T3-PKS. Of these, only one BGC terpene showed a low homology of 28% with a carotenoid molecule, while the rest of the BGCs do not exhibit similarity with any experimentally characterized gene encoding the synthesis of known chemical metabolites. Even though most of the gene clusters did not show a match with known molecules, the nucleotide translated sequences of those BGCs revealed that this genus, including *L. parthenopeia* Mr9^T, possesses multifunctional enzymes related to the synthesis of squalene, lycopene, fatty acids, flavonoids, and resveratrol; thus, they are probably synthesizing these types of molecules. In contrast to the rest of the species, only *L. parthenopeia* Mr9^T possesses NRPS and the significance of this type of BGC relies upon its well-known synthesis of antibiotic and antitumor compounds, thus the importance of *L. parthenopeia* Mr9^T with respect to the other species.

Additionally, the low homology of the core peptide with patented proteins confirms that *L. parthenopeia* Mr9^T does not synthesize the same proteins that those patented, suggesting its high metabolic capability to produce novel compounds with entirely different biotechnological applications.

Overall, the biosynthetic profile of species of *Leeuwenhoekiella* is very streamlined in comparison with the profile of the species of *Salinispora*, the abyssal marine bacteria producer of antitumor compounds, that harbor 176 distinct BGCs, of which only 24 have been linked to their products (Letzel et al., 2017). Despite this simplicity and given the novelty of these BGCs, they are probably synthesizing molecules with potentially different functionalities.

The similarity networking and evolutive clustering of the BGCs linked to enzyme phylogenies revealed that all species of *Leeuwenhoekiella* form four prominent gene cluster families (GCF) groups and according to this classification, species of this genus synthesize molecules such as flavine, lycopene, and squalene/phytoene. Moreover, the presence of uncharacterized conserved proteins and beta-ketoacyl synthase also suggests the synthesis of molecules of unknown function by species of this genus. Beyond these GCFs, also in evolutionary terms, *L. parthenopeia* Mr9^T remains again potentially more interesting than its peers, since the presence of NRPS, which feature domains AMP-binding enzyme, not shared with the other species, confirms that this bacterium may synthesize other compounds of biotechnological interest.

4.3. Biological activity: Lipidomic bioassay

Regarding the antiproliferative effect that exhibits the total lipid extract (TLE) from *L. parthenopeia* Mr9^T against prostate adenocarcinoma cell line, DU-145, and glioblastoma cell line, U-87 MG, it is important to take into account the nature of these cell lines, considered as moderate and high malignancy, respectively. In fact, the total lipid extract of *L. parthenopeia* proved to be more effective on prostate cells DU-145 than on glioblastoma cells U-87 MG. This result could be expected due to the intrinsic characteristic of the U-87 MG cell line, whose heterogeneous nature is associated with clonal plasticity that makes glioblastoma extremely resistant to current treatments (Osuka and Van Meir, 2017; Oliver et al., 2020). The antecedent is that glioblastoma is one of the most aggressive types of brain cancer, characterized by its rapid multiplication and invasion (Oraiopoulou et al., 2017), which could explain the low effect of the lipid at doses lower than 60 μg/mL at 24 and 48 h, as shown in the line graph of MTT results (**Figure 5A**). The primary cell line U-87 MG (ATCC HTB-14) has a double time of generation ranging from 24 to 29 h (Kato et al., 2011) and this faster multiplication was also observed under laboratory conditions, for which higher doses of extract and a long exposure time are required. The lipid extract was more effective on DU-145 causing a homogeneous decrease of cell viability from the first 24 h of treatment. In both tumor cell lines, DU-145 and U-87 MG, the cell viability gradually decreased in a time- and dose-dependent manner reaching a maximum efficiency at 72 h. The lower effect of the extract in immortalized human keratinocyte HaCaT, used as no tumor cell line, suggests a selective effect on tumor cells. Individual screenings with purified fractions will allow us to determine if the cell death is caused by a single compound or if it is a synergistic effect of individual compounds, then it will be necessary to identify the molecular mechanism that induces death. In any case, the lipid extract of the cell membrane of *L. parthenopeia* Mr9^T possesses an effective antiproliferative activity on tumor cell lines DU-145 and U-87 MG, not previously reported in rare bacteria.

4.4. Lipidome analysis

The chemical composition of the lipid extract from *L. parthenopeia* Mr9^T consisted of four putative compounds with known chemical structures, besides nine molecules without coincidences in the dictionary of natural products (DNP) identified after the dereplication. Among the known molecules are: sulfobacin A (Kamiyama et al., 1995), WB 3559A, WB-3559B (Yoshida et al., 1985), and topostin B-567 (Suzuki et al., 1990), all are derived from species of the class *Flavobacteriia*, except topostin B-567 isolated from a culture broth of *Flexibacter topostinus* belonging to the class *Cytophagia*. Regarding their biological activities, sulfobacin A is a sulfonolipid that was described as an antagonist of von Willebrand factor receptor (vWF), a blood glycoprotein involved in hemostasis and thus proposed as an antithrombosis agent (Kamiyama et al., 1995). Furthermore, sulfonolipid analogous (flavocristamide A and B) were found to have inhibitory activity against DNA polymerase α in eukaryotic cells (Kohayashi et al., 1995). Subsequently, the cytotoxic properties of this molecule were reported against four cancer cell lines with maximum activity against human mammary adenocarcinoma (Chaudhari et al., 2009). Concerning the compounds WB-3559 A and B, they were reported as potent fibrinolytic agents produced by *Flavobacterium* sp. No. 3559, which stimulates the euglobulin clot lysis time of rabbit plasma (Uchida et al., 1985; Yoshida et al., 1985). Docosenamide, also known as erucamide is a fatty acid amide belonging to a family of brain lipids that induce sleep (Cravatt et al., 1995). This molecule acts as a human metabolite, and it is more used in the neuroscience area. Finally, Topostin B-567 is an inhibitor of mammalian topoisomerase I and was isolated from a broth culture of *F. topostinus* B-572 (Ikegami et al., 1990; Suzuki et al., 1990; Brinkmann et al., 2022). Topoisomerase I is the molecular target for anticancer drugs since the activity of this enzyme is increased in tumor cells (Buzun et al., 2020). Topoisomerase suppressors constitute a family of antitumor agents with cytostatic effect, which mechanism of action is the inhibition of enzyme activity leading to the interruption of DNA strands and causing cell death (Bailly, 2000; Galluzzi et al., 2015). Considering the biological activities of the known compounds, we can hypothesize that the antiproliferative effect of DU-145 and U-87 MG observed *in vitro* in this study could be attributed to sulfobacin A and topostin B-567. The reduction of cell viability could be due to the action of both compounds since both inhibit DNA replication. However, the only one that does so selectively with tumor cells is topostin B-567, which being a specific suppressor of the enzyme topoisomerase I would explain the death of tumor cells and the low effect on epithelial tissue cells. Beyond the interesting

known molecules found in the lipid extract, the great potential of *L. parthenopeia* Mr9^T remains in the nine molecules without coincidences in the dictionary of natural products and still to be chemically elucidated.

Identifying the active fractions and their chemical structure will allow us to decipher the possible genes involved in the antiproliferative activity shown *in vitro*. Thus, based only on the *in silico* study, we cannot accurately demonstrate which of the BGCs present in the genome of *L. parthenopeia* Mr9^T can be responsible for this activity. Nonetheless, the simplicity of the biosynthetic profile of *L. parthenopeia* would make feasible the discrimination of the genes involved in the antitumor activity circumscribed to the four categories of BGC of its profile.

In conclusion, our study demonstrates that the new bacterium *L. parthenopeia* and the other species of the genus *Leeuwenhoekiella* contain biosynthetic genes with potential antiproliferative activities and encourage to carry out similar approaches to explore the marine rare biosphere as a promising source of bioactive compounds.

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

GG, RH, JM, FR, and PC: formal analysis. GG and PC: investigation. GG, RH, JM, FR, CS-P, AF, CZ, SG-F, NS-M, and PC: methodology. GG, RH, JM, SG-F, NS-M, and PC: software and visualization. GG: data curation and writing original draft preparation. RH, FR, CS-P, MV, ER, NS-M, AV, and PC: validation. RH, AV, and PC: writing – review and editing. CS-P, AV, and PC: funding acquisition, project administration, and resources. PC: conceptualization and supervision. All authors contributed to the manuscript revision, read, and approved the submitted version.

Funding

This study was achieved under the project Bluepharma grant 2018-PDR-00533 funded by Fondazione CON IL SUD (Italy) to PC and FEDER/Spanish Ministry of Science and Innovation-State Research Agency (grant PID2020-118136GB-I00) to AV and CS-P, and Junta de Andalucía, Spain (Grant P-20_01066) to AV.

Acknowledgments

We thank Prof. Aharon Oren for his help on the nomenclature of the new taxon, and special thanks to Jose M. Haro-Moreno and Prof. Francisco Rodríguez-Valera for their advice and contribution to relative abundance studies.

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.1090197/full#supplementary-material

References

Agrawal, S., Acharya, D., Adholeya, A., Barrow, C. J., and Deshmukh, S. K. (2017). Nonribosomal peptides from marine microbes and their antimicrobial and anticancer potential. *Front. Pharmacol.* 8:828. doi: 10.3389/fphar.2017.00828

Alonso, C., Warnecke, F., Amann, R., and Pernthaler, J. (2007). High local and global diversity of flavobacteria in marine plankton. *Environ. Microbiol.* 9, 1253–1266. doi: 10.1111/j.1462-2920.2007.01244.x

Angelini, R., Corral, P., Lopalco, P., Ventosa, A., and Corcelli, A. (2012). Novel ether lipid cardiolipins in archaeal membranes of extreme haloalkaliphiles. *Biochim. Biophys. Acta* 1818, 1365–1373. doi: 10.1016/j.bbamem.2012.02.014

Auch, A. F., von Jan, M., Klenk, H.-P., and Göker, M. (2010). Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. *Stand. Genomic Sci.* 2, 117–134. doi: 10.4056/sigs.531120

Bailly, C. (2000). Topoisomerase I poisons and suppressors as anticancer drugs. *CMC* 7, 39–58. doi: 10.2174/0929867003375489

Baker, B. J., Appler, K. E., and Gong, X. (2021). New microbial biodiversity in marine sediments. *Annu. Rev. Mar. Sci.* 13, 161–175. doi: 10.1146/annurev-marine-032020-014552

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. (2012). SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Computat. Biol.* 19, 455–477. doi: 10.1089/cmb.2012.0021

Barrow, G. I., and Feltham, R. K. A. (1993). "Characterization tests," in *Cowan and steel's manual for the identification of medical bacteria*, eds G. I. Barrow and R. K. A. Feltham (Cambridge: Cambridge University Press), 219–238. doi: 10.1017/CBO9780511527104.020

Baumas, C. M. J., Le Moigne, F. A. C., Garel, M., Bhairy, N., Guasco, S., Riou, V., et al. (2021). Mesopelagic microbial carbon production correlates with diversity across different marine particle fractions. *ISME J.* 15, 1695–1708. doi: 10.1038/s41396-020-00880-z

Belknap, K. C., Park, C. J., Barth, B. M., and Andam, C. P. (2020). Genome mining of biosynthetic and chemotherapeutic gene clusters in *Streptomyces* bacteria. *Sci. Rep.* 10:2003. doi: 10.1038/s41598-020-58904-9

Bernardet, J. (2015). "Flavobacteriales ord. nov," in *Bergey's manual of systematics of archaea and bacteria*, eds W. B. Whitman, F. Rainey, P. Kämpfer, M. Trujillo, J. Chun, P. DeVos, et al. (New York, NY: Wiley). doi: 10.1002/9781118960608.obm00033

Bligh, E. G., and Dyer, W. J. (1959). A rapid method of total lipid extraction and purification. *Can. J. Biochem. Physiol.* 37, 911–917. doi: 10.1139/o59-099

Blin, K., Shaw, S., Kloosterman, A. M., Charlop-Powers, Z., van Wezel, G. P., Medema, M. H., et al. (2021). antiSMASH 6.0: Improving cluster detection and comparison capabilities. *Nucleic Acids Res.* 49, W29–W35. doi: 10.1093/nar/gkab335

Boccellato, C., Kolbe, E., Peters, N., Juric, V., Fullstone, G., Verreault, M., et al. (2021). Marizomib sensitizes primary glioma cells to apoptosis induced by a latest-generation TRAIL receptor agonist. *Cell Death Dis.* 12:647. doi: 10.1038/s41419-021-03927-x

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

Bosch, H. F., and Rowland Taylor, R. (1973). Diurnal vertical migration of an estuarine cladoceran, *Podon polyphemoides*, in the Chesapeake Bay. *Mar. Biol.* 19, 172–181. doi: 10.1007/BF00353589

Brinkmann, S., Spohn, M. S., and Schäberle, T. F. (2022). Bioactive natural products from bacteroidetes. *Nat. Prod. Rep.* 39, 1045–1065. doi: 10.1039/D1NP00072A

Buzun, K., Bielawska, A., Bielawski, K., and Gornowicz, A. (2020). DNA topoisomerases as molecular targets for anticancer drugs. *J. Enzyme Inhib. Med. Chem.* 35, 1781–1799. doi: 10.1080/14756366.2020.1821676

Camarena-Gómez, M. T., Ruiz-González, C., Piiparinen, J., Lipsewers, T., Sobrino, C., Logares, R., et al. (2021). Bacterioplankton dynamics driven by interannual and spatial variation in diatom and dinoflagellate spring bloom communities in the Baltic Sea. *Limnol. Oceanogr.* 66, 255–271. doi: 10.1002/lno.11601

Chaudhari, P. N., Wani, K. S., Chaudhari, B. L., and Chincholkar, S. B. (2009). Characteristics of sulfobacin A from a soil isolate *Chryseobacterium gleum*. *Appl. Biochem. Biotechnol.* 158, 231–241. doi: 10.1007/s12010-008-8417-7

Chun, J., and Rainey, F. A. (2014). Integrating genomics into the taxonomy and systematics of the bacteria and archaea. *Int. J. Syst. Evol. Microbiol.* 64, 316–324. doi: 10.1099/ijs.0.054171-0

Chun, J., Oren, A., Ventosa, A., Christensen, H., Arahal, D. R., da Costa, M. S., et al. (2018). Proposed minimal standards for the use of genome data for the taxonomy of prokaryotes. *Int. J. Syst. Evol. Microbiol.* 68, 461–466. doi: 10.1099/ijsem.0.002516

Chwastek, G., Surma, M. A., Rizk, S., Grosser, D., Lavrynenko, O., Rucińska, M., et al. (2020). Principles of membrane adaptation revealed through environmentally induced bacterial lipidome remodeling. *Cell Rep.* 32:108165. doi: 10.1016/j.celrep.2020.108165

Corral, P., de la Haba, R. R., Infante-Domínguez, C., Sánchez-Porro, C., Amoozegar, M. A., Papke, R. T., et al. (2018). *Halorubrum chaoviator* Mancinelli et al. 2009 is a later, heterotypic synonym of *Halorubrum ezzemoulense* Kharroub et al. 2006. Emended description of *Halorubrum ezzemoulense* Kharroub et al. 2006. *Int. J. Syst. Evol. Microbiol.* 68, 3657–3665. doi: 10.1099/ijsem.0.003005

Cram, J. A., Chow, C.-E. T., Sachdeva, R., Needham, D. M., Parada, A. E., Steele, J. A., et al. (2015). Seasonal and interannual variability of the marine bacterioplankton community throughout the water column over ten years. *ISME J.* 9, 563–580. doi: 10.1038/ismej.2014.153

Cravatt, B. F., Prospero-Garcia, O., Siuzdak, G., Gilula, N. B., Henriksen, S. J., Boger, D. L., et al. (1995). Chemical characterization of a family of brain lipids that induce sleep. *Science* 268, 1506–1509. doi: 10.1126/science.7770779

de la Haba, R. R., Corral, P., Sánchez-Porro, C., Infante-Domínguez, C., Makkay, A. M., Amoozegar, M. A., et al. (2018). Genotypic and lipid analyses of strains from the archaeal genus *Halorubrum* reveal insights into their taxonomy, divergence, and population structure. *Front. Microbiol.* 9:512. doi: 10.3389/fmicb.2018.00512

de la Haba, R. R., López-Hermoso, C., Sánchez-Porro, C., Konstantinidis, K. T., and Ventosa, A. (2019). Comparative genomics and phylogenomic analysis of the genus *Salinivibrio*. *Front. Microbiol.* 10:2104. doi: 10.3389/fmicb.2019.02104

Díez-Vives, C., Nielsen, S., Sánchez, P., Palenzuela, O., Ferrera, I., Sebastián, M., et al. (2019). Delineation of ecologically distinct units of marine Bacteroidetes in the Northwestern Mediterranean Sea. *Mol. Ecol.* 28, 2846–2859. doi: 10.1111/mec.15068

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

Felsenstein, J. (1981). Evolutionary trees from DNA sequences: A maximum likelihood approach. *J. Mol. Evol.* 17, 368–376. doi: 10.1007/BF01734359

Felsenstein, J. (1985). Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* 39, 783–791. doi: 10.1111/j.1558-5646.1985.tb00420.x

Fenical, W., Jensen, P. R., Palladino, M. A., Lam, K. S., Lloyd, G. K., and Potts, B. C. (2009). Discovery and development of the anticancer agent salinosporamide A (NPI-0052). *Bioorg. Med. Chem.* 17, 2175–2180. doi: 10.1016/j.bmc.2008.10.075

Fernández-Gómez, B., Richter, M., Schüler, M., Pinhassi, J., Acinas, S. G., González, J. M., et al. (2013). Ecology of marine bacteroidetes: A comparative genomics approach. *ISME J.* 7, 1026–1037. doi: 10.1038/ismej.2012.169

Fitch, W. M. (1971). Toward defining the course of evolution: Minimum change for a specific tree topology. *Syst. Biol.* 20, 406–416. doi: 10.1093/sysbio/20.4.406

Fuhrman, J. A., Cram, J. A., and Needham, D. M. (2015). Marine microbial community dynamics and their ecological interpretation. *Nat. Rev. Microbiol.* 13, 133–146. doi: 10.1038/nrmicro3417

Galluzzi, L., Buqué, A., Kepp, O., Zitvogel, L., and Kroemer, G. (2015). Immunological effects of conventional chemotherapy and targeted anticancer agents. *Cancer Cell* 28, 690–714. doi: 10.1016/j.ccell.2015.10.012

García-Davis, S., Reyes, C. P., Lagunes, I., Padrón, J. M., Fraile-Nuez, E., Fernández, J. J., et al. (2021). Bioprospecting antiproliferative marine microbiota from submarine volcano Tagoro. *Front. Mar. Sci.* 8:687701. doi: 10.3389/fmars.2021.687701

Ghiglione, J. F., and Murray, A. E. (2012). Pronounced summer to winter differences and higher wintertime richness in coastal Antarctic marine bacterioplankton: Temporal variation in Southern Ocean coastal bacterioplankton. *Environ. Microbiol.* 14, 617–629. doi: 10.1111/j.1462-2920.2011.02601.x

Gilbert, J. A., Steele, J. A., Caporaso, J. G., Steinbrück, L., Reeder, J., Temperton, B., et al. (2012). Defining seasonal marine microbial community dynamics. *ISME J.* 6, 298–308. doi: 10.1038/ismej.2011.107

Giovannoni, S. J., and Vergin, K. L. (2012). Seasonality in ocean microbial communities. *Science* 335, 671–676. doi: 10.1126/science.1198078

Gómez-Pereira, P. R., Fuchs, B. M., Alonso, C., Oliver, M. J., van Beusekom, J. E. E., and Amann, R. (2010). Distinct flavobacterial communities in contrasting water masses of the North Atlantic Ocean. *ISME J.* 4, 472–487. doi: 10.1038/ismej.2009.142

Goris, J., Konstantinidis, K. T., Klappenbach, J. A., Coenye, T., Vandamme, P., and Tiedje, J. M. (2007). DNA–DNA hybridization values and their relationship to whole-genome sequence similarities. *Int. J. Syst. Evol. Microbiol.* 57, 81–91. doi: 10.1099/ijs.0.64483-0

Grzymski, J. J., Riesenfeld, C. S., Williams, T. J., Dussaq, A. M., Ducklow, H., Erickson, M., et al. (2012). A metagenomic assessment of winter and summer bacterioplankton from Antarctica Peninsula coastal surface waters. *ISME J.* 6, 1901–1915. doi: 10.1038/ismej.2012.31

Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013). QUAST: Quality assessment tool for genome assemblies. *Bioinformatics* 29, 1072–1075. doi: 10.1093/bioinformatics/btt086

Haro-Moreno, J. M., López-Pérez, M., and Rodriguez-Valera, F. (2021). Enhanced recovery of microbial genes and genomes from a marine water column using long-read metagenomics. *Front. Microbiol.* 12:708782. doi: 10.3389/fmicb.2021.708782

Haro-Moreno, J. M., López-Pérez, M., de la Torre, J. R., Picazo, A., Camacho, A., and Rodriguez-Valera, F. (2018). Fine metagenomic profile of the Mediterranean stratified and mixed water columns revealed by assembly and recruitment. *Microbiome* 6:128. doi: 10.1186/s40168-018-0513-5

Haro-Moreno, J. M., Rodriguez-Valera, F., López-García, P., Moreira, D., and Martin-Cuadrado, A.-B. (2017). New insights into marine group III Euryarchaeota, from dark to light. *ISME J.* 11, 1102–1117. doi: 10.1038/ismej.2016.188

Ikegami, Y., Takeuchi, N., Hanada, M., Hasegawa, Y., Ishii, K., Andoh, T., et al. (1990). Topostin, a novel inhibitor of mammalian DNA topoisomerase I from *Flexibacter topostinus* sp. nov. II. Purification and some properties of topostin. *J. Antibiot.* 43, 158–162. doi: 10.7164/antibiotics.43.158

Infante-Domínguez, C., de la Haba, R. R., Corral, P., Sanchez-Porro, C., Arahal, D. R., and Ventosa, A. (2020). Genome-based analyses reveal a synonymy among *Halorubrum distributum* Zvyagintseva and Tarasov 1989; Oren and Ventosa 1996, *Halorubrum terrestre* Ventosa et al. 2004, *Halorubrum arcis* Xu et al. 2007 and *Halorubrum litoreum* Cui et al. 2007. Emended description of *Halorubrum distributum* Zvyagintseva and Tarasov 1989; Oren and Ventosa 1996. *Int. J. Syst. Evol. Microbiol.* 70, 1698–1705. doi: 10.1099/ijsem.0.003956

Jones, D. T., Taylor, W. R., and Thornton, J. M. (1992). The rapid generation of mutation data matrices from protein sequences. *Bioinformatics* 8, 275–282. doi: 10.1093/bioinformatics/8.3.275

Jukes, T. H., and Cantor, C. R. (1969). "Evolution of protein molecules," in *Mammalian protein metabolism*, ed. H. N. Munro (Amsterdam: Elsevier), 21–132. doi: 10.1016/B978-1-4832-3211-9.50009-7

Kamiyama, T., Umino, T., Satoh, T., Sawairi, S., Shirane, M., Onshima, S., et al. (1995). Sulfobacins A and B, Novel von Willebrand factor receptor antagonists. I. Production, isolation, characterization and biological activities. *J. Antibiot.* 48, 924–928. doi: 10.7164/antibiotics.48.924

Kato, T. A., Tsuda, A., Uesaka, M., Fujimori, A., Kamada, T., Tsujii, H., et al. (2011). In vitro characterization of cells derived from chordoma cell line U-CH1 following treatment with X-rays, heavy ions and chemotherapeutic drugs. *Radiat. Oncol.* 6:116. doi: 10.1186/1748-717X-6-116

Kautsar, S. A., Blin, K., Shaw, S., Navarro-Muñoz, J. C., Terlouw, B. R., van der Hooft, J. J. J., et al. (2019). MIBiG 2.0: A repository for biosynthetic gene clusters of known function. *Nucleic Acids Res.* 48, D454–D458. doi: 10.1093/nar/gkz882

Kelly, L. W., Nelson, C. E., Haas, A. F., Naliboff, D. S., Calhoun, S., Carlson, C. A., et al. (2019). Diel population and functional synchrony of microbial communities on coral reefs. *Nat. Commun.* 10:1691. doi: 10.1038/s41467-019-09419-z

Kirchman, D. L. (2002). The ecology of *Cytophaga*–flavobacteria in aquatic environments. *FEMS Microbiol. Ecol.* 39, 91–100. doi: 10.1111/j.1574-6941.2002.tb00910.x

Kirchman, D. L., Yu, L., and Cottrell, M. T. (2003). Diversity and abundance of uncultured *Cytophaga* -like bacteria in the Delaware Estuary. *Appl. Environ. Microbiol.* 69, 6587–6596. doi: 10.1128/AEM.69.11.6587-6596.2003

Kohayashi, J., Mikami, S., Shigemori, H., Takao, T., Shimonishi, Y., Izuta, S., et al. (1995). Flavocristamides A and B, new DNA polymerase α inhibitors from a marine bacterium *Flavobacterium* sp. *Tetrahedron* 51, 10487–10490. doi: 10.1016/0040-4020(95)00631-H

Konstantinidis, K. T., and Tiedje, J. M. (2005). Genomic insights that advance the species definition for prokaryotes. *Proc. Natl. Acad. Sci. U.S.A.* 102, 2567–2572. doi: 10.1073/pnas.0409727102

Krzywinski, M., Schein, J., Birol, İ, Connors, J., Gascoyne, R., Horsman, D., et al. (2009). Circos: An information aesthetic for comparative genomics. *Genome Res.* 19, 1639–1645. doi: 10.1101/gr.092759.109

Letunic, I., and Bork, P. (2021). Interactive Tree Of Life (iTOL) v5: An online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* 49, W293–W296. doi: 10.1093/nar/gkab301

Letzel, A.-C., Li, J., Amos, G. C. A., Millán-Aguiñaga, N., Ginigini, J., Abdelmohsen, U. R., et al. (2017). Genomic insights into specialized metabolism in the marine actinomycete *Salinispora*: Genomic insights into specialized metabolism. *Environ. Microbiol.* 19, 3660–3673. doi: 10.1111/1462-2920.13867

Liu, C.-M., Wong, T., Wu, E., Luo, R., Yiu, S.-M., Li, Y., et al. (2012). SOAP3: Ultra-fast GPU-based parallel alignment tool for short reads. *Bioinformatics* 28, 878–879. doi: 10.1093/bioinformatics/bts061

Liu, Q., Li, J., Wei, B., Zhang, X., Zhang, L., Zhang, Y., et al. (2016). *Leeuwenhoekiella nanhaiensis* sp. nov., isolated from deep-sea water. *Int. J. Syst. Evol. Microbiol.* 66, 1352–1357. doi: 10.1099/ijsem.0.000883

López-Pérez, M., Haro-Moreno, J. M., Gonzalez-Serrano, R., Parras-Moltó, M., and Rodriguez-Valera, F. (2017). Genome diversity of marine phages recovered from Mediterranean metagenomes: Size matters. *PLoS Genet.* 13:e1007018. doi: 10.1371/journal.pgen.1007018

Maeda, H., and Khatami, M. (2018). Analyses of repeated failures in cancer therapy for solid tumors: Poor tumor-selective drug delivery, low therapeutic efficacy and unsustainable costs. *Clin. Transl. Med.* 7:11. doi: 10.1186/s40169-018-0185-6

Martín, J., Crespo, G., González-Menéndez, V., Pérez-Moreno, G., Sánchez-Carrasco, P., Pérez-Victoria, I., et al. (2014). MDN-0104, an antiplasmodial betaine lipid from *Heterospora chenopodii*. *J. Nat. Prod.* 77, 2118–2123. doi: 10.1021/np500577v

Meier-Kolthoff, J. P., Auch, A. F., Klenk, H.-P., and Göker, M. (2013). Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics* 14:60. doi: 10.1186/1471-2105-14-60

Mena, C., Reglero, P., Balbín, R., Martín, M., Santiago, R., and Sintes, E. (2020). Seasonal niche partitioning of surface temperate open ocean prokaryotic communities. *Front. Microbiol.* 11:1749. doi: 10.3389/fmicb.2020.01749

Mestre, M., Borrull, E., Sala, M. M., and Gasol, J. M. (2017). Patterns of bacterial diversity in the marine planktonic particulate matter continuum. *ISME J.* 11, 999–1010. doi: 10.1038/ismej.2016.166

Mestre, M., Höfer, J., Sala, M. M., and Gasol, J. M. (2020). Seasonal variation of bacterial diversity along the marine particulate matter continuum. *Front. Microbiol.* 11:1590. doi: 10.3389/fmicb.2020.01590

Mestre, M., Ruiz-González, C., Logares, R., Duarte, C. M., Gasol, J. M., and Sala, M. M. (2018). Sinking particles promote vertical connectivity in the ocean microbiome. *Proc. Natl. Acad. Sci. U.S.A.* 115, E6799–E6807. doi: 10.1073/pnas.1802470115

MIDI (2012). *Sherlock microbial identification system operating manual*. Newark, DE: MIDI.

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

Navarro-Muñoz, J. C., Selem-Mojica, N., Mullowney, M. W., Kautsar, S. A., Tryon, J. H., Parkinson, E. I., et al. (2020). A computational framework to explore large-scale biosynthetic diversity. *Nat. Chem. Biol.* 16, 60–68. doi: 10.1038/s41589-019-0400-9

Nedashkovskaya, O. I., Vancanneyt, M., Dawyndt, P., Engelbeen, K., Vandemeulebroecke, K., Cleenwerck, I., et al. (2005). Reclassification of *[Cytophaga] marinoflava* Reichenbach 1989 as *Leeuwenhoekiella marinoflava* gen. nov., comb. nov. and description of *Leeuwenhoekiella aequorea* sp. nov. *Int. J. Syst. Evol. Microbiol.* 55, 1033–1038. doi: 10.1099/ijs.0.63410-0

Oliver, L., Lalier, L., Salaud, C., Heymann, D., Cartron, P. F., and Vallette, F. M. (2020). Drug resistance in glioblastoma: Are persisters the key to therapy? *Cancer Drug Resist.* 3, 287–301. doi: 10.20517/cdr.2020.29

Omand, M. M., Steinberg, D. K., and Stamieszkin, K. (2021). Cloud shadows drive vertical migrations of deep-dwelling marine life. *Proc. Natl. Acad. Sci. U.S.A.* 118:e2022977118. doi: 10.1073/pnas.2022977118

Oraiopoulou, M.-E., Tzamali, E., Tzedakis, G., Vakis, A., Papamatheakis, J., and Sakkalis, V. (2017). in vitro/in silico study on the role of doubling time heterogeneity among primary glioblastoma cell lines. *Biomed. Res. Int.* 2017, 1–12. doi: 10.1155/2017/8569328

Oren, A., and Garrity, G. M. (2021). Valid publication of the names of forty-two phyla of prokaryotes. *Int. J. Syst. Evol. Microbiol.* 71:005056. doi: 10.1099/ijsem.0.005056

Osuka, S., and Van Meir, E. G. (2017). Overcoming therapeutic resistance in glioblastoma: The way forward. *J. Clin. Invest.* 127, 415–426. doi: 10.1172/JCI89587

Paoli, L., Ruscheweyh, H.-J., Forneris, C. C., Hubrich, F., Kautsar, S., Bhushan, A., et al. (2022). Biosynthetic potential of the global ocean microbiome. *Nature* 607, 111–118. doi: 10.1038/s41586-022-04862-3

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

Parte, A. C., Sardà Carbasse, J., Meier-Kolthoff, J. P., Reimer, L. C., and Göker, M. (2020). List of prokaryotic names with standing in nomenclature (LPSN) moves to the DSMZ. *Int. J. Syst. Evol. Microbiol.* 70, 5607–5612. doi: 10.1099/ijsem.0.004332

Pascoal, F., Magalhães, C., and Costa, R. (2020). The link between the ecology of the prokaryotic rare biosphere and its biotechnological potential. *Front. Microbiol.* 11:231. doi: 10.3389/fmicb.2020.00231

Pedrós-Alió, C. (2007). Dipping into the rare biosphere. *Science* 315, 192–193. doi: 10.1126/science.1135933

Pedrós-Alió, C. (2012). The rare bacterial biosphere. *Annu. Rev. Mar. Sci.* 4, 449–466. doi: 10.1146/annurev-marine-120710-100948

Pedrós-Alió, C. (2013). "Rare biosphere," in *Encyclopedia of biodiversity*, ed. S. Levi (Amsterdam: Elsevier), 345–352. doi: 10.1016/B978-0-12-384719-5.00407-X

Pesant, S., Not, F., Picheral, M., Kandels-Lewis, S., Le Bescot, N., Gorsky, G., et al. (2015). Open science resources for the discovery and analysis of Tara Oceans data. *Sci. Data* 2:150023. doi: 10.1038/sdata.2015.23

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

Reichenbach, H. (1992). Validation of the publication of new names and new combinations previously effectively published outside the IJSB: List no. 41. *Int. J. Syst. Bacteriol.* 42, 327–328. doi: 10.1099/00207713-42-2-327

Richter, M., and Rosselló-Móra, R. (2009). Shifting the genomic gold standard for the prokaryotic species definition. *Proc. Natl. Acad. Sci. U.S.A.* 106, 19126–19131. doi: 10.1073/pnas.0906412106

Rooney-Varga, J. N., Giewat, M. W., Savin, M. C., Sood, S., LeGresley, M., and Martin, J. L. (2005). Links between phytoplankton and bacterial community dynamics in a coastal marine environment. *Microb. Ecol.* 49, 163–175. doi: 10.1007/s00248-003-1057-0

Roth, P., Mason, W. P., Richardson, P. G., and Weller, M. (2020). Proteasome inhibition for the treatment of glioblastoma. *Exp. Opin. Invest. Drugs* 29, 1133–1141. doi: 10.1080/13543784.2020.1803827

Saitou, N., and Nei, M. (1987). The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4, 406–425. doi: 10.1093/oxfordjournals.molbev.a040454

Sanz-Sáez, I., Salazar, G., Sánchez, P., Lara, E., Royo-Llonch, M., Sà, E. L., et al. (2020). Diversity and distribution of marine heterotrophic bacteria from a large culture collection. *BMC Microbiol.* 20:207. doi: 10.1186/s12866-020-01884-7

Sasser, M. (1990). *Bacterial identification by gas chromatographic analysis of fatty acid methyl esters (GC-FAME)*. Newark, DE: MIDI.

Scherlach, K., and Hertweck, C. (2021). Mining and unearthing hidden biosynthetic potential. *Nat. Commun.* 12:3864. doi: 10.1038/s41467-021-24133-5

Shade, A., Jones, S. E., Caporaso, J. G., Handelsman, J., Knight, R., Fierer, N., et al. (2014). Conditionally rare taxa disproportionately contribute to temporal changes in microbial diversity. *mbio* 5:e01371–14. doi: 10.1128/mBio.01371-14

Shimodaira, H., and Hasegawa, M. (1999). Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol. Biol. Evol.* 16, 1114–1116. doi: 10.1093/oxfordjournals.molbev.a026201

Si, O.-J., Kim, S.-J., Jung, M.-Y., Choi, S.-B., Kim, J.-G., Kim, S.-G., et al. (2015). *Leeuwenhoekiella polynyae* sp. nov., isolated from a polynya in Western Antarctica. *Int. J. Syst. Evol. Microbiol.* 65, 1694–1699. doi: 10.1099/ijs.0.000160

Siegel, R. L., Miller, K. D., Fuchs, H. E., and Jemal, A. (2021). Cancer statistics, 2021. *CA Cancer J. Clin.* 71, 7–33. doi: 10.3322/caac.21654

Sogin, M. L., Morrison, H. G., Huber, J. A., Welch, D. M., Huse, S. M., Neal, P. R., et al. (2006). Microbial diversity in the deep sea and the underexplored "rare biosphere.". *Proc. Natl. Acad. Sci. U.S.A.* 103, 12115–12120. doi: 10.1073/pnas.0605127103

Stackebrandt, E., and Goebel, B. M. (1994). Taxonomic note: A place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. *Int. J. Syst. Evol. Microbiol.* 44, 846–849. doi: 10.1099/00207713-44-4-846

Sunagawa, S., Coelho, L. P., Chaffron, S., Kultima, J. R., Labadie, K., Salazar, G., et al. (2015). Structure and function of the global ocean microbiome. *Science* 348:1261359. doi: 10.1126/science.1261359

Suzuki, K., Yamaguchi, H., Miyazaki, S., Nagai, K., Watanabe, S.-I., Saito, T., et al. (1990). Topostin, a novel inhibitor of mammalian DNA topoisomerase I from *Flexibacter topostinus* sp. nov. I. Taxonomy, and fermentation of producing strain. *J. Antibiot.* 43, 154–157. doi: 10.7164/antibiotics.43.154

Tahon, G., Lebbe, L., De Troch, M., Sabbe, K., and Willems, A. (2020). *Leeuwenhoekiella aestuarii* sp. nov., isolated from salt-water sediment and first insights in the genomes of *Leeuwenhoekiella* species. *Int. J. Syst. Evol. Microbiol.* 70, 1706–1719. doi: 10.1099/ijsem.0.003959

Tatusova, T., DiCuccio, M., Badretdin, A., Chetvernin, V., Nawrocki, E. P., Zaslavsky, L., et al. (2016). NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res.* 44, 6614–6624. doi: 10.1093/nar/gkw569

Tavaré, S., and Miura, R. M. (1986). "Some probabilistic and statistical problems in the analysis of DNA sequences," in *Lectures on mathematics in the life sciences*, ed. R. M. Miura (Providence, RI: American Mathematical Society), 57–86.

Teeling, H., Fuchs, B. M., Bennke, C. M., Krüger, K., Chafee, M., Kappelmann, L., et al. (2016). Recurring patterns in bacterioplankton dynamics during coastal spring algae blooms. *Elife* 5:e11888. doi: 10.7554/eLife.11888

Uchida, I., Yoshida, K., Kawai, Y., Takase, S., Itoh, Y., Tanaka, H., et al. (1985). Structure and synthesis of WB-3559 A, B, C and D, new fibrinolytic agents isolated from *Flavobacterium* sp. *Chem. Pharm. Bull.* 33, 424–427. doi: 10.1248/cpb.33.424

van Heel, A. J., de Jong, A., Song, C., Viel, J. H., Kok, J., and Kuipers, O. P. (2018). BAGEL4: A user-friendly web server to thoroughly mine RiPPs and bacteriocins. *Nucleic Acids Res.* 46, W278–W281. doi: 10.1093/nar/gky383

Wenley, J., Currie, K., Lockwood, S., Thomson, B., Baltar, F., and Morales, S. E. (2021). Seasonal prokaryotic community linkages between surface and deep ocean water. *Front. Mar. Sci.* 8:659641. doi: 10.3389/fmars.2021.659641

Westram, R., Bader, K., Prüsse, E., Kumar, Y., Meier, H., Glöckner, F. O., et al. (2011). "ARB: A software environment for sequence data," in *Handbook of molecular microbial ecology I*, ed. F. J. de Bruijn (Hoboken, NJ: John Wiley & Sons, Inc.), 399–406. doi: 10.1002/9781118010518.ch46

Xue, C., Zhang, H., Lin, H., Sun, Y., Luo, D., Huang, Y., et al. (2020). Ancestral niche separation and evolutionary rate differentiation between sister marine flavobacteria lineages. *Environ. Microbiol.* 22, 3234–3247. doi: 10.1111/1462-2920.15065

Yoon, S.-H., Ha, S.-M., Kwon, S., Lim, J., Kim, Y., Seo, H., et al. (2017). 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

Yoshida, K., Iwami, M., Umehara, Y., Nishikawa, M., Uchida, I., Kohsaka, M., et al. (1985). Studies on WB-3559 A,B,C and D, new potent fibrinolytic agents. I. Discovery, identification, isolation and characterization. *J. Antibiot.* 38, 1469–1475. doi: 10.7164/antibiotics.38.1469

Zhou, Q., Hotta, K., Deng, Y., Yuan, R., Quan, S., and Chen, X. (2021). Advances in biosynthesis of natural products from marine microorganisms. *Microorganisms* 9:2551. doi: 10.3390/microorganisms9122551

Zuchegna, C., Di Zazzo, E., Moncharmont, B., and Messina, S. (2020). Dual-specificity phosphatase (DUSP6) in human glioblastoma: Epithelial-to-mesenchymal transition (EMT) involvement. *BMC Res. Notes* 13:374. doi: 10.1186/s13104-020-05214-y

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 13 April 2023
DOI 10.3389/fmicb.2023.1164632

OPEN ACCESS

EDITED BY
Utkarsh Sood,
University of Delhi,
India

REVIEWED BY
Paulina Corral,
Sevilla University,
Spain
Debarati Paul,
Amity University,
India

*CORRESPONDENCE
Gilman Kit-Hang Siu
✉ gilman.siu@polyu.edu.hk

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

RECEIVED 13 February 2023
ACCEPTED 14 March 2023
PUBLISHED 13 April 2023

CITATION
Lee AW-T, Chan CT-M, Wong LL-Y, Yip C-Y, Lui W-T, Cheng K-C, Leung JS-L, Lee L-K, Wong IT-F, Ng TT-L, Lao H-Y and Siu GK-H (2023) Identification of microbial community in the urban environment: The concordance between conventional culture and nanopore 16S rRNA sequencing.
Front. Microbiol. 14:1164632.
doi: 10.3389/fmicb.2023.1164632

Identification of microbial community in the urban environment: The concordance between conventional culture and nanopore 16S rRNA sequencing

Annie Wing-Tung Lee, Chloe Toi-Mei Chan, Lily Lok-Yee Wong, Cheuk-Yi Yip, Wing-Tung Lui, Kai-Chun Cheng, Jake Siu-Lun Leung, Lam-Kwong Lee, Ivan Tak-Fai Wong, Timothy Ting-Leung Ng, Hiu-Yin Lao and Gilman Kit-Hang Siu*

Department of Health Technology and Informatics, The Hong Kong Polytechnic University, Hong Kong, Hong Kong SAR, China

Introduction: Microbes in the built environment have been implicated as a source of infectious diseases. Bacterial culture is the standard method for assessing the risk of exposure to pathogens in urban environments, but this method only accounts for <1% of the diversity of bacteria. Recently, full-length 16S rRNA gene analysis using nanopore sequencing has been applied for microbial evaluations, resulting in a rise in the development of long-read taxonomic tools for species-level classification. Regarding their comparative performance, there is, however, a lack of information.

Methods: Here, we aim to analyze the concordance of the microbial community in the urban environment inferred by multiple taxonomic classifiers, including ARGpore2, Emu, Kraken2/Bracken and NanoCLUST, using our 16S-nanopore dataset generated by MegaBLAST, as well as assess their abilities to identify culturable species based on the conventional culture results.

Results: According to our results, NanoCLUST was preferred for 16S microbial profiling because it had a high concordance of dominant species and a similar microbial profile to MegaBLAST, whereas Kraken2/Bracken, which had similar clustering results as NanoCLUST, was also desirable. Second, for culturable species identification, Emu with the highest accuracy (81.2%) and F1 score (29%) for the detection of culturable species was suggested.

Discussion: In addition to generating datasets in complex communities for future benchmarking studies, our comprehensive evaluation of the taxonomic classifiers offers recommendations for ongoing microbial community research, particularly for complex communities using nanopore 16S rRNA sequencing.

KEYWORDS

urban microbiome, nanopore sequencing, 16S rRNA gene sequencing, taxonomic classifiers, conventional culture

Introduction

Built environment refers to man-made or modified structures that provide people with living, working, and recreational spaces. With the development of urbanization, people spend most of their time indoors. Microorganisms, especially the *Enterococcus faecium*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Acinetobacter baumannii, Pseudomonas aeruginosa*, and *Enterobacter* sp. (ESKAPE), in the built environment have become a common source of infection (Rice, 2008; Neiderud, 2015; Wang et al., 2022). Microbial profiling of high-touch surfaces is an effective way for the surveillance of infectious agents in public indoor environments (Danko et al., 2021).

While the conventional culture-based method effectively identifies microorganisms with anti-microbial resistance, it only reflects <1% of the bacterial diversity (Pedron et al., 2020; Sala-Comorera et al., 2020; Marshall et al., 2021). Next-generation sequencing (NGS) directly traces DNA from the urban samples and characterizes microbial communities in a higher resolution (Kerkhof, 2021). However, second-generation sequencing has a low phylogenetic resolution and could not accurately classify microorganisms below the genus level because of the limitation of short read length (Kerkhof, 2021). In recent years, nanopore sequencing, developed by Oxford Nanopore Technologies (ONTs), has become a dominant microbiome analysis platform, which could generate reads with lengths up to 2×10^6 bases (Rodriguez-Perez et al., 2021). This enables bacterial identification to the species level with a 92–94% accuracy in mock communities (Benitez-Paez et al., 2016; Lao et al., 2022). Different bioinformatic pipelines have been established for the analysis of nanopore sequencing data.

To date, MegaBLAST is the gold-standard alignment tool for classifying nucleotide sequences, but it can be computationally intensive (Ye et al., 2019). The classification speed of the 5.7 million-read dataset was estimated to be 4 h using MegaBLAST, while a variety of alignment-free classifiers completed the task in significantly less time (<10 min) (Ye et al., 2019). Kraken2 coupled with Bracken is a recent alignment-free classifier that provides a fast taxonomic classification of nanopore sequence data (1 min) (Wood et al., 2019; Lu and Salzberg, 2020). It maintains high accuracy and sensitivity in metagenomics analysis by using the *k*-mer-based approach. After that, three long-read classifiers were developed. NanoCLUST, which is the first tailor-made method for full-length 16S amplicon sequencing of the nanopore, was published in 2021 (Rodriguez-Perez et al., 2021). It is based on Uniform Manifold Approximation and Projection (UMAP), followed by the construction of a polished read and subsequent BLAST classification. Nevertheless, it might provide false positives because of the use of consensus sequences (Curry et al., 2022). Recently, Emu has been published to profile microbial communities using full-length 16S data (Curry et al., 2022). It claimed that the implementation of an expectation–maximization (EM)-algorithm could identify bacterial abundance more accurately than all the existing methods. Meanwhile, ARGpore2 was established to identify antibiotic resistance genes (ARGs) and their host populations from nanopore reads (Cheng et al., 2022). It combined the results of Centrifuge and MetaPhlan2 (Metagenomic Phylogenetic Analysis) markergene database to annotate the taxonomy of nanopore reads (Segata et al., 2012; Kim et al., 2016).

While most researchers analyzed several mock communities of simple microbiomes to evaluate the accuracy, sensitivity, and specificity of the established classifiers (Deshpande et al., 2019; Winand et al., 2019; Leidenfrost et al., 2020; Urban et al., 2021), an actual and complex microbial community from the urban environment provides a better assessment of the classifiers. In this study, a total of 46 environmental samples were collected from the mass transit system of a city populated with 7.4 million people. The species-level microbial communities were determined by bacterial culture and nanopore 16S rRNA sequencing. Based on our nanopore dataset produced by MegaBLAST and conventional culture results, we aim to assess the clustering and concordance of the microbial community in the urban environment inferred by various taxonomic classifiers, namely ARGpore2, Kraken2/Bracken, NanoCLUST, and Emu. We demonstrated that the sequencing-based method uncovered misclassification and the absence of species, especially the pathogenic species, in the culture-based method. This study also offered an unbiased and comprehensive assessment of taxonomic classifiers for the analysis of full-length 16S reads in complex microbial communities, allowing researchers to identify the most reliable bioinformatic tools for the study of the urban microbiome.

Materials and methods

Sample collection

A total of 46 urban samples were collected from 18 stations of the mass transit system of Hong Kong in the period of July to October (summer season). Enviroscreen Sponges (Technical Service Consultants) were used to collect bacteria from the high-touch environmental surfaces, including handrails, ticket kiosks, automated teller machines (ATMs), elevators, and escalators according to the manufacturer's protocol (Technical Service Consultants Ltd., 2014). The detail of the collection sites is shown in Supplementary Table S1. The sponges were filled with 50 mL of Milli-Q Water with 0.1% peptone water and 0.1% Tween 80 in a sealed bag and squeezed for 5 min. The extracts were collected and centrifuged for 15 min at 4,600 × *g*. The bacterial pellet was resuspended in 50 mL of peptone water and divided into two parts, culture (1 mL) and DNA extraction procedures (49 mL).

Species identification by culture-based methods

The resuspended bacterial pellet was cultured using the Brain Heart Infusion Agar (BHI Agar, BD BBL™) and CHROMagar Extended Spectrum β-lactamases (ESBL, Kanto Chemical), carbapenem-resistant Enterobacteriaceae (CRE, Kanto Chemical), multi-drug-resistant Acinetobacter (MDRA, Kanto Chemical), methicillin-resistant *Staphylococcus Aureus* (MRSA, Kanto Chemical), and vancomycin-resistant Enterococcus (VRE, Kanto Chemical). A total of 100 µL of the resuspended bacterial pellet was poured over the agar plates and spread evenly using the plate spreader. Six culture plates were placed in the incubator for 24 h at 37°C for each sample. Isolates of different morphology were picked and identified at the species level (score, 2.00) by the IVD MALDI Biotyper Microflex® with the database version BD-6763 (Bruker Daltonics, Bremen, Germany).

DNA extraction

DNA was extracted from the bacterial suspension using the QIAamp BiOstic Bacteremia DNA Kit (Qiagen) according to the manufacturer's protocol. In brief, DNA was released after bacterial cell

lysis *via* mechanical homogenization. DNA was then purified by column-based methods after inhibitor removal procedures and collected with an elution buffer. The DNA content was then quantified using the Qubit 2.0 fluorometer (Thermo Fisher Scientific) with the double-stranded DNA (dsDNA) HS assay kit (Thermo Fisher Scientific).

Taxonomic assignment by nanopore sequencing

Libraries were prepared using the 16S barcoding kit 1–24 (SQK-16S024) from ONT according to the manufacturer's protocol and quantified using the Qubit as described earlier. A total of 24 barcoded libraries were then pooled with equal concentrations. After adapter ligation, sequencing was performed using the Flow Cell (R9.4.1) with the GridION (Oxford Nanopore Technologies) for >24 h. Raw reads from the nanopore sequencing were trimmed by NanoFilt (De Coster et al., 2018). Reads with an average read quality score of <8.0 and of length below 1,000 base pairs were discarded. The data details are listed in Supplementary Table S2. Urban samples with a total read of less than 10,000 were eliminated. The filtered clean reads were then classified by MegaBLAST (Nucleotide BLAST 2.13.0+) (Morgulis et al., 2008; Camacho et al., 2009), ARGpore2 (Cheng et al., 2022), Kraken2/Bracken (Kraken version 2.1.2 and Bracken 2.5) (Wood et al., 2019; Lu and Salzberg, 2020), Emu (emu v3.2.0) (Wang et al., 2022), and NanoCLUST (Rodriguez-Perez et al., 2021) using default parameters and database. In the MegaBLAST classification, the output of the first good hit found in the database with an E-value of less than 0.001 was chosen for each read (Shah et al., 2019). Species of relative abundance of <0.1% in the four classifiers were discarded in four classifiers (ARGpore2, Emu, Kraken2/Bracken, and NanoCLUST). Due to limited computer RAM, ARGpore2, Emu, and NanoCLUST were unable to analyze some of the samples. ARGpore2, Emu, and NanoCLUST were able to analyze 43, 44, and 22 of the 46 samples, respectively.

Filtered 16S sequencing data processing

Microbial profiles of eight groups, namely ARGpore2, Emu, Kraken2/Bracken, NanoCLUST, MegaBLAST (46 sites), MegaBLAST_ARGpore2 (43 sites), MegaBLAST_Emu (44 sites), and MegaBLAST_NanoCLUST (22 sites), were generated by the average abundance of the overlapped species. Species present in only one classifier were eliminated. The mean abundance of MegaBLAST and the four classifiers was calculated by (no. of reads from each species/ total no. of reads × 100%) and (mean relative abundance/ no. of sites × 100%,) respectively. In parallel, 20 urban samples (Site4, Site5, Site7, Site11, Site12, Site18, Site20, Site21, Site23, Site24, Site26, Site29, Site31, Site32, Site35, Site40, Site41, Site43, Site45, and Site46) that could be analyzed by all four classifiers were selected for downstream clustering analysis.

Principal component analysis and hierarchical clustering

For principal component analysis (PCA), the R function pcromp was used to generate the graph. Utilizing the 20 urban samples indicated earlier, a three-dimensional figure with the top three principal components (PC1, PC2, and PC3) was created by R using the microbial profiles of the five groups (ARGpore2, Emu, Kraken2/Bracken, NanoCLUST, and MegaBLAST). Afterward, hierarchical clustering was performed using the complete linkage method. The resulting species of dendrograms were separated into the earlier eight groups.

Statistical analysis and data availability

The microbial communities were assessed using phyloseq coupled with R-based computational tools in R-studio to generate the graphs (McMurdie and Holmes, 2013; McMurdie and Holmes, 2014; McMurdie and Holmes, 2015; Callahan et al., 2016). Performances of the classifiers were calculated as Accuracy = (TP + TN)/(TP + FP + FN + TN), Precision = TP/(TP + FP), Sensitivity (Recall) = TP/(TP + FN), Specificity = TN/(TN + FP), and F1 score = 2 × (Recall × Precision) / (Recall + Precision); TP: true positive (species was found in both outcomes), TN: true negative (species was absent in both outcomes), FP: false positive (species was found in one of the outcomes), and FN: false negative (species was absent in one of the outcomes). The values were based on the species' total concordance between MegaBLAST and the four classifiers, as well as the species' total concordance between nanopore sequencing using various classifiers and conventional cultures. Sequence data were archived in the National Center for Biotechnology Information (NCBI) Short Read Archive (SRA) (PRJNA875407).

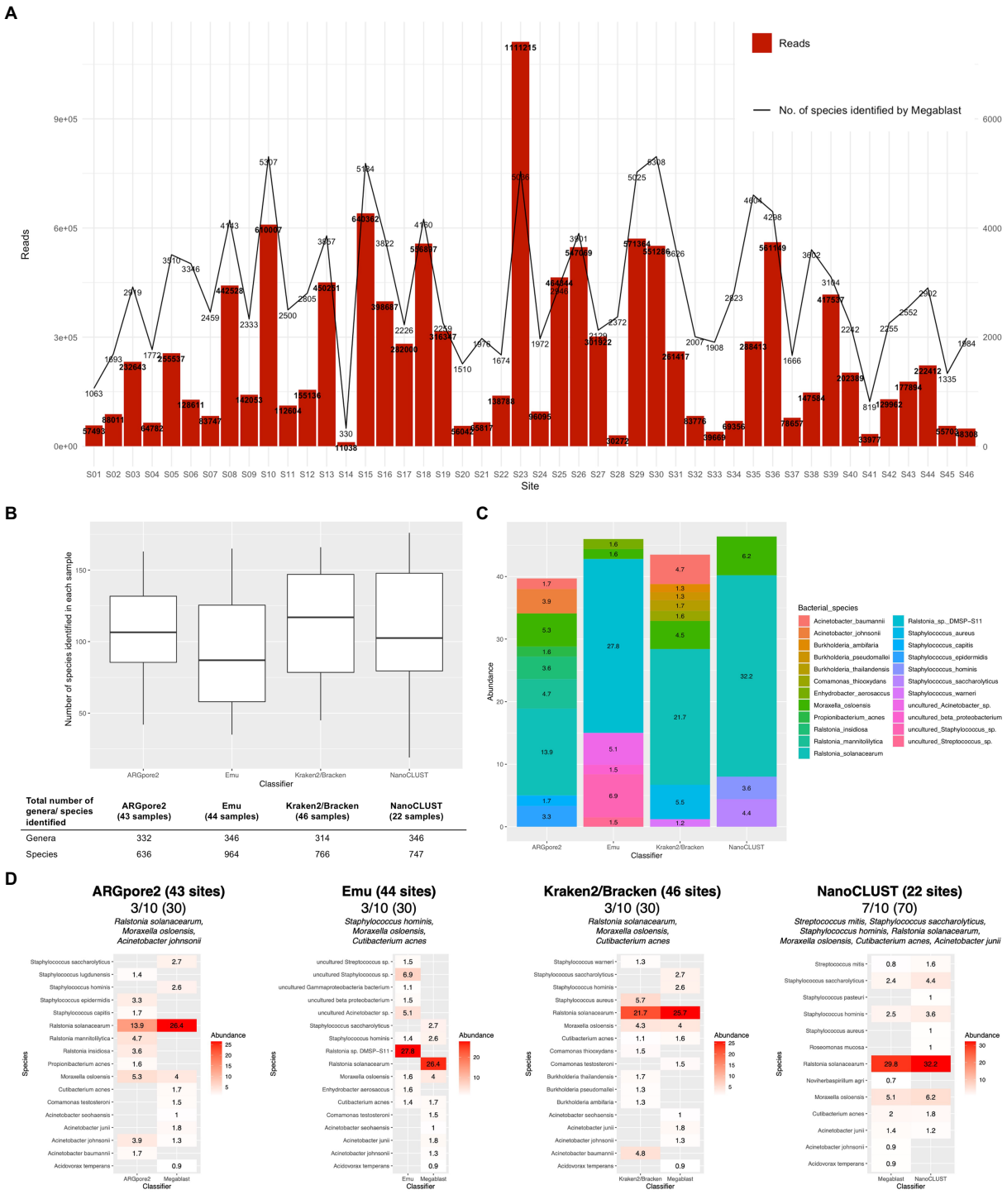
Results

Species-level taxonomic assignment of the urban samples using nanopore sequencing

Collectively, 2,854 ± 1,250 species were identified in 46 urban samples by nanopore sequencing (256,112 ± 229,162 raw reads) using MegaBLAST (Figure 1A; Supplementary Table S3). Based on nanopore sequencing data from the same set of samples, Kraken2/Bracken was able to reveal the presence of 766 species and 314 genera (Supplementary Table S4) from 46 urban samples. Meanwhile, ARGpore2, Emu, and NanoCLUST analyses were performed but they could only provide a taxonomic classification for 43 (636 species and 332 genera, Supplementary Table S5), 44 (964 species and 346 genera, Supplementary Table S6), and 22 urban samples (747 species and 346 genera, Supplementary Table S7) because of limited computational resources (Figure 1B).

Dominant species in each sample by the four classifiers

The overall microbiome profiles generated by the four classifiers are shown in Figure 1C (Supplementary Table S8). In ARGpore2 analysis, the most abundant species were *Ralstonia solanacearum* (13.9% ± 15.7%), *Moraxella osloensis* (5.3% ± 6.7%), *Ralstonia mannitolilytica* (4.7% ± 5.2%), and *Acinetobacter johnsonii* (3.9% ± 10.2%) (Figure 1C). For Emu, the dominant species included *Ralstonia* sp. *DMSP-S11* (27.8% ± 0.7%), *uncultured Staphylococcus* sp.



Total number of genera/ species identified	ARGpore2 (43 samples)	Emu (44 samples)	Kraken2/Bracken (46 samples)	NanoCLUST (22 samples)
Genera	332	346	314	346
Species	636	964	766	747

FIGURE 1
Species-level taxonomic assignment of the urban samples. **(A)** Total number of reads in each site and the number of species classified by MegaBLAST. **(B)** The number of bacterial species and genera in each sample identified by ARGpore2, Emu, Kraken2/Bracken, and NanoCLUST. **(C)** Average abundance of dominant species classified by ARGpore2, Emu, Kraken2/Bracken, and NanoCLUST. **(D)** Comparison of the top 10 dominant species between MegaBLAST and the four classifiers.

(6.9% ± 0.1%), *uncultured Acinetobacter* sp. (5.1% ± 0.3%), and *Enhydrobacter aerosaccus* (1.6% ± 0.0%), while in Kraken2/Bracken classification, *Ralstonia solanacearum* (21.7% ± 24.2%) was also the most abundant, followed by *Staphylococcus aureus* (5.5% ± 4.0%), *Acinetobacter baumannii* (4.7% ± 7.3%), and *Moraxella osloensis* (4.5% ± 5.9%) (Figure 1C). It is important to note that *Enhydrobacter aerosaccus* was distinctively detectable in Emu. In addition, NanoCLUST generated a similar result, in which *Ralstonia solanacearum* (32.2% ± 28.8%), *Moraxella osloensis* (6.2% ± 7.6%), and *Staphylococcus saccharolyticus* (4.4% ± 3.4%) were dominant

(Figure 1C). In addition, the concordance of the top 10 dominant species between MegaBLAST and the four classifiers was determined. NanoCLUST showed the highest concordance (7/10), followed by ARGpore2 (3/10), Emu (3/10), and Kraken2/Bracken (3/10) (Figure 1D). Among the concordant species, only *Moraxella osloensis* could be consistently identified by all classifiers, while *Ralstonia solanacearum* was identified by all classifiers except Emu (Figure 1D). *Cutibacterium acnes* could also be identified by all the classifiers except ARGpore2.

Comparisons of the microbiomes classified by MegaBLAST and the four classifiers

In addition to the dominant core microbiome, microbial profiles classified by MegaBLAST and the four classifiers were also compared based on the abundance and microbial taxa results. In total, 20 urban samples from each of the four classifiers were chosen for principal component analysis (PCA) and hierarchical clustering analysis to ensure a fair comparison. PCA in Figure 2A illustrates that microbial profiles generated by the four classifiers were distinctively different except Kraken2/Bracken and NanoCLUST. Of them, the microbiome created by NanoCLUST (blue) and Kraken2/Bracken (red) had the highest similarity to that of MegaBLAST (dark blue) (Figure 2A; Supplementary Table S8). This observation was supported by the hierarchical clustering in Figure 2B (Supplementary Table S8), where the profile of NanoCLUST was clustered with MegaBLAST, followed by Kraken2/Bracken, ARGpore2, and Emu. Furthermore, the abundance of MegaBLAST-positive and negative species classified by ARGpore2, Emu, Kraken2/Bracken, and NanoCLUST is shown in Figure 2C. Importantly, the mean abundance of the detectable and undetectable species was 0.478% ± 0.196 and 0.017% ± 0.014%, respectively. Therefore, the 0.1% relative abundance cutoff was believed to be acceptable for filtering true negative species. The accuracy, precision, sensitivity, specificity, and F1 score of the four classifiers were then evaluated based on the species detected by MegaBLAST (Figure 2D; Supplementary Table S9). No significant differences were observed in the accuracy of the four classifiers but F1 scores showed the highest in NanoCLUST (6.64%), followed by Kraken2/Bracken (4.17%), ARGpore2 (3.60%), and Emu (2.35%) (Figure 2D).

Comparisons of bacterial species identified by conventional culture and the four classifiers

Overall, 51 species and 19 genera were identified in 46 urban samples by the culture-based method. The dominant species belonged to *Staphylococcus hominis* ($n = 28$), followed by *Bacillus cereus* ($n = 18$), *Staphylococcus haemolyticus* ($n = 18$), and *Staphylococcus epidermidis* ($n = 17$) (Figure 3A; Supplementary Table S10). The identified isolates are listed in Supplementary Table S10. Afterward, we found that the mean abundance of the culture-based detectable species classified by ARGpore2, Emu, Kraken2/Bracken, and NanoCLUST was 0.640% ± 0.142% (Figure 3B). Notably, the mean abundance of the culture-based undetectable was 0.275% ± 0.129%, indicating that the 0.1% relative abundance cutoff, as proposed in previous studies (Zaura et al., 2009; Doan et al., 2016; Sun et al., 2021), might not be sufficient to filter true negative species. The mean abundance of the detectable species in Emu was the lowest (0.428%), followed by ARGpore2 (0.694%), NanoCLUST (0.733%), and Kraken2/Bracken (0.703%) (Figure 3B). This might explain the best performance in Emu, which has the highest accuracy (81.2%), specificity (84.8%), and F1 score (29%) (Figure 3Ci). Meanwhile, the accuracy (78.9%) and specificity (83.6%) of NanoCLUST were as high as Emu (Figure 3Ci). Although Kraken2/Bracken has lower accuracy and specificity, its sensitivity (55.9%) was much higher than others (Figure 3Ci; Supplementary Table S11). When the cutoff was changed to the mean abundance of the detectable species in MegaBLAST, i.e., 0.559, 0.663, 0.205, and 0.484 for ARGpore2, Kraken2/Bracken, Emu, and NanoCLUST, respectively, the accuracy and the specificity of all the four classifiers were > 80% (Figure 3Cii; Supplementary Table S11). Notably, Emu has the highest sensitivity, while its F1 score was still the highest (29.6%) (Figure 3Cii).

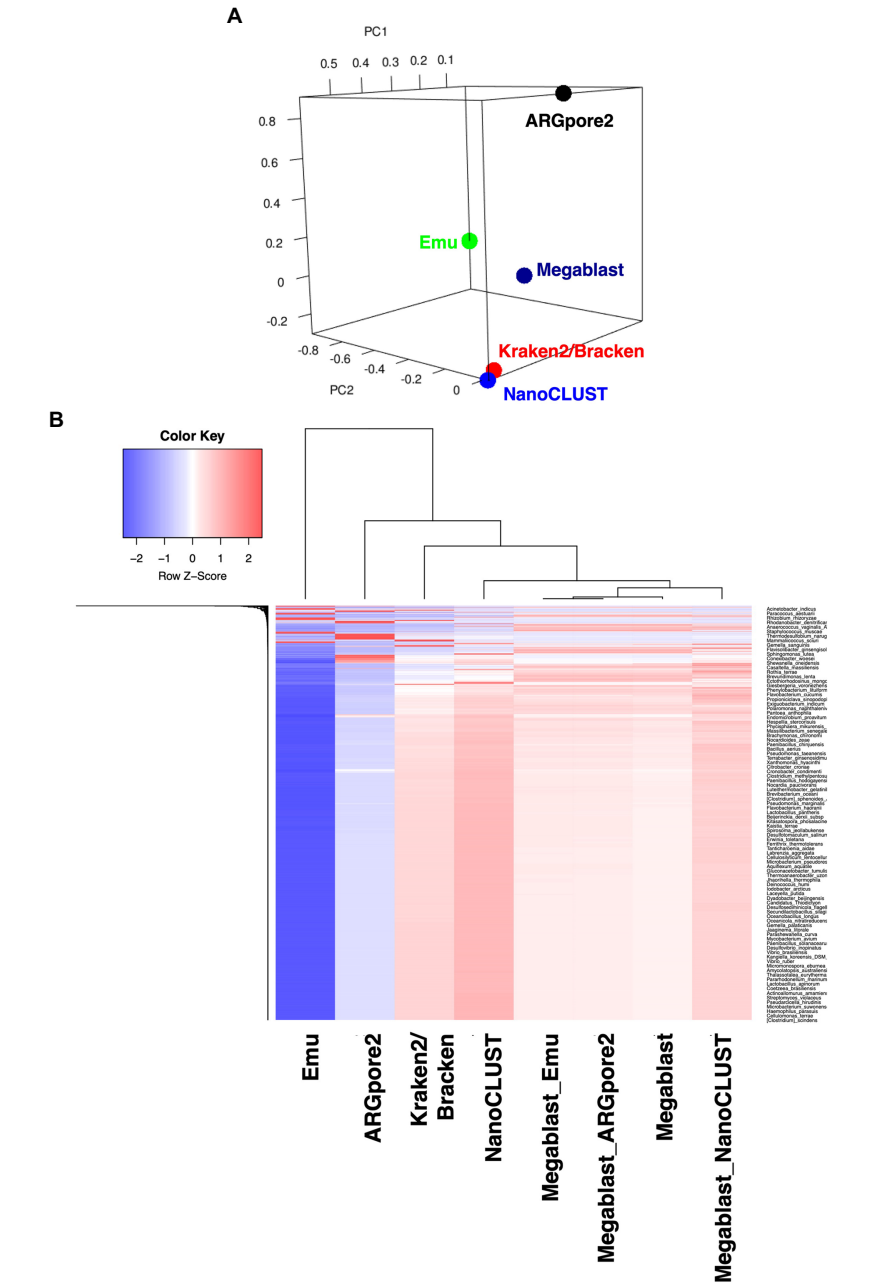
Nanopore 16S MegaBLAST failed to detect 13 culture-positive species

Four species, *Acinetobacter nosocomialis, Bacillus thuringiensis, Staphylococcus xylosus,* and *Stenotrophomonas maltophilia,* were detected by Kraken2/Bracken (orange dots) or ARGpore2 (black dots) but not by MegaBLAST (Figure 4A). Another nine cultured species were undetectable in all classifiers (blue dots), with five of them being *Bacillus spp.* (Figure 4A). Two *Bacilli* (*Oceanobacillus profundus* and *Paenibacillus* illinoisensis), *Empedobacter brevis,* and *Stenotrophomonas rhizophila* were also undetectable (Figure 4A). Interestingly, *Empedobacter brevis* was found in the conventional culture of Site 28 only (Figure 3A), whereas *Empedobacter falsenii (Wautersiella falsenii)* was found in its sequencing data (Supplementary Table S3). Additionally, MegaBLAST was able to detect other *Oceanobacillus* species at Site 1, Site 2, and Site 8 (Supplementary Table S3), which harbored *Oceanobacillus profundus* in conventional cultures (Figure 3A). Taken together, these species might be misclassified using nanopore sequencing.

Species that are culture- and MegaBLAST-positive but undetectable in the four classifiers

The number of cultured species detected by nanopore sequencing is shown in Figure 4A. Most of the cultured species (74.5%, 38/51) were detectable by MegaBLAST. Among the 38 species, 39.5% (15/38) could not be classified by the four classifiers, which are highlighted in red in Figure 4A, including *Wautersiella falsenii, Staphylococcus spps., Sphingobacterium multivorum, Rothia terrae, Pseudomonas* spp. *Kocuria kristinae, Corynebacterium aurimucosum, Bacillus spp., and Acinetobacter ursingii.* For the detectable species (23/38, 60.5%), all classifiers could detect *Bacillus cereus, Micrococcus luteus, Moraxella osloensis, Staphylococcus aureus*, and *Staphylococcus haemolyticus* (Figure 4A). Importantly, only Kraken2/Bracken could detect *Bacillus subtilis, Klebsiella pneumoniae, Staphylococcus caprae*, and *Staphylococcus pasteuri* (Figure 4A). Although Emu could not detect *Acinetobacter* species because most of them are classified as

10.3389/fmicb.2023.1164632



D

	ARGpore2	Emu	Kraken2/Bracken	NanoCLUST
Accuracy (%)	77.1	77.5	77.3	78.7
Precision (%)	76.1	91.6	80.7	97.1
Sensitivity (%)	1.8	1.2	2.1	3.4
Specificity (%)	99.8	100	99.8	100
F1 Score (%)	3.60	2.35	4.17	6.64

FIGURE 2
Comparisons of the microbiomes classified by MegaBLAST and the four classifiers. **(A)** Principal component analysis of microbiomes classified by MegaBLAST and the four classifiers, 20 urban examples that appeared in all of the classifiers were chosen. **(B)** Hierarchical clustering of the bacterial

(Continued)

FIGURE 2 (Continued)
species abundances classified by MegaBLAST, and the four classifiers **(C)** Species classified by ARGpore2, Emu, Kraken2/Bracken, and NanoCLUST were detectable (Pos) and undetectable (Neg) by MegaBLAST. The y-axis represents the abundance of the classified species, and the values were the abundance means. **(D)** The average values for accuracy, precision, sensitivity, specificity, and F1 score of the four classifiers based on the species detected by MegaBLAST.

Acinetobacter sp., it was the only classifier that can identify *Pseudomonas putida* and successfully found *Enterobacter asburiae, Staphylococcus capitis*, and *Staphylococcus saprophyticus*, while Kraken2/Bracken and NanoCLUST could not (Figure 4A). In addition, ARGpore2 exclusively detected *Pseudomonas stutzeri* and *Staphylococcus lugdunensis* (Figure 4A).

Complementation of nanopore 16S MegaBLAST and conventional culture results

The concordance of the culture and the MegaBLAST results is shown in Supplementary Table S11. It was highlighted that 60.4% of the species present in the MegaBLAST results were uncultivated, which were considered false negatives in the culture. To eliminate false positives in the outputs of the four classifiers, the MegaBLAST data were combined with standard culture results to create a new set of the occurrences of these 51 bacterial species (Supplementary Table S12). After removing these false positives, F1 scores of the four classifiers increased from 24.5% ± 3.7 to 46.7% ± 7.9%, while specificity increased from 78.9% ± 6.2 to 94.9% ± 4.1% (Figure 4B; Supplementary Table S11). The accuracy, however, decreased from 75.8% ± 5.0 to 51.4% ± 3.8% (Figure 4B; Supplementary Table S11).

Discussion

Herein we collected 46 samples from the mass transit system of Hong Kong and identified the bacterial community by conventional culture and nanopore full-length 16S rRNA sequencing. The performances of four taxonomic classifiers, namely ARGpore2, Emu, Kraken2/Bracken, and NanoCLUST, were evaluated based on a set of urban microbiome data generated by MegaBLAST. The ideal taxonomic classifier should demonstrate a high concordance of dominant species and a similar microbial profile as MegaBLAST, which was one of the most sensitive metagenomics alignment methods but computationally intensive (Ye et al., 2019). Second, we comprehensively compared the bacterial species identified by conventional culture and 16S nanopore sequencing and eventually provided recommendations for classifier selection.

Despite having a higher error rate and lower throughput than Illumina sequencing, nanopore sequencing can produce much longer reads, allowing for the study of complex microbial samples and the identification of rare taxa (Ciuffreda et al., 2021). Data analysis such as read mapping and *de novo* assembly for nanopore sequencing may be difficult due to specialized tools and expertise (Magi et al., 2018). Limited studies were performed to evaluate the performance of classifiers using 16S nanopore sequencing data (Ciuffreda et al., 2021; Curry et al., 2022). As such, our study looked into the use of nanopore sequencing for 16S microbial profiling. Kraken2/Bracken, which was not designed for nanopore sequencing, performed well in bacterial community classification, according to our findings. Following that, a comparison of Kraken2/Bracken to different nanopore classification tools (ARGpore2, Emu, and NanoCLUST) will be shown.

Previous studies demonstrated that the similarities between NanoCLUST/Emu and the expected mock communities were higher than that of Kraken2/Bracken (Rodriguez-Perez et al., 2021; Curry et al., 2022). While there were < 100 bacterial species in the mock communities used in NanoCLUST and Emu studies, our studies used actual communities with >700 distinct species. NanoCLUST, the first custom-built pipeline for analyzing nanopore 16S rRNA amplicon reads, had the highest concordance of dominant species and microbial profile similarity to MegaBLAST. In addition, NanoCLUST had the highest F1 score (6.64%). The underlying cause could be the use of subsequent BLAST classification in NanoCLUST (Rodriguez-Perez et al., 2021). Nevertheless, the time required for a 300,000-read dataset using NanoCLUST was 20 min, while MegaBLAST took >24 h in our computational analysis. In the clustering analysis, Kraken2/Bracken, which was the fastest taxonomic classifier (1 min) and could analyze all urban samples, had a similar microbial profile to NanoCLUST. Its F1 score (4.17%) was also higher than ARGpore2 (3.60%) and Emu (2.35%).

Curiously, the most prevalent species was *Ralstonia solanacearum* in all classifiers except Emu, where *Ralstonia* sp. *DMSP-S11* was the most abundant. In addition, Emu could not provide an exact *Acinetobacter* and *Staphylococcus* species and therefore reported hits at a higher taxonomic rank that was labeled as "*Acinetobacter sp.*" and "*Staphylococcus sp.*" Excluding the two species identified only at the genus level as "*Acinetobacter sp.*" and "*Staphylococcus sp.*", there were still a hundred other species that were classified at the genus level. This might explain its distinct microbial profile when compared with MegaBLAST. Taken together, NanoCLUST was the most desirable classifier for complex community profiling, followed by Kraken2/Bracken, ARGpore2, and Emu.

In addition to microbial profiling, the performances in culturable species detection of the four classifiers were also estimated. Overall, the concordance between culture and 16S sequencing was low (38.6%). A previous study indicated the challenge of cutoff selection for accurate microbial identification in complex community analysis (Sun et al., 2021). The relative abundance of 0.1% as the typical cutoff to filter true negatives is still inconclusive. To that end, we used the mean abundance of the MegaBLAST-positive species as the cutoff and thus lead to overall increases in accuracy, F1 score, and specificity of the four classifiers. Of note, Emu, which could not provide accurate microbial profiling based on a set of urban microbiome data generated by MegaBLAST, has the highest F1 score in culturable species detection. The combination of rrnDB version 5.635 (Stoddard et al., 2015) and NCBI 16S RefSeq (O'Leary et al., 2016) in the Emu 16S database might benefit the species-level classification process and effectively reduce the number of unclassified or misclassified reads (Curry et al., 2021). Interestingly, only Emu could detect *Enhydrobacter aerosaccus* among the four classifiers, which was found in 41 urban samples. It was revealed that *Enhydrobacter aerosaccus* was enriched

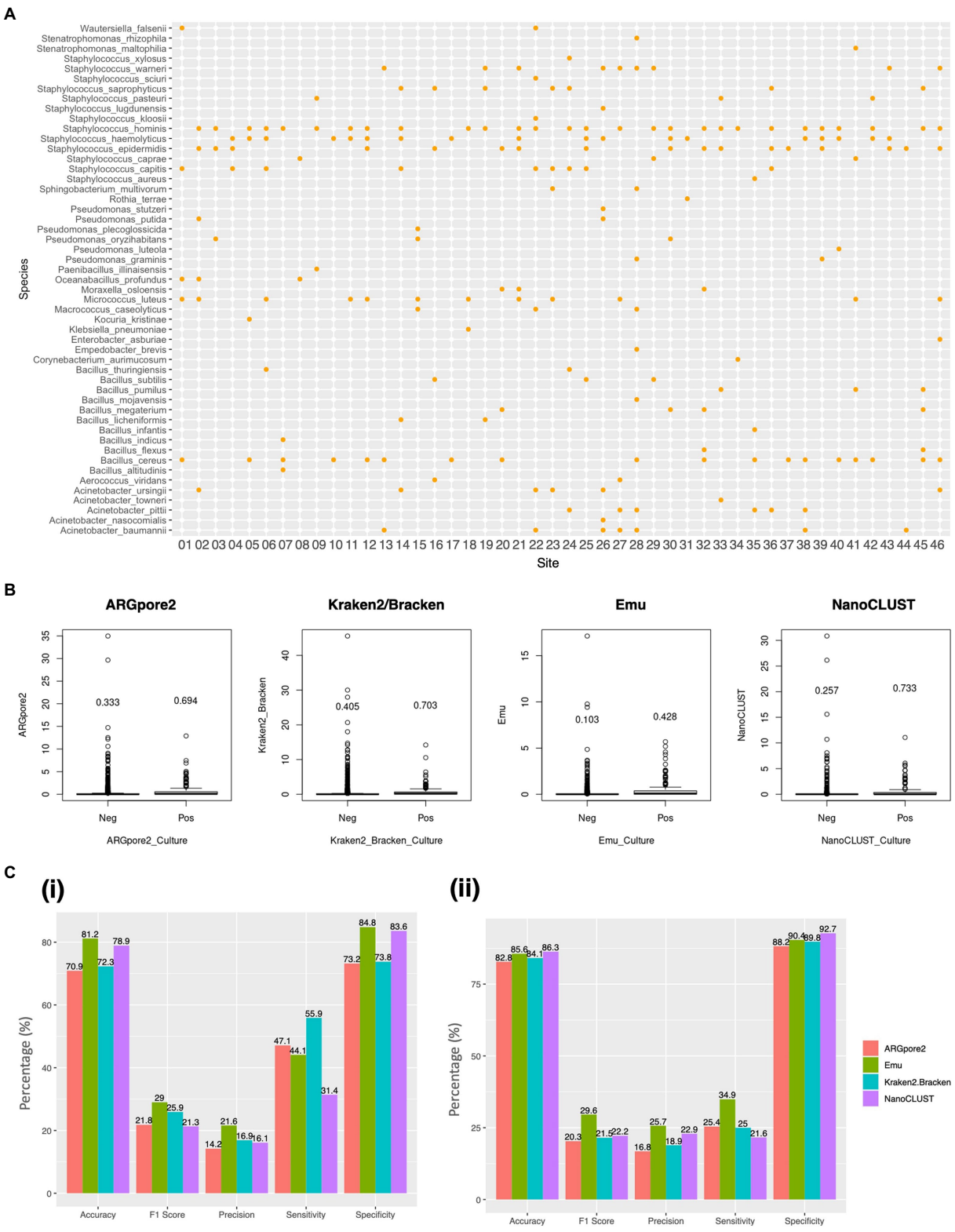


FIGURE 3
Comparisons of bacterial species identified by conventional culture and the four classifiers. **(A)** Bacterial presence in routine culture from the 46 urban samples. Each dot represents at least one colony that has been identified. **(B)** Species classified by ARGpore2, Emu, Kraken2/Bracken, and NanoCLUST were present (Pos) and absent (Neg) in culture. The y-axis represents the abundance of the classified species, and the values were the abundance means. **(C)** Bar plots of average accuracy, precision, sensitivity, specificity, and F1 score of the four classifiers based on the species present in culture using (i) 0.1% and (ii) mean abundance of the detectable species in MegaBLAST as the cutoff.

10.3389/fmicb.2023.1164632

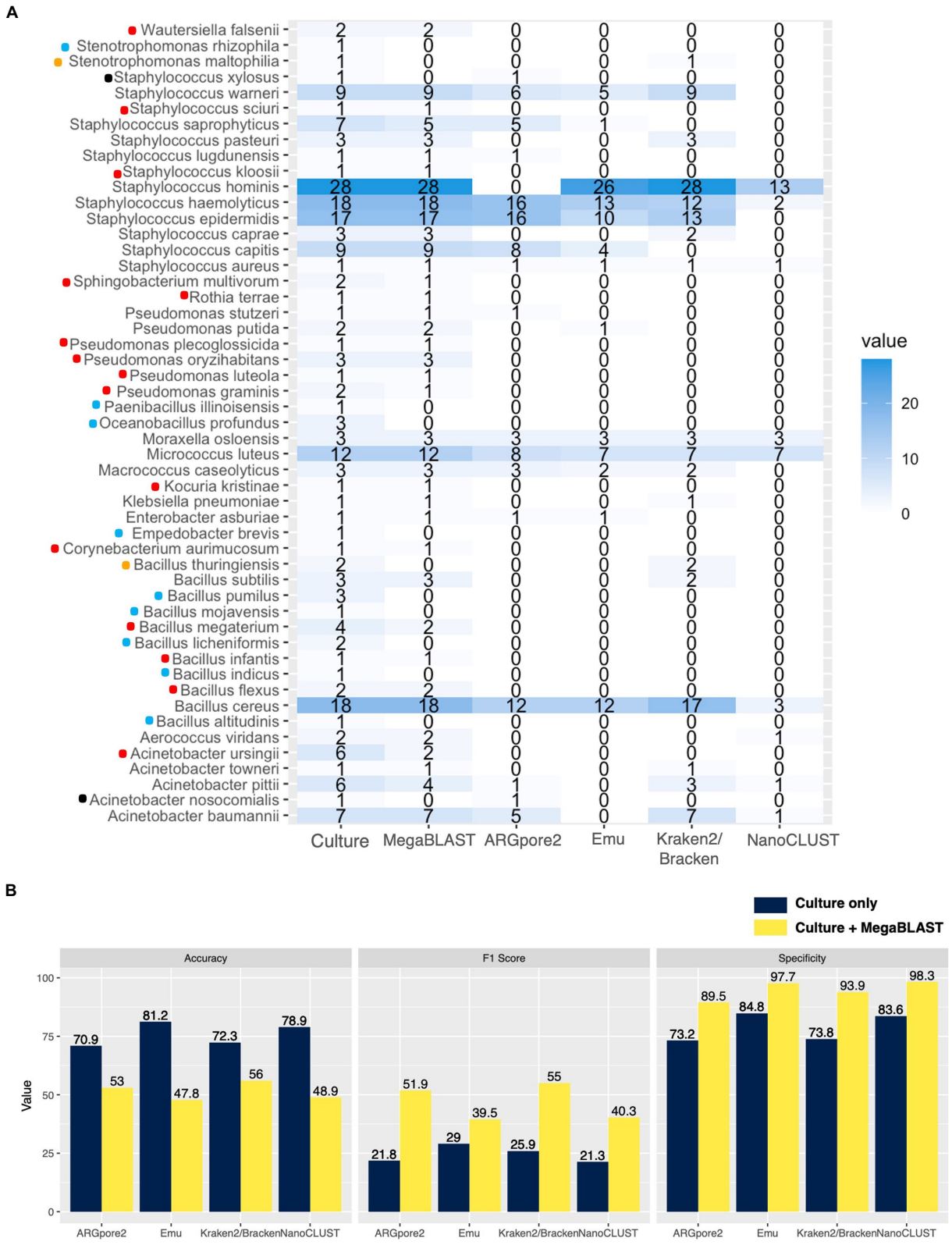


FIGURE 4
Concordance of the culturable species between MegaBLAST and the four classifiers. **(A)** Detectable and cultured species in the four classifiers. The value represents the number of urban samples with the corresponding bacterial species concordant with the culture results. Red dots indicate the detectable species in MegaBLAST, while blue dots represent the absence of the species in the sequencing data. Four species were not detected by MegaBLAST but could be identified by Kraken2/Bracken and ARGpore2, which were labeled with orange and black dots, respectively. **(B)** Average F1 score (%) and specificity (%) increased in the complement results of culture and MegaBLAST while accuracy (%) decreased.

in the public transit air microbiome of Hong Kong (Leung et al., 2021), and *Enhydrobacter sp.* was more prevalent in Asian individuals (Leung et al., 2015). Furthermore, the ability to eliminate the misclassification caused by nanopore sequencing through the expectation–maximization algorithm in Emu might explain the additional concordant species classified by Emu (Curry et al., 2021). Hence, the distinct microbial profiling by Emu is caused by the uncertainty of species-level identification. Only genus-level identification was given for *Staphylococcus sp.* and *Acinetobacter sp.* Therefore, Emu would be preferable if the uncertainty of species identification is improved. The inability of species-level classification, in addition to Emu, may account for the discrepancy between culture and 16S sequencing for all classifiers.

In the in-depth analysis of the concordance between culture and nanopore sequencing, 17.6% of culturable species were undetectable. Previous studies suggested that DNA isolation procedures, the main determinants of sequencing results, could result in some degree of DNA loss that might lower the final DNA quantities of some bacterial species and eventually cause undetectable species (Knudsen et al., 2016). Furthermore, misclassification is still a challenge to classify genomes from distinct species with a high genomic identity (Wood et al., 2019), especially for *Bacillus* spp. in our results. It has been indicated that *Bacillus cereus* was misclassified as *Bacillus thuringiensis*, owing to the 99.73% similarity in their 16S rRNA sequence (Rodriguez-Perez et al., 2021). Furthermore, MegaBLAST might misclassify *Empedobacter brevis* as *Empedobacter falsenii* (*Wautersiella falsenii*) due to its highest 16S rRNA gene sequence similarity (Schauss et al., 2015). Overall, the mismatch between culture and 16S sequencing could also be explained by the loss of bacterial DNA and misclassification.

Our results demonstrated that bacterial species of the highest abundance (*Ralstonia solanacearum*) in 16S sequencing were not present in the culture. Typically, 48–72 h of incubation at 28°C instead of 37°C is required for *R. solanacearum* (Pawaskar et al., 2014). Many bacterial species are unculturable in general growth conditions or survive before sample processing (Tringe and Rubin, 2005). In addition, it was reported that uncultured genera comprise 81% of bacterial species across urban environments (Lloyd et al., 2018). Meanwhile, some bacterial species may be missed in MALDI identification, leading to false negative results in conventional culture. The findings of combining MegaBLAST with conventional culture results showed that culture-negative but MegaBLAST-positive results significantly influenced the F1 scores of the four classifiers.

To conclude, it is suggested to use various tools depending on the applications. For 16S microbial profiling, NanoCLUST with a high concordance of dominant species and a similar microbial profile to MegaBLAST was desirable whereas Kraken2/Bracken, which had comparable clustering results to NanoCLUST, was also recommended. Second, Emu was preferred for recognizing culturable species accurately. Our unbiased and unsupervised assessment strategy could comprehensively evaluate newly developed classifiers, improved classifiers, and established classifiers of different step parameters for complex communities. It is recommended that the performance of the classifiers be investigated for both simple and complex communities. When using these classifiers, it will be beneficial to provide different parameters or settings, such as community complexity and sample types.

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

AL, CC, and GS conceived the original concept and designed the study. AL, CC, LW, C-YY, W-TL, and K-CC performed the experiments. AL analyzed the data and wrote the original manuscript. GS and AL amended and revised the writing. All authors read and approved the final version of the manuscript.

Funding

This manuscript was supported by Interdisciplinary Large External Project Application Scheme from the Faculty of Health and Social Sciences, The Hong Kong Polytechnic University (1-ZVZL).

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2023.1164632/full#supplementary-material

References

Benitez-Paez, A., Portune, K. J., and Sanz, Y. (2016). Species-level resolution of 16S rRNA gene amplicons sequenced through the MinION portable nanopore sequencer. *Gigascience* 5:4. doi: 10.1186/s13742-016-0111-z

Callahan, B. J., Sankaran, K., Fukuyama, J. A., McMurdie, P. J., and Holmes, S. P. (2016). Bioconductor workflow for microbiome data analysis: from raw reads to community analyses. *F1000Res* 5:1492. doi: 10.12688/f1000research.8986.2

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: architecture and applications. *BMC Bioinformat.* 10:421. doi: 10.1186/1471-2105-10-421

Cheng, H., Sun, Y., Yang, Q., Deng, M., Yu, Z., Liu, L., et al. (2022). An ultra-sensitive bacterial pathogen and antimicrobial resistance diagnosis workflow using Oxford Nanopore adaptive sampling sequencing method. *medRxiv*. doi: 10.1101/2022.07.03.22277093

Ciuffreda, L., Rodriguez-Perez, H., and Flores, C. (2021). Nanopore sequencing and its application to the study of microbial communities. *Comput. Struct. Biotechnol. J.* 19, 1497–1511. doi: 10.1016/j.csbj.2021.02.020

Curry, K. D., Wang, Q., Nute, M. G., Tyshaieva, A., Reeves, E., Soriano, S., et al. (2021). Emu: Species-level microbial community profiling for full-length Nanopore 16S reads. *bioRxiv*. doi: 10.1101/2021.05.02.442339

Curry, K. D., Wang, Q., Nute, M. G., Tyshaieva, A., Reeves, E., Soriano, S., et al. (2022). Emu: species-level microbial community profiling of full-length 16S rRNA Oxford Nanopore sequencing data. *Nat. Methods* 19, 845–853. doi: 10.1038/s41592-022-01520-4

Danko, D., Bezdan, D., Afshin, E. E., Ahsanuddin, S., Bhattacharya, C., Butler, D. J., et al. (2021). A global metagenomic map of urban microbiomes and antimicrobial resistance. *Cells* 184, 3376–3393.e17 e3317. doi: 10.1016/j.cell.2021.05.002

De Coster, W., D'Hert, S., Schultz, D. T., Cruts, M., and Van Broeckhoven, C. (2018). NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics* 34, 2666–2669. doi: 10.1093/bioinformatics/bty149

Deshpande, S. V., Reed, T. M., Sullivan, R. F., Kerkhof, L. J., Beigel, K. M., and Wade, M. M. (2019). Offline next generation Metagenomics sequence analysis using MinION detection software (MINDS). *Genes (Basel)* 10:578. doi: 10.3390/genes10080578

Doan, T., Akileswaran, L., Andersen, D., Johnson, B., Ko, N., Shrestha, A., et al. (2016). Paucibacterial microbiome and resident DNA virome of the healthy conjunctiva. *Invest. Ophthalmol. Vis. Sci.* 57, 5116–5126. doi: 10.1167/iovs.16-19803

Kerkhof, L. J. (2021). Is Oxford Nanopore sequencing ready for analyzing complex microbiomes? *FEMS Microbiol. Ecol.* 97:fiab001. doi: 10.1093/femsec/fiab001

Kim, D., Song, L., Breitwieser, F. P., and Salzberg, S. L. (2016). Centrifuge: rapid and sensitive classification of metagenomic sequences. *Genome Res.* 26, 1721–1729. doi: 10.1101/gr.210641.116

Knudsen, B. E., Bergmark, L., Munk, P., Lukjancenko, O., Prieme, A., Aarestrup, F. M., et al. (2016). Impact of sample type and DNA isolation procedure on genomic inference of microbiome composition. *mSystems* 1:16. doi: 10.1128/mSystems.00095-16

Lao, H. Y., Ng, T. T., Wong, R. Y., Wong, C. S., Lee, L. K., Wong, D. S., et al. (2022). The clinical utility of two high-throughput 16S rRNA gene sequencing workflows for taxonomic assignment of unidentifiable bacterial pathogens in matrix-assisted laser desorption ionization-time of flight mass spectrometry. *J. Clin. Microbiol.* 60:e0176921. doi: 10.1128/JCM.01769-21

Leidenfrost, R. M., Pöther, D.-C., Jäckel, U., and Wünschiers, R. (2020). Benchmarking the MinION: evaluating long reads for microbial profiling. *Sci. Rep.* 10:5125. doi: 10.1038/s41598-020-61989-x

Leung, M. H. Y., Tong, X., Boifot, K. O., Bezdan, D., Butler, D. J., Danko, D. C., et al. (2021). Characterization of the public transit air microbiome and resistome reveals geographical specificity. *Microbiome* 9:112. doi: 10.1186/s40168-021-01044-7

Leung, M. H., Wilkins, D., and Lee, P. K. (2015). Insights into the pan-microbiome: skin microbial communities of Chinese individuals differ from other racial groups. *Sci. Rep.* 5:11845. doi: 10.1038/srep11845

Lloyd, K. G., Steen, A. D., Ladau, J., Yin, J., and Crosby, L. (2018). Phylogenetically novel uncultured microbial cells dominate earth microbiomes. *mSystems* 3:18. doi: 10.1128/mSystems.00055-18

Lu, J., and Salzberg, S. L. (2020). Ultrafast and accurate 16S rRNA microbial community analysis using kraken 2. *Microbiome* 8:124. doi: 10.1186/s40168-020-00900-2

Magi, A., Semeraro, R., Mingrino, A., Giusti, B., and D'Aurizio, R. (2018). Nanopore sequencing data analysis: state of the art, applications and challenges. *Brief. Bioinform.* 19, 1256–1272. doi: 10.1093/bib/bbx062

Marshall, C. W., Kurs-Lasky, M., McElheny, C. L., Bridwell, S., Liu, H., and Shaikh, N. (2021). Performance of conventional urine culture compared to 16S rRNA gene amplicon sequencing in children with suspected urinary tract infection. *Microbiol Spectr* 9:e0186121. doi: 10.1128/spectrum.01861-21

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

McMurdie, P. J., and Holmes, S. (2014). Waste not, want not: why rarefying microbiome data is inadmissible. *PLoS Comput. Biol.* 10:e1003531. doi: 10.1371/journal.pcbi.1003531

McMurdie, P. J., and Holmes, S. (2015). Shiny-phyloseq: web application for interactive microbiome analysis with provenance tracking. *Bioinformatics* 31, 282–283. doi: 10.1093/bioinformatics/btu616

Morgulis, A., Coulouris, G., Raytselis, Y., Madden, T. L., Agarwala, R., and Schaffer, A. A. (2008). Database indexing for production MegaBLAST searches. *Bioinformatics* 24, 1757–1764. doi: 10.1093/bioinformatics/btn322

Neiderud, C. J. (2015). How urbanization affects the epidemiology of emerging infectious diseases. *Infect Ecol Epidemiol* 5:27060. doi: 10.3402/iee.v5.27060

O'Leary, N. A., Wright, M. W., Brister, J. R., Ciufo, S., Haddad, D., McVeigh, R., et al. (2016). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* 44, D733–D745. doi: 10.1093/nar/gkv1189

Pawaskar, J., Joshi, M., Navathe, S., and Agale, R. (2014). Physiological and biochemical characters of Ralstonia solanacearum. *Int. J. Res. Agricult. Sci.* 1, 2348–3997.

Pedron, J., Guyon, L., Lecomte, A., Blottiere, L., Chandeysson, C., Rochelle-Newall, E., et al. (2020). Comparison of environmental and culture-derived bacterial communities through 16S Metabarcoding: a powerful tool to assess media selectivity and detect rare taxa. *Microorganisms* 8:1129. doi: 10.3390/microorganisms8081129

Rice, L. B. (2008). Federal funding for the study of antimicrobial resistance in nosocomial pathogens: no ESKAPE. *J. Infect. Dis.* 197, 1079–1081. doi: 10.1086/533452

Rodriguez-Perez, H., Ciuffreda, L., and Flores, C. (2021). NanoCLUST: a species-level analysis of 16S rRNA nanopore sequencing data. *Bioinformatics* 37, 1600–1601. doi: 10.1093/bioinformatics/btaa900

Sala-Comorera, L., Caudet-Segarra, L., Galofre, B., Lucena, F., Blanch, A. R., and Garcia-Aljaro, C. (2020). Unravelling the composition of tap and mineral water microbiota: divergences between next-generation sequencing techniques and culture-based methods. *Int. J. Food Microbiol.* 334:108850. doi: 10.1016/j.ijfoodmicro.2020.108850

Schauss, T., Busse, H. J., Golke, J., Kampfer, P., and Glaeser, S. P. (2015). Empedobacter stercoris sp. nov., isolated from an input sample of a biogas plant. *Int. J. Syst. Evol. Microbiol.* 65, 3746–3753. doi: 10.1099/ijsem.0.000486

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

Shah, N., Nute, M. G., Warnow, T., and Pop, M. (2019). Misunderstood parameter of NCBI BLAST impacts the correctness of bioinformatics workflows. *Bioinformatics* 35, 1613–1614. doi: 10.1093/bioinformatics/bty833

Stoddard, S. F., Smith, B. J., Hein, R., Roller, B. R., and Schmidt, T. M. (2015). rrnDB: improved tools for interpreting rRNA gene abundance in bacteria and archaea and a new foundation for future development. *Nucleic Acids Res.* 43, D593–D598. doi: 10.1093/nar/gku1201

Sun, Z., Huang, S., Zhang, M., Zhu, Q., Haiminen, N., Carrieri, A. P., et al. (2021). Challenges in benchmarking metagenomic profilers. *Nat. Methods* 18, 618–626. doi: 10.1038/s41592-021-01141-3

Technical Service Consultants Ltd (2014). INSTRUCTION FOR USE Hygiene Sponge with Neutraliser. Biotrading Available at: http://biotrading.com/assets/productinformatie/tsc/ifu/ifu_ts15b.pdf

Tringe, S. G., and Rubin, E. M. (2005). Metagenomics: DNA sequencing of environmental samples. *Nat. Rev. Genet.* 6, 805–814. doi: 10.1038/nrg1709

Urban, L., Holzer, A., Baronas, J. J., Hall, M. B., Braeuninger-Weimer, P., Scherm, M. J., et al. (2021). Freshwater monitoring by nanopore sequencing. *elife* 10:61504. doi: 10.7554/eLife.61504

Wang, H., Liu, N., Chen, J., and Guo, S. (2022). The relationship between Urban renewal and the built environment: a systematic review and Bibliometric analysis. *J. Plan. Lit.* 37, 293–308. doi: 10.1177/08854122211058909

Winand, R., Bogaerts, B., Hoffman, S., Lefevre, L., Delvoye, M., Braekel, J. V., et al. (2019). Targeting the 16s Rrna gene for bacterial identification in complex mixed samples: comparative evaluation of second (Illumina) and third (Oxford Nanopore technologies) generation sequencing technologies. *Int. J. Mol. Sci.* 21:298. doi: 10.3390/ijms21010298

Wood, D. E., Lu, J., and Langmead, B. (2019). Improved metagenomic analysis with kraken 2. *Genome Biol.* 20:257. doi: 10.1186/s13059-019-1891-0

Ye, S. H., Siddle, K. J., Park, D. J., and Sabeti, P. C. (2019). Benchmarking Metagenomics tools for taxonomic classification. *Cells* 178, 779–794. doi: 10.1016/j.cell.2019.07.010

Zaura, E., Keijser, B. J. F., Huse, S. M., and Crielaard, W. (2009). Defining the healthy" core microbiome" of oral microbial communities. *BMC Microbiol.* 9, 1–12. doi: 10.1186/1471-2180-9-259

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 25 May 2023
DOI 10.3389/fmicb.2023.1199144

OPEN ACCESS

EDITED BY
Helianthous Verma,
University of Delhi, India

REVIEWED BY
Paulina Corral,
University of Seville, Spain
Utkarsh Sood,
University of Delhi, India

*CORRESPONDENCE
Huanhuan Liu
✉ lh_tust@tust.edu.cn
Xiaoping Liao
✉ liao_xp@tib.cas.cn

RECEIVED 03 April 2023
ACCEPTED 09 May 2023
PUBLISHED 25 May 2023

CITATION
Zhang Z, Cui M, Chen P, Li J, Mao Z, Mao Y, Li Z, Guo Q, Wang C, Liao X and Liu H (2023) Insight into the phylogeny and metabolic divergence of *Monascus* species (*M. pilosus*, *M. ruber*, and *M. purpureus*) at the genome level.
Front. Microbiol. 14:1199144.
doi: 10.3389/fmicb.2023.1199144

Insight into the phylogeny and metabolic divergence of *Monascus* species (*M. pilosus*, *M. ruber*, and *M. purpureus*) at the genome level

Zhiyu Zhang[1,2], Mengfei Cui[1,2], Panting Chen[1,2], Juxing Li[1,2], Zhitao Mao[3], Yufeng Mao[3], Zhenjing Li[1,2], Qingbin Guo[1,2], Changlu Wang[1,2], Xiaoping Liao[3,4]* and Huanhuan Liu[1,2]*

[1]State Key Laboratory of Food Nutrition and Safety, Tianjin University of Science and Technology, Tianjin, China, [2]State Key Laboratory of Food Nutrition and Safety, Tianjin University of Science and Technology, Ministry of Education, Tianjin, China, [3]Biodesign Center, Key Laboratory of Engineering Biology for Low-Carbon Manufacturing, Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences, Tianjin, China, [4]Haihe Laboratory of Synthetic Biology, Tianjin, China

Background: Species of the genus *Monascus* are economically important and widely used in the production of food colorants and monacolin K. However, they have also been known to produce the mycotoxin citrinin. Currently, taxonomic knowledge of this species at the genome level is insufficient.

Methods: This study presents genomic similarity analyses through the analysis of the average nucleic acid identity of the genomic sequence and the whole genome alignment. Subsequently, the study constructed a pangenome of *Monascus* by reannotating all the genomes and identifying a total of 9,539 orthologous gene families. Two phylogenetic trees were constructed based on 4,589 single copy orthologous protein sequences and all the 5,565 orthologous proteins, respectively. In addition, carbohydrate active enzymes, secretome, allergic proteins, as well as secondary metabolite gene clusters were compared among the included 15 *Monascus* strains.

Results: The results clearly revealed a high homology between *M. pilosus* and *M. ruber*, and their distant relationship with *M. purpureus*. Accordingly, all the included 15 *Monascus* strains should be classified into two distinctly evolutionary clades, namely the *M. purpureus* clade and the *M. pilosus-M. ruber* clade. Moreover, gene ontology enrichment showed that the *M. pilosus-M. ruber* clade had more orthologous genes involved with environmental adaptation than the *M. purpureus* clade. Compared to *Aspergillus oryzae*, all the *Monascus* species had a substantial gene loss of carbohydrate active enzymes. Potential allergenic and fungal virulence factor proteins were also found in the secretome of *Monascus*. Furthermore, this study identified the pigment synthesis gene clusters present in all included genomes, but with multiple nonessential genes inserted in the gene cluster of *M. pilosus* and *M. ruber* compared to *M. purpureus*. The citrinin gene cluster was found to be intact and highly conserved only among *M. purpureus* genomes. The monacolin K gene cluster was found only in the genomes of *M. pilosus* and *M. ruber*, but the sequence was more conserved in *M. ruber*.

Conclusion: This study provides a paradigm for phylogenetic analysis of the genus *Monascus*, and it is believed that this report will lead to a better understanding of these food microorganisms in terms of classification, metabolic differentiation, and safety.

KEYWORDS

***Monascus*, phylogenetic analysis, taxonomic divergence, secondary metabolite gene clusters, pan-genome**

1. Introduction

Monascus spp. is a highly valuable edible fungus that has been traditionally consumed in China and other Asian countries such as Japan, Republic of Korea, and the Philippines. It has high nutritional value due to the synthesis of a variety of beneficial secondary metabolites, including *Monascus* azaphilone pigments (MonAzPs), monacolin K (MK), aminobutyric acid, ergosterol, and Hong Qu polysaccharide (Wang et al., 2021), with MonAzPs and MK being of special concern. MonAzPs are a type of chromogenic chemical consisting of a chromophore with a polyketide structure and medium- or long-chain fatty acids. They are highly promising in the food, healthcare, and cosmetic industries due to their excellent coloring properties, good biological efficacy (antioxidant, anti-inflammatory, hypolipidemic, and anti-tumor properties), and non-toxic side effects as natural food additives (Chen et al., 2019). Another important metabolite, MK, is an inhibitor of 3-hydroxy-3-methylglutaryl-coenzyme A (HMG-CoA) reductase, a critical enzyme in endogenous cholesterol production (Zhang et al., 2020). It is a prescription drug with brand names Mevinoline, Lovastatin, or Mevalonate used to treat excessive cholesterol, coronary heart disease, and other disorders (Karthikeyan and Dharumadurai, 2023). However, concerns were raised regarding the safety of *Monascus* products when citrinin, a mycotoxin hazardous to both humans and animals, was detected in *Monascus* products as early as 1995 (Shao et al., 2014). Citrinin is nephrotoxic and can cause kidney enlargement, renal tubule expansion, and renal epithelial cell degeneration and necrosis. Fortunately, some citrinin-free *Monascus* strains and production techniques have been developed recently, including metabolic engineering to eliminate the production of citrinin by disrupting the polyketide synthase gene *pksCT* (Jia et al., 2010) or dehydrogenase gene *citE* (Ning et al., 2017), natural screening (Feng et al., 2016), mutagenesis (Kalaivani and Rajasekaran, 2014), genome shuffling (Ghosh and Dam, 2020), low pH (Kang et al., 2014), and genistein addition (Ouyang et al., 2021). With the increased safety of *Monascus*, people are becoming more interested in the health benefits of this edible fungus.

Despite the economic importance of *Monascus*, taxonomic research on this genus remains limited (Barbosa et al., 2017). *Monascus* spp. are categorized under the phylum Eumycota, subphylum Ascomycota, class Plectomycetes, order Eurotiales, and family *Monascaceae*. Morphologically, this genus generates non-porous perithecia at the top of the stem-like hypha, ascospores distributed throughout the hypha, virtually spherical to wide spherical ascospores, and transparent and oval ascospores detached from the closed capsule. In 1983 Hawksworth and Pitt updated the genus based on physiological and morphological criteria, reducing the number of recognized species to three: *M. pilosus*, *M. ruber*, and *M. purpureus* (Barbosa et al., 2017). Following the discovery of new species, the NCBI taxonomy database now contains more than twenty records on *Monascus* species.[1] Among these, *M. purpureus* has the most heterotypic synonyms, such as *M. albidus*, *M. anka*, *M. erroneous*, and *M. rubiginosus*.

More recently, DNA sequencing and molecular phylogenetics have become essential in microbiological taxonomy. ITS (Dai et al., 2021), LSU (He et al., 2020; Higa et al., 2020), β-tubulin (Tong et al., 2022), calmodulin (He et al., 2020; Ruiz and Radwan, 2021) are frequently used for accurate species identification of close relatives among *Monascus* spp. However, phylogenetic delineation of this genus was complicated by gene region inconsistency and low support for internal nodes. Phenotype-based identification schemes in *Monascus* have been difficult to reconcile with the results obtained by ITS, partial LSU and/or β-tubulin gene sequencing (Park and Jong, 2003; Park et al., 2004; Shao et al., 2011, 2014), and the genetic identities of *Monascus* species are still under debate or are even confusing. Back in Park and Jong (2003) investigated the phylogenetic relationships among the species using sequences from the D1/D2 region of the large subunit (LSU) rRNA genes. They found that *M. ruber*, *M. pilosus*, and *M. purpureus* were closely related and clustered into the same subgroup, but *M. ruber* and *M. pilosus* were unable to be distinguished from each other. In Park et al. (2004) amplified and sequenced the ITS and partial β-tubulin genes of 17 ATCC reference strains of *Monascus* species. Still, they found that *M. pilosus* and *M. ruber* could not be differentiated using these sequences (Park et al., 2004), implying that species boundaries in *Monascus* should be reexamined. In fact, although ITS is the formally recognized fungal barcode, it sometimes does not distinguish among closely related phylogenetic species (Ruiz and Radwan, 2021). Due to insufficient phylogenetic information and gene-specific noise, one or a few loci usually yield incongruent phylogenies, resulting in several weakly supported nodes.

To advance the understanding of phylogeny on these edible fungi, a deeper sampling of larger and identical gene sets across the genome is required (Binder et al., 2013; Ruiz and Radwan, 2021). Whole-genome sequencing (WGS) has provided a significant benefit in establishing phylogenetic relationships, genetic diversity, virulence-related components, and biotechnological features (Carrillo and Blais, 2021). The

1 https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?id=5097

widespread availability of WGS effectively eliminated data availability as a limiting factor for inferring phylogenetic trees, offering hundreds to thousands of loci for analysis, and eventually replacing molecular phylogenetics with phylogenomics (Nagy and Szollosi, 2017). As the available genomes of *Monascus* spp., the first WGS project (*M. ruber* NRRL 1597, SRR1800507, Illumina HiSeq 2000 platform) was published in 2015 as part of the JGI 1,000 Fungal Genomes Project to represent members of the ascomycete family Monascaceae, while the first sequence assembly was deposited in NCBI database in 2017 (*M. ruber* ASM90018405v1) (Chen et al., 2015), currently up to more than 10 deposits. Only a few have been assembled to the chromosomal level and functionally annotated (Yang et al., 2015).[2] In the research conducted by Yang et al. (2015) they used 2,053 single-copy orthologs across the genome of *M. purpureus* YY-1 and other 17 genomes from the class Eurotiomycetes and Sordariomycetes to create a phylogenetic tree by the maximum-likelihood method, revealing the *M. purpureus* YY-1's close relationship with the family Aspergillaceae. In another study, Higa et al. (2020) compared the biosynthetic gene clusters (BGCs) of MK, citrinin, and MonAzPs using WGS of *M. pilosus* NBRC 4520, *M. purpureus* NBRC 4478, and *M. ruber* NBRC 4483. They found that *M. pilosus* and *M. purpureus* are chemotaxonomically distinct while *M. ruber* has similar secondary metabolite BGCs to *M. pilosus*. The genome-scale approach to microbial taxonomy obviously offers improved resolution, stability, and reliability in evolutionary analyses compared to established methods of identifying physiologically and biochemically changeable and monogenic markers. Unfortunately, no phylogenomics studies on *Monascus* have been conducted within this genus.

On the other hand, WGS enables us to more fully comprehend the intrinsic mechanisms that underlie the phenotypic variations in nutritional profile, pathogenicity, host specificity, secondary metabolite synthesis and so on (San et al., 2019). Genomes within a species or all strains within a clade frequently have a core/conserved component as well as a variable set of genetic material that is referred to as a "pan-genome" among individuals or populations (Barber et al., 2021). The core genome comprises sequences present in all strains and is typically linked to biological functions and key phenotypic traits of the species. The variable/accessory/dispensable genomes contain sequences that are unique to one strain or subset of strains and are related to the adaptability of the species to specific environments or unique biological characteristics, reflecting the characteristics of the individuals. The size of the pan-genome is determined by the effective population size, lifestyle, and niche heterogeneity of the species. The gene pool of a species largely controls its ecological interactions and adaptive capacity, with core and variable genes contributing to the presence-absence variations (PAVs) (Siren et al., 2021). Therefore, a pan-genome of *Monascus* can allow us to better understand the relationship between individual characteristics and genetic variation within this species.

To expand the understanding of the phylogenetic relationships and metabolic divergence in *Monascus* species, this study collected 15 genome assemblies from the genus. They were compared at the whole-genome level using average nucleotide identity (ANI), whole-genome alignment (WGA), PAVs of the pan-genome, species tree inferred from all orthologous gene sequences (STAG), and species tree from single-copy orthologous genes (SCOG). Additionally, this study investigated noteworthy characteristics of *Monascus*, such as secondary metabolite biosynthetic gene clusters (BGCs), carbohydrate active enzymes, secretome, and pathogenicity that occur in this genus' genomes. This approach moves away from a single reference genome that may not necessarily represent the species as a whole, and allows for better understanding of its metabolic versatility, ultimately leading to better management of these food microorganisms.

2. Materials and methods

2.1. Collection of genome assemblies

All available sequenced *Monascus* genomes defined by the taxonomically united genome database in NCBI[3] were collected, resulting in a collection of 15 genomes (July 2022). The collection included a complete assembly (GCA_003184285.1 of *M. purpureus* YY-1) and 14 assemblies labeled as scaffold or contig. **Table 1** summarized several key features for the 15 *Monascus* genomes. The completeness of genome assemblies were assessed by BUSCO (version 5.4.3) (Manni et al., 2021) with a reference set of single-copy orthologs of fungi (fungi_odb10)[4] and default parameters.

2.2. ANI and WGA

Average nucleotide identity is a metric used to compare genetic relatedness of two genomes at the nucleotide level, particularly among strains that belong to the same species or a close phylogenetic clade (Yoon et al., 2017). The ANI values of *Monascus* genomes were calculated using fastANI (version 1.33) (Jain et al., 2018) with the parameter -fragLen set to 500 bp. The resulting ANI matrix was subjected to clustering analysis using the method of compete with squared Euclidean distance metrics in R. Minimap2 is a tool for fast and accurate pairwise alignment of nucleotide sequences. It can align short reads, long reads, assemble contigs, and complete genomes (Li, 2018). To perform assembly level alignments, we used minimap2 with the parameters -c, -cx, and asm5, and visualized the results using the R package pafr.

2.3. Genome annotation

Funannotate (version 1.8.14) (Palmer and Stajich, 2020) was used to perform genome cleaning, FASTA header sorting, repeat sequence masking, and gene prediction on *Monascus* assemblies. The reference protein sequences were collected by integrating the Funannotate protein models and the *Monascus* proteomes from UniProt. The weight of *ab initio* gene predictors, including Augustus, snap, glimmerHMM, and GeneMark-ES/ET, was set to 1:1:8:8.

2 https://www.ncbi.nlm.nih.gov/data-hub/genome/?taxon=5097

3 https://www.ncbi.nlm.nih.gov/genome/?term=Monascus

4 https://busco-data.ezlab.org/v4/data/lineages/

TABLE 1 Genome assembly information of *Monascus* spp. included in this study.

Strain	Accession	Num_contigs	Length (bp)	N50	L50	N90	L90	GC_content (%)
M. purpureus CSU M183	GCA_019320005.1	69	23,752,195	1,018,695	8	272,101	24	49.43
M. purpureus PF1702S	GCA_023624875.1	341	23,230,699	132,158	46	42,757	166	48.99
M. purpureus GB01	GCA_004359145.1	122	24,325,354	327,499	19	91,302	70	48.94
M. purpureus P7048 × 2	GCA_023624895.1	300	23,295,031	145,302	48	44,858	160	49.03
M. purpureus HQ1	GCA_006542485.1	578	23,216,438	90,089	77	25,331	252	49.02
M. purpureus YJX8	GCA_011319195.1	14	24,529,005	3,318,727	3	1,996,896	7	48.86
M. purpureus YY-1	GCA_003184285.1	8	24,147,356	2,821,034	4	2,248,774	7	45.23
M. purpureus RP2	GCA_023935125.1	24	24,403,261	3,297,839	4	975,033	8	48.84
M. ruber KACC 46666	GCA_024449045.1	13	25,909,023	3,185,132	4	1,567,545	9	48.84
M. ruber FWB13	GCA_002976275.1	23	26,287,179	3,364,862	4	1,128,020	9	48.91
M. ruber GA[a]	GCA_900184055.1	198	24,882,894	314,215	24	96,185	74	47.93
M. pilosus YDJ-1	GCA_018806905.1	11	26,137,741	3,368,137	4	2,535,095	7	48.9
M. pilosus YDJ-2	GCA_018806955.1	8	26,144,475	3,479,861	4	2,422,675	7	48.89
M. pilosus K104061	GCA_018806895.1	10	26,125,137	3,364,842	4	2,535,067	7	48.87
M. pilosus MS-1	GCA_018806995.1	11	26,196,030	3,510,661	4	2,260,563	8	48.89

[a]Original record as *Monascus ruber* genome assembly (GCA_900184055.1).

2.4. Construction of *Monascus* pan-genome

The construction of *Monascus* pan-genome was implemented as follows. OrthoFinder (version 2.5.4) (Emms and Kelly, 2019) was used to perform all-against-all sequence similarity searches using Blastp among the predicted protein sequences generated by Funannotate. In the OrthoFinder workflow, orthogroups were generated with Markov Cluster Algorithm clustering method. OrthoFinder output files (Orthogroups folder) were used to extract the pan-genome (the total orthogroups across strains), core genome (orthogroups present at all strains), and accessory genome (orthogroups present at more than one strain but not all). The pan-genome's gene presence-absence variation (PAV) matrix was then subjected to hierarchical clustering in R using the complete method and squared Euclidean distance metrics.

Two strategies were utilized to construct genome-wide phylogenetic trees based on orthologous proteins. The initial approach involved using the STAG algorithm, integrated within OrthoFinder (Emms and Kelly, 2018), with all orthologous protein sequences and default settings. For the second strategy, a species tree was constructed using SCOG protein sequences, which involved the following steps: (1) aligning the sequences in the OrthoFinder output folder (single_copy_orthologous) using muscle (version 5.1) (Edgar, 2004); (2) extracting conserved sequences with Gblocks[5]; (3) converted.fa format to.phy using MEGA (version 7) (Kumar et al., 2016); (4) predicting a suitable amino acid substitution model with ProtTest (version 3) (Darriba et al., 2011) and finally constructing the maximum likelihood phylogenetic tree with RAxML (version 8.2.12) (Stamatakis, 2014).

5 http://phylogeny.lirmm.fr/phylo_cgi/one_task.cgi?task_type=gblocks

2.5. Functional gene annotation

2.5.1. Carbohydrate-active enzymes (CAZymes)

Carbohydrate-active enzymes (Cantarel et al., 2009) describes the families of structurally related catalytic and carbohydrate-binding modules (or functional domains) of enzymes involved in the synthesis and degradation of complex carbohydrates and glycoconjugates. To perform the CAZy annotation, a local dbCAN (Yin et al., 2012) was employed using HMMER (version 3.2.2) searching against a hidden Markov model database of CAZyme domains derived from CDD and CAZy (Potter et al., 2018).

2.5.2. Prediction of secreted proteins

Secreted proteins are proteins exported from the cell by a specific pathway. A secreted protein was defined in this study as a protein with a secretory signal peptide but no transmembrane domains, mitochondrial localization, or chloroplast localization. The prediction of secreted protein was as follows. To predict secreted proteins, we used SignalP6 (Teufel et al., 2022) to identify the presence and location of signal peptides in protein sequences, TMHMM (Sonnhammer et al., 1998) to predict transmembrane domains, and TargetP (Emanuelsson et al., 2007) to determine the subcellular localization of proteins based on their N-terminal sequence features and to provide a potential cleavage site.

2.5.3. Virulence factor annotation

Fungal VFDF database (Lu et al., 2012) describes the virulence factors in fungal pathogens that contains information on the genes, proteins, functions, mechanisms, and pathways of virulence factors from various fungal species that cause human and plant diseases. To perform VFDF prediction, all the proteins in the DFVF database were downloaded and aligned them with the secreted protein sequences using Blastp program.

2.5.4. Prediction of the BGCs of secondary metabolites

The BGCs of citrinin, MK, and the MonAzPs was identified and annotated using the tool antisMASH (Blin et al., 2021). The visualization of genetic cluster was implemented with R package genes.

2.5.5. Gene Ontology (GO) annotation and enrichment analysis

Gene Ontology annotation was performed using the online webtool eggnog-mapper[6] (Cantalapiedra et al., 2021). GO enrichment and visualization were implement in Cytoscape (version 3.8.3) (Shannon et al., 2003) using BiNGO plugin (Maere et al., 2005).

Unless otherwise specified, a 40% sequence identity cutoff was used for protein functional annotation.

3. Results

3.1. Genome information and assessment of assembly quality

Fifteen representative genomic assemblies of the *Monascus* genus were downloaded from the NCBI assembly database[7] and classified into three species, namely, *M. ruber*, *M. pilosus*, and *M. purpureus* (**Table 1**). The sizes of the assemblies ranged from 23.2 to 26.3 Mb, with GC contents varying from 45.23 to 49.43%. The BUSCO evaluation indicated that more than 95% of the 758 single-copy gene homologs were completely assembled in these genomes (**Figure 1**), indicating high quality of genome assembly (Manni et al., 2021).

To reannotate these fungal genome assemblies, the Funannotate script (Palmer and Stajich, 2020), a well-streamlined annotation tool, was used. The resulting annotations (**Table 2**) revealed that each genome encodes 8,103–9,030 genes, with an average protein length of 484.33–495.67 amino acids. Interestingly, the total number of genes encoded by the *M. pilosus* or *M. ruber* genomes was higher than 8,800, more than those of *M. purpureus*.

3.2. ANI analysis

The whole-genome average nucleotide identity (ANI) is a reliable method to determine the genetic relatedness of two genomes and evaluate the boundaries between species with prokaryotic organisms from the same species typically showing 95% ANI among themselves (Jain et al., 2018). Recently, ANI analysis has also been used to assess relationships between eukaryotic genomes, such as yeasts (Lachance et al., 2020), microsporidia (de Albuquerque and Haag, 2023), and plankton species (Delmont et al., 2022). In this study, ANI analysis revealed that the 15 *Monascus* strains could be delineated into two distinct clades, the *M. purpureus* clade and the *M. ruber-M. pilosus* clade (**Figure 1B**). ANI values within each clade were greater than 99.86%, while those from different clades were less than 94.85%. Although 95% ANI is not yet accepted as the species boundary in eukaryotes, an ANI close to 100% suggested a high degree of overall genomic sequence identity and indicates that two genomes share a large proportion of similar DNA sequences.

3.3. Whole-genome alignment (WGA)

In contrast to ANI, which measures similarity of query genome fragments are to their homologous counterparts in the subject genome (Yoon et al., 2017), WGA focuses on predicting evolutionarily related sequence positions and identifying large-scale structural changes, such as duplications and rearrangements (Dewey, 2012). For WGA analysis in this study, the less fragmented genomes from three species with high assembly integrity were chosen, i.e., *M. purpureus* YY-1, *M. pilosus* YDJ2, and *M. ruber* KACC 46666. The remaining WGA analysis outputs were included in the **Supplementary Material 1**. **Figure 2** illustrated the homologous regions that had undergone DNA translocation, rearrangement, or recombination, resulting in them being scrambled or inverted. Missing genome regions were represented by the gaps in the alignments, and **Figures 2B, C** clearly showed chromosome rearrangements. Genomes of *M. ruber* KACC 46666 and *M. pilosus* YDJ2 exhibited an incredibly strong collinearity, as shown in **Figure 2A**, with no significant genomic insertions, conversions, or translocations found in either strain, except for 5 fragment deletions and inversions. Conversely, numerous fragment inversions and DNA translocations were found between *M. purpureus* YY-1 and *M. pilosus* YDJ2, as well as *M. purpureus* YY-1 and *M. ruber* KACC 46666, resulting in low similarity and poor linear match. Distinct chromosomal rearrangement events were detected in chromosomes 1, 2, 5, 6, and 8 between *M. purpureus* YY-1 and *M. ruber* KACC 46666, while chromosome 1 of *M. purpureus* YY-1 was a fusion of the inverted Ct.4 of *M. ruber* KACC 46666 and the translocated Ct.10. Overall, WGA revealed that *M. ruber* KACC 46666 had better collinearity with *M. pilosus* YDJ2, and they were more closely related.

3.4. *Monascus*' pan-genome

The pan-genome of *Monascus* spp. was constructed using OrthoFinder (Emms and Kelly, 2019). All proteins identified from the 15 genomes were used to infer orthologous protein clusters (orthogroups) (**Figure 3A**). With the addition of more genome assemblies, the core genome size decreased while the pan-genome size increased. The "open form" idea of Heap law (Kadiri et al., 2023) was reflected as more orthologous families were included in the analysis (**Figure 3B**). On the whole, the pan-genome encompassed 9,539 orthogroups, of which 6,683 (70.06%) had been converged as the core genome, while the remaining were variable and exhibited the PAV feature, which was an important source of genetic divergence and diversity, as well as having profound effects on phenotypic variations (Wang et al., 2014). Visualization of PAV and hierarchical clustering revealed that the 15 *Monascus* strains could be divided into two major groups, the *M. purpureus* clade

6 http://eggnog-mapper.embl.de/

7 https://www.ncbi.nlm.nih.gov/data-hub/taxonomy/tree/?taxon=5097

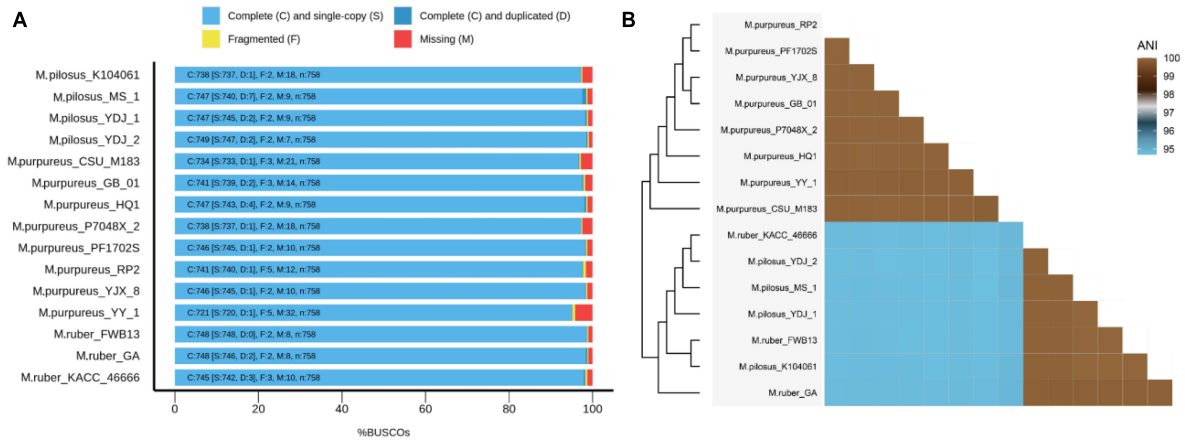


FIGURE 1
BUSCO evaluation on the genome assembly quality **(A)** and ANI-based genomic relatedness among *Monascus* spp. **(B)**.

TABLE 2 *Monascus* annotation information.

Strain	Num_genes	Num_mRNA	Num_tRNA	Avg_gene length	Total exons	Avg_exon length	Avg_protein length
M. purpureus CSU M183	8,498	8,365	133	1603.21	24,434	442.64	488.7
M. purpureus PF1702S	8,391	8,255	136	1608.37	24,303	441.44	490.7
M. purpureus GB01	8,480	8,339	141	1627.85	25,039	438.75	495.4
M. purpureus P7048 × 2	8,377	8,250	127	1629.75	24,845	437.07	495.67
M. purpureus HQ1	8,441	8,320	121	1598.75	24,280	441.17	487.08
M. purpureus YJX8	8,633	8,482	151	1602.5	24,831	442.07	489.54
M. purpureus YY-1	8,103	7,984	119	1629.66	23,986	438.32	491.45
M. purpureus RP2	8,462	8,311	151	1629.64	25,178	437.68	495.51
M. ruber KACC 46666	8,896	8,745	151	1595.31	25,763	436.85	484.79
M. ruber FWB13	9,030	8,867	163	1591.81	26,269	436.52	485.13
M. ruber GA	8,807	8689	118	1602.66	25,568	437.2	485.78
M. pilosus YDJ-1	8,954	8,814	140	1589.74	25,599	440.86	484.33
M. pilosus YDJ-2	8,838	8,704	134	1618.85	25,975	436.7	491.18
M. pilosus K104061	8,931	8,771	160	1592.98	25,610	440.66	486.15
M. pilosus MS-1	8,853	8,705	148	1619.35	26,023	437.84	492.41

and the *M. ruber-M. pilosus* clade (Figure 3C), consistent with ANI-based clustering.

Furthermore, 277 orthogroups were found only in *M. purpureus* strains, while 546 were found only in *M. pilosus-M. ruber* strains marked by asterisk in Figure 3C. Using the unique orthogroups from the *M. ruber-M. pilosus* clade and the *M. purpureus* clade, we conducted GO enrichment analysis on each strain separately. However, the unique orthogroups associated with the *M. purpureus* clade did not enrich into any GO categories (corrected $P > 0.05$), indicating a more stochastic and discrete occurrence of these genes (Supplementary Material 2). In contrast, the *M. ruber-M. pilosus* clade's unique genes were mainly involved in three categories: as shown in Figure 4 (1) Biological processes, such as transport and localization, stimulus response, cellular component organization, and regulation of cellular homeostasis; (2) Molecular functions, such as transport activities; (3) Cellular components, such as plasma membrane. Based on these findings, it can be concluded that the strains from *M. ruber-M. pilosus* clade had a stronger ability to transport and maintain cellular homeostasis than the strains from the *M. purpureus* clade, enabling them to better adapt to changing living environments.

3.5. Phylogenetic analysis at the genome level

The phylogenetic relationships among the 15 strains in *Monascus* species were investigated using two different approaches (Figure 5). *A. oryzae* RIB40 was the only outgroup species included because both *Aspergillus* and *Monascus* belong to the Aspergillaceae family, and more distant outgroup taxa would likely further reduce the percent coverage of orthogroups present

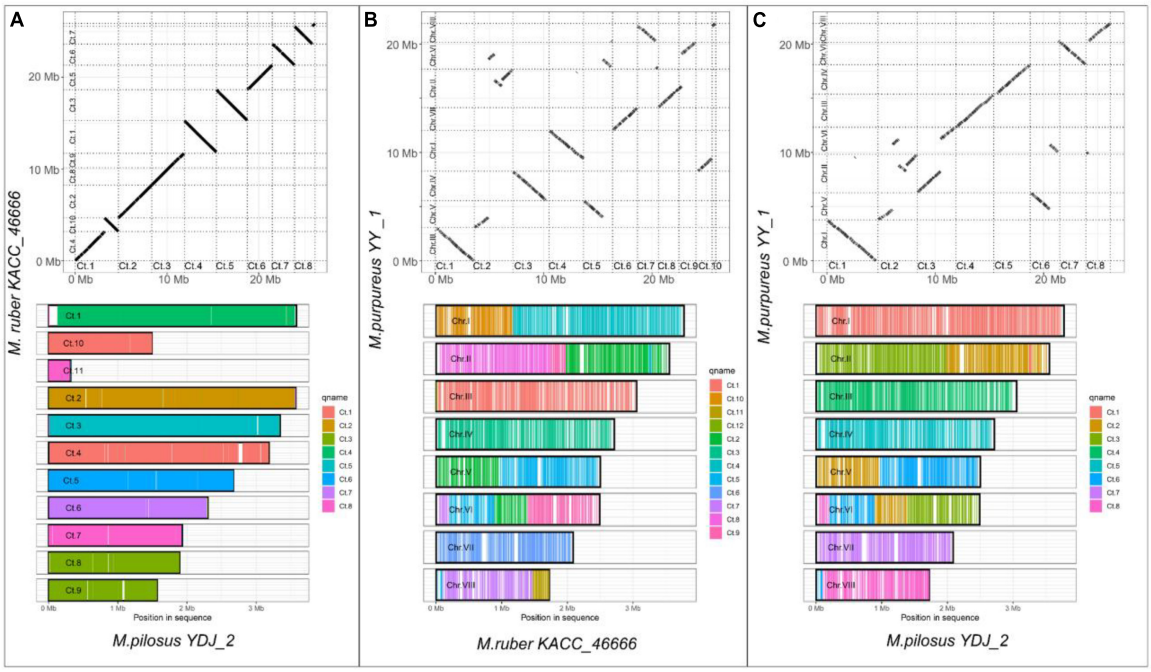


FIGURE 2
Whole-genome alignment on *Monascus ruber* KACC 46666, *M. pilosus* YDJ2 and *M. purpureus* YY1. **(A)** *M. ruber* KACC 46666 vs. *M. pilosus* YDJ2; **(B)** *M. ruber* KACC 46666 vs. *M. purpureus* YY1; **(C)** *M. pilosus* YDJ2 vs. *M. purpureus* YY1. The upper panel is the scatter diagram of genome collinearity and the lower panel describes the chromosome coverage. Other whole-genome alignments are supplied in **Supplementary Material 1**.

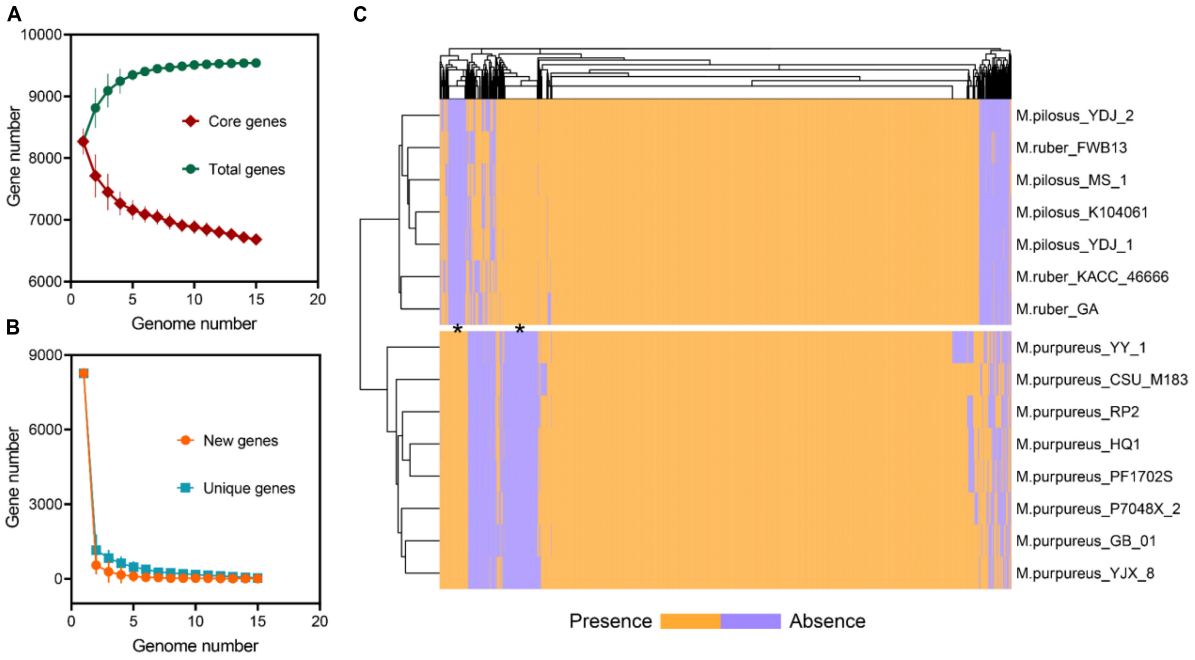


FIGURE 3
Pan-genome of *Monascus spp*. **(A)** Total and core orthogroups along with the increase of *Monascus* genome number; **(B)** new and unique orthogroups; **(C)** visualization of PAVs and hierarchical clustering.

in all species. The first approach, STAG created a rooted phylogenomic tree using OrthoFinder, which utilized 5.565 single- and multi-copy orthologous protein sequences found in all the genomes (**Figure 5A**). The support values for each bipartition in a consensus STAG tree are the proportion of times that the bipartition is seen in each of the individual species tree estimates (Emms and Kelly, 2018). Meanwhile, the second approach, SCOG, employed RAxML to construct another phylogenetic tree using

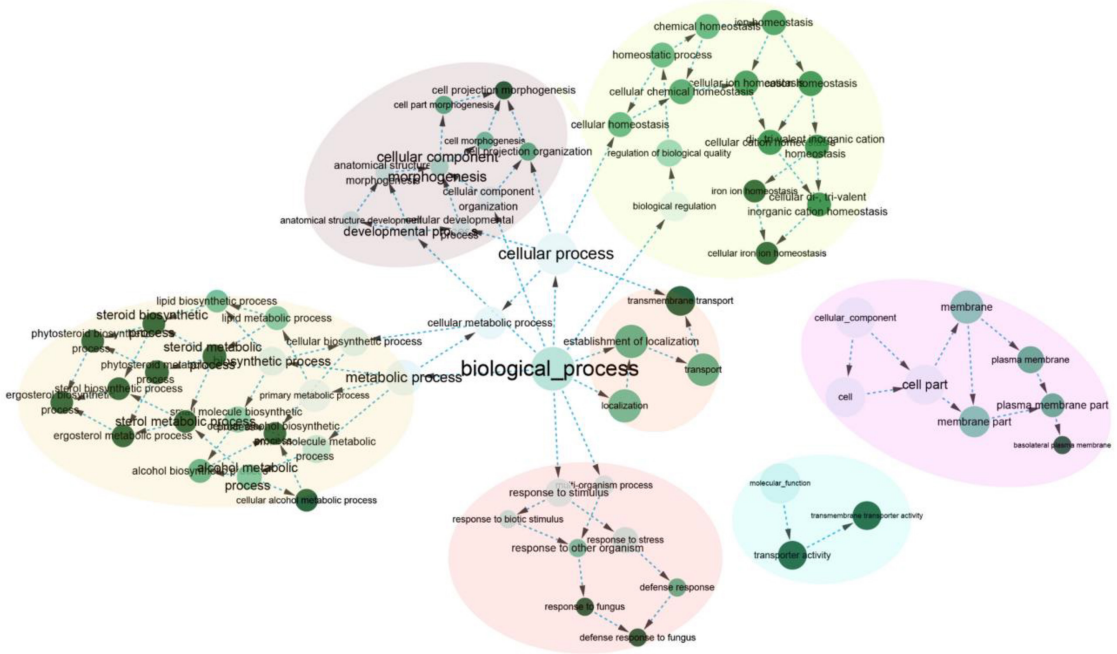


FIGURE 4
GO enrichment of the unique orthogroups in *M. ruber-M. pilosus* clade. The analysis was based on the genome of *M. ruber* FWB13 (GCA_002976275.1) as a representative. Additional GO enrichment outputs of the other strains were available in **Supplementary Material 2**.

4,589 single-copy orthologs (**Figure 5B**) by JTT + I + G + F amino acid substitution model. Node support was estimated with 1,000 bootstrap replicates.

The STAG phylogenetic tree identified two distinct binary clades with a support value of 100%, the *M. purpureus* clade and the *M. ruber-M. pilosus* clade (**Figure 5A**). Similarly, the SCOG tree demonstrated a comparable topology with a 100% support for these two clades (**Figure 5B**), indicating that species trees inferred from all orthologous proteins were as accurate as those inferred from single-copy orthologs. However, the SCOG tree showed significant advantages in supporting sub-branches due to higher support values. In the STAG tree, most of the nodes within both the *M. purpureus* and *M. ruber-M. pilosus* clades had low supports (<70%), which made the topology of the sub-branches unreliable. In contrast, only one node within the *M. purpureus* clade in the SCOG tree was unsound. Moreover, *M. ruber* FWB13 was a unique presence in the *M. ruber-M. pilosus clade*, forming a monophyletic group with four *M. pilosus* species but a paraphyletic group with the other two *M. ruber* species. This topology was also evident in the STAG tree, where *M. ruber* FWB13 occupied a similar phylogenetic position.

Overall, the phylogenomic trees based on these two approaches confirmed the results of ANI and PAV clustering as well as verifying the deduction from genome collinearity analysis.

3.6. Comparative functional genome within the *Monascus* species

To further compare the metabolic divergences among the 15 strains, the protein sequences of all strains were annotated with CAZy, and secretome. And the BGCs of second metabolite in the genome were identified and characterized.

3.6.1. CAZyme annotation

A total of 2,542 coding genes in these 15 genomes were annotated as CAZymes, including 246 auxiliary activity proteins (AAs), 172 carbohydrate-binding modules (CBMs), 15 carbohydrate esterases (CEs), 1,241 glycoside hydrolases (GHs), 846 glycosyltransferases (GTs), and 22 polysaccharide lyases (PLs). Compared to *A. oryzae*, *Monascus* showed a weaker carbohydrate utilization capacity due to the absence or fewer copies of CAZymes. Although *Monascus* species are able to use starch substrates, such as rice, for growth and metabolism, only α-amylase of GH13 family (EC 3.2.1.1, splitting the α-1,4 glycosidic linkages in amylose to yield maltose and glucose) was found, while β-amylases, which cleave β-maltose at the non-reducing end of starch, were absent in all genomes, as shown in **Figure 6**.

Moreover, *M. purpureus* genomes contained more copies of endoglucanases (EC 3.2.1.4, GH5∼GH10) and β-1,3-glucosidase (EC 3.2.1.-, GH132) than *M. ruber* or *M. pilosus*. Endoglucanase is a cellulase family member that has a higher affinity for cellulose and also acts on xylan and mixed β- (1–3, 1–4)-glucan, while β-1,3-glucosidase catalyzes the hydrolysis of β(1→3)-glucosidic linkages in β(1→3)-d-glucan, which is the main constituent of fungal cell walls (Ramos and Malcata, 2011).

Auxiliary activity proteins family members, such as AA4 vanillin oxidase (VAO, EC 1.1.3.38) were only found in the *M. purpureus* genome. VAO is a fungal flavoenzyme that converts a wide range of para-substituted phenols and is the only known fungal member of the 4-phenol oxidizing subgroup of the VAO/PCMH flavoprotein family (Gygli et al., 2018).

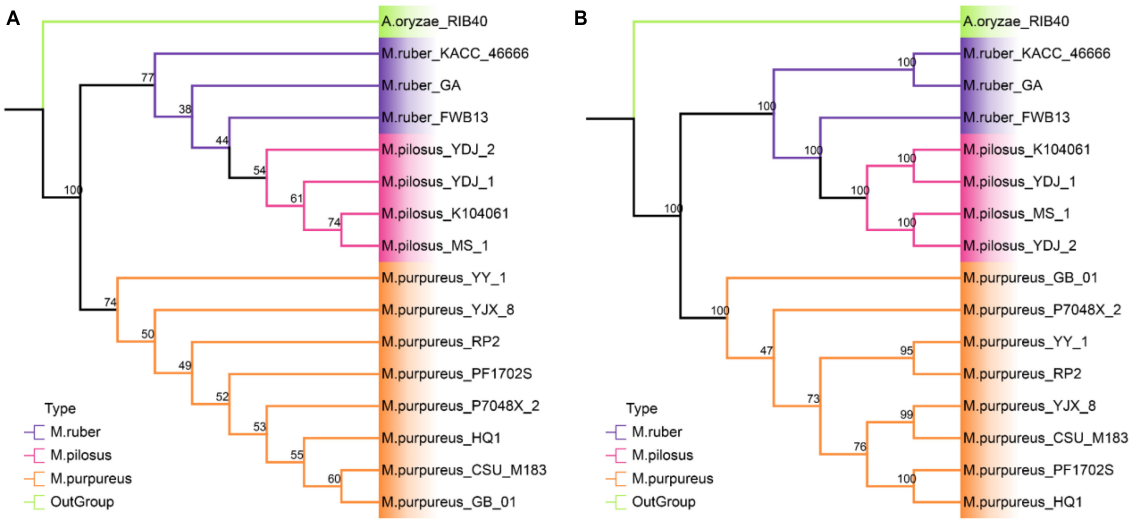


FIGURE 5
Phylogenetic tree at genome level. **(A)** STAG tree with 5565 orthologous protein sequences; **(B)** SCOG tree with 4589 single-copy orthologous sequences.

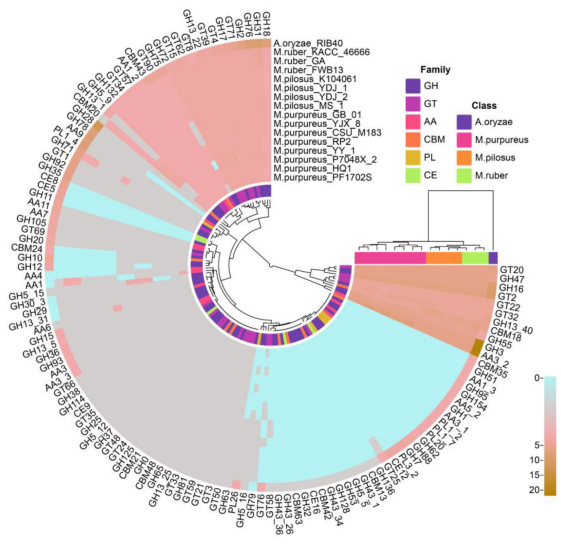


FIGURE 6
CAZY annotation of *Monascus* proteomes. Proteins of sequence identity < 40% were filtered.

In addition, GH5_16, GH78, and GH79 family members were only found in *M. ruber-M. pilosus* genomes. GH5_16 is an endo-1,6-β-galactanase (EC 3.2.1.164) that hydrolyzes 1,6-β-D-galactooligosaccharides with a polymerization degree of more than three and their acidic derivatives with 4-*O*-methylglucosyluronate or glucosyluronate groups at the non-reducing ends (Zhang et al., 2022). GH78 glycoside hydrolases hydrolyze α-L-rhamnosides (EC 3.2.1.40) and degrade flavonoid glycosides that are common in human diets and have important applications in food and medicine industries (O'Neill et al., 2015). GH79 glycoside hydrolases are widely distributed in eukaryotes such as fungi, plants, and animals as well as bacteria and their known members include β-glucuronidase (EC 3.2.1.31), baicalin β-glucuronidase (EC 3.2.1.167), and heparanase (EC 3.2.1.166) (Zhu and Tang, 2021).

3.6.2. *Monascus*' secretome and allergen proteins

The "secretome" refers to the complete collection of proteins secreted by microorganisms that perform various functions such as digestion, signaling, and defense (Caccia et al., 2013). Notably, the top-ranked annotation items for *Monascus* covered carbohydrate transport and metabolism (G), post-translational modification, protein turnover, chaperones (O), and function unknown (S) (**Figure 7A**), including lipase, acid phosphatase, glycoside hydrolases, aspartic-type endopeptidase activity (GO:0004190), and protein hydrolysis serine-type endopeptidase activity (GO:0006508). Some secreted proteins allergenic, such as allergen (orthologous to CADAFLAP00008692), allergen Asp (orthologous to CADAFUBP00002039, Asp F4), allergen Asp F7 (orthologous to V5GFQ9). Of them, CADAFLAP00008692 is a putative allergen from *A. flavus*. Asp F4 is associated with allergic bronchopulmonary aspergillosis (Ramachandran et al., 2004), while Asp F7 from *A. fumigatus* is a peroxiredoxin, a major fungal allergen known for its function as a virulence factor candidate vaccine and reactive oxygen scavenger (Blanco-Ulate et al., 2013). Additionally, all 15 genomes had a defensive secreted β-lactamase, an enzyme that confers resistance to penicillins, cephalosporins, and monobactams (Livermore, 1998).

Given the presence of allergens in secreted proteins and the food safety of *Monascus* products, these proteins were further annotated for fungal pathogens using the DFVF database. It's worth noting that 35 secreted proteins (7.8% of total secretome) in *Monascus* were predicted to be involved in virulence and pathogenicity (**Figure 7B**). Among them, CARP_ASPFU had high identity ≥ 80% present in the genomes of *M. pilosus*, *M. ruber* (except KACC 46666), *M. purpureus* HQ1, *M. purpureus* YY-1, and *M. purpureus* YJX8. This protein is a secreted vacuolar aspartic endopeptidase with broad specificity for peptide bonds

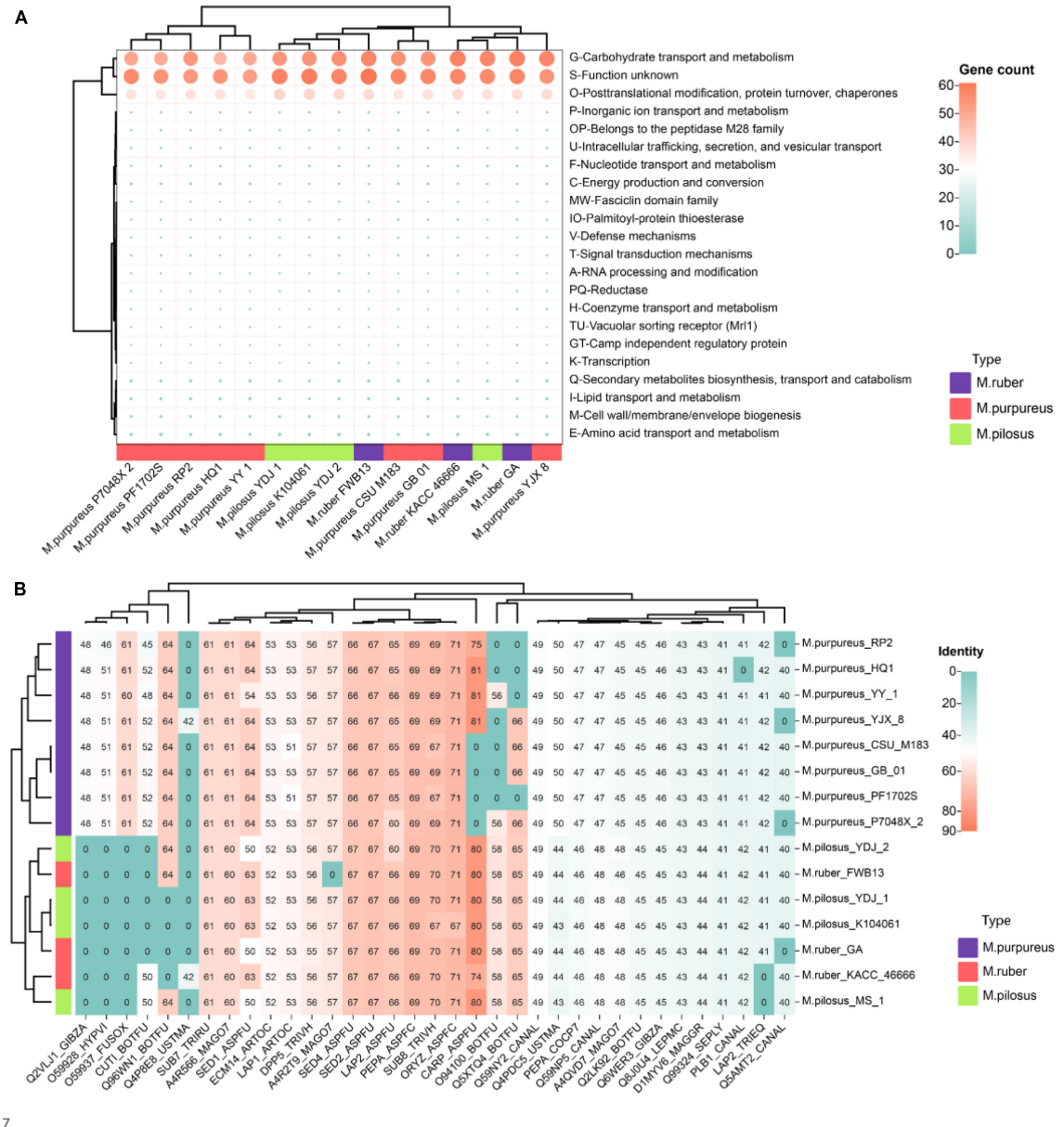


FIGURE 7
Monascus' secretome **(A)** and fungal virulence proteins **(B)**. Proteins of sequence identity < 40% were filtered.

protein hydrolysis, and has an important function in allergen processing and causing various rare human infections such as lung aspergillosis and mycotic keratitis (Jehangir and Ahmad, 2004).

3.6.3. Secondary metabolic synthetic genetic cluster

Monascus azaphilone pigments, MK and citrinin are the most concerned metabolites produced by *Monascus* spp. and their primary structures are synthesized by type I polyketide synthase (T1PKS). According to antiSMASH search results, the BGC responsible for producing MonAzPs was found present in each genome (reference BGC: BGC0000027.1) (Chen et al., 2019; **Figure 8A**). Despite this, the MonAzPs BGC in *M. purpureus* was approximately 56 kb in length, whereas those in *M. pilosus* and *M. ruber* were 65 kb and had seven non-essential genes inserted into the gene cluster. The sequences of the 15 core genes from each BGC did not differ considerably from the reference sequence (identity > 90%, **Supplementary Material 3**).

The reference citrinin BGC (BGC0001338) from *M. ruber* M7 contains 16 genes, including the essential genes *citS* (polyketide synthase), *citA* (serine hydrolase), *citB* [Fe(II) oxidoreductase], *citC* (oxidoreductase), *citD* (aldehyde dehydrogenase), and *citE* (short-chain dehydrogenase) (He and Cox, 2016). In this study, only eight *M. purpureus* strains had the complete citrinin gene cluster sequence (**Figure 8B**). Except for *M. purpureus* HQ1, the citrinin gene cluster size was approximately 42 kb, and the core gene cluster (total length of the key sequences responsible for citrinin synthesis) was around 20 kb. The gene cluster from *M. purpureus* HQ1 was split into two fragments, with *citS-citA-citB-mrr3-citD-mrr5-citE-citC* present in the contig of VIFY01000224.1 and *mrr-8-mrr7-mrr6-mrr5-mrr-4-mrr3-mrr2* in VIFY01000166.1. Nevertheless, the sequence between these eight gene clusters and

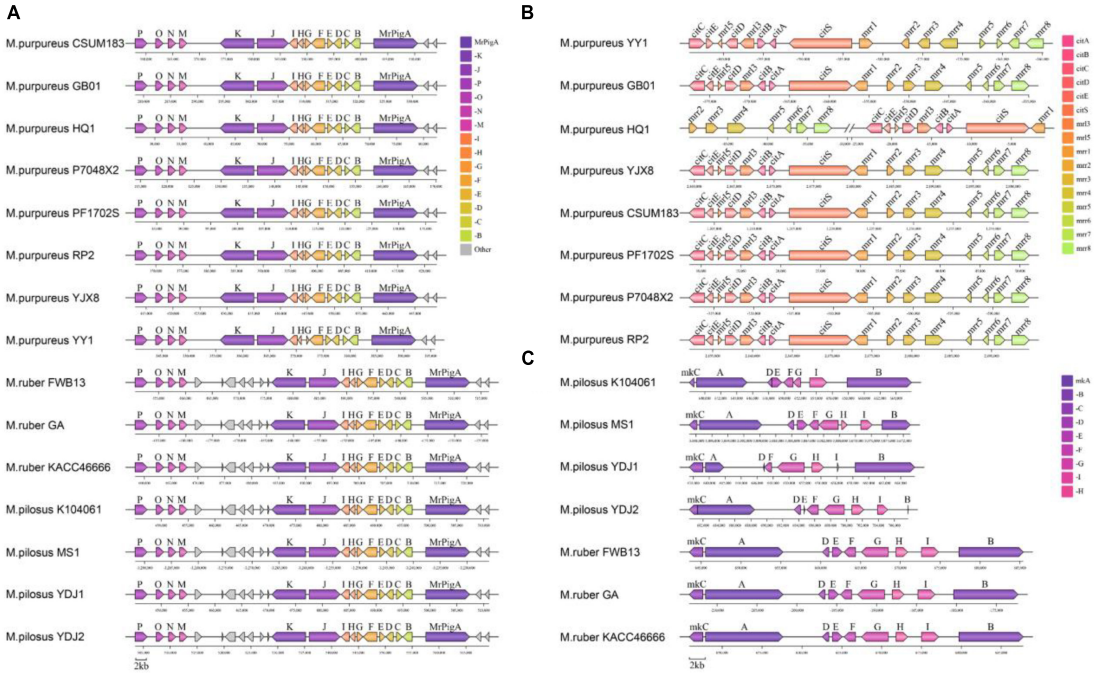


FIGURE 8
Organization of the secondary metabolic synthetic genetic cluster. **(A)** MonAzPs BGCs; **(B)** citrinin BGCs; **(C)** MK BGCs. Negative coordinate values represent the reversed redirection of BGCs.

the reference gene cluster was highly conservative (identity > 95% for pairwise homologous genes). Additionally, several non-core genes, including *mrr2*, *mrr3*, *mrr4*, *mrr5*, *mrr6*, *mrr7*, and *mrr8*, were discovered in *M. ruber* and *M. pilosus* genomes, but all the core genes were absent (**Supplementary Material 3**).

The MK BGC, which typically contains nine genes (*mokA* to *mokI*) identified in BGC0000098 from *M. pilosus* based on sequence similarity (Zhang et al., 2020). In this study, the MK BGC was identified in seven genomes of *M. ruber* and *M. pilosus*, with a size of approximately 41 kb in *M. ruber* and 27 kb in *M. pilosus*. The gene sequences of the BGCs in *M. ruber* FWB13, *M. ruber* KACC 46666, *M. ruber* GA, and *M. pilosus* MS-1 were found to be highly similar to the reference BGC. *M. pilosus* YDJ1, *M. pilosus* YDJ2, and *M. pilosus* K104061 were isolated from commercial MK products and found to be capable of producing MK (Dai et al., 2021), but antiSMASH or BLAST searches revealed incomplete BGC sequences in their genomes. *MokH* was absent in *M. pilosus* K104061, and *MokE* was missing from *M. pilosus* YDJ1's BGC. Furthermore, *MokB* and *MokE* from *M. pilosus* YDJ2, as well as *MokD* and *MokI* from *M. pilosus* YDJ1, were truncated compared to the other six homologous genes. These findings implied that MK synthesis might have diverged in different strains, and additional evidence for critical enzymes in MK synthesis was required in *Monascus*.

4. Discussion

Monascus species are widely found in various habitats such as soil, starch, grain, dried fish, surface sediments of rivers, and roots of pine trees (Mohan Kumari, 2009). Due to the production of MonAzPs and MK, these filamentous fungi are frequently used in food and medicine. However, the taxonomy of *Monascus* has long been a matter of confusion. The phenotype-based identification schemes in *Monascus* were difficult to match with the results obtained by ITS, partial LSU and/or β-tubulin gene sequencing (Patakova, 2013), especially in the case of single-gene locus-based phylogenetics. For example, Dai et al. (2021) inferred a Neighbor-Joining tree using ITS sequences, which revealed that several isolates from *M. pilosus*, *M. fuliginosus*, *M. barkeri*, *M. paxii*, *M. albidulus*, *M. ruber*, *M. purpureus*, and *M. fumeus* were evolutionarily close in the same clade with high bootstrap values. In another clade, several strains from *M. purpureus*, *M. rutilus*, *M. aurantiacus*, and *M. kaoliang* were clustered. Within these two main clades, the strain-level division was still vague due to the low support level.

To overcome inappropriate taxonomy of fungi caused by a single locus, the genealogical concordance phylogenetic species recognition concept (GCPSR) was proposed as an empirical method for recognizing cryptic speciation (Taylor et al., 2000). GCPSR involves sequencing multiple genes that are then combined in phylogenetic analyses. In a report conducted by He et al. (2020) a phylogenetic tree was constructed using concatenated sequences of five protein genes (*BenA*, *CaM*, *RPB2*, *pksKS*, and *MAT1-1*) and two ribosomal RNA genes (ITS and LSU) with a total length of 6,983 bp. Strains from *M. ruber* and *M. purpureus* were clustered into different species clades, respectively, with a high Bayesian analysis/bootstrap percentage. Unfortunately, *M. pilous* was not included in their phylogenetic analysis. In 2017, Patakova divided the *Monascus* spp. into section Floridani and Section Rubri by concatenated phylogeny based on the sequences of ITS + BenA + CaM + LSU + RPB2 with Bayes/RAXML method

(Patakova, 2013). *M. ruber*, *M. pilous* and *M. purpureus* were all in the section Rubri, and *M. purpureus* was located in a separate subclade from *M. ruber* and *M. pilous* with high Bayes/RAXML support. *M. ruber* and *M. pilous* were clustered into the same evolutionary branch.

Recently, modern phylogenetic analyses utilize hundreds to thousands of sites from throughout the genome, which are orders of magnitude larger than traditional sequencing datasets. Thus, the size of these datasets significantly reduces the impact of random error and data availability, making them promising for addressing historically recalcitrant nodes in the tree of life. In this study, using the *Monascus* assembly deposits from the NCBI genome database, a genome-level phylogenetic analysis was conducted. These 15 genomes were predominantly descended from two clades, the *M. purpureus* clade and the *M. ruber-M. pilous* clade, from either the SCOG tree according to the concatenated phylogeny based on 4,589 single copy protein orthologs or the STAG analysis with all the 5,565 single-/multiple- copy protein orthologs. Both of the phylogenomic analysis strategies were reliable with 100% support values for the two major clades. Furthermore, evidence from the nucleic acid level was proposed to support this conclusion, including analyses of average nucleotide identity and genomic collinearity. In particular, there was considerable genomic collinearity between *M. ruber* and *M. pilous*, indicating a high degree of similarity between these two species, rather than with *M. purpureus*. However, genome collinearity analysis, which frequently relies on genome assembly quality, makes it difficult to distinguish between strains when using fragmented genome assemblies. Hence, it is essential to employ a combination of several methods from various perspectives for evaluating the similarity between genomes.

Through identification and characterization of the BGCs of *Monascus*, it was found that all genomes contain the MonAzPs BGCs. However, the citrinin BGCs were only discovered in *M. purpureus*, while the BGCs of MK were only present in *M. pilosus* and *M. ruber*. The production of the mycotoxin citrinin, was originally described in *M. purpureus* and *M. ruber* in Blanc et al. (1995). It was subsequently shown that the *M. ruber* used in that study was, in fact, *M. purpureus* (Chen et al., 2008) because the *pksCT* gene for citrinin polyketide synthase was only present in *M. purpureus* and *M. kaoliang* (a synonym for *M. purpureus*), but not in *M. pilosus*, *M. ruber*, *M. floridanus*, *M. sanguineus*, *M. barkeri*, or *M. lunisporas*. Despite this, citrinin production in *M. ruber* was later demonstrated by other authors (Li et al., 2010, 2021). However, whether or not this was due to incorrect strain classification requires additional validation. Another possibility is that this strain-specific citrinin synthesis might originate from a horizontal gene transfer of the BGC among fungi, because the citrinin pathway belongs to the general pathway shared by many *Penicillium*, *Aspergillus*, and *Monascus* species (Higa et al., 2020).

Additionally, when compared to the reference BGCs, the BGC sequences of MonAzPs and citrinin revealed the highest degree of conservation. Interestingly, seven additional inserted genes in the MonAzPs BGCs could well differentiate *M. purpureus* apart from *M. pilosus* or *M. ruber*. Furthermore, the BGCs of MK were more varied in *M. pilosus* but were conserved in *M. ruber*. Although this suggested that the three strain classes might be distinguished by divergence from the MK gene cluster sequence, further sequencing evidence from more strains is necessary to support this.

According to the findings of this study, the investigated *Monascus* species can be classified into two groups: the *M. pilosus-M. ruber* clade and the *M. purpureus* clade. This classification may have significant implications in *Monascus*-related industries. Typically, commercial *Monascus* products are divided into two categories, those intended for the production of MonAzPs for food coloring, and those for the production of MK, a secondary metabolite to lower cholesterol and treat hypolipidemia. Among them, *M. purpureus* is a prominent red-colored mold species (Yang et al., 2015), but a number of strains including *M. purpureus*, *M. pilosus, M. sanguineus* and *M. ruber* were reported to produce MK (Wen et al., 2020). This is confusing because it is unclear whether the synthesis of MK is due to individual differences among *Monascus* strains or incorrect classification, as the genome of all *M. purpureus* strains used in this study lacked the complete gene cluster for MK biosynthesis. Therefore, further identification of these strains at the phenotypic and genotypic levels is necessary. For example, in a recent study by Higa et al. (2020) the metabolite analysis based on liquid chromatography-mass spectrometry revealed significant differences in the MonAzPs and related metabolites produced by the three species (*M. pilosus*, *M. ruber*, and *M. purpureus*) in liquid media, despite *M. ruber* had similar biosynthetic and secondary metabolite BGCs to *M. pilosus*.

Moreover, genome-level analysis provides insights into metabolic differences among individual strains. Comparative genomic analyses showed that the genome size of *M. purpureus* was smaller than that of *M. pilosus/M. ruber* and had undergone significant gene losses, particularly in cellular transport and maintaining homeostasis, as a result of specialized adaptation to the environment. Compared to *A. oryzae*, *Monascus* also displayed gene losses in both enzyme species and quantities involved in carbohydrate metabolism, which might be due to strain degradation resulting from prolonged domestication of the starch-rich matrix (Dufossé, 2018). Furthermore, *Monascus*' genomes were predicted to contain a number of secreted proteins that could act as allergens or be involved in virulence and pathogenicity, although they exhibit interindividual variability. One such protein is CARP_ASPFU, which is present in most of the genomes and has a high identity $\geq$ 80%, and may play a crucial role in processing or signaling to allergens, causing rare human infections such as lung aspergillosis and mycotic keratitis. Given the importance of food safety, it is necessary to confirm whether the toxins produced by particular *Monascus* strains are actually produced by gene expression or just exist in the genome.

5. Conclusion

In conclusion, *Monascus* is a widely consumed commodity strain due to its good coloring and health care efficacy. However, its taxonomic profile remains under debate and is difficult to classify using a single locus sequence alignment and comparison, as is the case with other fungi. This study utilized phylogenomics, which draws information from comparing entire or large portions of genomes, to gain more detailed insights into evolutionary relationships, as compared to traditional phylogenetic methods that rely on a smaller number of genetic markers. Our findings clearly demonstrate differences between *M. purpureus* and

M. pilosus/*M. ruber*, as well as a high degree of similarity between the genomes of *M. pilosus* and *M. ruber*. By comparing the genomes of different *Monascus* strains, we were able to identify differences in metabolism for environmental adaptation, carbohydrate-active enzymes, secretome, fungal pathogens, as well as in secondary metabolite gene clusters. Genome-level research provides insights into the species classification of *Monascus*, and as the cost of genome sequencing decreases, this high-resolution phylogenetic method will become an important means for evaluating the safety of *Monascus* and other edible fungi.

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

ZZ, HL, and XL conceived and designed the study. ZZ, MC, PC, JL, ZM, YM, ZL, QG, and CW performed the data analysis. All authors wrote the manuscript and approved the final manuscript.

Funding

This research was supported by the National Key Research and Development Program of China (No. 2020YFA0908300), Tianjin Synthetic Biotechnology Innovation Capacity Improvement Projects (TSBICIP-PTJS-001, TSBICIP-PTJJ-007, and TSBICIP-KJGG-006), Innovation Fund of Haihe Laboratory of Synthetic Biology (No. 22HHSWSS00021), and the National Natural Science Foundation of China (No. 31800072).

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2023.1199144/full#supplementary-material

SUPPLEMENTARY MATERIAL 1
Whole-genome alignment on all the genomes from *Monascus* spp.

SUPPLEMENTARY MATERIAL 2
GO enrichment of the unique orthogroups in each *Monascus* genome.

SUPPLEMENTARY MATERIAL 3
Secondary metabolite synthetic genetic clusters in *Monascus* genome.

References

Barber, A. E., Sae-Ong, T., Kang, K., Seelbinder, B., Li, J., Walther, G., et al. (2021). *Aspergillus fumigatus* pan-genome analysis identifies genetic variants associated with human infection. *Nat. Microbiol.* 6, 1526–1536. doi: 10.1038/s41564-021-00993-x

Barbosa, R. N., Leong, S. L., Vinnere-Pettersson, O., Chen, A. J., Souza-Motta, C. M., Frisvad, J. C., et al. (2017). Phylogenetic analysis of *Monascus* and new species from honey, pollen and nests of stingless bees. *Stud. Mycol.* 86, 29–51.

Binder, M., Justo, A., Riley, R., Salamov, A., Lopez-Giraldez, F., Sjokvist, E., et al. (2013). Phylogenetic and phylogenomic overview of the *Polyporales*. *Mycologia* 105, 1350–1373. doi: 10.3852/13-003

Blanc, P. J., Laussac, J. P., Le Bars, J., Le Bars, P., Loret, M. O., Pareilleux, A., et al. (1995). Characterization of monascidin A from *Monascus* as citrinin. *Int. J. Food Microbiol.* 27, 201–213. doi: 10.1016/0168-1605(94)00167-5

Blanco-Ulate, B., Rolshausen, P., and Cantu, D. (2013). Draft genome sequence of *Neofusicoccum parvum* isolate ucr-np2, a fungal vascular pathogen associated with grapevine cankers. *Genome Announc.* 1:e00339-13. doi: 10.1128/genomeA.00339-13

Blin, K., Shaw, S., Kloosterman, A. M., Charlop-Powers, Z., van Wezel, G. P., Medema, M. H., et al. (2021). antiSMASH 6.0: Improving cluster detection and comparison capabilities. *Nucleic Acids Res.* 49, W29–W35. doi: 10.1093/nar/gkab335

Caccia, D., Dugo, M., Callari, M., and Bongarzone, I. (2013). Bioinformatics tools for secretome analysis. *Biochim. Biophys. Acta* 1834, 2442–2453. doi: 10.1016/j.bbapap.2013.01.039

Cantalapiedra, C. P., Hernandez-Plaza, A., Letunic, I., Bork, P., and Huerta-Cepas, J. (2021). eggNOG-mapper v2: Functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol. Biol. Evol.* 38, 5825–5829. doi: 10.1093/molbev/msab293

Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009). The carbohydrate-active EnZymes database (CAZy): An expert resource for glycogenomics. *Nucleic Acids Res.* 37, D233–D238. doi: 10.1093/nar/gkn663

Carrillo, C., and Blais, B. (2021). Whole-genome sequence datasets: A powerful resource for the food microbiology laboratory toolbox. *Front. Sustain. Food Syst.* 5:754988. doi: 10.3389/fsufs.2021.754988

Chen, W., Feng, Y., Molnar, I., and Chen, F. (2019). Nature and nurture: confluence of pathway determinism with metabolic and chemical serendipity diversifies *Monascus* azaphilone pigments. *Nat. Prod. Rep.* 36, 561–572. doi: 10.1039/c8np00060c

Chen, W., He, Y., Zhou, Y., Shao, Y., Feng, Y., Li, M., et al. (2015). Edible filamentous fungi from the species *Monascus*: Early traditional fermentations, modern molecular biology, and future genomics. *Compreh. Rev. Food Sci. Food Saf.* 14, 555–567. doi: 10.1111/1541-4337.12145

Chen, Y. P., Tseng, C. P., Chien, I. L., Wang, W. Y., Liaw, L. L., and Yuan, G. F. (2008). Exploring the distribution of citrinin biosynthesis related genes among *Monascus* species. *J. Agric. Food Chem.* 56, 11767–11772. doi: 10.1021/jf802371b

Dai, W., Shao, Y., and Chen, F. (2021). Production of monacolin K in *Monascus pilosus*: Comparison between industrial strains and analysis of its gene clusters. *Microorganisms* 9:747. doi: 10.3390/microorganisms9040747

Darriba, D., Taboada, G. L., Doallo, R., and Posada, D. (2011). ProtTest 3: Fast selection of best-fit models of protein evolution. *Bioinformatics* 27, 1164–1165. doi: 10.1093/bioinformatics/btr088

de Albuquerque, N. R. M., and Haag, K. L. (2023). Using average nucleotide identity (ANI) to evaluate microsporidia species boundaries based on their genetic relatedness. *J. Eukaryot. Microbiol.* 70, e12944. doi: 10.1111/jeu.12944

Delmont, T. O., Gaia, M., Hinsinger, D. D., Fremont, P., Vanni, C., Fernandez-Guerra, A., et al. (2022). Functional repertoire convergence of distantly related eukaryotic plankton lineages abundant in the sunlit ocean. *Cell Genom.* 2:100123. doi: 10.1016/j.xgen.2022.100123

Dewey, C. (2012). "Whole-Genome Alignment," in *Evolutionary genomics: Statistical and computational methods*, ed. M. Anisimova (Totowa, NJ: Humana Press).

Dufossé, L. (2018). "Microbial pigments from bacteria, yeasts, fungi, and microalgae for the food and feed industries," in *Natural and Artificial Flavoring Agents and Food Dyes*, eds A. M. Grumezescu and A. M. Holban (Cambridge, MA: Academic Press).

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

Emanuelsson, O., Brunak, S., von Heijne, G., and Nielsen, H. (2007). Locating proteins in the cell using TargetP, SignalP and related tools. *Nat. Protoc.* 2, 953–971. doi: 10.1038/nprot.2007.131

Emms, D. M., and Kelly, S. (2018). STAG: Species tree inference from all genes. *bioRxiv* [Preprint]. doi: 10.1101/267914

Emms, D. M., and Kelly, S. (2019). OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biol.* 20:238. doi: 10.1186/s13059-019-1832-y

Feng, Y., Chen, W., and Chen, F. A. (2016). *Monascus pilosus* MS-1 strain with high-yield monacolin K but no citrinin. *Food Sci. Biotechnol.* 25, 1115–1122. doi: 10.1007/s10068-016-0179-3

Ghosh, S., and Dam, B. (2020). Genome shuffling improves pigment and other bioactive compound production in *Monascus purpureus*. *Appl. Microbiol. Biotechnol.* 104, 10451–10463. doi: 10.1007/s00253-020-10987-0

Gygli, G., de Vries, R. P., and van Berkel, W. J. H. (2018). On the origin of vanillyl alcohol oxidases. *Fungal Genet. Biol.* 116, 24–32. doi: 10.1016/j.fgb.2018.04.003

He, Y., and Cox, R. J. (2016). The molecular steps of citrinin biosynthesis in fungi. *Chem. Sci.* 7, 2119–2127. doi: 10.1039/c5sc04027b

He, Y., Liu, J., Chen, Q., Gan, S., Sun, T., and Huo, S. (2020). *Monascus sanguineus* may be a natural nothospecies. *Front. Microbiol.* 11:614910. doi: 10.3389/fmicb.2020.614910

Higa, Y., Kim, Y. S., Altaf-Ul-Amin, M., Huang, M., Ono, N., and Kanaya, S. (2020). Divergence of metabolites in three phylogenetically close *Monascus* species (*M. pilosus*, *M. ruber*, and *M. purpureus*) based on secondary metabolite biosynthetic gene clusters. *BMC Genomics* 21:679. doi: 10.1186/s12864-020-06864-9

Jain, C., Rodriguez, R. L., Phillippy, A. M., Konstantinidis, K. T., and Aluru, S. (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat. Commun.* 9:5114. doi: 10.1038/s41467-018-07641-9

Jehangir, M., and Ahmad, S. F. (2004). Structural studies of aspartic endopeptidase pep2 from *Neosartorya fisherica* using homolgy modeling techniques. *Int. J. Bioinform. Biosci.* 3, 7–20. doi: 10.5121/ijbb.2013.310

Jia, X. Q., Xu, Z. N., Zhou, L. P., and Sung, C. K. (2010). Elimination of the mycotoxin citrinin production in the industrial important strain *Monascus purpureus* SM001. *Metab. Eng.* 12, 1–7. doi: 10.1016/j.ymben.2009.08.003

Kadiri, M., Sevugapperumal, N., Nallusamy, S., Ragunathan, J., Ganesan, M. V., Alfarraj, S., et al. (2023). Pan-genome analysis and molecular docking unveil the biocontrol potential of *Bacillus velezensis* VB7 against *Phytophthora infestans*. *Microbiol. Res.* 268:127277. doi: 10.1016/j.micres.2022.127277

Kalaivani, M., and Rajasekaran, A. (2014). Improvement of monacolin K/citrinin production ratio in *Monascus purpureus* using UV mutagenesis. *Nutrafoods* 13, 79–84.

Kang, B., Zhang, X., Wu, Z., Wang, Z., and Park, S. (2014). Production of citrinin-free *Monascus* pigments by submerged culture at low pH. *Enzyme Microb. Technol.* 55, 50–57. doi: 10.1016/j.enzmictec.2013.12.007

Karthikeyan, M., and Dharumadurai, D. (2023). "Production and entrepreneurship plan for red pigment from *Monascus sp.*," in *Food Microbiology Based Entrepreneurship*, eds N. Amaresan, D. Dharumadurai, and O. O. Babalola (Singapore: Springer Nature Singapore).

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

Lachance, M. A., Lee, D. K., and Hsiang, T. (2020). Delineating yeast species with genome average nucleotide identity: a calibration of ANI with haplontic, heterothallic *Metschnikowia* species. *Antonie Van Leeuwenhoek.* 113, 2097–2106. doi: 10.1007/s10482-020-01480-9

Li, H. (2018). Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* 34, 3094–3100. doi: 10.1093/bioinformatics/bty191

Li, L., Xu, N., and Chen, F. (2021). Inactivation of *mrpigH* gene in *Monascus ruber* M7 results in increased *Monascus* pigments and decreased citrinin with *mrpyrG* selection marker. *J. Fungi* 7:1094. doi: 10.3390/jof7121094

Li, Y., Image, I., Xu, W., Image, I., Tang, Y., and Image, I. (2010). Classification, prediction, and verification of the regioselectivity of fungal polyketide synthase product template domains. *J. Biol. Chem.* 285, 22764–22773. doi: 10.1074/jbc.M110.128504

Livermore, D. M. (1998). Beta-lactamase-mediated resistance and opportunities for its control. *J. Antimicrob. Chemother.* 41(Suppl D), 25–41. doi: 10.1093/jac/41.suppl_4.25

Lu, T., Yao, B., and Zhang, C. (2012). DFVF: Database of fungal virulence factors. *Database.* 2012:bas032. doi: 10.1093/database/bas032

Maere, S., Heymans, K., and Kuiper, M. (2005). BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* 21, 3448–3449. doi: 10.1093/bioinformatics/bti551

Manni, M., Berkeley, M. R., Seppey, M., and Zdobnov, E. M. (2021). BUSCO: Assessing genomic data quality and beyond. *Curr Protoc.* 1:e323.

Mohan Kumari, H. (2009). *Monascus purpureus in relation to statin and sterol production and mutational analysis*. Ph.D. thesis. Mysore: Central Food Technological Research Institute.

Nagy, L. G., and Szollosi, G. (2017). Fungal phylogeny in the age of genomics: Insights into phylogenetic inference from genome-scale datasets. *Adv. Genet.* 100, 49–72. doi: 10.1016/bs.adgen.2017.09.008

Ning, Z. Q., Cui, H., Xu, Y., Huang, Z. B., Tu, Z., and Li, Y. P. (2017). Deleting the citrinin biosynthesis-related gene, ctnE, to greatly reduce citrinin production in *Monascus aurantiacus* Li AS3.4384. *Int. J. Food Microbiol.* 241, 325–330. doi: 10.1016/j.ijfoodmicro.2016.11.004

O'Neill, E. C., Stevenson, C. E., Paterson, M. J., Rejzek, M., Chauvin, A. L., Lawson, D. M., et al. (2015). Crystal structure of a novel two domain GH78 family alpha-rhamnosidase from *Klebsiella oxytoca* with rhamnose bound. *Proteins* 83, 1742–1749. doi: 10.1002/prot.24807

Ouyang, W., Liu, X., Wang, Y., Huang, Z., and Li, X. (2021). Addition of genistein to the fermentation process reduces citrinin production by *Monascus* via changes at the transcription level. *Food Chem.* 343:128410. doi: 10.1016/j.foodchem.2020.128410

Palmer, J., and Stajich, J. (2020). Funannotate v1. 8.1: Eukaryotic genome annotation. *Zenodo* 2020:4054262. doi: 10.5281/zenodo.4054262

Park, H., Stamenova, E., and Jong, S. (2004). Phylogenetic relationships of *Monascus* species inferred from the ITS and the partial b-tubulin gene. *Bot. Bull. Acad. Sin.* 45:325. doi: 10.7016/BBAS.200410.0325

Park, H. G., and Jong, S. (2003). Molecular characterization of *Monascus* strains based on the D1/D2 regions of *LSU* rRNA genes. *Mycoscience* 44, 25–32. doi: 10.1007/s10267-002-0077-9

Patakova, P. (2013). *Monascus* secondary metabolites: production and biological activity. *J. Ind. Microbiol. Biotechnol.* 40, 169–181. doi: 10.1007/s10295-012-1216-8

Potter, S. C., Luciani, A., Eddy, S. R., Park, Y., Lopez, R., and Finn, R. D. (2018). HMMER web server: 2018 update. *Nucleic Acids Res.* 46, W200–W204. doi: 10.1093/nar/gky448

Ramachandran, H., Banerjee, B., Greenberger, P. A., Kelly, K. J., Fink, J. N., and Kurup, V. P. (2004). Role of C-terminal cysteine residues of *Aspergillus fumigatus* allergen Asp f 4 in immunoglobulin E binding. *Clin. Diagn. Lab. Immunol.* 11, 261–265. doi: 10.1128/cdli.11.2.261-265.2004

Ramos, O. S., and Malcata, F. X. (2011). "Food-grade enzymes," in *Comprehensive Biotechnology*, ed. M. Moo-Young (Burlington: Academic Press).

Ruiz, O. N., and Radwan, O. (2021). Difficulty in assigning fungal identity based on DNA sequences. *Microbiol. Resour. Announc.* 10:e0046021.

San, J. E., Baichoo, S., Kanzi, A., Moosa, Y., Lessells, R., Fonseca, V., et al. (2019). Current affairs of microbial genome-wide association studies: Approaches, bottlenecks and analytical pitfalls. *Front. Microbiol.* 10:3119. doi: 10.3389/fmicb.2019.03119

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

Shao, Y., Lei, M., Mao, Z., Zhou, Y., and Chen, F. (2014). Insights into *Monascus* biology at the genetic level. *Appl. Microbiol .Biotechnol.* 98, 3911–3922. doi: 10.1007/s00253-014-5608-8

Shao, Y., Xu, L., and Chen, F. (2011). Genetic diversity analysis of *Monascus* strains using SRAP and ISSR markers. *Mycoscience* 52, 224–233. doi: 10.1007/s10267-010-0087-y

Siren, J., Monlong, J., Chang, X., Novak, A. M., Eizenga, J. M., Markello, C., et al. (2021). Pangenomics enables genotyping of known structural variants in 5202 diverse genomes. *Science* 374:abg8871. doi: 10.1126/science.abg8871

Sonnhammer, E. L., von Heijne, G., and Krogh, A. (1998). A hidden Markov model for predicting transmembrane helices in protein sequences. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* 6, 175–182.

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

Taylor, J. W., Jacobson, D. J., Kroken, S., Kasuga, T., Geiser, D. M., Hibbett, D. S., et al. (2000). Phylogenetic species recognition and species concepts in fungi. *Fungal Genet. Biol.* 31, 21–32. doi: 10.1006/fgbi.2000.1228

Teufel, F., Almagro Armenteros, J. J., Johansen, A. R., Gislason, M. H., Pihl, S. I., Tsirigos, K. D., et al. (2022). SignalP 6.0 predicts all five types of signal peptides using protein language models. *Nat. Biotechnol.* 40, 1023–1025. doi: 10.1038/s41587-021-01156-3

Tong, A., Lu, J., Huang, Z., Huang, Q., Zhang, Y., Farag, M. A., et al. (2022). Comparative transcriptomics discloses the regulatory impact of carbon/nitrogen fermentation on the biosynthesis of *Monascus kaoliang* pigments. *Food Chem. X.* 13:100250. doi: 10.1016/j.fochx.2022.100250

Wang, J., Huang, Y., and Shao, Y. (2021). From traditional application to genetic mechanism: Opinions on *Monascus* research in the new milestone. *Front. Microbiol.* 12:659907. doi: 10.3389/fmicb.2021.659907

Wang, Y., Lu, J., Chen, S., Shu, L., Palmer, R. G., Xing, G., et al. (2014). Exploration of presence/absence variation and corresponding polymorphic markers in soybean genome. *J. Integr. Plant Biol.* 56, 1009–1019. doi: 10.1111/jipb.12208

Wen, Q., Cao, X., Chen, Z., Xiong, Z., Liu, J., Cheng, Z., et al. (2020). An overview of *Monascus* fermentation processes for monacolin K production. *Open Chem.* 18, 10–21. doi: 10.1515/chem-2020-0006

Yang, Y., Liu, B., Du, X., Li, P., Liang, B., Cheng, X., et al. (2015). Complete genome sequence and transcriptomics analyses reveal pigment biosynthesis and regulatory mechanisms in an industrial strain, *Monascus purpureus* YY-1. *Sci. Rep.* 5:8331. doi: 10.1038/srep08331

Yin, Y., Mao, X., Yang, J., Chen, X., Mao, F., and Xu, Y. (2012). dbCAN: A web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* 40, W445–W451. doi: 10.1093/nar/gks479

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

Zhang, X., Wang, Y., Liu, J., Wang, W., Yan, X., Zhou, Y., et al. (2022). Cloning, expression, and characterization of endo-beta-1,6-galactanase PoGal30 from *Penicillium oxalicum*. *Appl. Biochem. Biotechnol.* 194, 6021–6036. doi: 10.1007/s12010-022-04093-2

Zhang, Y., Chen, Z., Wen, Q., Xiong, Z., Cao, X., Zheng, Z., et al. (2020). An overview on the biosynthesis and metabolic regulation of monacolin K/lovastatin. *Food Funct.* 11, 5738–5748. doi: 10.1039/d0fo00691b

Zhu, X., and Tang, S. (2021). Enzymatic properties of alpha-L-rhamnosidase and the factors affecting its activity: A review. *Chin. J. Biotechnol.* 37, 2623–2632. doi: 10.13345/j.cjb.200565

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 05 July 2023
DOI 10.3389/fmicb.2023.1217727

OPEN ACCESS

EDITED BY
Utkarsh Sood,
University of Delhi, India

REVIEWED BY
Mani Abdul Karim,
Periyar University, India
Dejian Yu,
Nanjing Audit University, China

*CORRESPONDENCE
Rachid Daoud
✉ rachid.daoud@um6p.ma
Achraf El Allali
✉ achraf.elallali@um6p.ma

RECEIVED 05 May 2023
ACCEPTED 20 June 2023
PUBLISHED 05 July 2023

CITATION
Ascandari A, Aminu S, Safdi NEH, El Allali A and Daoud R (2023) A bibliometric analysis of the global impact of metaproteomics research. *Front. Microbiol.* 14:1217727. doi: 10.3389/fmicb.2023.1217727

A bibliometric analysis of the global impact of metaproteomics research

AbdulAziz Ascandari[1], Suleiman Aminu[1,2], Nour El Houda Safdi[1], Achraf El Allali[1]* and Rachid Daoud[1]*

[1]African Genome Center, Mohammed VI Polytechnic University, Ben Guerir, Morocco, [2]Department of Biochemistry, Ahmadu Bello University, Zaria, Nigeria

Background: Metaproteomics is a subfield in meta-omics that is used to characterize the proteome of a microbial community. Despite its importance and the plethora of publications in different research area, scientists struggle to fully comprehend its functional impact on the study of microbiomes. In this study, bibliometric analyses are used to evaluate the current state of metaproteomic research globally as well as evaluate the specific contribution of Africa to this burgeoning research area. In this study, we use bibliometric analyses to evaluate the current state of metaproteomic research globally, identify research frontiers and hotspots, and further predict future trends in metaproteomics. The specific contribution of Africa to this research area was evaluated.

Methods: Relevant documents from 2004 to 2022 were extracted from the Scopus database. The documents were subjected to bibliometric analyses and visualization using VOS viewer and Biblioshiny package in R. Factors such as the trends in publication, country and institutional cooperation networks, leading scientific journals, author's productivity, and keywords analyses were conducted. The African publications were ranked using Field-Weighted Citation Impact (FWCI) scores.

Results: A total of 1,138 documents were included and the number of publications increased drastically from 2004 to 2022 with more publications (170) reported in 2021. In terms of publishers, *Frontiers in Microbiology* had the highest number of total publications (62). The United States of America (USA), Germany, China, and Canada, together with other European countries were the most productive. Institution-wise, the Helmholtz Zentrum für Umweltforschung, Germany had more publications while Max Plank Institute had the highest total collaborative link strength. Jehmlich N. was the most productive author whereas Hettich RL had the highest h-index of 63. Regarding Africa, only 2.2% of the overall publications were from the continent with more publication outputs from South Africa. More than half of the publications from the continent had an FWCI score≥1.

Conclusion: The scientific outputs of metaproteomics are rapidly evolving with developed countries leading the way. Although Africa showed prospects for future progress, this could only be accelerated by providing funding, increased collaborations, and mentorship programs.

KEYWORDS

metaproteomics, metagenomics, microbiome, bibliometric analyses, field weighted citation impact

1. Introduction

The development of "omics" technologies has revolutionized the field of molecular biology by enabling the analysis of biological systems at various molecular levels. The omics approaches which include genomics, transcriptomics, proteomics, and metabolomics focus on a specific aspect of molecular biology, allowing researchers to gain a comprehensive understanding of biological systems. Of course, these have contributed to numerous technological advancements across a wide range of fields (Seyhan and Carini, 2019).

One special field of omics named "meta-omics" has enabled the understanding of the microbial world. It involves the use of high-throughput sequencing technologies to investigate the collective genomic and functional potential of microbial communities. Using meta-omics, complex microbial communities (which cannot be cultured and studied in isolation) can be explored (Zhang et al., 2019; Mauger et al., 2022).

The different subfields of meta-omics (metagenomics, metatranscriptomics, metaproteomics, and metabolomics) work hand-in-hand to fully characterize the microbial population. In brief, metagenomics involves sequencing of DNA fragments in a sample, without the need for isolation and cultivation of individual microbes while metatranscriptomics focuses on the study of the RNA transcripts expressed by the microbial community, providing insights into the active functional pathways and metabolic processes of the community (Shakya et al., 2019; Maghini et al., 2021). Metaproteomics and metabolomics, on the other hand, involve the study of proteins and small molecules produced by microbial communities, providing insights into the metabolic, functional, and biochemical pathways of the community (Nephali et al., 2020; Armengaud, 2023). Hence, new biological molecules of significant importance can be identified, thereby unlocking new opportunities for biotechnology and environmental sciences (Zhang et al., 2019; Tiwari and Taj, 2020).

While metagenomics provides information on the potential functional capabilities of microbial communities, metaproteomics can confirm which functions are actively being performed (Schiebenhoefer et al., 2019). As many studies are focusing on metagenomics, coupling these studies with metaproteomics can address some of the limitations of metagenomics (Issa Isaac et al., 2019; Zheng et al., 2020). For example, metagenomics can be limited by the quality and completeness of the genomic data obtained from environmental samples, while metaproteomics can provide direct evidence of the proteins that are being expressed by microbial communities in the sample (Nowrotek et al., 2019; Renu et al., 2019).

Promising metaproteomics research has been published in recent years. The increase in the publication output could be associated with the technological advancements made in instrumentation (Bailey et al., 2019). One of the advancements is the development of Nano-Liquid chromatography (LC) techniques that are more advantageous than the conventional LC techniques, as well as the advent of high-resolution mass spectrometry (MS) that enables the identification and quantification of tens of thousands of peptides and proteins per sample (Vargas Medina et al., 2020). These have contributed to the understanding of proteins and their significance in different conditions (Hardouin et al., 2021).

However, despite the remarkable advancements, metaproteomics is still in its infancy, and its full potential has not been realized, particularly in developing nations (Hamdi et al., 2021). For instance, few studies in omics were conducted in Africa, most of which are collaborative with the developed world. Although these could be attributed to the lack of funding from the government and private sectors, the limited expertise in use of the advanced technologies, bioinformatics analysis, and interpretation of the generated data could be a limiting factor (El Jaddaoui et al., 2020; Iskandar et al., 2021).

Interestingly, the field of metaproteomics has existed for so long, and its impact on the functional characterization of the microbial population is far-reaching. Many publications have widely used the technology to determine the abundance, diversity, and activity of proteins in different research domains (Chandran et al., 2020; Rane et al., 2022). To mention a few, metaproteomics has been used in studies such as examining the microbiome composition and function (Kumar et al., 2021), comparative proteomics in health and disease states (Calabrese et al., 2022), evaluating the effects of environmental toxins (Jiao et al., 2022), and insights into metabolic pathways (Armengaud, 2023). Furthermore, the technique has been applied in the development of novel biomarkers for medical diagnostics as well as a better understanding of the function proteins play in host-pathogen interactions (Moreira et al., 2021). It is important to note that several reviews have discussed improvements made in metaproteomics (Schiebenhoefer et al., 2019; Andersen et al., 2021; Hardouin et al., 2021; Salvato et al., 2021; Bahule et al., 2022). However, there is scanty information concerning the state-of-art in the field of metaproteomics compared to other omics.

In this article, we conducted a bibliometric analysis in the field of metaproteomics. This kind of analysis has been conducted in other fields to have a full grasp of the achievements made (Md Khudzari et al., 2018; Yu et al., 2018; Kushairi and Ahmi, 2021; Wang et al., 2021). The specific objectives are to assess the current state of metaproteomic research on a global scale, evaluate its significance and potential impact, and specifically examine the contribution of Africa in the research field. The study aims to provide insights into the overall trends, collaboration patterns, and key research areas in metaproteomics globally while highlighting Africa's research output and its potential role in advancing the field. By understanding the global landscape and Africa's contributions, this study can inform future research directions, identify potential collaborations, and promote the growth of metaproteomics research in Africa.

2. Materials and methods

2.1. Data source and search strategy

The bibliometric analysis was conducted by searching for publications on metaproteomics in the Scopus database. Initially, a search was conducted on the Web of Science and Scopus databases to draw a comparison between their outputs. It was discovered that the Scopus database contained the majority of publications. The Scopus database is a comprehensive abstract and indexing database, offering extensive coverage across a wide range of subjects. It provides access to a significant number of international journals, ensuring global coverage and facilitating the inclusion of research from different geographic regions (Baas et al., 2020; Zhang et al., 2020). Keywords were selected by initially conducting a review of the relevant literature to identify commonly used terms related to our research topic. The MeSH (Medical Subject Headings) term of "Metaproteomics," thus

"Metaproteogenomic" was identified and the keywords "Metaproteomic" and "Metaproteogenomic" were used to search the titles, abstracts, and keywords of the Scopus database in combination with the Boolean operator "OR." The wild card "*" was used at the end of the search term to cater to all other variants of the term, that is, "metaproteomic, metaproteomics, metaproteogenomic, and metaproteogenomics." Only English publications were considered for this work. Hence, the query string used is as follows: TITLE-ABS-KEY (metaproteomic* OR metaproteogenomic*) AND [LIMIT-TO (LANGUAGE, "English")] AND [EXCLUDE (PUBYEAR, 2023)]. Consequently, the results starting from 2004 to 2022 were collected. Altogether, 1,144 publications of all literature types were retrieved and included in order to ensure the overall comprehensiveness of the metaproteomics research. The completeness of the bibliographic metadata was determined using Biblioshiny (Supplementary Figure 1). In the database, information on journals, publication date, authors' institutions, countries, publication sources, abstract, citation frequency, keywords, and bibliographies was selected and subsequently downloaded in a CSV file. In order to clean the data, duplicates were searched and removed, and the remaining literature was used for subsequent analysis. To obtain more information, the author's h-index and Scopus ID, in addition to the journal cite scores, were retrieved from the database. It is important to mention that, in order to eliminate possible bias caused by database updates, data searching and gathering were conducted on the same day (January 17, 2023).

2.2. Bibliometric analyses and data visualization

The bibliometric analysis was conducted according to previous studies (Md Khudzari et al., 2018; Ejaz et al., 2022). VOSviewer (due to its user-friendly interface, ability to handle large datasets, and intuitive visualization options) and the Biblioshiny package in R (Aria and Cuccurullo, 2017) were used for the analysis. The VOSviewer has been known to be useful in developing more sophisticated bibliometric maps (Yu et al., 2018, 2020). The combination of the software and the R package enables in-depth analyses that aid in a more holistic understanding of scholarly collaboration, research impact, and citation dynamics in specific studies (Ragazou et al., 2022).

Using VOSviewer (Version 1.6.18), Country cooperation network, Institutional collaboration network, and Keyword analysis were conducted. Briefly, after importing the CSV file into the software, the co-authorship option was chosen, and the unit of analysis was designated as "countries." The maximum number of writers per document and the counting method were left as defaults, while threshold criteria were set to a minimum of five (5) papers per country and a minimum of one (1) citation. For the Institutional collaboration network, a thesaurus file was developed and imported into VOSviewer in order to harmonize and re-label synonymic institutional names. Thereafter, co-authorship was selected, and organizations were chosen as the unit of analysis, while the counting method and the maximum number of organizations per document were left as defaults. The threshold criteria were set at a minimum of three (3) documents and zero (0) citations per organization. The type of analysis chosen for the keyword analysis was co-occurrence, and the unit of analysis was Author keywords, with a minimum number of keyword occurrences of five (5). The default counting method was used. A thesaurus file was developed and imported into VOSviewer to re-label synonymic single words and congeneric phrases. Overall, VOSviewer was used to construct network maps for keyword analysis, as well as collaboration between countries and institutions.

For the Biblioshiny, the CSV file was uploaded for analysis of trends in publication, major document types, peer-reviewed scientific journals, country productivity, and distributions of the author's productivity. The analyses were performed by clicking the tabs on the web interface as follows:

Trends in publication: To access information in relation to the trend in publication, the *Biblioshiny*→ *Overview*→ *Annual Scientific Production*→ *Table* tabs were clicked. The annual scientific production is based on the total number of publications within each year.

Major document types: The *Biblioshiny* → *Overview*→ *Main information* tabs were clicked to have the information on the document types and their quantity.

Peer-reviewed scientific journals: Information related to the Leading Scientific Journals was accessed by clicking *Biblioshiny* → *Sources* → *Most relevant Sources* tabs.

Country Productivity: For the Country's Productivity, *Biblioshiny* → *Authors* → *Countries' Scientific Production* tabs were selected. Information on the total citation count and the average article citations of the countries was obtained by clicking *Biblioshiny* → *Authors* → *Most Cited Countries* tabs. The map of international collaboration between the countries was obtained by clicking *Biblioshiny* → *Social Structure* → *Countries' Collaboration World Map* tabs.

Distribution of Author's Productivity: Information concerning the leading Authors in Metaproteomics research was accessed by clicking *Biblioshiny* → *Authors* → *Most Relevant Authors* tabs. The author's productivity period was ascertained by clicking *Biblioshiny* → *Authors* → *Author's Production Over-Time* → *Plots* tabs. The analysis of the Author's productivity using Lotka's law was performed by clicking the tabs Biblioshiny → Authors → Lotka's law.

Trending topics: To analyze trending terms based on the Keywords, the trend topics plot was generated by clicking the *Biblioshiny* → *Documents* → *Trend topics* → *Plot* tabs.

2.3. Sub-analysis of African publication trends and ranking using FWCI parameter

The emergence of African contributions to the field of metaproteomics was investigated by assessing the distribution and the overall trends in publication activities. To achieve this, data from publications related to Africa were extracted and analyzed using Excel. The ranking of the publications was conducted using the FWCI parameter from Scopus (Zanotto and Carvalho, 2021).

3. Results

3.1. Publications retrieval and screening process

From the Scopus database, 1,144 publications were collected. After screening for duplicates, the number of articles was reduced to 1,138. A bibliometric analysis of the 1,138 publications on

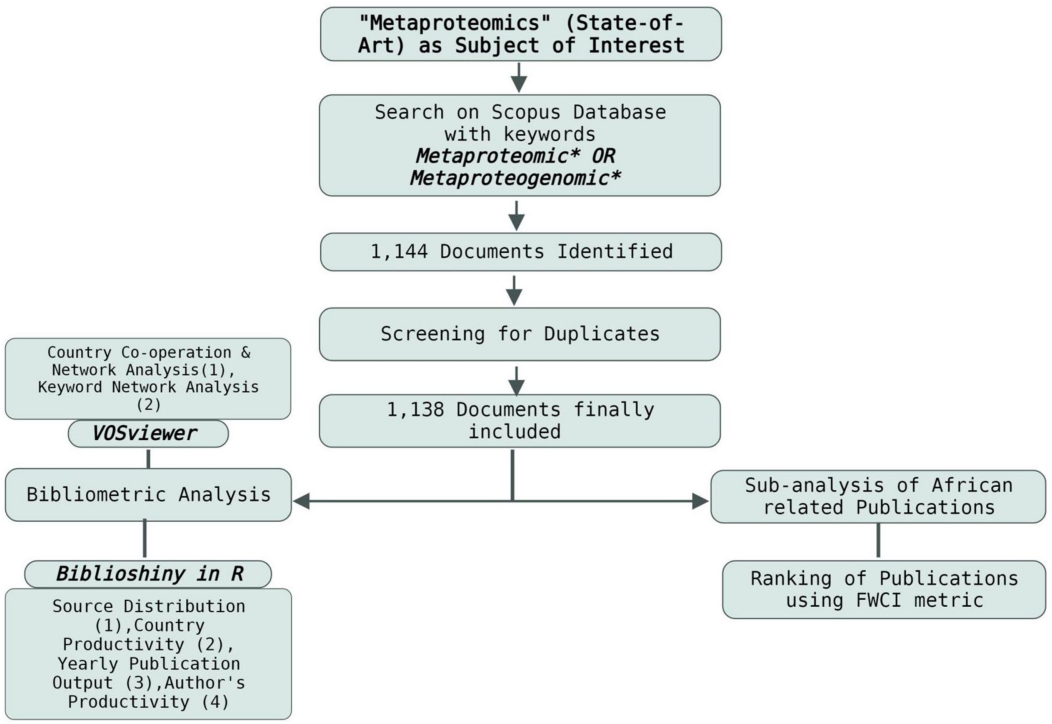


FIGURE 1
Workflow of documents retrieval process and analyses.

metaproteomics obtained from the database was conducted to understand the global impacts of metaproteomics research. The publications included research articles, reviews, conference papers, editorials, book chapters, and others. The workflow of the retrieval process is indicated in Figure 1.

3.2. The trend in publication output and major document types

The analysis revealed that the number of publications on metaproteomics has been increasing steadily since the early 2000s. Initially, studies on the topic started in 2004 with a single study. Afterwards, a slow increase in the number of publications was observed from 2006 to 2009. A significant rise in the number of publications was observed from 2010 onwards (Figure 2A). From 2010 onwards, there was a steady rise in the number of publications. Although a high number of publications (164; 14.4%) were recorded in 2022, there was more publication output in 2021 (170; 14.9%). Moreover, there was no publication record found in 2005. Overall, the estimated annual growth rate (EAGR) of the publications computed from the Biblioshiny package was 37.25%.

In terms of the document types (Figure 2B), the majority of the publications were research articles (772; 68%), followed by reviews (203; 18%) and book chapters (94; 8%). Others, including conference papers, editorials, and notes, account for 3.95%. Information gathered from the Scopus database revealed that the majority of the publications are all open access (678; 59.6%).

3.3. Leading subject areas and peer-reviewed scientific journals

The publications on metaproteomics included in the Scopus database cover 23 subject areas. The Major areas with the most densely distributed publications are Biochemistry, Genetics, and Molecular Biology with 525 published documents representing 46.13% of the total number of scholarly works (Figure 2C). Other subject areas with publications include Immunology and Microbiology (412; 36.20%), Medicine (272; 23.90%), Agricultural and Biological science (264; 23.19%), Environmental science (180; 15.82%), Chemistry (158; 13.88%), Chemical Engineering (82; 7.21%), Computer Science (68; 5.98%), Multidisciplinary (62; 5.45%), and Engineering (59; 5.19%) (Figure 2C).

The present finding investigated the major Journal houses that published the metaproteomic research, and among the top 10, *Springer Nature* (86; 7.56%) and *Wiley-Blackwell* (77; 6.77%) were the top publishers (Table 1). The *American Chemical Society (ACS)*, *Elsevier*, *Frontiers Media S.A.*, and the *Public Library of Science* were also among them (Table 1).

In terms of the top 10 journals, *Frontiers in Microbiology* has the highest number of publications (62; 5.45%), followed by the *Journal of Proteome Research* (50; 4.39%), *Proteomics* (49; 4.31%), *Microbiome* (37; 3.25%), and *ISME Journal* (32; 2.88%). Conversely, while assessing the Journal's CiteScore from the Scopus database (2021 report), *Microbiome* has the highest score of 24.5, followed by *Nature Communications* (23.2), and *ISME Journal* (18.2) (Table 1).

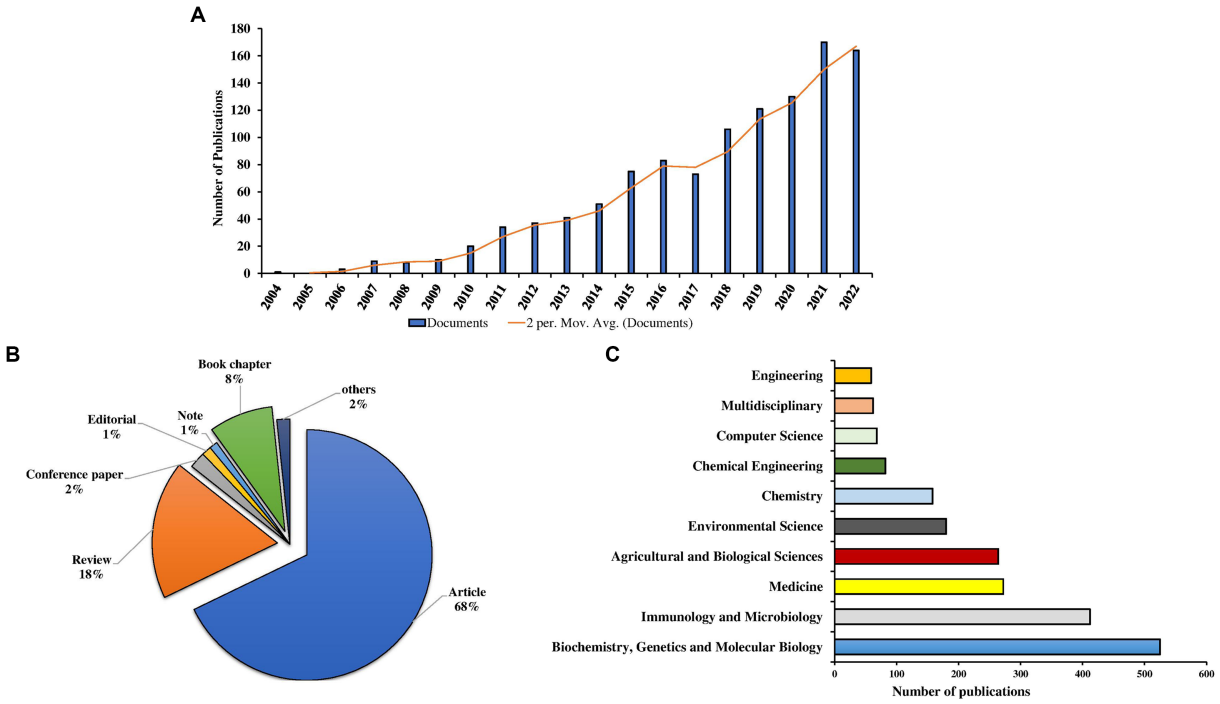


FIGURE 2
Global publication trend **(A)**, document types **(B)**, and global leading subject areas **(C)** in metaproteomics research.

3.4. Country productivity and cooperation network

Looking at the country's productivity, the top 20 most productive countries (in terms of publications) are the USA (359), Germany (257), China (172), and Canada (90). These countries have the highest total citations (>1,504), in addition to other countries like Spain (1,728), Sweden (1,620), and the United Kingdom (1,414). Contrariwise, Switzerland, Finland, and the UK have high average article citations of 147.3, 58.0, and 56.6, respectively (Table 2).

The map and network of international and country cooperation (Figures 3A,B) demonstrate close collaboration among countries or regions involved in metaproteomics research. In the first cluster, the USA is the most affiliated country, linked to 38 countries or territories with a total link strength of 330, while China, as the second, is affiliated with 28 countries with a link strength of 119. Germany also has a total of 30 links and a total link strength of 334, followed by the UK in the second cluster. The final cluster includes Italy and Canada with more collaboration networks.

In terms of African countries, Morocco had the highest number of collaborations (14 links), followed by Egypt (13 links) and South Africa (11 links), with total link strengths of 20, 18, and 28, respectively.

3.5. Top institutions and distribution of author's productivity

Among the institutions, Helmholtz Zentrum für Umweltforschung, Germany (83; 7.29%), Oak Ridge National Laboratory, USA (59; 5.18%), Otto von Guericke University of Magdeburg, Germany (46; 4.04%), and Universität Greifswald, Germany (42; 3.69%) are the top institutions (Table 3). In terms of institutional collaborations, the Max Planck Institute, Oak Ridge National Laboratory, and Helmholtz Centre for Environmental Research have the highest total link strength of collaborations (Supplementary Figure 2). Figure 4 depicts the collaboration network among these institutions.

In terms of authorship, the major authors (top 20) are from Germany, Canada, the United States of America (USA), France, Spain, Belgium, and Luxembourg, in that order. Among them, the author "Jehmlich N" from Helmholtz Centre for Environmental Research, Germany, has the highest publication record with 61 articles. His productive period spans from 2010, and he currently holds an h-index of 41 (Table 4; Figure 5). "Von Bergen M." and "Hettich R.L.", affiliated with the Helmholtz Centre for Environmental Research, Germany, and the Oak Ridge National Laboratory, USA, respectively, are ranked second and third. Similar to Jehmlich N., "Von Bergen M." started his productive period in 2010 and currently holds an h-index of 56, while "Hettich R.L." has an index of 63 with his first publication in 2008. Other authors with their h-index and productivity period are shown in Table 4 and Figure 5, respectively. An analysis of Lotka's law reveals a skewed distribution of productivity among the authors contributing to metaproteomics research (Supplementary Figure 3).

3.6. Keyword analysis and current trending topics

Among 1,138 publications, a total of 2,459 keywords were identified. After re-labeling synonymous single words and consolidating similar phrases, 82 keywords met the threshold of a

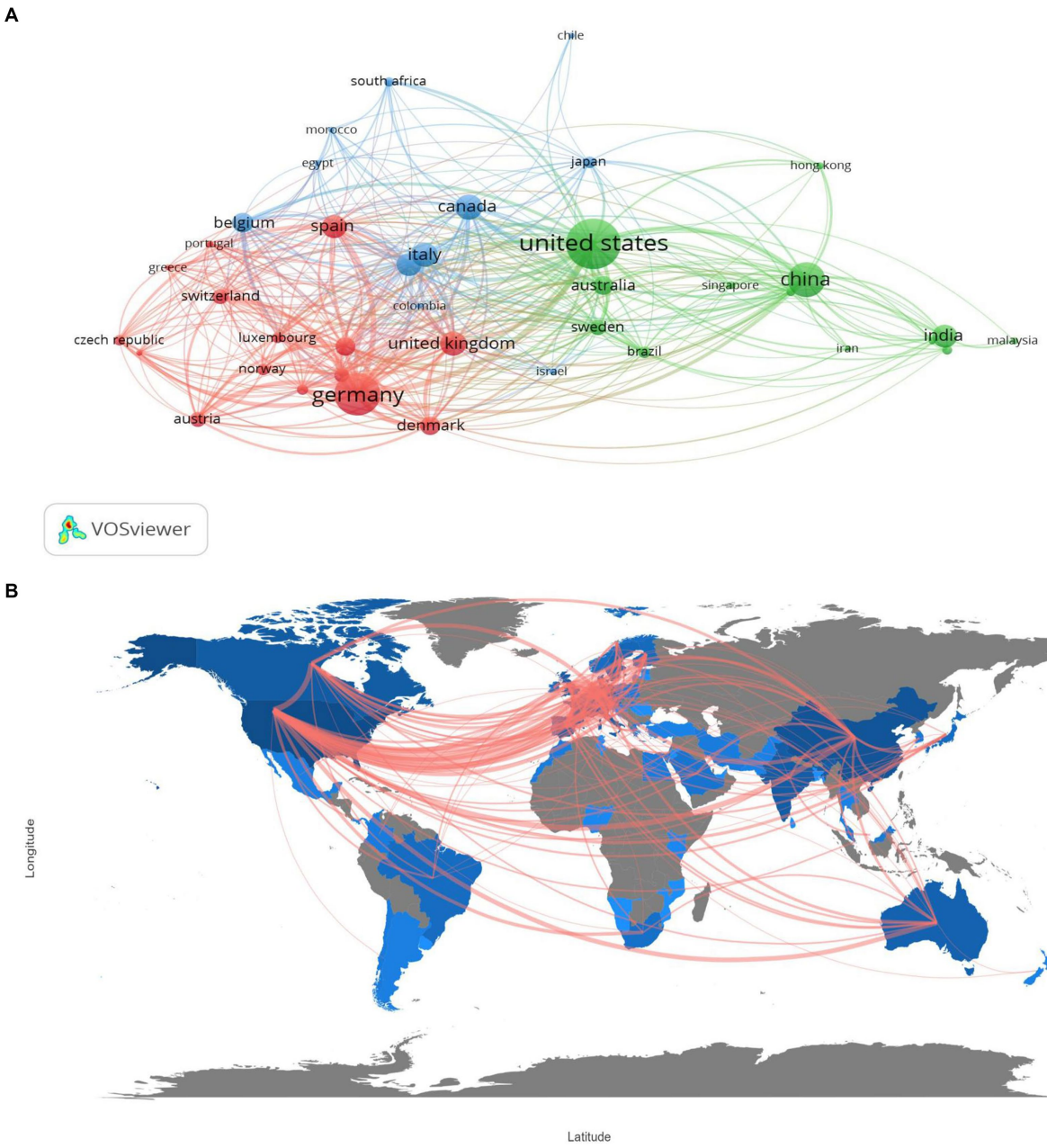


FIGURE 3
Network of country co-operation **(A)**, and worldwide map indicating international collaborations **(B)** in metaproteomics research.

minimum of 5 occurrences for mapping in VOS viewer (Version 1.6.18). As shown in Figures 6A,B, nine clusters were formed and categorized in Table 5 to clearly indicate the selected and leading keywords in each cluster. Clusters 1 to 3 contained more than 10 keywords, while clusters 4 to 9 had fewer than 10 keywords. Cluster 1 had the highest number of selected keywords (14), followed by Cluster 2 (13) and Cluster 3 (12).

Some leading keywords encountered in each cluster, based on their degree of occurrence, included metaproteome (Cluster 1; 41 occurrences; 27 links; 51 total link strength), microbiome (Cluster 2; 115 occurrences; 60 links; total link strength 274), metagenomics (Cluster 3; 173 occurrences; 61 links; 447 total link strength), bacteria (Cluster 4; 15 occurrences; 19 links; 42 total link strength), microbial community (Cluster 5; 93 occurrences; 46 links; 212 total link strength), mass spectrometry (Cluster 6; 72 occurrences; 44 links; 207 total link strength), metaproteomics (Cluster 7; 480 occurrences; 79 links; 937 total link strength), Protein-SIP (Cluster 8; 14 occurrences; 13 links; 28 total link strength), and saliva (Cluster 9; 13 occurrences; 17 links; 41 total link strength).

In order to analyze the trending terms based on the keywords, the Trend topic plot was constructed (Figure 7). The size of the circles shows the frequency of the term, and the length of the lines shows how long it has been studied (Abafe et al., 2022). From the plot, wastewater, omics, and DNA sequencing are the trending topics in recent years. The three most commonly used terms are proteomics ($f = 1110$), metagenomics ($f = 564$), and metaproteomics ($f = 412$) (Figure 7).

TABLE 1 The top 10 most productive journals that published metaproteomics research.

Sources	No. of publications	CiteScore (2021)	Publisher
Frontiers in Microbiology	62	8.2	Frontiers Media S.A.
Journal of Proteome Research	50	7.7	American Chemical Society
Proteomics	49	8.1	Wiley-Blackwell
Microbiome	37	24.5	Springer Nature
ISME Journal	32	18.2	Springer Nature
Environmental Microbiology	28	8.2	Wiley-Blackwell
Journal of Proteomics	25	7.0	Elsevier
Applied and Environmental Microbiology	24	7.8	American Society for Microbiology
PLoS One	20	5.6	Public Library of Science
Nature Communications	17	23.2	Springer Nature

3.7. Publication analyses from Africa

The Scopus database retrieved 13 articles, 9 reviews, 1 book chapter, 1 editorial, and 1 short survey related to metaproteomics research in Africa (Figure 8A). Accordingly, the major research areas that utilized metaproteomics include Immunology and Microbiology (10 publications), Biochemistry, Genetics, and Molecular Biology (9 publications), as well as Medicine (8 publications) (Figure 8B).

Looking at the trend of publication, it was observed that metaproteomics research in Africa began in 2013 (Figure 8C). From 2013 onwards, there has been an exponential growth in the publication outputs, with an EAGR of 19.58% (Supplementary Figure 4).

The countries in Africa, including South Africa (14; 56%) and Egypt (4; 16%), have the highest number of publications, while Nigeria and Morocco have 2 publications, accounting for 8% each. Other countries with a single publication in metaproteomics are Kenya, Mozambique, and Tunisia (Figure 8D). Supplementary Figure 4 provides additional information about the general characteristics of the African publications. Furthermore, we previously highlighted the country cooperation networks of three African countries: Morocco, Egypt, and South Africa, in Figure 3A. The top 10 journals for African publications are presented in Supplementary Table 1, with *Frontiers in Cellular and Infection Microbiology* and *Microbiome* emerging as the leading journals.

For the keyword analyses, a minimum threshold of two (2) occurrences was set, resulting in the identification of 10 keywords that met the criteria. The majority of the keywords are metaproteomics, metagenomics, metatranscriptomics, and microbiome. In terms of the network, the leading keywords are metabolites (Cluster 1), metaproteomics (Cluster 2), and metagenomics (Cluster 3). According to the Trend topic plot, it is evident that non-human and human research, along with review writing, are currently trending topics in the field. In concordance with the global trend, proteomics ($f = 17$) and metagenomics ($f = 17$) are the most commonly used terms (Figure 9).

3.8. Ranking publications from Africa using the field-weighted citation impact (FWCI) scores

The African publications were ranked using FWCI (Zanotto and Carvalho, 2021). About 52% of publications had an FWCI score of 1 or higher, indicating a positive impact (Table 6). For instance, most of the publications from South Africa, Egypt, Morocco, Kenya, Tunisia, and Nigeria have FWCI scores ≥1. One study from Morocco and a study from Mozambique have scores <1, respectively.

4. Discussion

Metaproteomics is a rapidly growing field of research that seeks to understand the complex microbial communities that inhabit the world (Kumar et al., 2021). To better understand the impact of this research area, tracking and analyzing its scientific outputs will be paramount in identifying emerging trends and key contributors in the field. To achieve this goal, we conducted a bibliometric analysis of more than 1,100 publications on metaproteomics obtained from the Scopus database. As the Web of Science is also commonly utilized for this type of study, we started our study with a comparative analysis, which revealed that 97.3% of the total extracted publications from the Scopus database are shared with the Web of Science. Only 2.7% of the publications (31 in total) are unique to the Web of Science database (statistics of these 31 publications have been included in the Supplementary material). From the analysis, information on the publication patterns and collaboration structures, among others was studied. A similar statistical analysis was adopted recently in order to understand the trends and the focus of the link between gut microbiota and type 1 diabetes (Guo et al., 2023).

Since 2004, the yearly trend in publication outputs on metaproteomics has been increasing with a steady growth rate starting from 2010. This increase, in part, could be attributed to the increased interest of scientists in identifying novel proteins for several biotechnological applications from microbial communities (Purohit et al., 2020; Mishra et al., 2021). Although there is not much difference in publication output between 2021 and 2022, the higher numbers in 2021 could be attributed to several notable advancements in metaproteomics tools within the year. For instance, the use of single-cell proteomics to study microbial communities and the development of methods for analyzing post-translational modifications in metaproteomics samples (Taylor et al., 2021; Armengaud, 2023), in addition to COVID-19 pandemic which spurred interest in using the technique to study the human microbiome and its role in infectious diseases (He et al., 2021; Sun et al., 2022). Overall, the global EAGR suggests that, under normal circumstances, the number of publications in metaproteomics is expected to increase by 37% over time. The abundance of research articles in this field suggests that many authors are actively engaged in experimental research and eager to advance our understanding by conducting thorough analyses of their findings.

TABLE 2 The top 20 most productive countries involved in metaproteomics research.

Country	Total number of publication	Total number citation	Average number of article citations
United States	359	8,201	36.45
Germany	257	5,575	38.45
China	172	4,202	30.45
Canada	90	1,504	26.3
United Kingdom	86	1,414	56.56
Italy	83	1,363	25.72
India	81	622	14.14
Spain	77	1,728	37.57
France	76	1,086	27.85
Belgium	57	751	28.88
Australia	52	1,330	42.90
Denmark	51	390	28.14
Netherlands	49	197	32.50
Sweden	35	1,620	18.45
Austria	34	577	48.08
Switzerland	32	203	147.27
Finland	27	522	58.00
Brazil	21	294	16.73
Japan	21	266	42.00
Norway	19	184	26.60

Their contributions are likely to shed new light on the intricate workings of protein chemistry in a particular phenomenon, potentially leading to significant breakthroughs and progress.

Based on the leading subject areas, most of the publications are related to Biochemistry, Genetics, and Molecular Biology as well as Immunology and Microbiology. This indicates that metaproteomics is a promising field that helps scientists gain a more comprehensive understanding of biological systems. Using the techniques of metaproteomics, an enhanced understanding of genetics, as well as possible changes that might occur in disease states, could be achieved (Henry et al., 2022). Also, some publications were on Agriculture and Environmental sciences, enabling the design of strategies that ensures agricultural productivity and environmental sustainability (Chandran et al., 2020; Bahule et al., 2022).

Analysis of the top 10 journals reveals that all of them are of high quality, with both Q1 or Q2 rankings and impact factors above 4. When examining the journals' CiteScore, it becomes apparent that both Microbiome and Nature Communications have a substantial impact on the metaproteomics research community. The CiteScore could influence the decision of some authors when selecting a journal for their work (Roldan-Valadez et al., 2019). Overall, based on the findings, it could be asserted that researchers in the field of metaproteomics will be inclined to publish their findings in these top 10 journals.

In terms of country productivity and cooperation network, Europe and America have the highest total number of publications. Individually, the USA has the highest number of total citations followed by Germany and China, indicating that these countries have a high overall impact on the research output. Contrary to the total citation counts, Switzerland, Finland, and the UK exhibit high average

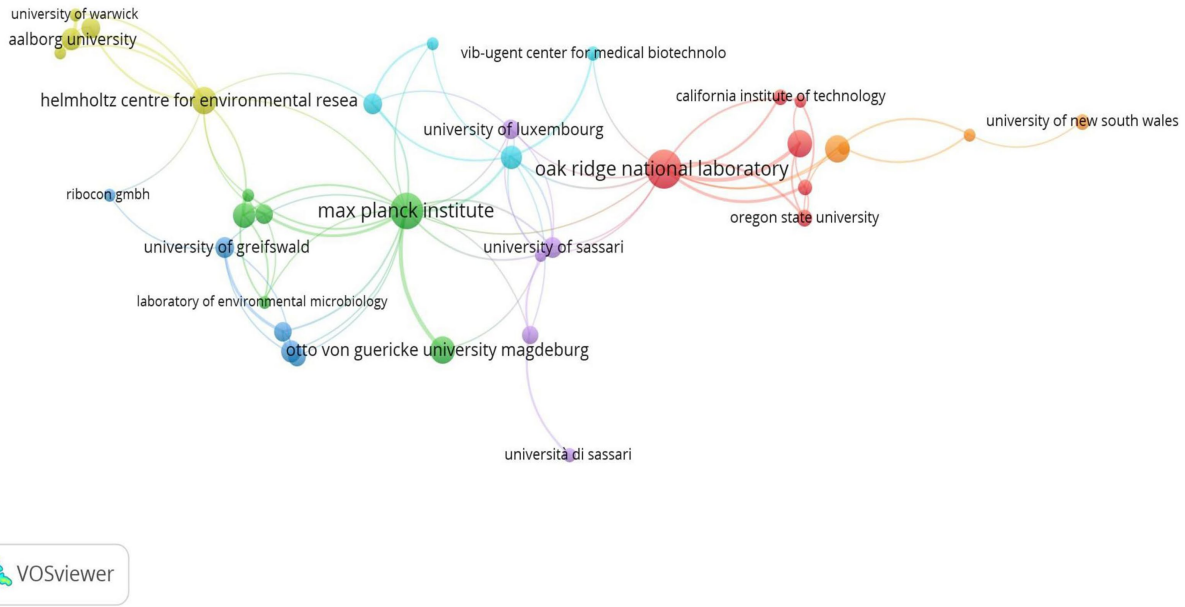


FIGURE 4
Institutional collaboration network in metaproteomics research.

TABLE 3 The top 10 most productive institutions involved in metaproteomics research.

Institution	Documents published	Country
Helmholtz Zentrum für Umweltforschung	83	Germany
Oak Ridge National Laboratory	59	United States
Otto von Guericke University of Magdeburg	46	Germany
Universität Greifswald	42	Germany
Max Planck Institute for Dynamics of Complex Technical Systems	40	Germany
Universiteit Gent	39	Belgium
Pacific Northwest National Laboratory	37	United States
The University of Tennessee, Knoxville	35	United States
University of Ottawa	34	Canada
Universität Leipzig	31	Germany

article citations, indicating the production of high-quality research by these countries (Roldan-Valadez et al., 2019). As the total citation measures quantity (Tahamtan and Bornmann, 2019), the average citation measures quality and determines scientific contribution as well as plays a role in advancing a specific field. For collaboration, it appears that the USA, Germany, China, Italy, and Canada are the most connected countries in addition to some other European countries. The collaboration networks observed could be attributed to scientific advancements and investment in research and development (R&D) in these countries (El Jaddaoui et al., 2020; Guo et al., 2023).

Consistent with the above analysis, the institutions from Europe and America have the highest publication outputs, with German institutions taking the lead. Generally, German institutions have a long history of scientific excellence, and they have been at the forefront of many scientific breakthroughs. Specifically, in the case of metaproteomics research, German institutions have been able to leverage their expertise in proteomics and microbiology to develop methods and techniques in metaproteomic analyses (Schiebenhoefer et al., 2019). In terms of collaboration networks, there are more links among German institutions, with the Max Planck Institute having the highest collaboration. This observation is not surprising, as the institution has been observed to be frequently involved with metaproteomic research. To sum it up, the high publication outputs in the aforementioned top 10 institutions from Europe (Germany) and America may be due to a variety of factors, including access to funding, a high level of expertise and specialized knowledge, successful collaborations with other researchers and institutions, and strategic research priorities that are in line with the institution's goals.

The diversity of research partners, a high proportion of foreign postgraduates and visiting academics, and robust research funding are all possible contributors to the dynamism of international collaboration and increased publication outputs (Roldan-Valadez et al., 2019). The top 20 authors with the highest publication records are primarily from Europe and America. The foremost author among them serves as the group leader for Microbiome Biology at the Helmholtz-Center for Environmental Research in Germany. The key areas of interest for this author's research are the study of microbiome biology in terrestrial microbial communities, metaproteomic studies, and both qualitative and quantitative microbial proteomics. In particular, the author is known for their work in identifying key microbial players using protein-SIP, a cutting-edge technique in the field of microbiology. The author's first publication, titled "Phylogenetic and proteomic analysis of an anaerobic toluene-degrading community," examined, for the first time, a sulfate-reducing community grown with toluene as a carbon source by a combination of molecular genetics and proteomic techniques in order to uncover the physiological interplay in this microbial anaerobic community (Jehmlich et al., 2010). The author achieved the highest publication output in 2016, resulting in a total citation count of 87.5, and subsequently, in 2020, the author's publication output amassed a total citation count of 30.75. As well, it is worth highlighting that one of the top 20 authors is Wilmes P., who is a notable figure in the field. In collaboration with Bond P.L., Wilmes P. proposed the term "metaproteomics" in their 2004 publication entitled "The application of two-dimensional polyacrylamide gel electrophoresis and downstream analysis to a mixed community of prokaryotic microorganisms" (Wilmes and Bond, 2004). This contribution to the scientific community has had a significant impact on the field of metaproteomics and serves as a testament to Wilmes P.'s expertise and innovative thinking. In terms of h-index, Hettich R.L. (the leader of the Bioanalytical Mass Spectrometry group at Oak Ridge National Laboratory) has the highest value of 63 among the top 20. The h-index used here measures the productivity and impact of a researcher's publications (Roldan-Valadez et al., 2019). It also takes into account engagements in secondary research, which could influence its value. An assessment of the author's productivity in this field reveals that it adheres to Lotka's law, which claims that a small percentage of authors produce the majority of output in a given field (Kushairi and Ahmi, 2021). This information could be useful for developing strategies to support and encourage the productivity of all authors in the field, regardless of their current level of output.

Keyword analyses offer a comprehensive overview of the research area's trajectory and themes (Donthu et al., 2021). Based on the keyword analyses, Cluster 1 had the highest number of selected keywords, implying that it contains keywords frequently used in the context of metaproteomic research (Donthu et al., 2021). In terms of leading keywords, metaproteomics had the highest occurrence, followed by metagenomics and other omics. This is unsurprising as the omics fields are often coupled together in order to have a thorough understanding of the microbial world (Kumar et al., 2021; Jiang et al., 2022).

Further the keyword clusters generated give an overview of the thematic research hotspots. For instance Cluster 1 sheds light on the interest in understanding microbial processes and diversity in biotechnology and the environment while the terms in Cluster 2 are related to metaproteomics research aimed at understanding the importance of gut microbiota in human health. Clusters 3 and 4 on the other hand indicate interest in the concept of molecular components and interactions within biological systems as well as understanding the role of metaproteomics in investigating the soil microbiome and biodegradation and bioremediation processes.

TABLE 4 Top 20 prolific authors with high publication outputs in metaproteomics research.

Authors	Publications	Scopus ID	Current h index	Current affiliated institution	Country
Jehmlich N	61	24366846200	41	Helmholtz Zentrum für Umweltforschung	Germany
Von Bergen M	52	6603254363	56	Helmholtz Zentrum für Umweltforschung	Germany
Hettich RL	47	7006786519	63	Oak Ridge National Laboratory	United States
Benndorf D	43	55933825600	23	Otto von Guericke University of Magdeburg	Germany
Zhang X	36	57192503794	28	Université d'Ottawa, Faculté de Médecine	Canada
Figeys D	34	7005139451	57	Université d'Ottawa, Faculté de Médecine	Canada
Seifert J	34	35230690700	39	Universität Hohenheim, Stuttgart	Germany
Heyer R	29	55263328800	15	Leibniz-Institut für Analytische Wissenschaften	Germany
Reichl U	28	6602727388	49	Otto von Guericke University of Magdeburg	Germany
Li L	27	57194857863	16	Université d'Ottawa, Faculté de Médecine	Canada
Ning Z	27	25642980100	27	University of Ottawa,Faculté de Médecine	Canada
Mayne J	24	7004815175	26	Université d'Ottawa, Faculté de Médecine	Canada
Bastida F	23	14057630300	38	CEBAS- CSIC, Centro de Edafología y Biología Aplicada del Segura	Spain
Riedel K	23	57202754539	15	Universität Greifswald, Greifswald	Germany
Griffin TJ	21	7202249196	43	University of Minnesota Twin Cities	United States
Schallert K	21	57193336359	7	Leibniz-Institut für Analytische Wissenschaften	Germany
Armengaud J	20	6603746656	42	Université Paris-Saclay, Gif-sur-Yvette	France
Martens L	20	15923262500	56	Universiteit Gent, Ghent	Belgium
Tanca A	20	34769263100	25	Università degli Studi di Sassari, Sassari	Italy
Wilmes P	20	57207607143	47	University of Luxembourg	Luxembourg

Further the other clusters concentrated on techniques for analyzing the protein expressions interactions and functional roles of microorganisms in the soil and gut microbiome in diseases such as COVID-19 as well as researching microbial communities and their functional proteins in activated sludge and wastewater systems. Some important diseases that have been researched using the metaproteomics approach in humans include obesity (Zhong et al., 2019; Biemann et al., 2021; Calabrese et al., 2022) cystic fibrosis (Debyser et al., 2019; Saralegui et al., 2022) inflammatory bowel disease (Moon et al., 2018; Zhang et al., 2018; Lehmann et al., 2019) and the oral microbiome (Jagtap et al., 2012; Bostanci et al., 2021). In addition to the research conducted on human diseases metaproteomics has also been extensively applied in the field of soil ecosystems – soil metaproteomics (Bastida et al., 2016b, 2018; Chourey and Hettich, 2018; Liu et al., 2019) – with specific focus on the rhizosphere (Renu et al., 2019; White et al., 2021). Other environmental phenomena such as bioremediation (Bastida et al., 2016a; Aishwarya et al., 2022) and biodegradation (Chang et al., 2018; Gunasekaran et al., 2022; Xie et al., 2022) has been explored

To further determine the metabolically active players in microbial communities, Protein SIP experiments are increasingly being conducted in the area of environmental microbiology. In addition, studies concerning fermentation or anaerobic digestion of samples from the environment such as biofilms or from wastewater treatment plants which leads to the production of biogas such as methane have been investigated (Joyce et al., 2018; Heyer et al., 2019; Lam et al., 2021; Chen et al., 2022).

Overall, the main instrumentation used in the analysis is tandem mass spectrometry. The peptides are initially separated using liquid chromatography, particularly in the LC–MS/MS procedure. Shotgun proteomics and quantitative proteomics are popular approaches employed to identify and quantify proteins (Balotf et al., 2022). The generated data are analyzed using bioinformatics tools in order to identify the proteomes in the studies.

Sub-analysis of African contributions in metaproteomics revealed that the continent contributed only 25 publications of the total 1,138. Among the 54 countries in Africa, only 7 countries engage in metaproteomic research. As previously mentioned, a lack of infrastructure, funding, and expertise (El Jaddaoui et al., 2020) could contribute to these low statistics. Therefore, considering the significant scientific impact of metaproteomics and omics in general, it is essential to redirect more research efforts toward Africa to characterize the microbial diversity on the continent (Cowan et al., 2022). Remarkably, the positive EAGR observed suggests that research efforts in the continent are increasing, although more has to be done in order to accelerate the growth. In terms of country collaboration, the collaboration network in the continent is relatively weak compared to global networking. Among the countries, only South Africa was noticed to establish research partnerships with the advanced countries.

Interestingly, in comparison with the major publishers, the publications from the African continent are also found in high-impact journals. Likewise, the top keywords used in Africa are comparable to those used globally, with metaproteomics ranking first. This word has links with the other omics terms. The trend plot portrays trending research topics related to non-human studies as well as human studies. Additionally, it is worth noting that review writing was highly predominant. To increase research output beyond review writing, it is

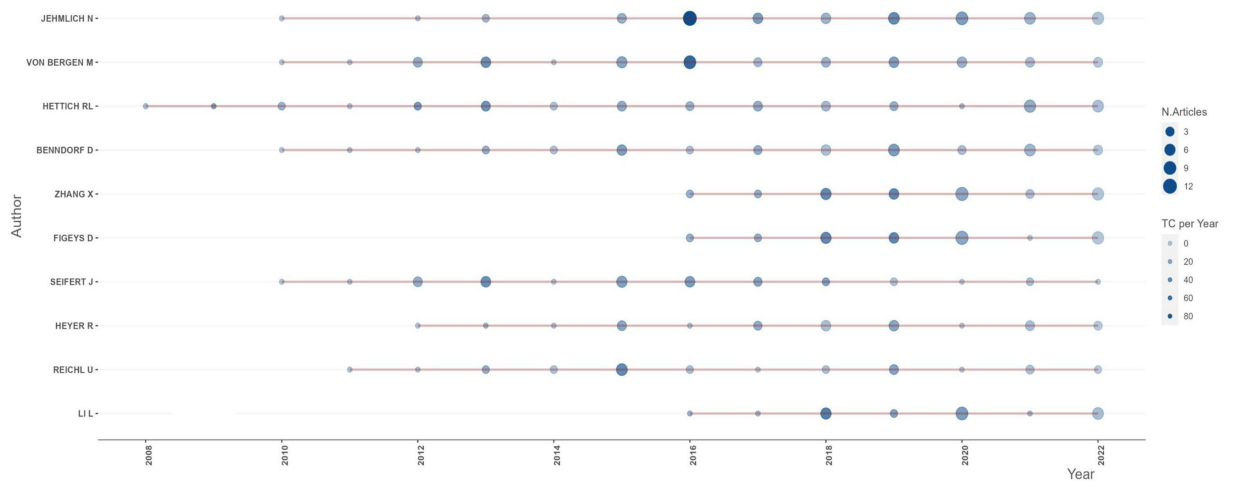


FIGURE 5
Yearly productivity of top authors involved in metaproteomics research.

recommended to establish more laboratories and research centers specifically dedicated to metaproteomics.

The FWCI calculates how frequently a given publication is cited relative to others in the same field. Given its normalization, it can be used to directly compare an article's performance against those of other articles (even those in different subject areas). The FWCI ≥1 in some of the African publications indicates that the output performs exactly as expected by the global average, while the values <1 in some of the publications indicate underperformance in comparison to the global average (Zanotto and Carvalho, 2021). Hence, there is a need for African authors to develop means to improve their performance in order to increase the impact and visibility of their work.

To further emphasize, the insights gained from metaproteomics research have wide-ranging applications in fields such as medicine, agriculture, and environmental science. For instance, metaproteomics enables the taxonomic and functional analysis of complex microbial communities in various samples, allowing the quantification of proteins and providing a detailed understanding of cellular phenotypes at the molecular level. By leveraging metaproteomics, researchers can gain comprehensive insights into these complex biological systems and pave the way for advancements in various disciplines. Fernanda Salvato recently discussed, in detail, the extensive examples of metaproteomics-based approaches used to address a wide range of questions in diverse areas of biological research (Salvato et al., 2021).

5. Conclusion

In conclusion, the bibliometric analyses of more than 1,100 publications on metaproteomics from the Scopus database indicated an increasing trend of publications with a high percentage of open-access research articles. The majority of these publications fall within the field of Biochemistry, Genetics, and Molecular Biology, with *Microbiome* and *Nature Communications* emerging as the top publishers based on their CiteScore rankings. With respect to the author's contribution and country productivity, Jehmlich N. was the most productive author, while the United States, Germany, and China were the most productive countries, collectively contributing to 69% of the total publications. Keyword analyses revealed metaproteomics and metagenomics as the highest co-occurring keywords in addition to other omics. Zooming in on Africa, only 2.2% of total publications came from the continent, with South Africa producing more publications. Based on the EAGR and the FWCI, more studies are required in the region in order to meet the global scale and contribute to the advancement of metaproteomics. To encourage this research area in Africa, some initiatives have begun to emerge. For instance, in Morocco, the Sharifian Phosphate Office (OCP) has funded a new proteomic platform at Mohammed VI Polytechnic University (RD is the principal investigator), which will help to advance the field of proteomics.

5.1. Recommendations

With the strides made by this emerging field, its global challenges cannot be overlooked. Some of the drawbacks in this field are in the areas of sample complexity and preparation, accurate protein database construction for microbial communities, false discovery rate assessments, annotation, and software integration (Issa Isaac et al., 2019; Rechenberger et al., 2019; Saito et al., 2019; Duong and Lee, 2023). Each of these challenges can affect the quality and reproducibility of metaproteomic results and addressing them will require collaborative efforts between researchers and software developers. By doing so, standardized protocols and tools that will improve data quality, accuracy, and reproducibility could be developed. In addition, the current bibliometric analysis has revealed a significant merit of metaproteomics in studying the spatiotemporal characterization of microbial communities at the functional level. Recently, Manuel Kleiner published a very interesting and detailed review that illustrates the diversity of questions that can be addressed solely through metaproteomics in the study of microbial communities, including those associated with plants and animals (Kleiner, 2019). This emerging area holds significant potential for advancing scientific knowledge, addressing biotechnological

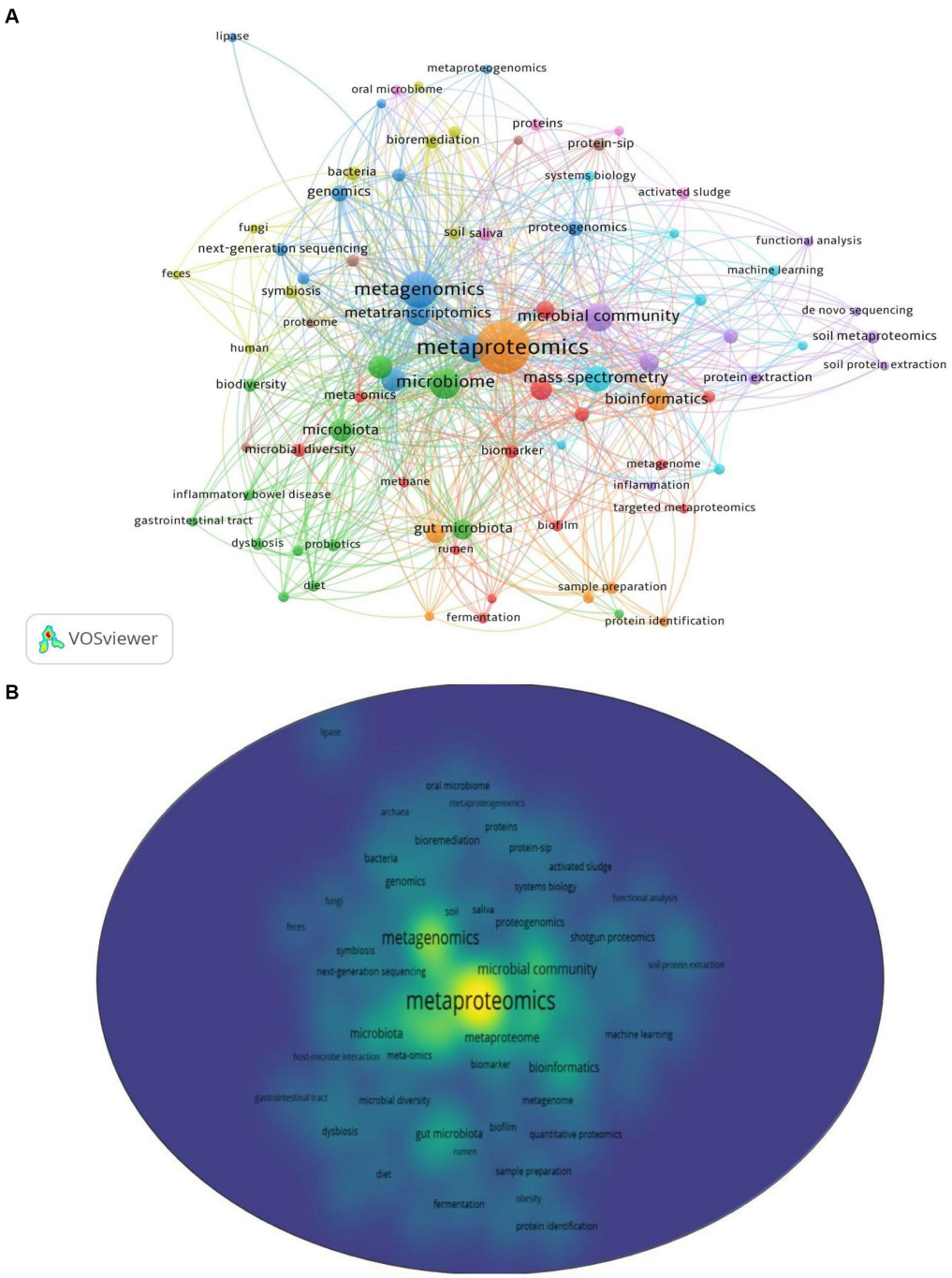


FIGURE 6
Analyses of keywords used in metaproteomics research **(A)** with their density visualization **(B)**.

challenges, improving human health, and informing environmental management strategies. Further, to advance the field of metaproteomics in Africa, several steps can be taken. This includes prioritizing metaproteomics research by providing funding, equipment, and training opportunities. Also, encouraging collaboration between African and international researchers through joint projects, exchange programs, and workshops is essential. For instance, the International Metaproteomics Initiative (Van Den Bossche et al., 2021) is a commendable step toward advancing the field and promoting knowledge sharing. Lastly, support should be provided for publications in high-impact journals, travel grants, and mentoring programs to enhance the quality of research on the continent.

Overall, the findings of this study provide valuable insights that can be used to advocate for increased support, funding, and resources for metaproteomics research on a global scale, as well as specifically in Africa. These findings serve as a basis for informed decision-making by policymakers, enabling them to drive policies, allocate funds, and distribute resources in a manner that promotes the growth and impact of metaproteomics research. By harnessing the potential of metaproteomics, policymakers can address complex societal

TABLE 5 Keyword clusters and leading keywords in metaproteomic research.

Cluster	Number of selected keywords	Selected keywords	Leading keywords
1	14	16S rRNA gene sequencing, anaerobic digestion, biofilm, biogas, biomarker, fermentation, lc–ms/ms, meta-omics, metagenome, metaproteome, methane, microbial diversity, rumen	Metaproteome (41 occurrences, 27 links, 51 total link strength)
2	13	Biodiversity, cystic fibrosis, diet, dysbiosis, gastrointestinal tract, gut microbiota, inflammatory bowel disease, microbiome, microbiota, obesity, omics, probiotics, short-chain fatty acids	Microbiome (115 occurrences, 60 links, 274 total link strength)
3	12	Genomics, lipase, metabolomics, metagenomics, metaproteogenomics, microorganisms, next-generation sequencing, proteogenomics, proteomics, rhizosphere, transcriptomics	Metagenomics (173 occurrences, 61 links, 447 total link strength)
4	9	Archaea, bacteria, biodegradation, bioremediation, feces, fungi, human, soil, symbiosis	Bacteria (15 occurrences, 19 links, 42 total link strength)
5	9	*de novo* sequencing, functional analysis, inflammation, microbial community, microbiology, protein extraction, shotgun proteomics, soil metatranscriptomics, soil protein extraction	Microbial community (93 occurrences, 46 links, 212 total link strength)
6	8	Human gut microbiome, liquid chromatography, machine learning, mass spectrometry, peptide identification, quantitative proteomics, system biology, tandem mass spectrometry	Mass spectrometry (72 occurrences, 44 links, 207 total link strength)
7	7	Covid-19, bioinformatics, gut microbiome, human microbiome, metaproteomics, protein identification, sample preparation	Metaproteomics (480 occurrences, 79 links, 937 total link strength)
8	5	Environmental microbiology, host–microbe interaction, metabolism, Protein-SIP, proteome	Protein-SIP (14 occurrences, 13 links, 28 total link strength)
9	5	Activated sludge, oral microbiome, proteins, saliva, wastewater treatment	Saliva (13 occurrences, 17 links, 41 total link strength)

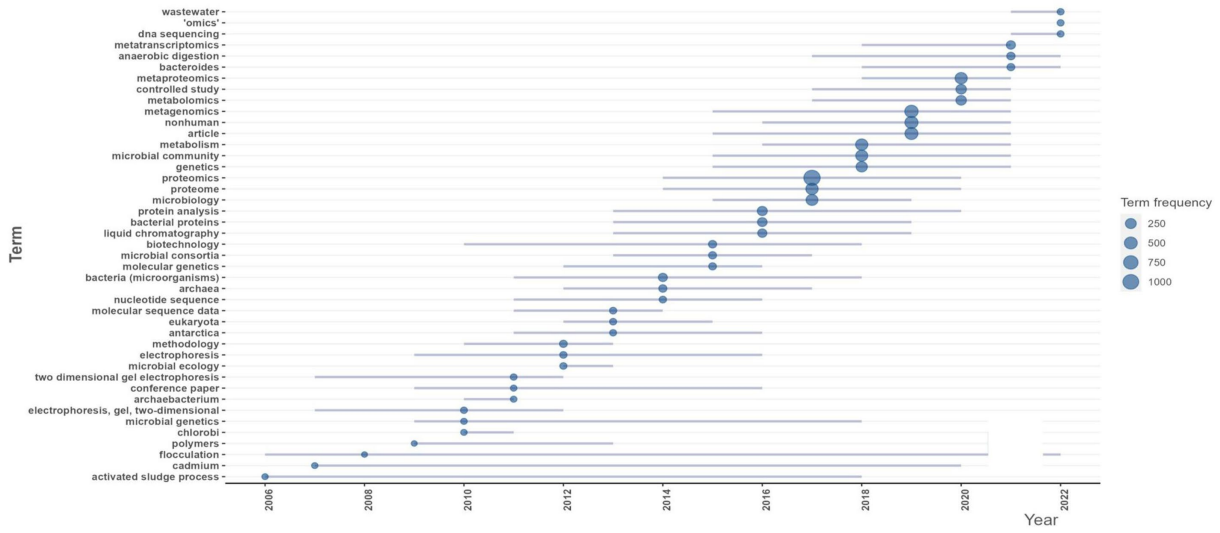


FIGURE 7
Global trending topics in metaproteomics research.

challenges by studying the intricate dynamics of microbial communities and their functions in various fields. Further, the application of metaproteomics in studying soil microbiomes, bioremediation, and biodegradation, for instance, can contribute to the development of policies related to sustainable agriculture, effective waste management, and environmental conservation.

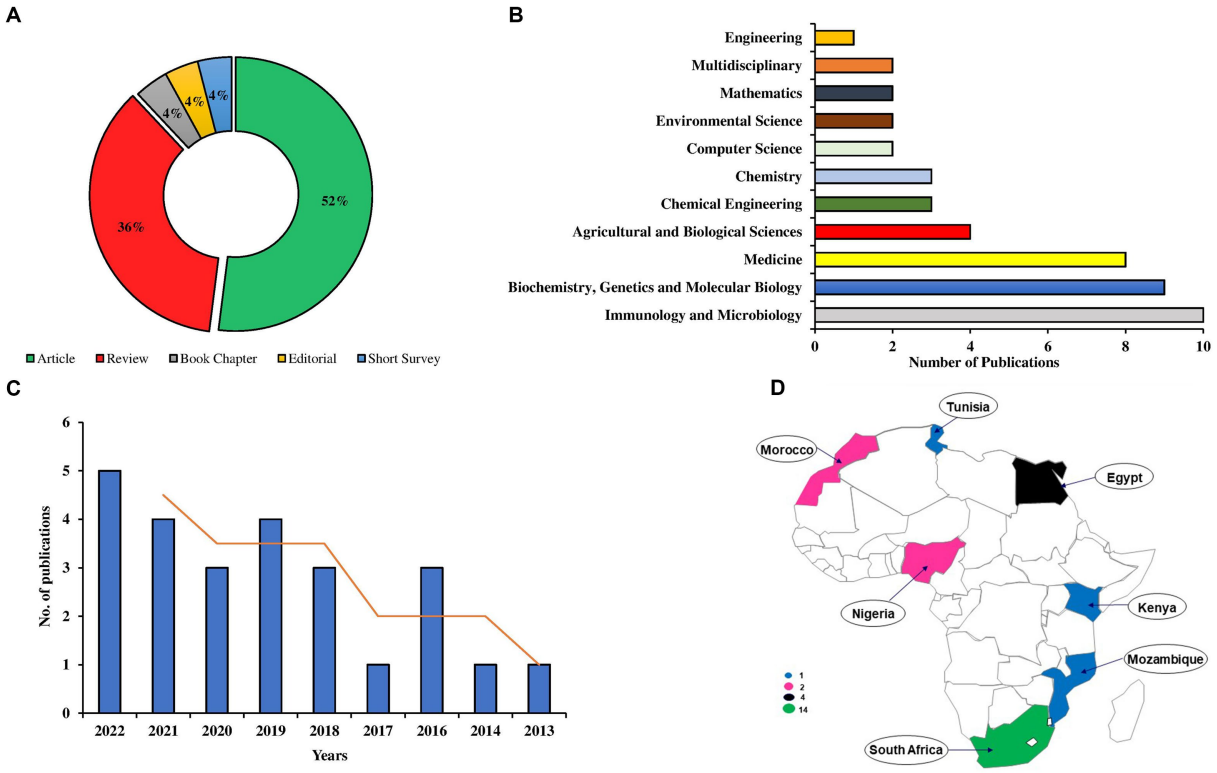


FIGURE 8
Analyses of the African metaproteomics publication including the document types **(A)**, leading subject areas **(B)**, yearly publication trend **(C)**, and geographical distribution of publication output **(D)**.

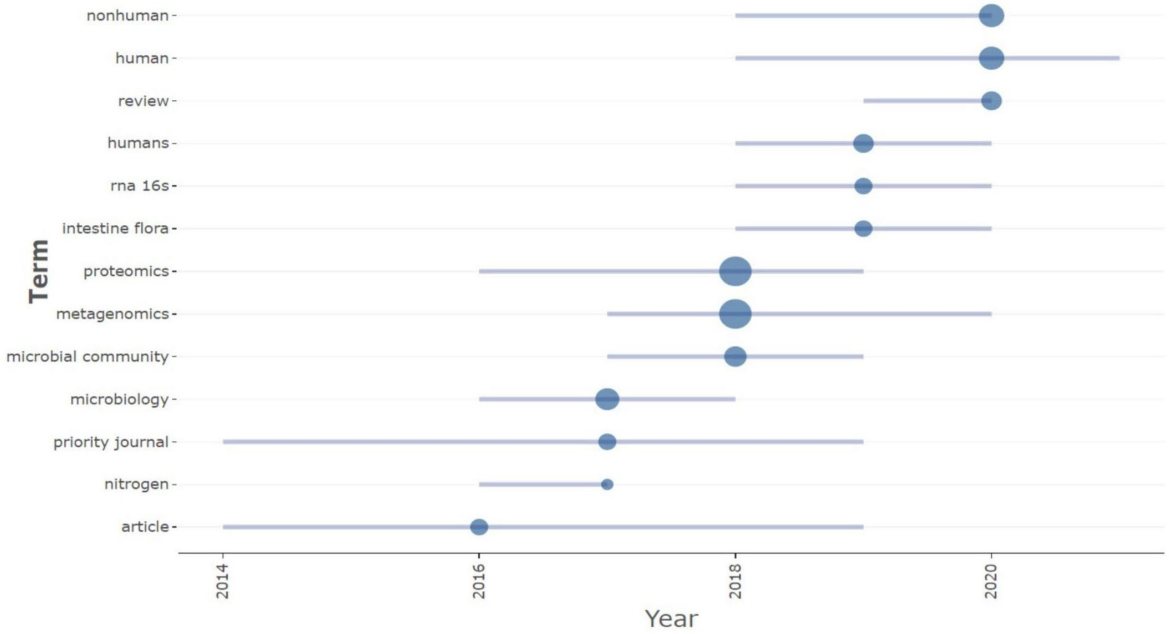


FIGURE 9
Trending topics in metaproteomics research in Africa.

TABLE 6 Field weighted citation impact (FWCI) of African-affiliated publications in metaproteomics.

Ranking	Author and their publication year	Total citations	Principal affiliated African institution	Country	Field weighted citation impact (FWCI)
1	Blackburn and Martens (2016)	14	Institute of Infectious Disease and Molecular Medicine, Department of Integrative Biomedical Sciences, University of Cape Town, Cape Town, South Africa	South Africa	3.5
2	Lau et al. (2016)	117	Department of Microbial Biochemical, and Food Biotechnology, University of Free State, Bloemfontein, 9301, South Africa	South Africa	3.2
3	Malan-Muller et al. (2018)	89	Department of Psychiatry, Faculty of Medicine and Health Sciences, Stellenbosch University, Tygerberg, 7600, South Africa	South Africa	2.4
4	Liu et al. (2019)	25	International Livestock Research Institute (ILRI), Mazingira Centre for Environmental Research and Education, Box 30709, Nairobi, 00100, Kenya	Kenya	1.9
5	Moussa et al. (2022)	4	Department of Clinical Pathology, School of Medicine, Mansoura University, Mansoura, Egypt	Egypt	1.8
6	Andrés-Barrao et al. (2016)	36	Agricultural Genetic Engineering Research Institute (AGERI), Agricultural Research Center (ARC), Giza, Egypt	Egypt	1.5
7	Abdool Karim et al. (2019)	28	Centre for the AIDS Programme of Research in South Africa (CAPRISA), University of KwaZulu-Natal, Durban, South Africa	South Africa	1.4
8	Kantor et al. (2017)	29	Centre for Bioprocess Engineering Research, Department of Chemical Engineering, University of Cape Town, Rondebosch, 7701, South Africa	South Africa	1.3
9	Ugya et al. (2019)	8	Department of Environmental Management, Kaduna State University, Kaduna, Nigeria	Nigeria	1.2
10	Ekwanzala et al. (2021)	1	Department of Environmental, Water and Earth Sciences, Tshwane University of Technology, Pretoria, South Africa	South Africa	1.2
11	Daffonchio et al. (2013)	24	Laboratory of Microorganisms and Active Biomolecules, University of Tunis El Manar, Tunis, Tunisia	Tunisia	1.2
12	Magnabosco et al. (2018)	15	Department of Microbial, Biochemical and Food Biotechnology, University of the Free State Bloemfontein, Free State, 9300, South Africa	South Africa	1.1
13	Lahlali et al. (2021)	7	Plant Pathology Unit, Department of Plant Protection, Ecole Nationale d'Agriculture de Meknes, BP S/40, Meknes, 50001, Morocco	Morocco	1.0
14	Alisoltani et al. (2020)	9	Division of Medical Virology, Department of Pathology, University of Cape Town, Cape Town, 7925, South Africa	South Africa	0.8
15	Guo et al. (2018)	13	DST/NRF Centre of Excellence for Biomedical Tuberculosis Research, SAMRC Centre for Tuberculosis Research, Division of Molecular Biology and Human Genetics, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, 7505, South Africa	South Africa	0.8
16	Moosa et al. (2020)	19	Africa Health Research Institute, Durban, South Africa	South Africa	0.6

(Continued)

TABLE 6 (Continued)

Ranking	Author and their publication year	Total citations	Principal affiliated African institution	Country	Field weighted citation impact (FWCI)
17	Okeke et al. (2021)	6	Department of Biochemistry, Faculty of Biological Sciences, University of Nigeria, Nsukka, 410001, Enugu State, Nigeria	Nigeria	0.5
18	Gunnigle et al. (2014)	15	Centre for Microbial Ecology and Genomics (CMEG), Department of Genetics, University of Pretoria, South Africa	South Africa	0.5
19	Lukhele et al. (2020)	10	Nanotechnology and Water Sustainability Research Unit, College of Science Engineering and Technology, University of South Africa, Science Campus, Johannesburg, South Africa	South Africa	0.4
20	Bahule et al. (2022)	1	Center of Studies in Science and Technology (NECET), Universidade Rovuma, Niassa branch, Lichinga, Mozambique	Mozambique	0.3
21	Sehli et al. (2021)	3	Department of fundamental sciences, School of Medicine, Mohammed VI University of Health Sciences, Casablanca, Morocco	Morocco	0.3
22	Ezzeldin et al. (2019)	4	Proteomics and Metabolomics Unit, Department of Basic Research, Children's Cancer Hospital Egypt, Cairo, 57357, Egypt	Egypt	0.1
23	Delgado-Diaz et al. (2022)	0	Division of Medical Virology, Department of Pathology, University of Cape Town, Cape Town, 7925, South Africa	South Africa	0.0
24	Chigorimbo-Murefu et al. (2022)	0	Divisions of Medical Virology, Institute of Infectious Diseases and Molecular Medicine, University of Cape Town, Cape Town, South Africa	South Africa	0.0
25	Hirtz et al. (2022)	0	Higher Institute of Engineering and Technology, Alexandria, New Borg AlArab City, 21934, Egypt	Egypt	0.0

5.2. Limitations of the study

Although the Scopus and Web of Science databases have been recognized as the largest searchable collection of citations and abstracts in the field of literature research, it would be interesting to incorporate other databases such as PubMed, and Cochrane into the present analysis subsequently. Also, language bias introduced in the selection of documents may inadvertently exclude contributions from non-English-published articles. Moreover, refining and expanding the search strategy to include a broader range of relevant terms, synonyms, and variations could help in capturing more documents for the bibliometric analysis.

Data availability statement

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

Author contributions

RD, AE, and AA: conceptualization. AA and SA: methodology. AA, SA, NS, AE, and RD: formal analysis, investigation, and writing—review and editing. AA, SA, and RD: writing—original draft preparation. RD, AA, AE, and SA: resources. RD: supervision. All authors contributed to the article and approved the submitted version.

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

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2023.1217727/full#supplementary-material

References

Abafe, E. A., Bahta, Y. T., and Jordaan, H. (2022). Exploring biblioshiny for historical assessment of global research on sustainable use of water in agriculture. *Sustainability* 14:10651. doi: 10.3390/su141710651

Abdool Karim, S. S., Baxter, C., Passmore, J. S., McKinnon, L. R., and Williams, B. L. (2019). The genital tract and rectal microbiomes: their role in HIV susceptibility and prevention in women. *J. Int. AIDS Soc.* 22:e25300. doi: 10.1002/jia2.25300

Aishwarya, S., Rajalakshmi, S., and Veena Gayathri, K. (2022). "The proteome mapping—metabolic modeling, and functional elucidation of the microbiome in the remediation of dyes and treating industrial effluents" in *Metagenomics to bioremediation: applications, cutting edge tools, and future outlook*, eds Vineet K, Muhammad B, Sushil Kumar S, and Vinod Kumar G (Academic Press), 311–328.

Alisoltani, A., Manhanzva, M. T., Potgieter, M., Balle, C., Bell, L., Ross, E., et al. (2020). Microbial function and genital inflammation in young South African women at high risk of HIV infection. *Microbiome* 8:165. doi: 10.1186/s40168-020-00932-8

Andersen, T. O., Kunath, B. J., Hagen, L. H., Arntzen, M. Ø., and Pope, P. B. (2021). Rumen metaproteomics: closer to linking rumen microbial function to animal productivity traits. *Methods* 186, 42–51. doi: 10.1016/j.ymeth.2020.07.011

Andrés-Barrao, C., Saad, M. M., Cabello Ferrete, E., Bravo, D., Chappuis, M.-L., Ortega Pérez, R., et al. (2016). Metaproteomics and ultrastructure characterization of *Komagataeibacter* spp. involved in high-acid spirit vinegar production. *Food Microbiol.* 55, 112–122. doi: 10.1016/j.fm.2015.10.012

Aria, M., and Cuccurullo, C. (2017). Bibliometrix: an R-tool for comprehensive science mapping analysis. *J. Informet.* 11, 959–975. doi: 10.1016/j.joi.2017.08.007

Armengaud, J. (2023). Metaproteomics to understand how microbiota function: the crystal ball predicts a promising future. *Environ. Microbiol.* 25, 115–125. doi: 10.1111/1462-2920.16238

Baas, J., Schotten, M., Plume, A., Côté, G., and Karimi, R. (2020). Scopus as a curated, high-quality bibliometric data source for academic research in quantitative science studies. *Quant. Sci. Studies* 1, 377–386. doi: 10.1162/qss_a_00019

Bahule, C. E., Da Silva Martins, L. H., Chaúque, B. J. M., and Lopes, A. S. (2022). Metaproteomics as a tool to optimize the maize fermentation process. *Trends Food Sci. Technol.* 129, 258–265. doi: 10.1016/j.tifs.2022.09.017

Bailey, M., Thomas, A., Francis, O., Stokes, C., and Smidt, H. (2019). The dark side of technological advances in analysis of microbial ecosystems. *J. Anim. Sci. Biotechnol.* 10:49. doi: 10.1186/s40104-019-0357-2

Balotf, S., Wilson, R., Tegg, R. S., Nichols, D. S., and Wilson, C. R. (2022). Shotgun proteomics as a powerful tool for the study of the proteomes of plants, their pathogens, and plant-pathogen interactions. *Proteomes* 10:5. doi: 10.3390/proteomes10010005

Bastida, F., Jehmlich, N., Lima, K., Morris, B. E. L., Richnow, H. H., Hernández, T., et al. (2016a). The ecological and physiological responses of the microbial community from a semiarid soil to hydrocarbon contamination and its bioremediation using compost amendment. *J. Proteome* 135, 162–169. doi: 10.1016/j.jprot.2015.07.023

Bastida, F., Jehmlich, N., Torres, I. F., and García, C. (2018). The extracellular metaproteome of soils under semiarid climate: a methodological comparison of extraction buffers. *Sci. Total Environ.* 619-620, 707–711. doi: 10.1016/j.scitotenv.2017.11.134

Bastida, F., Torres, I. F., Moreno, J. L., Baldrian, P., Ondoño, S., Ruiz-Navarro, A., et al. (2016b). The active microbial diversity drives ecosystem multifunctionality and is physiologically related to carbon availability in Mediterranean semi-arid soils. *Mol. Ecol.* 25, 4660–4673. doi: 10.1111/mec.13783

Biemann, R., Buß, E., Benndorf, D., Lehmann, T., Schallert, K., Püttker, S., et al. (2021). Fecal metaproteomics reveals reduced gut inflammation and changed microbial metabolism following lifestyle-induced weight loss. *Biomol. Ther.* 11:726. doi: 10.3390/biom11050726

Blackburn, J. M., and Martens, L. (2016). The challenge of metaproteomic analysis in human samples. *Expert Rev. Proteomics* 13, 135–138. doi: 10.1586/14789450.2016.1135058

Bostanci, N., Grant, M., Bao, K., Silbereisen, A., Hetrodt, F., Manoil, D., et al. (2021). Metaproteome and metabolome of oral microbial communities. *Periodontol.* 85, 46–81. doi: 10.1111/prd.12351

Calabrese, F. M., Porrelli, A., Vacca, M., Comte, B., Nimptsch, K., Pinart, M., et al. (2022). Metaproteomics approach and pathway modulation in obesity and diabetes: a narrative review. *Nutrients* 14:47. doi: 10.3390/nu14010047

Chandran, H., Meena, M., and Sharma, K. (2020). Microbial biodiversity and bioremediation assessment through omics approaches. *Front. Environ. Chem.* 1:570326. doi: 10.3389/fenvc.2020.570326

Chang, B.-V., Fan, S.-N., Tsai, Y.-C., Chung, Y.-L., Tu, P.-X., and Yang, C.-W. (2018). Removal of emerging contaminants using spent mushroom compost. *Sci. Total Environ.* 634, 922–933. doi: 10.1016/j.scitotenv.2018.03.366

Chen, Y., Ping, Q., Li, D., Dai, X., and Li, Y. (2022). Comprehensive insights into the impact of pretreatment on anaerobic digestion of waste active sludge from perspectives of organic matter composition, thermodynamics, and multi-omics. *Water Res.* 226:119240. doi: 10.1016/j.watres.2022.119240

Chigorimbo-Murefu, N. T. L., Potgieter, M., Dzanibe, S., Gabazana, Z., Buri, G., Chawla, A., et al. (2022). A pilot study to show that asymptomatic sexually transmitted infections alter the foreskin epithelial proteome. *Front. Microbiol.* 13:928317. doi: 10.3389/fmicb.2022.928317

Chourey, K., and Hettich, R. L. (2018). "Utilization of a detergent-based method for direct microbial cellular lysis/proteome extraction from soil samples for metaproteomics studies" in *Microbial proteomics: methods and protocols*. ed. D. Becher (New York, NY: Springer), 293–302.

Cowan, D., Lebre, P., Amon, C., Becker, R., Boga, H., Boulangé, A., et al. (2022). Biogeographical survey of soil microbiomes across sub-Saharan Africa: structure, drivers, and predicted climate-driven changes. *Microbiome* 10:131. doi: 10.1186/s40168-022-01297-w

Daffonchio, D., Ferrer, M., Mapelli, F., Cherif, A., Lafraya, Á., Malkawi, H. I., et al. (2013). Bioremediation of southern Mediterranean oil polluted sites comes of age. *New Biotechnol.* 30, 743–748. doi: 10.1016/j.nbt.2013.05.006

Debyser, G., Aerts, M., Van Hecke, P., Mesuere, B., Duytschaever, G., Dawyndt, P., et al. (2019). "A method for comprehensive proteomic analysis of human faecal samples to investigate gut dysbiosis in patients with cystic fibrosis" in *Emerging sample treatments in proteomics*. ed. J.-L. Capelo-Martínez (Cham: Springer International Publishing), 137–160.

Delgado-Diaz, D. J., Jesaveluk, B., Hayward, J. A., Tyssen, D., Alisoltani, A., Potgieter, M., et al. (2022). Lactic acid from vaginal microbiota enhances cervicovaginal epithelial barrier integrity by promoting tight junction protein expression. *Microbiome* 10:141. doi: 10.1186/s40168-022-01337-5

Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., and Lim, W. M. (2021). How to conduct a bibliometric analysis: an overview and guidelines. *J. Bus. Res.* 133, 285–296. doi: 10.1016/j.jbusres.2021.04.070

Duong, V.-A., and Lee, H. (2023). Bottom-up proteomics: advancements in sample preparation. *Int. J. Mol. Sci.* 24:5350. doi: 10.3390/ijms24065350

Ejaz, H., Zeeshan, H. M., Ahmad, F., Bukhari, S. N. A., Anwar, N., Alanazi, A., et al. (2022). Bibliometric analysis of publications on the omicron variant from 2020 to 2022 in the Scopus database using R and VOSviewer. *Int. J. Environ. Res. Public Health* 19:12407. doi: 10.3390/ijerph191912407

Ekwanzala, M. D., Budeli, P., and Unuofin, J. O. (2021). "Chapter 8—application of metatranscriptomics in wastewater treatment processes" in *Wastewater treatment*. eds. M. P. Shah, A. Sarkar and S. Mandal (Cham: Elsevier), 187–204.

El Jaddaoui, I., Allali, I., Sehli, S., Ouldim, K., Hamdi, S., Al Idrissi, N., et al. (2020). Cancer omics in Africa: present and prospects. *Front. Oncol.* 10:606428. doi: 10.3389/fonc.2020.606428

Ezzeldin, S., El-Wazir, A., Enany, S., Muhammad, A., Johar, D., Osama, A., et al. (2019). Current understanding of human metaproteome association and modulation. *J. Proteome Res.* 18, 3539–3554. doi: 10.1021/acs.jproteome.9b00301

Graf, A. C., Striesow, J., Pané-Farré, J., Sura, T., Wurster, M., Lalk, M., et al. (2021). An innovative protocol for metaproteomic analyses of microbial pathogens in cystic fibrosis sputum. *Front. Cell. Infect. Microbiol.* 11:724569. doi: 10.3389/fcimb.2021.724569

Gunasekaran, V., Canela, N., and Constantí, M. (2022). Comparative proteomic analysis of an ethyl tert-butyl ether-degrading bacterial consortium. *Microorganisms* 10, 10:2331. doi: 10.3390/microorganisms10122331

Gunnigle, E., Ramond, J.-B., Frossard, A., Seeley, M., and Cowan, D. (2014). A sequential co-extraction method for DNA, RNA and protein recovery from soil for future system-based approaches. *J. Microbiol. Methods* 103, 118–123. doi: 10.1016/j.mimet.2014.06.004

Guo, K., Li, J., Li, X., Huang, J., and Zhou, Z. (2023). Emerging trends and focus on the link between gut microbiota and type 1 diabetes: A bibliometric and visualization analysis. *Front. Microbiol.* 14:1137595. doi: 10.3389/fmicb.2023.1137595

Guo, X., Li, Z., Yao, Q., Mueller, R. S., Eng, J. K., Tabb, D. L., et al. (2018). Sipros ensemble improves database searching and filtering for complex metaproteomics. *Bioinformatics* 34, 795–802. doi: 10.1093/bioinformatics/btx601

Hamdi, Y., Zass, L., Othman, H., Radouani, F., Allali, I., Hanachi, M., et al. (2021). Human OMICs and computational biology research in Africa: current challenges and prospects. *OMICS* 25, 213–233. doi: 10.1089/omi.2021.0004

Hardouin, P., Chiron, R., Marchandin, H., Armengaud, J., and Grenga, L. (2021). Metaproteomics to decipher CF host-microbiota interactions: overview, challenges and future perspectives. *Genes* 12:892. doi: 10.3390/genes12060892

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

Henry, C., Bassignani, A., Berland, M., Langella, O., Sokol, H., and Juste, C. (2022). Modern metaproteomics: a unique tool to characterize the active microbiome in health and diseases, and pave the road towards new biomarkers—example of Crohn's disease and ulcerative colitis flare-ups. *Cells* 11:1340. doi: 10.3390/cells11081340

Heyer, R., Schallert, K., Siewert, C., Kohrs, F., Greve, J., Maus, I., et al. (2019). Metaproteome analysis reveals that syntrophy, competition, and phage-host interaction shape microbial communities in biogas plants. *Microbiome* 7:69. doi: 10.1186/s40168-019-0673-y

Hirtz, C., Mannaa, A. M., Moulis, E., Pible, O., O'Flynn, R., Armengaud, J., et al. (2022). Deciphering black extrinsic tooth stain composition in children using metaproteomics. *ACS Omega* 7, 8258–8267. doi: 10.1021/acsomega.1c04770

Iskandar, K., Molinier, L., Hallit, S., Sartelli, M., Hardcastle, T. C., Haque, M., et al. (2021). Surveillance of antimicrobial resistance in low- and middle-income countries: a scattered picture. *Antimicrobial Resistance Infection Control* 10:63. doi: 10.1186/s13756-021-00931-w

Issa Isaac, N., Philippe, D., Nicholas, A., Raoult, D., and Eric, C. (2019). Metaproteomics of the human gut microbiota: challenges and contributions to other OMICS. *Clin. Mass Spectr.* 14, 18–30. doi: 10.1016/j.clinms.2019.06.001

Jagtap, P., McGowan, T., Bandhakavi, S., Tu, Z. J., Seymour, S., Griffin, T. J., et al. (2012). Deep metaproteomic analysis of human salivary supernatant. *Proteomics* 12, 992–1001. doi: 10.1002/pmic.201100503

Jehmlich, N., Kleinsteuber, S., Vogt, C., Benndorf, D., Harms, H., Schmidt, F., et al. (2010). Phylogenetic and proteomic analysis of an anaerobic toluene-degrading community. *J. Appl. Microbiol.* 109, 1937–1945. doi: 10.1111/j.1365-2672.2010.04823.x

Jiang, N., Wu, R., Wu, C., Wang, R., Wu, J., and Shi, H. (2022). Multi-omics approaches to elucidate the role of interactions between microbial communities in cheese flavor and quality. *Food Rev. Intl.*, 1–13. doi: 10.1080/87559129.2022.2070199

Jiao, H., Huang, Z., Chen, Z., Wang, H., Liu, H., and Wei, Z. (2022). Lead removal in flue gas from sludge incineration by denitrification: insights from metagenomics and metaproteomics. *Ecotoxicol. Environ. Saf.* 244:114059. doi: 10.1016/j.ecoenv.2022.114059

Joyce, A., Ijaz, U. Z., Nzeteu, C., Vaughan, A., Shirran, S. L., Botting, C. H., et al. (2018). Linking microbial community structure and function during the acidified anaerobic digestion of grass. *Front. Microbiol.* 9:540. doi: 10.3389/fmicb.2018.00540

Kantor, R. S., Huddy, R. J., Iyer, R., Thomas, B. C., Brown, C. T., Anantharaman, K., et al. (2017). Genome-resolved meta-omics ties microbial dynamics to process performance in biotechnology for thiocyanate degradation. *Environ. Sci. Technol.* 51, 2944–2953. doi: 10.1021/acs.est.6b04477

Kleiner, M. (2019). Metaproteomics: much more than measuring gene expression in microbial communities. *mSystems* 4, 1–6. doi: 10.1128/msystems.00115-19

Kumar, V., Singh, K., Shah, M. P., Singh, A. K., Kumar, A., and Kumar, Y. (2021). "Chapter 1—application of Omics Technologies for Microbial Community Structure and Function Analysis in contaminated environment" in *Wastewater treatment*. eds. M. P. Shah, A. Sarkar and S. Mandal (Cham: Elsevier), 1–40.

Kushairi, N., and Ahmi, A. (2021). Flipped classroom in the second decade of the Millenia: A Bibliometrics analysis with Lotka's law. *Educ. Inf. Technol.* 26, 4401–4431. doi: 10.1007/s10639-021-10457-8

Lahlali, R., Ibrahim, D. S. S., Belabess, Z., Kadir Roni, M. Z., Radouane, N., Vicente, C. S. L., et al. (2021). High-throughput molecular technologies for unraveling the mystery of soil microbial community: challenges and future prospects. *Heliyon* 7:e08142. doi: 10.1016/j.heliyon.2021.e08142

Lam, T.-K., Yang, J.-T., Lai, S.-J., Liang, S.-Y., and Wu, S.-H. (2021). Meta-proteomics analysis of microbial ecosystem during the anaerobic digestion of chicken manure in biogas production farm. *Bioresour. Technol. Rep.* 13:100643. doi: 10.1016/j.biteb.2021.100643

Lau, M. C. Y., Kieft, T. L., Kuloyo, O., Linage-Alvarez, B., van Heerden, E., Lindsay, M. R., et al. (2016). An oligotrophic deep-subsurface community dependent on syntrophy is dominated by sulfur-driven autotrophic denitrifiers. *Proc. Natl. Acad. Sci.* 113, E7927–E7936. doi: 10.1073/pnas.1612244113

Lehmann, T., Schallert, K., Vilchez-Vargas, R., Benndorf, D., Püttker, S., Sydor, S., et al. (2019). Metaproteomics of fecal samples of Crohn's disease and ulcerative colitis. *J. Proteome* 201, 93–103. doi: 10.1016/j.jprot.2019.04.009

Liu, D., Keiblinger, K. M., Leitner, S., Wegner, U., Zimmermann, M., Fuchs, S., et al. (2019). Response of microbial communities and their metabolic functions to drying–rewetting stress in a temperate forest soil. *Microorganisms* 7:5. doi: 10.3390/microorganisms7050129

Lukhele, T., Selvarajan, R., Nyoni, H., Mamba, B. B., and Msagati, T. A. M. (2020). Acid mine drainage as habitats for distinct microbiomes: current knowledge in the era of molecular and omic technologies. *Curr. Microbiol.* 77, 657–674. doi: 10.1007/s00284-019-01771-z

Maghini, D. G., Moss, E. L., Vance, S. E., and Bhatt, A. S. (2021). Improved high-molecular-weight DNA extraction, nanopore sequencing and metagenomic assembly from the human gut microbiome. *Nat. Protoc.* 16, 458–471. doi: 10.1038/s41596-020-00424-x

Magnabosco, C., Timmers, P. H. A., Lau, M. C. Y., Borgonie, G., Linage-Alvarez, B., Kuloyo, O., et al. (2018). Fluctuations in populations of subsurface methane oxidizers in coordination with changes in electron acceptor availability. *FEMS Microbiol. Ecol.* 94:fiy089. doi: 10.1093/femsec/fiy089

Malan-Muller, S., Valles-Colomer, M., Raes, J., Lowry, C. A., Seedat, S., and Hemmings, S. M. J. (2018). The gut microbiome and mental health: implications for anxiety- and trauma-related disorders. *OMICS* 22, 90–107. doi: 10.1089/omi.2017.0077

Mauger, S., Monard, C., Thion, C., and Vandenkoornhuyse, P. (2022). Contribution of single-cell omics to microbial ecology. *Trends Ecol. Evol.* 37, 67–78. doi: 10.1016/j.tree.2021.09.002

Md Khudzari, J., Kurian, J., Tartakovsky, B., and Raghavan, G. S. V. (2018). Bibliometric analysis of global research trends on microbial fuel cells using Scopus database. *Biochem. Eng. J.* 136, 51–60. doi: 10.1016/j.bej.2018.05.002

Mishra, S., Lin, Z., Pang, S., Zhang, W., Bhatt, P., and Chen, S. (2021). Recent advanced technologies for the characterization of xenobiotic-degrading microorganisms and microbial communities. *Front. Bioeng. Biotechnol.* 9:632059. doi: 10.3389/fbioe.2021.632059

Moon, C., Stupp, G. S., Su, A. I., and Wolan, D. W. (2018). Metaproteomics of colonic microbiota unveils discrete protein functions among colitic mice and control groups. *Proteomics* 18:1700391. doi: 10.1002/pmic.201700391

Moosa, Y., Kwon, D., de Oliveira, T., and Wong, E. B. (2020). Determinants of vaginal microbiota composition. *Front. Cell. Infect. Microbiol.* 10:467. doi: 10.3389/fcimb.2020.00467

Moreira, M., Schrama, D., Farinha, A. P., Cerqueira, M., Raposo de Magalhães, C., Carrilho, R., et al. (2021). Fish pathology research and diagnosis in aquaculture of farmed fish; a proteomics perspective. *Animals* 11:125. doi: 10.3390/ani11010125

Moussa, D. G., Ahmad, P., Mansour, T. A., and Siqueira, W. L. (2022). Current state and challenges of the global outcomes of dental caries research in the meta-omics era. *Front. Cell. Infect. Microbiol.* 12:887907. doi: 10.3389/fcimb.2022.887907

Nephali, L., Piater, L. A., Dubery, I. A., Patterson, V., Huyser, J., Burgess, K., et al. (2020). Biostimulants for plant growth and mitigation of abiotic stresses: a metabolomics perspective. *Meta* 10:505. doi: 10.3390/metabo10120505

Nowrotek, M., Jałowiecki, Ł., Harnisz, M., and Płaza, G. A. (2019). Culturomics and metagenomics: in understanding of environmental resistome. *Front. Environ. Sci. Eng.* 13:40. doi: 10.1007/s11783-019-1121-8

Okeke, E. S., Ita, R. E., Egong, E. J., Udofia, L. E., Mgbechidinma, C. L., and Akan, O. D. (2021). Metaproteomics insights into fermented fish and vegetable products and associated microbes. *Food Chem.* 3:100045. doi: 10.1016/j.fochms.2021.100045

Purohit, J., Chattopadhyay, A., and Teli, B. (2020). Metagenomic exploration of plastic degrading microbes for biotechnological application. *Curr. Genomics* 21, 253–270. doi: 10.2174/1389202921999200525155711

Ragazou, K., Passas, I., Garefalakis, A., and Dimou, I. (2022). Investigating the research trends on strategic ambidexterity, agility, and open innovation in SMEs: perceptions from Bibliometric analysis. *J. Open Innov. Technol. Market Complexity* 8:118. doi: 10.3390/joitmc8030118

Rane, N. R., Tapase, S., Kanojia, A., Watharkar, A., Salama, E.-S., Jang, M., et al. (2022). Molecular insights into plant–microbe interactions for sustainable remediation of contaminated environment. *Bioresour. Technol.* 344:126246. doi: 10.1016/j.biortech.2021.126246

Rechenberger, J., Samaras, P., Jarzab, A., Behr, J., Frejno, M., Djukovic, A., et al. (2019). Challenges in clinical metaproteomics highlighted by the analysis of acute leukemia patients with gut colonization by multidrug-resistant enterobacteriaceae. *Proteomes* 7:2. doi: 10.3390/proteomes7010002

Renu, , Gupta, S. K., Rai, A. K., Sarim, K. M., Sharma, A., Budhlakoti, N., et al. (2019). Metaproteomic data of maize rhizosphere for deciphering functional diversity. *Data Brief* 27:104574. doi: 10.1016/j.dib.2019.104574

Roldan-Valadez, E., Salazar-Ruiz, S. Y., Ibarra-Contreras, R., and Rios, C. (2019). Current concepts on bibliometrics: a brief review about impact factor, Eigenfactor score, CiteScore, SCImago journal rank, source-normalised impact per paper, H-index, and alternative metrics. *Irish J. Med. Sci.* 188, 939–951. doi: 10.1007/s11845-018-1936-5

Saito, M. A., Bertrand, E. M., Duffy, M. E., Gaylord, D. A., Held, N. A., Hervey, W. J. I., et al. (2019). Progress and challenges in ocean metaproteomics and proposed best practices for data sharing. *J. Proteome Res.* 18, 1461–1476. doi: 10.1021/acs.jproteome.8b00761

Salvato, F., Hettich, R. L., and Kleiner, M. (2021). Five key aspects of metaproteomics as a tool to understand functional interactions in host-associated microbiomes. *PLoS Pathog.* 17:e1009245. doi: 10.1371/journal.ppat.1009245

Saralegui, C., García-Durán, C., Romeu, E., Hernáez-Sánchez, M. L., Maruri, A., Bastón-Paz, N., et al. (2022). Statistical evaluation of metaproteomics and 16S rRNA amplicon sequencing techniques for study of gut microbiota establishment in infants with cystic fibrosis. *Microbiol. Spectr.* 10:e0146622. doi: 10.1128/spectrum.01466-22

Schiebenhoefer, H., Van Den Bossche, T., Fuchs, S., Renard, B. Y., Muth, T., and Martens, L. (2019). Challenges and promise at the interface of metaproteomics and genomics: an overview of recent progress in metaproteogenomic data analysis. *Expert Rev. Proteomics* 16, 375–390. doi: 10.1080/14789450.2019.1609944

Sehli, S., Allali, I., Chahboune, R., Bakri, Y., Al Idrissi, N., Hamdi, S., et al. (2021). Metagenomics approaches to investigate the gut microbiome of COVID-19 patients. *Bioinform. Biol. Insights* 15:1177932221999428. doi: 10.1177/1177932221999428

Seyhan, A. A., and Carini, C. (2019). Are innovation and new technologies in precision medicine paving a new era in patients centric care? *J. Transl. Med.* 17:114. doi: 10.1186/s12967-019-1864-9

Shakya, M., Lo, C.-C., and Chain, P. S. G. (2019). Advances and challenges in metatranscriptomic analysis. *Front. Genet.* 10:904. doi: 10.3389/fgene.2019.00904

Sun, Z., Song, Z.-G., Liu, C., Tan, S., Lin, S., Zhu, J., et al. (2022). Gut microbiome alterations and gut barrier dysfunction are associated with host immune homeostasis in COVID-19 patients. *BMC Med.* 20:24. doi: 10.1186/s12916-021-02212-0

Tahamtan, I., and Bornmann, L. (2019). What do citation counts measure? An updated review of studies on citations in scientific documents published between 2006 and 2018. *Scientometrics* 121, 1635–1684. doi: 10.1007/s11192-019-03243-4

Taylor, M. J., Lukowski, J. K., and Anderton, C. R. (2021). Spatially resolved mass spectrometry at the single cell: recent innovations in proteomics and metabolomics. *J. Am. Soc. Mass Spectrom.* 32, 872–894. doi: 10.1021/jasms.0c00439

Tiwari, A., and Taj, G. (2020). "Applications of advanced omics technology for harnessing the high altitude agriculture production" in *Microbiological advancements for higher altitude agro-ecosystems & sustainability*. eds. R. Goel, R. Soni and D. C. Suyal (New York, NY: Springer), 447–463.

Ugya, A. Y., Hua, X., Agamuthu, P., and Ma, J. (2019). Molecular approach to uncover the function of bacteria in petrochemical refining wastewater: a mini review. *Appl. Ecol. Environ. Res.* 17, 3645–3665. doi: 10.15666/aeer/1702_36453665

Van Den Bossche, T., Arntzen, M. Ø., Becher, D., Benndorf, D., Eijsink, V. G. H., Henry, C., et al. (2021). The metaproteomics initiative: a coordinated approach for propelling the functional characterization of microbiomes. *Microbiome* 9:243. doi: 10.1186/s40168-021-01176-w

Vargas Medina, D. A., Maciel, E. V. S., de Toffoli, A. L., and Lanças, F. M. (2020). Miniaturization of liquid chromatography coupled to mass spectrometry: 2. Achievements on modern instrumentation for miniaturized liquid chromatography coupled to mass spectrometry. *TrAC Trends Anal. Chem.* 128:115910. doi: 10.1016/j.trac.2020.115910

Wang, Q., Li, R., and Zhan, L. (2021). Blockchain technology in the energy sector: from basic research to real world applications. *Comput. Sci. Rev.* 39:100362. doi: 10.1016/j.cosrev.2021.100362

White, R. A., Rosnow, J., Piehowski, P. D., Brislawn, C. J., and Moran, J. J. (2021). In situ non-destructive temporal measurements of the rhizosphere microbiome 'hot-spots' using metaproteomics. *Agronomy* 11:2248. doi: 10.3390/agronomy11112248

Wilmes, P., and Bond, P. L. (2004). The application of two-dimensional polyacrylamide gel electrophoresis and downstream analyses to a mixed community of prokaryotic microorganisms. *Environ. Microbiol.* 6, 911–920. doi: 10.1111/j.1462-2920.2004.00687.x

Xie, X., Zheng, H., Zhang, Q., Fan, J., Liu, N., and Song, X. (2022). Co-metabolic biodegradation of structurally discrepant dyestuffs by *Klebsiella* sp. KL-1: a molecular mechanism with regards to the differential responsiveness. *Chemosphere* 303:135028. doi: 10.1016/j.chemosphere.2022.135028

Yu, D., Wang, W., Zhang, W., and Zhang, S. (2018). A bibliometric analysis of research on multiple criteria decision making. *Curr. Sci.* 114, 747–758. doi: 10.18520/cs/v114/i04/747-758

Yu, D., Xu, Z., and Pedrycz, W. (2020). Bibliometric analysis of rough sets research. *Appl. Soft Comput.* 94:106467. doi: 10.1016/j.asoc.2020.106467

Yu, D., Xu, Z., and Wang, W. (2018). Bibliometric analysis of fuzzy theory research in China. *Knowl. Based Syst.* 141, 188–199. doi: 10.1016/j.knosys.2017.11.018

Zanotto, E. D., and Carvalho, V. (2021). Article age- and field-normalized tools to evaluate scientific impact and momentum. *Scientometrics* 126, 2865–2883. doi: 10.1007/s11192-021-03877-3

Zhang, X., Deeke, S. A., Ning, Z., Starr, A. E., Butcher, J., Li, J., et al. (2018). Metaproteomics reveals associations between microbiome and intestinal extracellular vesicle proteins in pediatric inflammatory bowel disease. *Nat. Commun.* 9:2873. doi: 10.1038/s41467-018-05357-4

Zhang, X., Li, L., Butcher, J., Stintzi, A., and Figeys, D. (2019). Advancing functional and translational microbiome research using meta-omics approaches. *Microbiome* 7:154. doi: 10.1186/s40168-019-0767-6

Zhang, Y., Liu, H., Kang, S.-C., and Al-Hussein, M. (2020). Virtual reality applications for the built environment: research trends and opportunities. *Autom. Constr.* 118:103311. doi: 10.1016/j.autcon.2020.103311

Zheng, Q., Wang, Y., Lu, J., Lin, W., Chen, F., and Jiao, N. (2020). Metagenomic and metaproteomic insights into photoautotrophic and heterotrophic interactions in a *Synechococcus* culture. *mBio* 11:e03261-19. doi: 10.1128/mBio.03261-19

Zhong, H., Ren, H., Lu, Y., Fang, C., Hou, G., Yang, Z., et al. (2019). Distinct gut metagenomics and metaproteomics signatures in prediabetics and treatment-naïve type 2 diabetics. *EBioMedicine* 47, 373–383. doi: 10.1016/j.ebiom.2019.08.048

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 20 November 2023
DOI 10.3389/fmicb.2023.1226945

OPEN ACCESS

EDITED BY
Janmeajy Pandey,
Central University of Rajasthan, India

REVIEWED BY
Zothanpuia,
Pachhunga University College, India
Sanjay Kumar Singh Patel,
Konkuk University, Republic of Korea

*CORRESPONDENCE
Yuki Inahashi
✉ y-ina@lisci.kitasato-u.ac.jp
Bungonsiri Intra
✉ bungonsiri.int@mahidol.edu

RECEIVED 22 May 2023
ACCEPTED 23 October 2023
PUBLISHED 20 November 2023

CITATION
Ngamcharungchit C, Matsumoto A, Suriyachadkun C, Panbangred W, Inahashi Y and Intra B (2023) *Nonomuraea corallina* sp. nov., isolated from coastal sediment in Samila Beach, Thailand: insights into secondary metabolite synthesis as anticancer potential.
Front. Microbiol. 14:1226945.
doi: 10.3389/fmicb.2023.1226945

Nonomuraea corallina sp. nov., isolated from coastal sediment in Samila Beach, Thailand: insights into secondary metabolite synthesis as anticancer potential

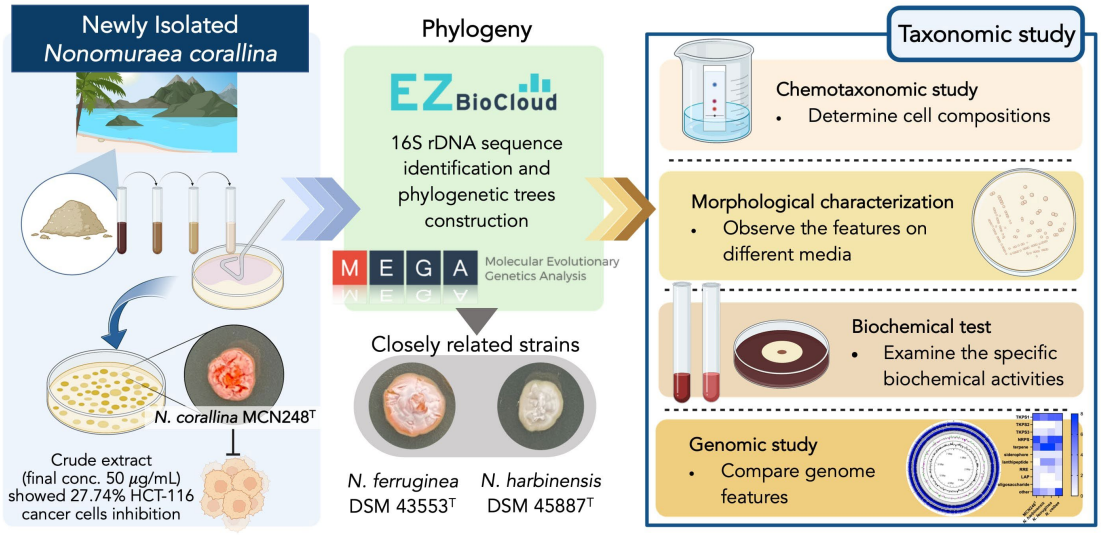
Chananan Ngamcharungchit[1,2], Atsuko Matsumoto[3,4], Chanwit Suriyachadkun[5], Watanalai Panbangred[6], Yuki Inahashi[3,4]* and Bungonsiri Intra[1,2]*

[1]Department of Biotechnology, Faculty of Science, Mahidol University, Bangkok, Thailand, [2]Mahidol University and Osaka Collaborative Research Center on Bioscience and Biotechnology, Bangkok, Thailand, [3]Graduate School of Infection Control Sciences, Kitasato University, Tokyo, Japan, [4]Kitasato Institute for Life Sciences (Ōmura Satoshi Memorial Institute), Kitasato University, Tokyo, Japan, [5]Thailand Bioresource Research Center (TBRC), National Science and Technology Development Agency, Pathumthani, Thailand, [6]Research, Innovation and Partnerships Office – RIPO (Office of the President), King Mongkut's University of Technology Thonburi, Bangkok, Thailand

A novel marine actinomycete, designated strain MCN248^T, was isolated from the coastal sediment in Songkhla Province, Thailand. Based on the 16S rRNA gene sequences, the new isolate was closely related to *Nonomuraea harbinensis* DSM45887^T (99.2%) and *Nonomuraea ferruginea* DSM43553^T (98.6%). Phylogenetic analyzes based on the 16S rRNA gene sequences showed that strain MCN248^T was clustered with *Nonomuraea harbinensis* DSM45887^T and *Nonomuraea ferruginea* DSM43553^T. However, the digital DNA–DNA hybridization analyzes presented a low relatedness of 40.2% between strain MCN248^T and the above closely related strains. This strain contained meso-diaminopimelic acid. The acyl type of the peptidoglycan was acetyl, and mycolic acids were absent. The major menaquinones were MK-9(H_2) and MK-9(H_4). The whole cell sugars consisted of madurose, ribose, mannose, and glucose. Diphosphatidylglycerol, hydroxyl-phosphatidylethanolamine, phosphatidylethanolamine, phosphatidylinositol, and phosphatidylglycerol were detected as the major phospholipids. The predominant cellular fatty acids were *iso*-$C_{16:0}$ (40.4%), 10-methyl-$C_{17:0}$ (22.1%), and $C_{17:1}\,\omega$8c (10.9%). The DNA G + C content of the genomic DNA was 71.7%. With *in silico* analyzes, the antiSMASH platform uncovered a diverse 29 secondary metabolite biosynthesis arsenal, including non-ribosomal peptide synthetase (NRPS) and polyketide synthase (PKS) of strain MCN248^T, with a high prevalence of gene cluster encoding pathways for the production of anticancer and cytotoxic compounds. Consistently, the crude extract could inhibit colorectal HCT-116 cancer cells at a final concentration of 50 μg/mL. Based on the polyphasic approach, strain MCN248 was designated as a novel species of the genus *Nonomuraea*, for which the name *Nonomuraea corallina* sp. nov. is proposed. The type strain of the type species is MCN248^T (=NBRC115966^T = TBRC17110^T).

KEYWORDS

actinomycetes, novel marine taxa, phylogenetic analysis, *Nonomuraea*, anti-colorectal cancer



GRAPHICAL ABSTRACT
"Created with BioRender.com."

Introduction

Marine habitats have proven to be significant reservoirs of rare actinomycetes, including the genus *Nonomuraea*, in recent decades. These marine rare actinomycetes are renowned bioactive compound producers, since a total of 267 novel compounds were reported to be produced by them between 2007 and 2017 (Subramani and Sipkema, 2019). The genus *Nonomuraea* produces diverse bioactivity of secondary metabolites; for example, the antibacterial compound actinotiocin was first isolated from *Nonomuraea pusilla* IFO 14684^T in 1973 (Tamura et al., 1973). In addition, anticancer, antipsychotic, and biocatalytic compounds, as well as pigments, have been found to be produced by *Nonomuraea* spp. (Sungthong and Nakaew, 2015).

The genus *Nonomuraea* was included in the family Streptosporangiaceae by Quintana et al. (2003). The description of the family was subsequently emended by Stackebrandt et al. (1997), Zhi et al. (2009), Nouioui et al. (2018), and Ay et al. (2020), on the basis of the 16S rRNA gene sequence analysis and chemotaxonomic characteristics. At the time of writing this article, the genus *Nonomuraea* comprises 66 species with validly published names (LPSN, http://www.bacterio.net/nonomuraea.html). These species can be distinguished using a combination of chemotaxonomic, genomic, morphological, and phylogenetic criteria (Lechevalier and Lechevalier, 1970; Embley et al., 1998). They generally produce extensively branching aerial and substrate mycelia. Chains of aerial spores can be hooked, spiral, or straight, and the cell wall is composed of meso-diaminopimelic acid (meso-DAP) (Goodfellow et al., 1990). The major menaquinones are MK-9(H_4), MK-9(H_2), and MK-9(H_0), while the major phospholipids are diphosphatidylglycerol, phosphatidylethanolamine, hydroxylated phosphatidylethanolamine, and ninhydrin-positive phosphoglycolipids. The genomic DNA contains 64.0–73.0 mol% of G+C content (Kämpfer, 2012). The polyphasic approach combines various data sources, including morphological characteristics, DNA sequences, and ecological niches, to enhance the precision of species identification and delimitation (Yang et al., 2022). However, genomics has become a promising methodology as it provides a reproducible, reliable, highly informative means to infer phylogenetic relationships among prokaryotes, which allows the continuation of our tradition toward natural classification (Chun et al., 2018). In this study, we aimed to determine the taxonomic position of an actinomycete isolate that is a novel species of the genus *Nonomuraea* by using a polyphasic taxonomy.

Cancer possesses a significant global challenge to life expectancy. The worldwide cancer burden is projected to increase by 47% to 28.4 million cases by 2040, with transitioning countries facing an even greater increase, ranging from 64 to 95% (Sung et al., 2021). Anticancer drugs undergo metabolism through phase I and phase II metabolizing enzymes, involving oxidation/hydroxylation and hydrolysis (Muhamad and Na-Bangchang, 2020). For instance, curcumin regulates various cellular signaling pathways, including apoptosis, inhibiting progression, and blocking angiogenesis in pancreatic cancer (Kumar et al., 2023). Resistomycin inhibits Pellino-1, an E3 ubiquitin ligase that is reported to have an important role for lymphoid and several solid tumorigeneses, and the inhibition also leads to the downregulation of expression of transcription factors, SNAIL and SLUG, that contribute to tumor weight and lung metastasis in MDA-MB-231 cells (Liu et al., 2020).

Materials and methods

Isolation of actinomycetes and 2D anti-cancer screening

Strain MCN248^T was isolated from coastal sediment from Samila Beach, Songkhla Province, Thailand (GPS data 7.192455, 100.590469), using a dilution plating technique. A 10-fold dilution of sediment suspension was spread onto starch casein agar (g/L: soluble starch, 10; casein, 0.3; KNO_3, 2; $MgSO_4.7H_2O$, 0.05; K_2HPO_4, 2; NaCl, 2; $CaCO_3$, 0.02; $FeSO_4.7H_2O$, 0.01; and agar, 18) supplemented with

cycloheximide (50 mg ml^{-1}) and nalidixic acid (50 mg ml^{-1}) (Wattanasuepsin et al., 2017). The colonies were picked up after incubation for 6 weeks at 28°C. The pure culture was maintained on yeast extract–malt extract agar (International *Streptomyces* Project, ISP 2 medium) at 28°C and stored in 20% (v/v) glycerol at −80°C for long-term preservation.

For the 2D anti-colorectal cancer assay, the crude extract cultivated in A-3 M production medium (g/L: soluble starch, 20; glycerol, 20; pharmamedia, 15; dianion HP-20, 10; glucose, 5; and yeast extract, 3, pH 7) was prepared using ethanol/ethyl acetate extraction and then diluted in DMSO to prepare a stock concentration of 10 mg/mL. To conduct testing against cancer cell line, the crude extract was added to cell plates to obtain a final concentration of 50 µg/mL. High-throughput liquid handling and high-throughput detection system were used for screening compounds through the 3-(4, 5-dimethyl-thiazol-2-yl)-2, 5-diphenyl tetrazolium bromide (MTT) (colorimetric assay) method. The absorbance of this colored solution was measured at a wavelength of 570 nm by a Multi-Mode Microplate Reader (ENVISION). The optical density (OD) was used to calculate the percentage of cell inhibition.

16S rRNA sequence identification

PCR amplification was performed using a Perkin Elmer GeneAmp 2400 PCR system and a pair of 11F′ (5′-AGTTTG ATCATGGCTCAG-3′) and 1540R′ (5′-AAGGAGGTGATCCA ACCGCA-3′) universal primers. Sequencing of the 16S rRNA gene of strain MCN248^T was carried out by Macrogen, Inc, Korea. The values of sequence similarities among the most closely related strains were computed using the EzBioCloud server (Yoon et al., 2017a,b).[1] A nearly complete 16S rRNA gene sequence (1471 bp) of strain MCN248T was aligned with multiple sequences of available type strains in the genus *Nonomuraea* on the EZBioCloud database. Phylogenetic trees were constructed using neighbor-joining (Saitou and Nei, 1987) and maximum-likelihood (Felsenstein, 1981) tree-making algorithms in the software package MEGA (version 11) (Tamura et al., 2021). Evolutionary distance matrices were generated according to Kimura's two-parameter model (Kimura, 1980). The robustness of the tree topologies was assessed by performing bootstrap analysis with 1,000 replicates (Felsenstein, 1985).

Draft genomic sequencing and *in silico* analyzes

The extraction and purification of chromosomal DNA for DNA G + C content analysis were performed according to the method by Saito and Miura (1963). The draft genome of strain MCN248^T was sequenced using the paired-end method and the Illumina HiSeq platform. The G + C content of the genomic DNA of strain MCN248^T was calculated from the draft genome sequences. The genomic DNA sequence of *Nonomuraea harbinensis* DSM45887^T (GenBank accession no. JAHKRN000000000.1) was obtained from the NCBI database, while that of *Nonomuraea ferruginea* DSM43553^T (GenBank accession no. JAPNUD000000000) was first deposited at NCBI in this study. According to the phylogenomic tree construction (Meier-Kolthoff and Göker, 2019), based on genome data, a TYGS-genome blast distance phylogeny (GBDP) was generated using MCN248T and all the available genome data of *Nonomuraea* type strains in the TYGS database. The genomic similarities between strain MCN248^T and the above closely related strains were investigated using the average nucleotide identity (ANI) algorithm with the OrthoANIu tool from EZBioCloud software (Yoon et al., 2017a,b) and JSpeciesWS online services (Richter et al., 2015). Additionally, digital DDH (dDDH) analysis was conducted using the Genome-to-Genome Distance Calculator (GGDC) 2.1 platform (Meier-Kolthoff et al., 2013). Analysis of secondary metabolite biosynthesis gene clusters for rapid genome-wide identification was carried out using antiSMASH version 7.0.0beta1 (Medema et al., 2011).

Chemotaxonomic characterization

Biomass for chemical and molecular studies was harvested by centrifugation after cultivation in tryptic soy broth (TSB) (g/L: tryptone, 17; soytone, 3; sodium chloride, 5; dipotassium phosphate, 2.5; and glucose, 2.5, pH 7.3) at 28°C for 5 days. The purified cell wall was prepared according to the procedure described by Také et al. (2016). The whole-cell sugar composition was examined on cellulose plates using the TLC technique, which was performed following the procedures described by Staneck and Roberts (1974). The presence of mycolic acid was investigated using TLC according to the method by Tomiyasu (1982). Isoprenoid quinones were extracted and subsequently analyzed by liquid chromatography/mass spectrometry (JMS-T100LP, JEOL) with a CAPCELL PAK C18 UG120 column (OSAKA SODA) using methanol/isopropanol (7,3, v/v) and UV detection at 270 nm (Collins et al., 1977; Meyer, 1979; Kroppenstedt et al., 1990; Wang et al., 2014). Phospholipids in cells were extracted and analyzed by two-dimensional TLC according to the method by Minnikin et al. (1977). Methyl esters of cellular fatty acids were prepared by direct transmethylation with methanolic hydrochloride using cells grown in TSB broth for 7 days at 28°C. Cellular fatty acid compositions were identified using a GLC system (HP 6890; Hewlett Packard) assisted by the ACTIN 6 database, according to the instructions for the Sherlock Microbial Identification System (Microbial ID; MIDI, version 6.0) (Kämpfer and Kroppenstedt, 1996).

Phenotypic characterization

Cultural characteristics were observed after cultivation at 28°C for 14–21 days on International *Streptomyces* Project (ISP) media 2, 3, 4, 5, 6, and 7 (Difco or Nihon Pharmaceutical) (Shirling and Gottlieb, 1966). To specify the colors of aerial and substrate mycelia and diffusible pigment, the Color Harmony Manual was used for color designation (Jacobson et al., 1958). Morphological characteristics of strain MCN248 were observed using scanning electron microscopy (model JSM-5610, JEOL) after incubation on ISP 3 at 28°C for 21 days. Spore-producing samples of the strain were prepared according to previously described methods (Intra et al., 2013).

The temperature range (4–50°C) (temperature gradient incubator TN-2148, Advantec) and pH range (pH 3.0–11.0, at 1 pH unit

1 http://eztaxon-e.ezbiocloud.net

intervals) for growth and salinity (NaCl) tolerance [0–5% (w/v)] were examined on ISP 2 basal medium after 14 days of incubation. The carbon-source utilization was determined using the methods described by Shirling and Gottlieb (1966) and Williams et al. (1989). The capacity of the strain for starch hydrolysis, hydrogen sulfide production, melanin production, nitrate reduction, gelatin degradation, and casein hydrolysis was investigated using the following media: ISP 4, ISP 6, ISP 7, ISP 8 (0.5% peptone, 0.3% beef extract, and 0.5% KNO_3, pH 7.0), peptone-glucose-gelatin (2.0% glucose, 0.5% peptone, and 20% gelatin, pH 7.0), and skim milk agar (10% skim milk (Difco) and 1.5% agar, pH 7.0). For the coagulation and peptonization of milk, 10% skim milk (Difco) was used (Kovacs, 1956). A commercial kit API ZYM system (BioMérieux) was used to determine the enzymatic features of the strain according to the manufacturer's instructions.

Results and discussion

Phylogeny of Strain MCN248^T

Based on pairwise comparison of the almost-complete 16S rRNA gene sequences, strain MCN248^T displayed a close association between members of the genus *Nonomuraea*. The highest similarity value to strain MCN248^T was *N. harbinensis* DSM45887^T (99.2%). Congruent with these results, the neighbor-joining phylogenetic tree indicated that the strain MCN248^T was positioned within the genus *Nonomuraea*, where it formed a clade with *N. harbinensis* DSM45887^T and *N. ferruginea* DSM43553^T (Figure 1). The phylogenetic relationships of 40 *Nonomuraea* species were also supported by the neighbor-joining and maximum-likelihood trees, as shown in Supplementary Figures S1, S2, respectively.

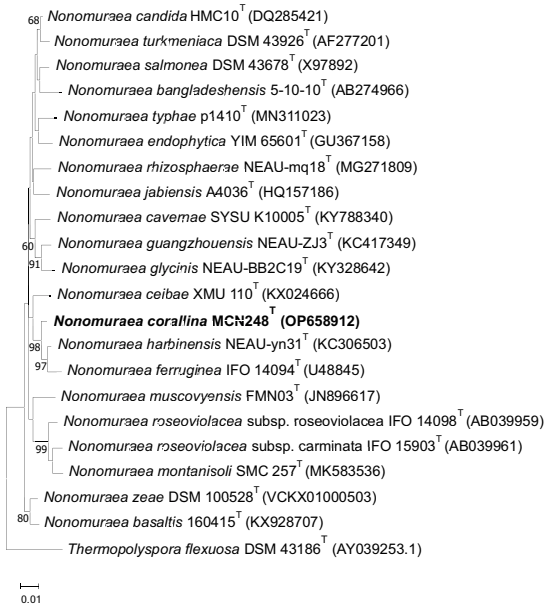


FIGURE 1
Neighbor-joining phylogenetic tree derived from 16S rRNA nucleotide sequences displaying the genetic relatedness between the MCN248^T isolate and other strains in the genus *Nonomuraea*, with *Thermopolyspora flexuosa* as the outgroup. Bootstrap values of 50% or higher are indicated at the branch points (percentages are based on 1,000 resamplings). The scale bar represents 0.01 nucleotide substitutions per site.

Draft genomic characterization and *in silico* secondary metabolite cluster profiles

The draft genome sequencing of strain MCN248^T yielded a genome of 7,557,982 bp (Figure 2) in length after assembly, which produced 370 contigs with an N50 value of 34,089 bp. The G + C content of strain MCN248^T, calculated from the draft genome sequence, was determined to be 71.7 mol% (Supplementary Table S1).

The dDDH value for comparing strain MCN248^T with both *N. harbinensis* DSM45887^T and *N. ferruginea* DSM43553^T was found to be 40.0 and 40.2%, respectively. The ANI values were in the range of 90.0 to 89.9% for these strains. Similarly, the ANIb values were 90.8% for both *N. harbinensis* and *N. ferruginea*, as determined by JSpeciesWS online services. All these values were below the threshold for bacterial species demarcation (Chun et al., 2018).

Phylogenomic tree construction based on the TYGS-genome blast distance phylogeny (GBDP) (Meier-Kolthoff and Göker, 2019) using MCN248^T and the available genome data of *Nonomuraea* type strains in the TYGS database supported the phylogenetic position of strain MCN248^T (Figure 3). Analysis with the antiSMASH tool revealed differences in the number and types of annotated biosynthetic gene clusters (BGCs) between *Nonomuraea corallina* MCN248^T and the closely related strains (*N. harbinensis*, *N. ferruginea*, and *N. ceibae*), as shown in the heatmap (Figure 4) and the table of numbers (Supplementary Table S2). The analyzed regions were mainly non-ribosomal peptide synthetase (NRPS), terpene, type I polyketide synthase (T1PKS), and other types. BGC numbers of MCN248^T obviously differ from related strains in lanthipeptide, PKS groups, and NRPS. Predicted genes involved in anticancer biosynthesis were identified in thr strain MCN248^T using the AntiSMASH database, highlighting its potential as a source of bioactive compounds. Notably, the gene cluster for the lipopeptides icosalide A/icosalide B showed 100% similarity, followed by the polyketide cytorhodin (37% similarity). Icosalides A and B were initially isolated from a fungal culture and reported for their cytotoxic activity against Madin-Darby canine kidney (MDCK) cells (Boros et al., 2006). Cytorhodins, belonging to the rhodomycin-antitumor compound group, were originally discovered in *Streptomyces* sp. HPL-Y11427 (Reddy et al., 1985; Hedtmann et al., 1992). These compounds have not yet been reported from *Nonomuraea* species. The discovery of these anticancer compound BGCs could lead to the discovery of new targets. For example, the novel glycopeptide A59926 was isolated from *Nonomuraea coxensis* DSM45129 based on a prior genome mining study (Yushchuk et al., 2021). Accordingly, based on a high-throughput screening of anti-colorectal cancer activity against HCT-116 cell lines, the crude extract of strain MCN248 showed partial inhibition (27.74%) at a final concentration of 50 µg/mL. This provides evidence that the strain can be further studied for the production of anticancer or other metabolites in future research endeavors.

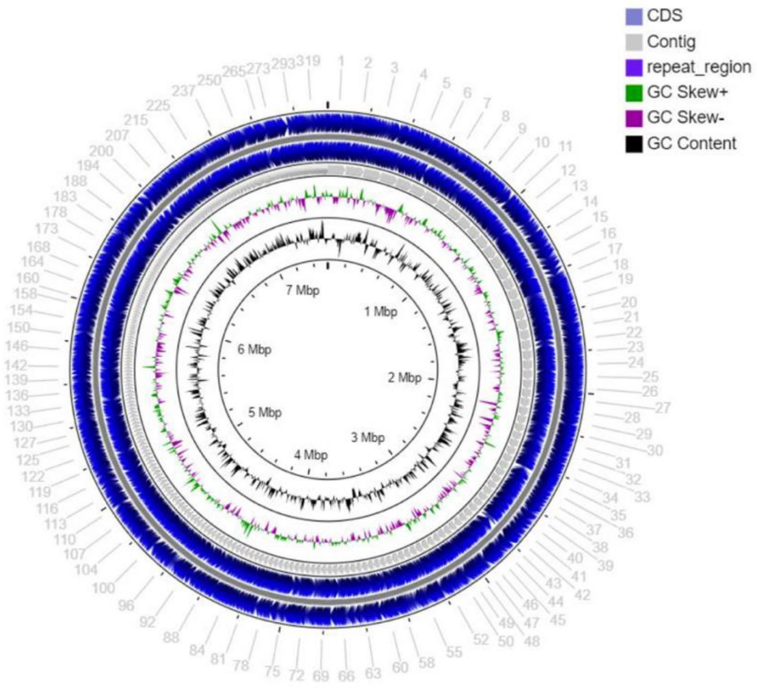


FIGURE 2
Circular map of the total length 7,557,982 bp draft genome visualized by CG viewer server, showing coding sequence (CDS), contig, repeat region, GC skew+, GC skew-, and GC content of stain MCN248.

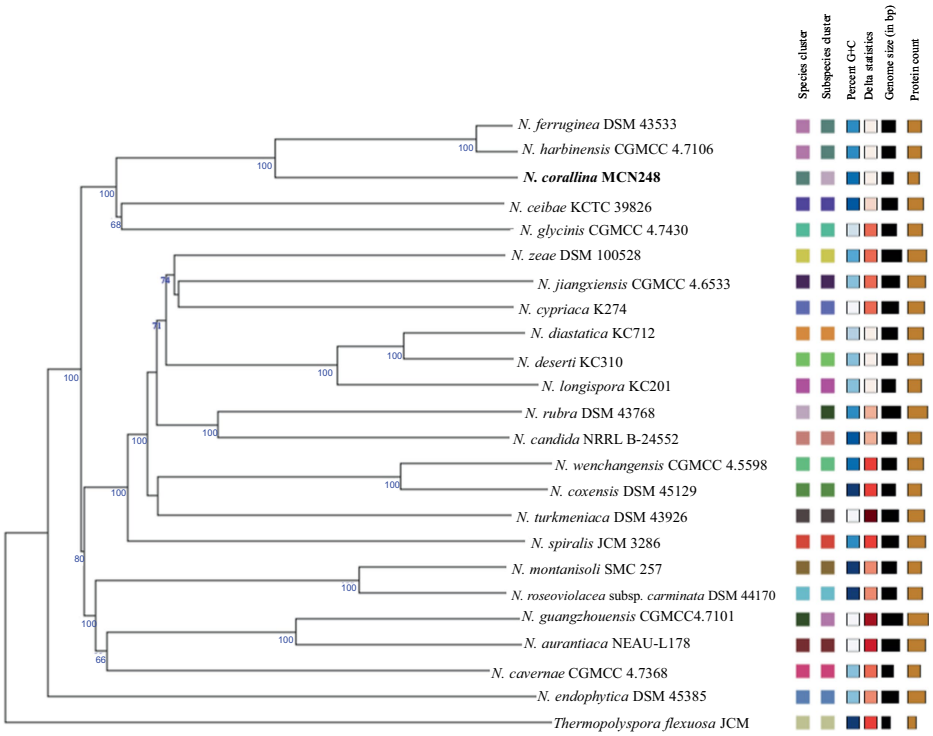


FIGURE 3
TYGS-genome blast distance phylogeny (GBDP) based on genome data of MCN248^{T}, including the available genome data of *Nonomuraea* type strains in the TYGS database. *Thermopolyspora flexuosa* JCM 3056^{T} was used as the outgroup.

Chemotaxononic analyzes of strain MCN248^T

Madurose, ribose, mannose, and glucose were detected as diagnostic whole-cell sugars. The type of muramic acid was acetyl, and mycolic acids were absent. The menaquinones found in strain MCN248 were MK-9(H2), with 50%, and MK-9(H4), with 44%. Phospholipids consisted of diphosphatidylglycerol, hydroxyl-phosphatidylethanolamine, phosphatidylethanolamine, phosphatidylinositol, phosphatidylglycerol, three unidentified glycolipids, two unidentified phospholipids, and two unidentified lipids (Supplementary Figure S3). The fatty acid profiles of strain MCN248^T were compared with those of the two reference strains, as shown in Table 1. The significant differences in the fatty acid patterns between MCN248^T and the closely related strains were found in the amount of iso-C16:0, iso-C15:0, and C16:0. The cellular fatty acids of strain MCN248^T that comprised more than 10% of total fatty acids were iso-C16:0 (40.4%), 10-methyl- C17:0 (22.0%), and C17:1ω8c (10.9%) (Table 1). The chemotaxonomic traits were consistent with those of other species in the genus (Saygin et al., 2021; Lin et al., 2022), as reported in Bergey's Manual of Systematics of Archaea and Bacteria (Kämpfer, 2012). Madurose is the major sugar, and MK-9(H_4), MK-9(H_2), and MK-9(H_0) are mainly found as the menaquinones. Iso-C16:0 and 10-methyl- C17:0 are major types of fatty acids. Phenotypic characteristic of strain MCN248T and MCN248^T exhibited good growth on ISP 3, ISP 4, and ISP 6. The color of the substrate mycelium varied from light coral red to burnt orange. No soluble pigment was observed in any media (see Supplementary Table S3). After 7 days of incubation, the pictures of strain MCN248 on 301 agar (24.0 g/L starch, 5.0 g/L yeast extract, 4.0 g/L $CaCO_3$, 3.0 g/L peptone, 3.0 g/L meat extract, and 1.0 g/L glucose) were demonstrated in comparison with the related strains, as shown in Supplementary Figure S4. Aerial mass color was white on all the media used for cultural characterization. Vegetative mycelia were branched but not fragmented. Straight and flexuous long chain spores were produced on substrate and aerial hyphae. Sporangia were not observed. The surface of the spores was rough, and the spores were 0.5–0.6 × 0.7–1.0 µm in size (Figure 5). Growth occurred at 24–42°C, with an optimal temperature range of 28–36°C and pH of 5.0–11.0 (optimum 7–11). Strain MCN248^T could grow on 5% (w/v) NaCl-containing medium, while reference strains could not tolerate 5% NaCl concentration and a pH above 9.0. Several physiological traits indicated differences between the test strain and the closely related species. All degradation tests—starch hydrolysis, nitrate reduction, gelatin liquefaction, milk peptonization, and hydrogen sulfide production—yielded negative results in strain MCN248^T, while tyrosinase activity was detected. On the other hand, *N. ferruginea* DSM43553^T was positive for starch hydrolysis and nitrate reduction, and *N. harbinensis* DSM45887^T was positive for only nitrate reduction (Table 2). The carbohydrate utilization capacity of the test strain—including D-mannitol, rhamnose, D-xylose, L-arabinose, mannose, and galactose—is the property that distinguishes it from the reference strains. Other physiological characteristics are shown in Table 2.

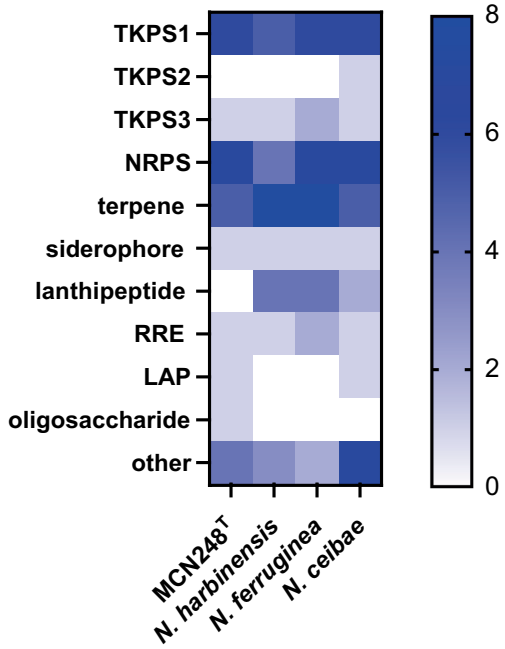


FIGURE 4
Heatmap demonstrating the numbers of biosynthetic regions in genome of strain MCN248^T and its closely related strains analyzed using antiSMASH software version 7.0.0beta1.

TABLE 1 Fatty acid compositions (%) of strain MCN248^T and closely related *Nonomuraea* species.

Fatty acids[a]	*N. corallina* MCN248^T	*N. ferruginea* DSM43553^T	*N. harbinensis* DSM45887^T
16:0 iso	40.35	30.21	32.76
17:0 10-methyl	22.06	28.17	30.28
17:1 ω8c	10.92	6.54	5.96
16:1 iso G	4.99	2.32	4.20
17:0	4.57	3.66	4.49
14:0 iso	2.34	0.97	0.76
15:0 iso	2.22	5.41	5.25
18:0 iso	1.89	2.92	3.02
16:0	1.25	1.35	1.02
17:0 iso	0.90	0.81	1.37
18:0 10-methyl, TBSA	0.89	2.64	1.58
17:0 iso 3OH	0.88	2.62	2.56
18:1 ω9c	0.79	1.78	1.00
18:1 iso H	0.43	–	–
12:0 iso	0.34	–	–
15:1 ω6c	0.33	1.04	0.81
17:0 2OH	0.31	–	0.94
13:0	0.29	0.30	0.35
14:0	0.27	0.88	0.64
17:0 anteiso	0.17	0.60	0.55
15:0 2OH	0.16	0.28	0.34
18:1 2OH	0.14	0.34	0.42
13:0 iso	0.10	–	–

[a]Fatty acid contents of < 0.1% are omitted.

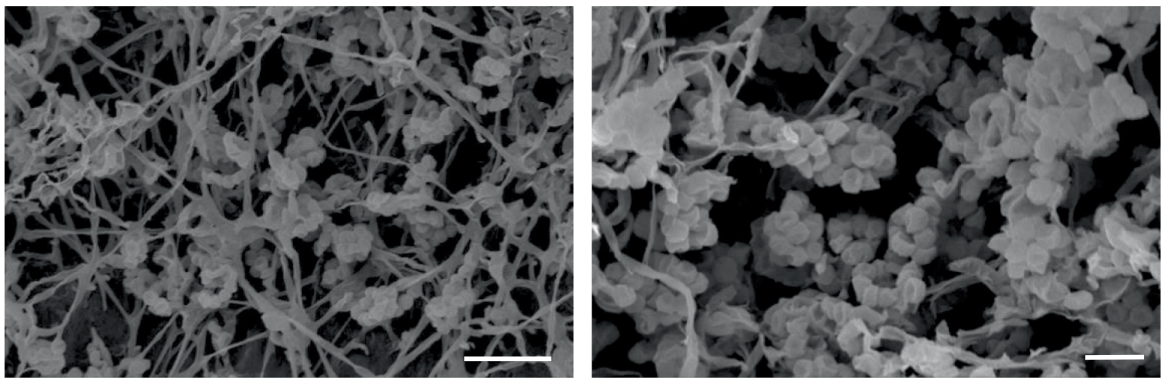


FIGURE 5
Scanning electron micrographs of spores of strain MCN248^T grown on ISP 3 medium for 21 days at 28°C. Bar, 5 μm (left); 2 μm (right).

TABLE 2 Differential physiological and biochemical properties of strain MCN248^T and closely related *Nonomuraea* species.

Characteristics	1	2	3
Colony color	Light coral red to burnt orange	Sand to burnt orange	Pale yellow to camel
Growth in NaCl (%, w/v)	0–5	0–3	0–4
Growth temperature (°C)	24–42 (28–36)	28–38	20–34
Growth pH	5.0–11.0	6.8–7.8	6.0–9.0
Biochemical test			
Starch hydrolysis	–	+	–
Nitrate reduction	–	+	+
Utilization of carbohydrate			
D-mannitol	+	–	–
Rhamnose	+	+	–
D-xylose	+	+	–
L-arabinose	+	–	–
Mannose	+	–	–
Galactose	+	+	–
APIZYM			
Naphthol-AS-BI-phosphohydrolase	–	–	+
β-glucuronidase	–	–	–
α-mannosidase	–	+	+
α-fucosidase	–	–	+

Strain: 1, MCN248^T; 2, *N. ferruginea* DSM43553^T [39, 40]; 3, *N. harbinensis* DSM45887^T [41]. Optimum temperatures are shown in parentheses. +, positive; –, negative; w, weakly positive; ND, not determined.

Description of *Nonomuraea corallina* sp. nov.

Nonomuraea corallina (co.ral.li'na. L. fem. Adj. *corallina* coral red, referring to the color of the substrate mycelium) is aerobic, gram-positive, and mesophilic actinomycetes. The colonies are light coral red to burnt orange in color. White aerial mycelium is produced, and a single spore develops on the substrate mycelia. The spores are non-motile. Vegetative mycelia are branched and not fragmented, and no sporangia are observed. The acyl type of the peptidoglycan is acetyl, and the cell wall contains meso-diaminopimelic acid. Madurose, ribose, mannose, and glucose are detected as whole-cell sugars. Major phospholipids are diphosphatidylglycerol, hydroxyl-phosphatidylethanolamine, phosphatidylethanolamine, phosphatidylinositol, phosphatidylglycerol, and unidentified lipids. The major menaquinones are MK-9(H_2) and MK-9(H_4), and mycolic acids are not detected. The major cellular fatty acid is iso-$C_{16:0}$. The growth of the type strain is observed at temperatures between 24°C and 42°C, with an optimum range of 28–36°C and pH of 5–11, with an optimum pH of 7–11. The maximum NaCl concentration for growth is 3% (w/v). Strain MCN248^T is capable of utilizing D-mannitol, rhamnose, sucrose, D-xylose, L-arabinose, mannose, and galactose as sole carbon sources, but it shows limited growth with maltose and sorbitol. Casein hydrolysis, gelatin liquefaction, nitrate reduction, starch hydrolysis, H_2S production, and melanin production are negative for the strain. Enzymatic activity of the API ZYM system is positive for alkaline phosphatase, α-chymotrypsin, α-glucosidase, *β*-galactosidase, *β*-glucosidase, cystine arylamidase, esterase (C4), esterase lipase (C8), leucine arylamidase, N-acetyl-*β*-glucosaminidase, trypsin, α-galactosidase, acid phosphatase, and valine arylamidase, while it is negative for lipase (C14), naphthol-AS-BI-phosphohydrolase, *β*-glucuronidase, α-mannosidase, and α-fucosidase. The G + C content of the genomic DNA is 71.7%. The type strain, MCN248^T (=NBRC115966^T = TBRC17110^T), was isolated from Songkhla Province, Thailand. The GenBank accession numbers are OP658912 (16S rRNA gene) and JAPNNL000000000 (draft genome), respectively.

Conclusion

Strain MCN248^T should be classified as a member of the genus *Nonomuraea* on the basis of both the phylogenetic analysis and chemotaxonomic characterization. However, strain MCN248^T was clearly distinguishable from the closely related strains by its biochemical and physiological properties (e.g., carbon utilization, starch hydrolysis, nitrate reduction, growth temperature and pH, and enzymatic activity) (Table 2). Furthermore, the dDDH and ANI values between MCN248^T and the closely related strains were lower than the bacterial species thresholds. Based on phenotypic and phylogenetic evidence and whole

genomic data, it is proposed that strain MCN248^T represents a novel species of the genus *Nonomuraea*, namely, *Nonomuraea corallina* sp. nov. With a 27.74% inhibition of HCT-116 cells at a 50-μg/ml concentration of the crude extract, pure compounds may exhibit higher activity, being devoid of inactive contaminants. Nevertheless, strain MCN248^T holds the potential for future research on anticancer compounds or metabolite production, as it possesses biosynthetic gene clusters encoding the lipopeptides icosalide A/icosalide B (100% similarity) and the polyketide cytorhodin (37% similarity).

Data availability statement

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

Ethics statement

Ethical approval was not required for the studies on humans in accordance with the local legislation and institutional requirements because only commercially available established cell lines were used.

Author contributions

CN, BI, and YI developed the ideas and designed the experimental plans. WP, AM, and CS supervised the research. CN and YI performed experiments. CN and BI analyzed the data and prepared the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This research project is supported by Mahidol University (Fundamental Fund: fiscal year 2023 by National Science Research and Innovation Fund (NSRF)), partially supported by the CIF and CNI Grant, Faculty of Science, Mahidol University.

Acknowledgments

The research scholarship (of CN) was supported by the Faculty of Science, Mahidol University. The authors are grateful to Mahidol University and the Osaka Collaborative Research Center for Bioscience and Biotechnology; the Faculty of Science, Mahidol University; and the Kitasato Institute for Life Sciences, Tokyo, Japan, for supporting materials and facilities. The authors also thank M. Iskander for critically proofreading the manuscript.

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2023.1226945/full#supplementary-material

References

Ay, H., Saygin, H., and Sahin, N. (2020). Phylogenomic revision of the family *Streptosporangiaceae*, reclassification of *Desertactinospora gelatinilytica* as *Spongiactinospora gelatinilytica* comb. nov. and a taxonomic home for the genus *Sinosporangium* in the family *Streptosporangiaceae*. *Int. J. Syst. Evol. Microbiol.* 70, 2569–2579. doi: 10.1099/ijsem.0.004073

Boros, C., Smith, C., Vasina, Y., Che, Y., Dix, A. B., Darveaux, B., et al. (2006). Isolation and identification of the Icosalides—cyclic Peptolides with selective antibiotic and cytotoxic activities. *J. Antibiot.* 59, 486–494. doi: 10.1038/ja.2006.68

Chun, J., Oren, A., Ventosa, A., Christensen, H., Arahal, D. R., da Costa, M. S., et al. (2018). Proposed minimal standards for the use of genome data for the taxonomy of prokaryotes. *Int. J. Syst. Evol. Microbiol.* 68, 461–466. doi: 10.1099/ijsem.0.002516

Collins, M. D., Pirouz, T., Goodfellow, M., and Minnikin, D. E. (1977). Distribution of menaquinones in actinomycetes and corynebacteria. *J. Gen. Microbiol.* 100, 221–230. doi: 10.1099/00221287-100-2-221

Embley, M. T., Smida, J., and Stackebrandt, E. (1998). The phylogeny of mycolateless wall chemotype IV actinomycetes and description of *Pseudonocardiaceae* fam. Nov. *Syst. Appl. Microbiol.* 1, 44–52. doi: 10.1016/S0723-2020(88)80047-X

Felsenstein, J. (1981). Evolutionary trees from DNA sequences: a maximum likelihood approach. *J. Mol. Evol.* 17, 368–376. doi: 10.1007/BF01734359

Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39, 783–791. doi: 10.2307/2408678

Goodfellow, M., Stanton, L. J., Simpson, K. E., and Minnikin, D. E. (1990). Numerical and chemical classification of *Actinoplanes* and some related *actinomycetes*. *J. Gen. Microbiol.* 136, 19–36. doi: 10.1099/00221287-136-1-19

Hedtmann, U., Fehlhaber, H. W., Sukatsch, D. A., Weber, M., Hoffmann, D., and Kraemer, H. P. (1992). The new cytotoxic antibiotic cytorhodin x, an unusual anthracyclinone-9α-glycoside. *J. Antibiot.* 45, 1373–1375. doi: 10.7164/antibiotics.45.1373

Intra, B., Matsumoto, A., Inahashi, Y., Omura, S., Takahashi, Y., et al. (2013). *Actinokineospora bangkokensis* sp. nov., isolated from rhizospheric soil. *Int. J. Syst. Evol. Microbiol.* 63, 2655–2660. doi: 10.1099/ijs.0.047928-0

Jacobson, E., Grauville, W. C., and Fogs, C. E. (1958). *Colour harmony manual*, *4th*. IL: Container Corporation of America.

Kämpfer, P. (2012). "Genus VI. Nonomuraea corrig. Zhang, Wang and Ruan 1998b, 419VP" in *Bergey's Manual of Systematic Bacteriology*. eds. M. Goodfellow, P. Kämpfer, H.-J. Busse, M. E. Trujillo, K. Suzuki and W. Ludwig (New York: Springer)

Kämpfer, P., and Kroppenstedt, R. M. (1996). Numerical analysis of fatty acid patterns of coryneform bacteria and related taxa. *Can. J. Microbiol.* 42, 989–1005. doi: 10.1139/m96-128

Kimura, M. A. (1980). Simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* 16, 111–120. doi: 10.1007/BF01731581

Kovacs, N. (1956). Identification of *pseudomonas pyocyanea* by the oxidase reaction. *Nature* 178, 703–704. doi: 10.1038/178703a0

Kroppenstedt, R. M., Stackebrandt, E., and Goodfellow, M. (1990). Taxonomic revision of the actinomycete genera *Actinomadura* and *Microtetraspora*. *Syst. Appl. Microbio.* 13, 148–160. doi: 10.1016/S0723-2020(11)80162-1

Kumar, L., Kumar, S., Sandeep, K., and Patel, S. K. S. (2023). Therapeutic approaches in pancreatic cancer: recent updates. *Biomedicine* 11:1611. doi: 10.3390/biomedicines11061611

Lechevalier, M. P., and Lechevalier, H. (1970). Chemical composition as a criterion in the classification of aerobic actinomycetes. *Int. J. Syst. Bacteriol.* 20, 435–443. doi: 10.1099/00207713-20-4-435

Lin, J., Zhang, L., Qian, L., Yang, Y., Xiang, W., Zhao, J., et al. (2022). *Nonomuraea aurantiaca* sp. nov., a novel cellulase-producing actinobacterium isolated from soil. *Int. J. Syst. Evol. Microbiol.* 72. doi: 10.1099/ijsem.0.005411

Liu, S. S., Qi, J., Teng, Z. D., Tian, F. T., Lv, X. X., Li, K., et al. (2020). Resistomycin attenuates triple-negative breast cancer progression by inhibiting E3 ligase Pellino-1 and inducing SNAIL/SLUG degradation. *Signal Transduct. Target. Ther.* 5:133. doi: 10.1038/s41392-020-00255-y

Medema, M. H., Blin, K., Cimermancic, P., de Jager, V., Zakrzewski, P., Fischbach, M. A., et al. (2011). antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters. *Nucleic Acids Res.* 39, W339–W346. doi: 10.1093/nar/gkr466

Meier-Kolthoff, J. P., Auch, A. F., Klenk, H. P., and Göker, M. (2013). Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinform.* 14, 60–78. doi: 10.1186/1471-2105-14-60

Meier-Kolthoff, J. P., and Göker, M. (2019). TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. *Nat. Commun.* 10:2182. doi: 10.1038/s41467-019-10210-3

Meyer, J. (1979). New species of the genus *Actinomadura*. *Z Allg. Microbio.* 19, 37–44. doi: 10.1002/jobm.19790190106

Minnikin, D. E., Patel, P. V., Alshamaony, L., and Goodfellow, M. (1977). Polar lipid composition in the classification of *nocardia* and related bacteria. *Int. J. Syst. Bacteriol.* 27, 104–117. doi: 10.1099/00207713-27-2-104

Muhamad, N., and Na-Bangchang, K. (2020). Metabolite profiling in anticancer drug development: a systematic review. *Drug Des. Devel. Ther.* 14, 1401–1444. doi: 10.2147/DDDT.S221518

Nouioui, I., Carro, L., García-López, M., Meier-Kolthoff, J. P., Woyke, T., Kyrpides, N. C., et al. (2018). Genome-based taxonomic classification of the phylum actinobacteria. *Front. Microbiol.* 9:22. doi: 10.3389/fmicb.2018.02007

Quintana, E., Maldonado, L., and Goodfellow, M. (2003). *Nonomuraea terrinata* sp. nov., a novel soil actinomycete. *Antonie Van Leeuwenhoek* 84, 1–6. doi: 10.1023/A:1024474719786

Reddy, C. S., Sahai, R., Fehlhaber, H. W., and Ganguli, B. N. (1985). Isolation and structure of a new ε-rhodomycin compound produced by *streptomyces* species HPL Y-11472. *J. Antibiot.* 38, 1423–1425. doi: 10.7164/antibiotics.38.1423

Richter, M., Rosselló-Móra, R., Glöckner, F. O., and Peplies, J. (2015). JSpeciesWS: a web server for prokaryotic species circumscription based on pairwise genome comparison. *Bioinformatics* 32, 929–931. doi: 10.1093/bioinformatics/btv681

Saito, H., and Miura, K. I. (1963). Preparation of transforming deoxyribonucleic acid by phenol treatment. *Biochim. Biophys. Acta* 72, 619–629. doi: 10.1016/0926-6550(63)90386-4

Saitou, N., and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4, 406–425. doi: 10.1093/oxfordjournals.molbev.a040454

Saygin, H., Ay, H., Guven, K., Cetin, D., and Sahin, N. (2021). Comprehensive genome analysis of a novel actinobacterium with high potential for biotechnological applications, *Nonomuraea aridisoli* sp. nov., isolated from desert soil. *Antonie Van Leeuwenhoek* 114, 1963–1975. doi: 10.1007/s10482-021-01654-z

Shirling, E. B., and Gottlieb, D. (1966). Methods for characterization of *streptomyces* species. *Int. J. Syst. Bacteriol.* 16, 313–340. doi: 10.1099/00207713-16-3-313

Stackebrandt, E., Rainey, F. A., and Ward-Rainey, N. L. (1997). Proposal for a new hierarchic classification system, *actinobacteria* classis nov. *Int. J. Syst. Bacteriol.* 47, 479–491. doi: 10.1099/00207713-47-2-479

Staneck, J. L., and Roberts, G. D. (1974). Simplified approach to identification of aerobic actinomycetes by thin-layer chromatography. *Appl. Microbiol.* 28, 226–231. doi: 10.1128/am.28.2.226-231.1974

Subramani, R., and Sipkema, D. (2019). Marine rare actinomycetes: a promising source of structurally diverse and unique novel natural products. *Mar. Drugs* 17:249. doi: 10.3390/md17050249

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., et al. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* 71, 209–249. doi: 10.3322/caac.21660

Sungthong, R., and Nakaew, N. (2015). The genus *Nonomuraea*: a review of a rare actinomycete taxon for novel metabolites. *J. Basic Microbiol.* 55, 554–565. doi: 10.1002/jobm.201300691

Také, A., Nakashima, T., Inahashi, Y., Shiomi, K., Takahashi, Y., Ōmura, S., et al. (2016). Analyses of the cell-wall peptidoglycan structures in three genera *Micromonospora*, *Catenuloplanes*, and *Couchioplanes* belonging to the family *Micromonosporaceae* by derivatization with FDLA and PMP using LC/MS. *J. Gen. Appl. Microbiol.* 62, 199–205. doi: 10.2323/jgam.2016.02.007

Tamura, A., Furuta, R., Naruto, S., and Ishii, H. (1973). Actinotiocin, a new sulfur-containing peptide antibiotic from *Actinomadura pusilla*. *J. Antibiot.* 26, 343–350. doi: 10.7164/antibiotics.26.343

Tamura, K., Stecher, G., and Kumar, S. (2021). MEGA11: molecular evolutionary genetics analysis version 11. *Mol. Biol. Evol.* 38, 3022–3027. doi: 10.1093/molbev/msab120

Tomiyasu, I. (1982). Mycolic acid composition and thermally adaptative changes in *Nocardia asteroides*. *J. Bacteriol.* 151, 828–837. doi: 10.1128/jb.151.2.828-837.1982

Wang, S., Liu, C., Zhang, Y., Zhao, J., Zhang, X., Yang, L., et al. (2014). *Nonomuraea guangzhouensis* sp. nov., and *Nonomuraea harbinensis* sp. nov., two novel actinomycetes isolated from soil. *Antonie Van Leeuwenhoek* 105, 109–118. doi: 10.1007/s10482-013-0058-3

Wattanasuepsin, W., Intra, B., Také, A., Inahashi, Y., Euanorasetr, J., Ōmura, S., et al. (2017). *Saccharomonospora colocasiae* sp. nov., an actinomycete isolated from the rhizosphere of *Colocasia esculenta*. *Int. J. Syst. Evol. Microbiol.* 67, 4572–4577. doi: 10.1099/ijsem.0.002336

Williams, S. T., Goodfellow, M., and Alderson, G. (1989). "Genus *streptomyces* Waksman and Henrici 1943, 339AL" in *Bergey's manual of systematic bacteriology*. eds. S. T. Williams, M. E. Sharpe and J. G. Holt, vol. *4* (Baltimore: Williams & Wilkins), 2452–2492.

Yang, B., Zhang, Z., Yang, C. Q., Wang, Y., Orr, M. C., Wang, H., et al. (2022). Identification of species by combining molecular and morphological data using convolutional neural networks. *Syst. Biol.* 71, 690–705. doi: 10.1093/sysbio/syab076

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

Yushchuk, O., Vior, N. M., Andreo-Vidal, A., Berini, F., Rückert, C., Busche, T., et al. (2021). Genomic-led discovery of a novel glycopeptide antibiotic by *Nonomuraea coxensis* DSM 45129. *ACS Chem. Biol.* 16, 915–928. doi: 10.1021/acschembio.1c00170

Zhi, X. Y., Li, W. J., and Stackebrandt, E. (2009). An update of the structure and 16S rRNA gene sequence-based definition of higher ranks of the class *actinobacteria*, with the proposal of two new suborders and four new families and emended descriptions of the existing higher taxa. *Int. J. Syst. Evol. Microbiol.* 59, 589–608. doi: 10.1099/ijs.0.65780-0

Frontiers in Microbiology

Explores the habitable world and the potential of microbial life

The largest and most cited microbiology journal which advances our understanding of the role microbes play in addressing global challenges such as healthcare, food security, and climate change.

Discover the latest Research Topics

See more →

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact