

OPEN ACCESS

EDITED BY

Wenyi Zhang,
Inner Mongolia Agricultural University,
China

REVIEWED BY

Henok Tegegne,
Baylor College of Medicine,
United States
Ke Shen,
Sichuan University, China

*CORRESPONDENCE

Bachir Balech
✉ bachir.balech@cnr.it

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

RECEIVED 10 October 2025

REVISED 29 January 2026

ACCEPTED 30 January 2026

PUBLISHED 16 February 2026

CITATION

Maisto L, Telegrafo C, Rubino F,
Santamaria M, Traka MH, Tullo A,
Bouwman J, Sbisà E and Balech B (2026)
Multifaceted human gut microbiome
data associated with health and
nutrition.
Front. Microbiol. 17:1722500.
doi: 10.3389/fmicb.2026.1722500

COPYRIGHT

© 2026 Maisto, Telegrafo, Rubino,
Santamaria, Traka, Tullo, Bouwman,
Sbisà and Balech. This is an open-access
article distributed under the terms of the
[Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is
permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication
in this journal is cited, in accordance
with accepted academic practice. No
use, distribution or reproduction is
permitted which does not comply with
these terms.

Multifaceted human gut microbiome data associated with health and nutrition

Lucia Maisto^{1†}, Claudia Telegrafo^{1†}, Francesco Rubino²,
Monica Santamaria³, Maria H. Traka⁴, Apollonia Tullo¹,
Jildau Bouwman⁵, Elisabetta Sbisà² and Bachir Balech^{1*}

¹Institute of Biomembranes, Bioenergetics and Molecular Biotechnologies, Consiglio Nazionale delle Ricerche (IBIOM-CNR), Bari, Italy, ²Institute for Biomedical Technologies, Consiglio Nazionale delle Ricerche (ITB-CNR), Bari, Italy, ³Department of Soil, Plant and Food Sciences (Di.S.S.P.A.), University of Bari, Bari, Italy, ⁴Food and Nutrition National Bioscience Research Infrastructure, Quadram Institute Bioscience, Norwich Research Park, Norwich, United Kingdom, ⁵Netherlands Organisation for Applied Scientific Research (TNO), Microbiology and Systems Biology, Leiden, Netherlands

The microbiome, also considered the hidden organ, is a fundamental ecosystem directly associated with the disease and health status of the human body. With the availability of high-throughput DNA sequencing technologies, a growing number of studies from clinical and experimental (observation and intervention) samples are constantly revealing new findings on the relationship between human organs and their microbiomes. In such a context, diet and nutrition are among the key factors influencing microbiome composition, richness, and functional behavior. In this review, we illustrate how microbiome-related data and associated metadata are in recent times scattered across primary and specialized databases with different levels of curation, annotation, and standardization, limiting, to some extent, the possibility of deep data discovery, reuse, alignment, and harmonization. Therefore, we describe the way Findable, Accessible, Interoperable, and Reusable (FAIR) data principles would enhance the onset of novel scientific hypotheses and potential microbiome-targeted therapies by improving the standardization policies in data sources. Accordingly, using advanced semantic classification and data mining technologies based on suitable and comprehensive ontologies, annotations of studies present in source databases or in scientific literature would further improve the data and metadata enrichment, integration and alignment relevant to microbiome data associated with health, disease and nutrition.

KEYWORDS

database, diet, disease, dysbiosis, FAIR, metagenomics

1 Introduction

The microbiota is a complex and dynamic ecosystem essential for health as it interacts with nearly every aspect of human physiology (Khalil et al., 2024; Origüela and Lopez-Zaplana, 2025). It is often referred to as the “hidden organ” within an organ, highlighting its critical roles in health and disease (Hou et al., 2022). Growing evidence links the human microbiota to mental and physical wellbeing through the gut–brain axis, a bidirectional network involving neural, immune, and endocrine pathways regulating both immunity and inflammation (Martin et al., 2018; Morais et al., 2021; Origüela and Lopez-Zaplana, 2025). Similarly, the skin, lung, vaginal, and oral microbiota, which vary by niche, regularly influence directly and indirectly

the response to several human body disorders (Yamashita and Takeshita, 2017; Mithradas et al., 2024; Kreouzi et al., 2025; Zhou et al., 2025).

Although a great number of scientific studies have been carried out, multiple aspects of human health in relation to food and its correlation with microbiome composition and function remain to be discovered. Such complexity can only be addressed through the availability and reusability of a suitable amount of high-quality microbiome-derived data, including nucleic acids (DNA/RNA sequences), capable of enhancing, maximizing, and standardizing scientific research and data elaboration and modeling. Accordingly, to achieve such a purpose, public DNA/RNA sequence databases adopting appropriate policies for sustainable data management, including structure, sharing protocols, accessibility, and reusability [i.e., (Zhao et al., 2023), Findable, Accessible, Interoperable, Reusable (FAIR) Wilkinson et al., 2016],¹ have become a fundamental aspect.

In this study, we explore the main features describing data type and availability in public databases related to the human gut microbiome in association with general health and food intake. The main microbiome data sources were explored, including their adherence to standards, open science, and FAIR data principles. Recommendations on the proper use of such data and the possibility of aligning studies and datasets across resources are also provided.

2 Eubiosis and dysbiosis

The gut microbiome significantly influences the health and disease status of each individual as it contributes to several interactions with near or distant organs. Scientific evidence links the effect of microbiome state to numerous important human diseases, such as cardiovascular diseases, type 2 diabetes, neurological development disorders, autoimmune diseases, and colon and colorectal cancer. The absence or the severity of such diseases/disorders is determined by two main classes of microbiome features called eubiosis or dysbiosis (Iebba et al., 2016; Yu et al., 2022; Goudman et al., 2024; Hamjane et al., 2024; Díez-Madueño et al., 2025).

Eubiosis is essential to maintain a healthy gut. Despite its importance, eubiosis remains partially understood, with ongoing research exploring the microbiome–host relationship and its therapeutic potential (Hou et al., 2022). It is a balanced gut microbiota state where diverse microbial communities interact harmoniously to support the host's health (Chen et al., 2024). This equilibrium is achieved through the interplay of key bacterial taxa, including *Firmicutes*, *Bacteroides*, *Actinobacteria*, and *Proteobacteria*, which collectively regulate essential processes such as nutrient metabolism and immune modulation (Afzaal et al., 2022). For instance, the small intestine hosts *Enterobacteriaceae* for nutrient absorption, while the colon, dominated by *Bacteroidetes* and *Firmicutes*, is the center of microbial activity (Rinninella et al., 2019). The gut microbiota shapes immune responses by helping immune cells distinguish commensals from pathogens, for instance, by activating NF-κB signaling via Toll-like receptors and promoting tolerance through Treg cell stimulation (Al-Rashidi, 2022; Brockmann et al., 2023). Moreover, a healthy microbiota prevents pathogen colonization by depleting essential

nutrients and forming physical biofilm barriers, while simultaneously using specialized enzymes to metabolize complex carbohydrates reserved for their prebiotic potential (Ma et al., 2015; Bedu-Ferrari et al., 2024). Additionally, commensals secrete antimicrobials such as bacteriocins and metabolites (e.g., reuterin produced by *Lactobacillus reuteri*) that inhibit pathogens and potentially prevent colon tumorigenesis (Bell et al., 2022). Recent advances have highlighted the impact of gut microbiota and oral probiotics supplementation on defense responses also against viral infections, including respiratory viruses, through the modulation of the pulmonary immune response of the gut–lung axis (Andrade et al., 2022; Chen N. et al., 2025; Liu et al., 2025).

On the other hand, dysbiosis, referring to an imbalanced gut microbiota, is characterized by reduced microbial diversity, promoting pathogenic overgrowth and leading to metabolic and immune dysfunctions (Chen et al., 2024). Emerging evidence links dysbiosis to colorectal cancer, as increased permeability allows pro-inflammatory metabolites to promote inflammation, DNA damage, and tumorigenesis (Artemev et al., 2022). For instance, the overgrowth of specific bacteria such as *Escherichia coli* and *Bacteroides fragilis* contributes to genotoxin production (Li et al., 2021; Zhao et al., 2023). A dysbiotic microbiota, marked by a higher Firmicutes-to-Bacteroidetes ratio, alters lipid and glucose metabolism and compromises intestinal barrier integrity and function, leading to systemic inflammation and insulin resistance, both of which are associated with obesity and type 2 diabetes (Brunkwall and Orho-Melander, 2017; Gomes et al., 2018), and a reduction in butyrate production, contributing to glucose intolerance (Gomes et al., 2018). Chronic intestinal disorders are also associated with reduced short-chain fatty acid (SCFA) production and disrupted tight junctions, which compromise epithelial integrity and trigger excessive immune activation that drives inflammatory disease progression (Qiu et al., 2022), while also negatively influencing the correct gut–brain axis communication (Mehta et al., 2025; Yassin et al., 2025). Restoring SCFA-producing bacterial genera, species, or strains may support epithelial repair and consequently dampen inflammatory responses (Effendi et al., 2022) and fix brain functions (Mehta et al., 2025; Yassin et al., 2025).

The proof of association between pathological conditions and microbiome composition and function is expected to grow continuously as related data are becoming increasingly available in an open-science context. As stated above, potential discoveries on specific taxonomic or functional profiles—including, eubiotic or dysbiotic taxa fingerprints connected to certain health status—can be recovered from relevant scientific studies. However, similar assumptions should be drawn only from comprehensive and combined microbiome datasets supported with accurate metadata annotation and curation, based on controlled vocabularies or specific ontologies, permitting an advanced level of harmonization across data sources.

3 Microbiome-derived DNA/RNA sequence data sources

Advances in sequencing technologies [e.g., next-generation sequencing (NGS)] have promoted deeper exploration of microbial communities and their relative functional profiles in environmental samples (i.e., human organs) through the application of metagenomics, metatranscriptomics, and metabarcoding (Franzosa et al., 2014;

¹ <https://faircookbook.elixir-europe.org/content/home.html>

Arkan and Muth, 2023). This advancement has resulted in a vast amount of sequence data being stored in public primary, specialized, or specific project-dedicated databases with different levels of analysis layers, data curation, metadata enrichment, and data sustainability plans.

Primary public databases such as the *Sequence Read Archive (SRA)* (Katz et al., 2022), managed by the *National Center for Biotechnology Information (NCBI)*, and the *European Nucleotide Archive (ENA)* (O’Cathail et al., 2025), managed by EBI, primarily serve as storage and organization facilities for raw sequencing data, including microbiome-related NGS data (i.e., metagenomics and metabarcoding). In parallel, specialized databases have been developed to offer additional services such as advanced analysis, data curation, and annotation. These databases make use of the data provided by the users or those available in primary databases or in related scientific papers and provide state-of-the-art bioinformatics tools and user-friendly interfaces for microbial communities profiling. In this context, adherence to FAIR data principles, ensuring accessibility, interoperability, and reusability, provides unprecedented benefits. However, FAIRness fulfillment does not guarantee enough information background for scientific use, as the inclusion of enriched metadata is not always mandatory.

In terms of data volume, annotation, wide content, and FAIR compliance, it is worthy to mention three main specialized data sources, namely (i) *MGNify*, developed by EBI-Metagenomic (Richardson et al., 2023), (ii) *MG-RAST* (Meyer et al., 2008), developed by the University of Chicago (Keegan et al., 2016), and (iii) *JGI IMG/M – /VR*, developed by the Joint Genome Institute of California (Chen et al., 2023; Mukherjee et al., 2025).

Although many other microbiomic data sources are of pronounced importance, their use might be more limited due to their specific, clear-cut content objectives, their reduced annotation level, or their outdated metadata, as they are related to closed scientific projects. Nevertheless, one of the most influential but archived initiatives is the *Human Microbiome Project* with its HMP data portal (Turnbaugh et al., 2007), which generated, over several years, a large volume of data and important scientific papers regarding the characterization of the human microbiome and its role in health and disease (Integrative HMP (iHMP) Research Network Consortium, 2019). Several databases make use of the data produced by HMP and integrate it into their own datasets.

As illustrated in Table 1, some resources target microbiomes from multiple organs or diseases [GIMICA (Tang et al., 2021), DISBIOME (Janssens et al., 2018), gcMeta (Shi et al., 2019), Microbiome database (MDB),² Qiita (Gonzalez et al., 2018), curatedMetagenomicData (Pasolli et al., 2017), and Phenotype Database (van Ommen et al., 2010)], while others focus on single organ mainly the gut such as gutMDisorder (Qi et al., 2022), GMrepo (Dai et al., 2022), Human Gut Microbiome Atlas,³ NIBN JMD (Chen Y.-A. et al., 2025). A majority of these databases offer both raw data (or links to raw data) and associated metadata, taxonomic and functional microbiome profiles, and an integrated framework of scientific research results elaborated according to a specific bioinformatics pipeline and experimental parameters. Moreover, some resources, such as Qiita and curated-MetagenomicData, provide ready-to-use reference backbone data in

standard formats (R or Python objects), which can be incorporated into an in-house data analysis routine.

Although all of the above resources have a significant level of data curation, integration, and annotation, studies with structured metadata related to food, diet, and their nutritional characteristics associated with health are only present in a few databases (Phenotype Database, NIBN JMD, and gutMDisorder) as such information are usually poorly highlighted and represented as free text format in experimental design description or in the accompanying scientific publications whenever present. Lacking some essential structured information denotes a clear limit for data interoperability and accessibility purposes (see Table 1 “notes” column). Accordingly, essential metadata should be enriched and included at the data submission step through pre-designed templates. Such templates should consult specific ontologies describing the samples comprehensively and their belonging to the study. Enriched metadata studies, when incorporated into frameworks with FAIR characteristics, would offer a concrete and important possibility for data reuse, studies alignment, and advanced scientific discoveries.

4 Main databases characteristics

Content characteristics and volume, data and metadata formats/standards, bioinformatics analysis availability, and FAIR-compliance were the criteria to select and describe three databases in the context of microbiome related to food and health. Accordingly, in the following, the main features of Mgnify, MG-RAST, and JGI databases are illustrated.

All three databases provide enhanced accessibility through either a Graphical User Interface (GUI) or through a dedicated Application Programming Interface (API), allowing the incorporation of the resource into standardized data analysis pipelines.

Data retrieval allows access to different metadata categories that contextualize sequencing experiments and describe various types of information about the sample’s origin using ontological terms, sequencing protocols, analytical parameters, and the underlying bioinformatic workflows used for data elaboration. The outputs are provided primarily in JavaScript Object Notation (JSON) format, easily extensible and transformable into widely used standard formats (e.g., csv—comma-separated values).

MGNify (formerly EBI-Metagenomics) is a freely accessible platform for assembling, analyzing, and storing metagenomic data from a wide range of environmental and host-associated samples. It supports the analysis of amplicon sequencing (metabarcoding), metagenomics, and metatranscriptomics through standardized, versioned bioinformatic pipelines optimized for different sample types. Multiple pipeline versions are available in a public repository on GitHub,⁴ allowing data reprocessing according to the latest update and straightforward alignment and comparison across different studies’ results. MGNify integrates with external repositories, such as the European Nucleotide Archive (ENA), to enhance data interoperability and reusability and studies metadata tracking, including study unique accession, experiment description (free text), the samples’ biome of origin and links to geographical coordinates, related scientific papers,

² <https://db.cngb.org/microbiome/>

³ <https://www.microbiomeatlas.org/>

⁴ <https://github.com/EBI-Metagenomics>

TABLE 1 Main features of human microbiome molecular data sources. Database type, content, data, and metadata formats, and the last update are shown.

Database name	Type	Data retrieval protocol	Link	Data content	Data and metadata format/standard	Last update	Notes	References
NCBI SRA	P	Api/Web/ftp	https://www.ncbi.nlm.nih.gov/sra	Global Multiple organs	Fastq, BAM, JSON	2025	Important to obtain raw data and metadata. Easy to include in data analysis pipelines.	Katz et al. (2022)
ENA	P	Api/Web/ftp	https://www.ebi.ac.uk/ena	Global Multiple organs	Fastq, BAM, CRAM, JSON	2025	Important to obtain raw data and metadata. Easy to include in data analysis pipelines.	O'Cathail et al. (2025)
Mgnify	S	Api/Web	https://www.ebi.ac.uk/about/teams/microbiome-informatics/mgnify/	Multiple organs Health—disease Diet	Fasta, JSON, csv	2025	Wide range of studies. Easy to use in standardized bioinformatics pipelines. Metadata is sometimes missing or not complete.	Richardson et al. (2023)
MG-RAST	S	Api/Web	https://mg-rast.org	Multiple organs Health—disease Diet	Fastq, Fasta, csv, BIOM, JSON	2025	Wide range of studies. Basic ontology annotations. Basic biodiversity metrics are available. Metadata are sometimes missing or not exhaustive.	Meyer et al. (2008)
JGI—IMG/M	S	Api/Web	https://genome.jgi.doe.gov/portal/	Multiple organs Health—disease Diet	Fastq, Fasta, GFF, csv, JSON	2025	Wide range of studies. Basic metadata annotations. Registration is needed for API access.	Chen et al. (2023) and Mukherjee et al. (2023)
gutMDisorder	S	Web	https://bio-computing.hrbmu.edu.cn/gutMDisorder/	Gut Health—disease Diet	csv	2022	Important to obtain the combined effect of disease and nutrients on gut microbiome composition. API is not available.	Qi et al. (2022)
GMrepo	S	Api/Web	https://gmrepo.humangut.info/home	Gut Health—disease	Fastq, JSON, csv	2020	Important to reveal the pangenome and marker taxa of the gut microbiome associated with phenotypes/diseases of other organs. API is not available.	Dai et al. (2022)
Human Gut Microbiome Atlas	S	Web	https://www.microbiomeatlas.org/	Gut/Oral Health—Disease	csv	2022	Possibility of data elaboration integrating microbiome taxonomy composition, disease, and geography. Results are visible only on the GUI, but not available for download.	NA
NIBN JMD	S	Web	https://jmd.nibn.go.jp/	Gut Health—diet	GUI visualization	2025	Fully accessible only through an email request. The results are visible only on the GUI but are not available for download.	Chen Y.-A. et al. (2025)
GIMICA	S	Web	https://gimica.idrblab.net/ttd/	Multiple organs Health—disease	csv	2020	Valuable resource for human genetic and immune factors regulating the microbiome. Data integration is not foreseen.	Tang et al. (2021)
DISBIOME	S	Api/Web	https://disbiome.ugent.be/home	Multiple organs Health—disease	JSON	2018	All embedded data are structured in JSON format. Easy to use in routine bioinformatics pipelines. The data are outdated.	Janssens et al. (2018)

(Continued)

TABLE 1 (Continued)

Database name	Type	Data retrieval protocol	Link	Data content	Data and metadata format/standard	Last update	Notes	References
gcMeta	S	Web	https://gcmeta.wdcm.org/	Multiple organs (gut, skin, vaginal)	csv	2025	Contains MAGs from different environments. Results within the DB can be compared and integrated. Data retrieval is challenging from the GUI as the API is absent. Not all data can be downloaded (e.g., functional features).	Shi et al. (2019)
Microbiome database (MDB)	S	Web	https://db.cngb.org/microbiome/	Multiple organs Health—disease	Fasta, csv	2025	It provides genes, MAGs, taxonomic, and functional profiles associated with different biomes. API is not available.	NA
Qiita	S	Api/Web/ Qiime2	https://qiita.ucsd.edu/	Multiple organs Health—disease Diet	BIOM, csv	2025	Accessible through pre-defined tools (Qiime2, redbiom). It needs advanced technical skills. Data integration and comparison are possible.	Gonzalez et al. (2018)
curatedMetagenomicData	S	R-Package	https://waldronlab.io/curatedMetagenomicData/	Multiple organs Health—disease	R/python objects	2021	Accessible through pre-defined tools (R, Python, Docker). It needs advanced technical skills. Data integration and comparison are possible.	Pasolli et al. (2017)
Phenotype database	S	API/Web	https://dashin.eu/interventionstudies/	Multiple organs Health—disease Diet	JSON, csv	2025	Important architecture for food ontology annotation and data integration. It would need automation for data/metadata upload and annotation.	van Ommen et al. (2010)

MAG, metagenome-associated genome; ftp, file transfer protocol; GUI, graphical user interface; JSON, JavaScript Object Notation; csv, comma-separated values; BIOM, Biological Observation Matrix format; NA, not available; P, primary; S, specialized.

analyses results, and download pages (example of data retrieval output is available in [Supplementary Table 1](#)).

Similarly, MG-RAST is an open-source platform for metagenomics data functional and taxonomic profiling, with additional statistical metrics (alpha diversity index). It contains information regarding *MIxS* data standards compliance (Yilmaz et al., 2011) and metadata classification based on Environmental Ontology (ENVO) (Buttigieg et al., 2013), such as *biome_id* and *material_id*, offering the possibility of studies/datasets reuse and potential cross-alignment with other databases (for more details and a complete list of metadata, see [Supplementary Table 2](#)). It also uses the information in the Metagenome Annotation Information Resource database (M5nr, Wilke et al., 2012) to harmonize functional annotations and biochemical pathways from multiple sources (i.e., GenBank, UniProt, KEGG, and SEED).

JGI Data Portal is a centralized repository for storing raw (also connected to SRA) and annotated genomic and metagenomic data from microbial communities, plasmids, viruses, and fungi (typical retrieval output details in [Supplementary Table 3](#)). Through the Genomes Online Database (GOLD) (Mukherjee et al., 2023), JGI organizes structured *MIxS*-compliant metadata, as well as *MIGS*/*MIMS* and *MIxE* standards (Field et al., 2008; Yilmaz et al., 2011). A key component is the *Integrated Microbial Genomes & Microbiomes (IMG/M)* system, which enables comparative genomic analysis and functional annotation of microbial communities through the JGI metagenome workflow and the DOE-JGI Metagenome annotator pipeline (Huntemann et al., 2016; Chen et al., 2023).

5 Discussion

The integration and alignment of metagenomic data related to diet, nutrition, and health present a critical challenge in human microbiomics research. While specific projects and their dedicated data resources (i.e., HMP, GMrepo, GIMICA, DISBIOME, and gcMeta) are focused on a defined scientific area, larger databases (i.e., Mgnify, MG-RAST, and JGI) aggregate large volumes of data across multiple scientific disciplines, including diet and/or health targeting the microbiome of multiple organs ([Table 1](#)). These encompass either observation (e.g., microbiome profiling of obese or diabetic subjects, healthy individuals, disease-related functionally/differentially expressed biome, and geographically tagged microbiome) or intervention studies (including time-series experiments on specific food intake or medicinal treatment). Structure and comprehensiveness of metadata related to such studies are crucial to understand the experiment itself and to provide insights into the scientific area of interest, helping the improvement and evolution of experimental designs with similar objectives and avoiding replicating research.

Accordingly, the adherence to specific data and metadata standards [e.g., *MIxS* (Field et al., 2008; Yilmaz et al., 2011)] and the unification of bioinformatics pipelines, as suggested by FAIR data principles (Wilkinson et al., 2016), (see text footnote 1) would offer an important asset toward data discovery, harmonization across platforms, interoperability, accurate reuse, and effective metadata management (Vitali et al., 2018; Balech et al., 2022). In this context, the use of ontologies (disease and/or food) as a unified and standardized

language is a vital step toward the implementation of FAIR principles, ensuring a consistent representation of complex data and facilitating the application of novel technologies (i.e., AI-driven data mining and classification) to connect various systems and platforms. For instance, linking Mgnify, MG-RAST, and JGI studies through unique identifiers or ontology terms, such as *biosampleID*, *bioprojectID*, or ENVO classes, would offer a concrete possibility to merge the corresponding data to obtain broader scientific research hypotheses.

The Ontology Lookup Service (OLS, Côté et al., 2010),⁵ of EBI or the ontologies available through BioPortal (Noy et al., 2009),⁶ provide a framework where it is possible to use multiple ontologies that could facilitate the classification of metagenomic studies data and associated metadata based on biological, ecological, functional, nutritional, and health parameters. Leveraging these ontologies and recognizing their importance to maximize research outputs can lead to their further development, which will promote consistent and innovative changes in shaping diet and microbiome research by proposing new experimental designs.

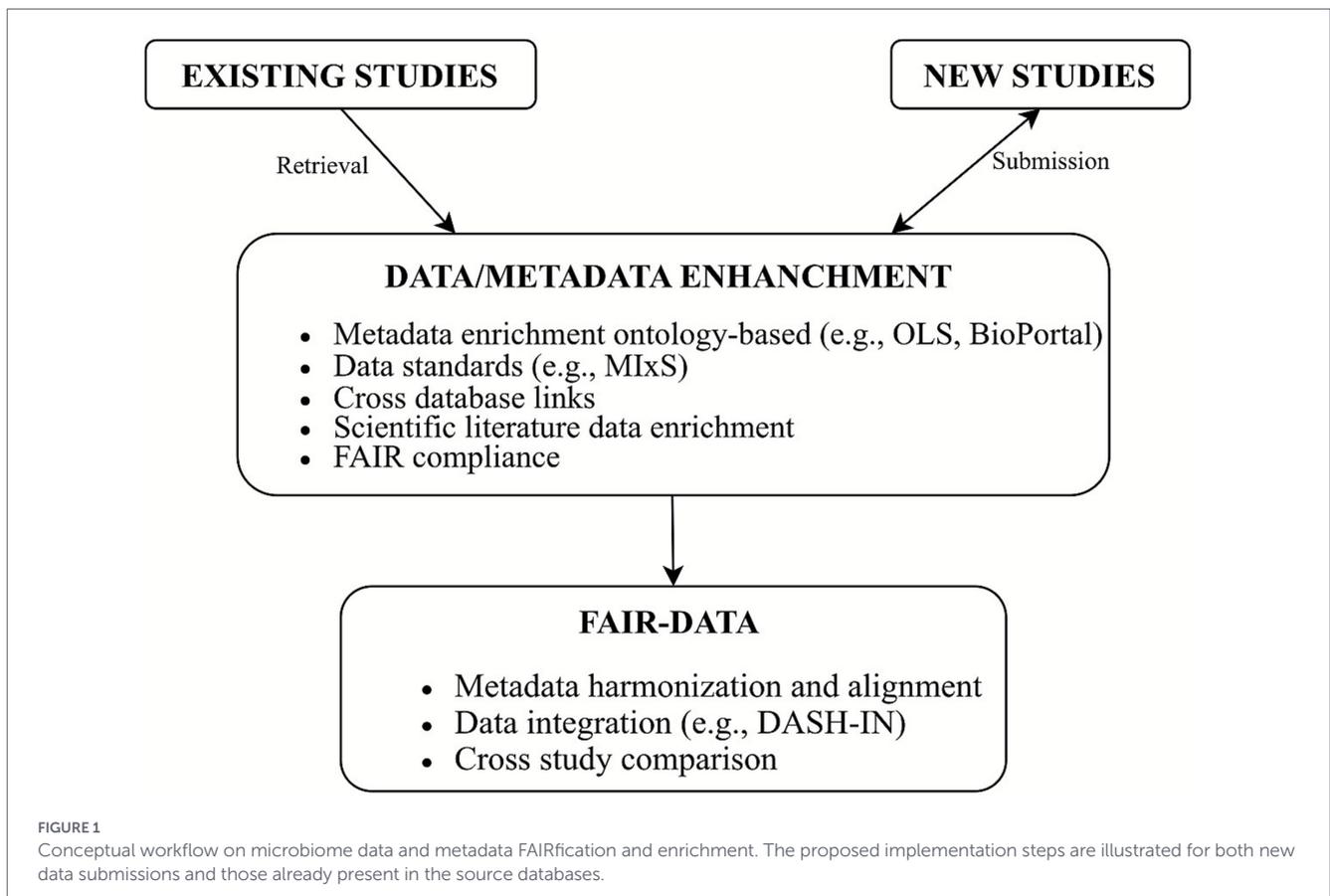
Another key aspect in this context is the integration of sequencing data with dietary patterns and disease conditions. Traditional microbiome studies often emphasize broad taxonomic classifications, but recent advances in functional metagenomics highlight the need to link microbiome data with specific biochemical pathways and metabolic functions. As seen in [Table 1](#), databases such as GIMICA (Tang et al., 2021) or DISBIOME (Janssens et al., 2018) provide an enhanced view of disease or genetic-immunological factors that can shape a healthy microbiome in different organs. However, such data would be additionally comprehensive while linking those aspects to lifestyle and dietary habits. By associating microbiome datasets with dietary interventions based on nutritional intake and health biomarkers, we gain deeper insights into how microbial communities influence metabolic pathways, immune responses, and disease progression. A valuable example of such an approach is the DASH-IN initiative (Data Sharing in Nutrition)⁷ (van Ommen et al., 2010), associated with Phenotype Database, that promotes data integration and sharing in nutritional research across platforms, while offering the possibility to annotate studies metadata using ontological terms to capture the complexity of experimental designs and fulfill the requirements of FAIR data principles. Similarly, the availability of structured and analyzed cohort studies of gut microbiome data with extensive metadata (diet, nutrient intake, physical activity) in databases such as gutM-Disorder (Qi et al., 2022) and NIBN JMD (Chen Y.-A. et al., 2025) represents a promising starting point toward building ready-to-use data integration and harmonization schemas spanning across other organs or disease-related factors. Therefore, comprehensive ontologies that annotate diets at different levels (dietary pattern, food group, food components) are important to understand the true relationships between diet and health mediated by the gut microbiome.

As stated above, the perspectives to obtain rich microbiome data and metadata resources are summarized in [Figure 1](#), which describes shortly the main workflow to be implemented as an integral part of the new data submission or enrichment system of

⁵ <https://www.ebi.ac.uk/ols4/>

⁶ <https://bioportal.bioontology.org/>

⁷ <https://dashin.eu/interventionstudies/>



studies already present in the source databases. Similar concepts, including technical information technology steps, were also proposed recently, illustrating potential actions to be fulfilled at the submission level (Hug et al., 2025; Speir et al., 2025). Although similar, the approach proposed in this review (Figure 1) highlights the potential strength of open science concept assessment and advanced technology annotations to exploit both existing and new microbiome studies.

An urgent need, however, lies in the implementation of a combined methodology exploiting state-of-the-art technologies to mine information embedded not only in primary or specialized databases but also in scientific literature. This has become increasingly possible with the use of data science and natural language processing for semantic annotation and searching of relevant metadata information. Such approaches, if adopted, introduce scientific research into the new era of technology enhanced by information management and strengthen decision-making dealing with microbiome-targeted therapies for health and nutrition.

Author contributions

LM: Writing – review & editing, Investigation, Writing – original draft, Validation, Data curation, Visualization. CT: Writing – original draft, Visualization, Data curation, Validation, Writing – review & editing, Investigation. FR: Visualization, Data curation, Writing – review & editing, Validation, Investigation. MS: Visualization, Writing – review & editing, Investigation.

MHT: Investigation, Visualization, Validation, Writing – review & editing. AT: Visualization, Investigation, Funding acquisition, Writing – review & editing. JB: Investigation, Visualization, Writing – review & editing. ES: Investigation, Validation, Writing – review & editing, Visualization, Funding acquisition. BB: Investigation, Methodology, Conceptualization, Writing – review & editing, Supervision, Formal analysis, Resources, Visualization, Funding acquisition, Writing – original draft.

Funding

The author(s) declared that financial support was received for this work and/or its publication. This research was supported by EU funding within the NextGenerationEU-MUR PNRR Extended Partnership initiative on Emerging Infectious Diseases (Project no. PE00000007, INF ACT).

Acknowledgments

This work was granted by the Italian Ministry of University and Research PNRR Project ELIXIRxNextGenIT - ELIXIR x NextGenerationIT: consolidation of the Italian Infrastructure for Omics Data and Bioinformatics - code n. IR0000010. The authors are thankful to Luigi Boccaccio, Maria Rosa Mirizzi and Barbara De Marzo for their administrative support. We also acknowledge the

members of ELIXIR Food & Nutrition community for their valuable scientific comments on the manuscript.

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declared that Generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy,

including review by the authors wherever possible. If you identify any issues, please contact us.

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.2026.1722500/full#supplementary-material>

References

- Afzaal, M., Saeed, F., Shah, Y. A., Hussain, M., Rabail, R., Socol, C. T., et al. (2022). Human gut microbiota in health and disease: unveiling the relationship. *Front. Microbiol.* 13:999001. doi: 10.3389/fmicb.2022.999001
- Al-Rashidi, H. E. (2022). Gut microbiota and immunity relevance in eubiosis and dysbiosis. *Saudi J Biol Sci* 29, 1628–1643. doi: 10.1016/j.sjbs.2021.10.068
- Andrade, B. G. N., Cuadrat, R. R. C., Tonetti, F. R., Kitazawa, H., and Villena, J. (2022). The role of respiratory microbiota in the protection against viral diseases: respiratory commensal bacteria as next-generation probiotics for COVID-19. *Biosci Microbiota Food Health* 41, 94–102. doi: 10.12938/bmfh.2022-009
- Arikan, M., and Muth, T. (2023). Integrated multi-omics analyses of microbial communities: a review of the current state and future directions. *Mol. Omics* 19, 607–623. doi: 10.1039/D3MO00089C
- Artemev, A., Naik, S., Pougno, A., Honnavar, P., and Shanbhag, N. M. (2022). The association of microbiome dysbiosis with colorectal cancer. *Cureus* 14:e22156. doi: 10.7759/cureus.22156
- Balech, B., Brennan, L., Pau, E., Cavalieri, D., Coort, S., D'Elia, D., et al. (2022). The future of food and nutrition in ELIXIR. *FI000Res* 11:978. doi: 10.12688/fi000research.51747.1
- Bedu-Ferrari, C., Biscarrat, P., Pepke, F., Vati, S., Chaudemanche, C., Castelli, F., et al. (2024). In-depth characterization of a selection of gut commensal bacteria reveals their functional capacities to metabolize dietary carbohydrates with prebiotic potential. *mSystems* 9:e01401-23. doi: 10.1128/msystems.01401-23
- Bell, H. N., Rebernick, R. J., Goyert, J., Singhal, R., Kuljanin, M., Kerk, S. A., et al. (2022). Reuterin in the healthy gut microbiome suppresses colorectal cancer growth through altering redox balance. *Cancer Cell* 40, 185–200.e6. doi: 10.1016/j.ccell.2021.12.001
- Brockmann, L., Tran, A., Huang, Y., Edwards, M., Ronda, C., Wang, H. H., et al. (2023). Intestinal microbiota-specific Th17 cells possess regulatory properties and suppress effector T cells via c-MAF and IL-10. *Immunity* 56, 2719–2735.e7. doi: 10.1016/j.immuni.2023.11.003
- Brunkwall, L., and Orho-Melander, M. (2017). The gut microbiome as a target for prevention and treatment of hyperglycaemia in type 2 diabetes: from current human evidence to future possibilities. *Diabetologia* 60, 943–951. doi: 10.1007/s00125-017-4278-3
- Buttigieg, P. L., Morrison, N., Smith, B., Mungall, C. J., and Lewis, S. E. (2013). The environment ontology: contextualising biological and biomedical entities. *J. Biomed. Semantics* 4:43. doi: 10.1186/2041-1480-4-43
- Chen, I.-M. A., Chu, K., Palaniappan, K., Ratner, A., Huang, J., Huntemann, M., et al. (2023). The IMG/M data management and analysis system v.7: content updates and new features. *Nucleic Acids Res.* 51, D723–D732. doi: 10.1093/nar/gkac976
- Chen, Y.-A., Kawashima, H., Park, J., Mohsen, A., Hosomi, K., Nakagata, T., et al. (2025). NIBN Japan microbiome database, a database for exploring the correlations between human microbiome and health. *Sci. Rep.* 15:19640. doi: 10.1038/s41598-025-04339-z
- Chen, N., Li, L., Han, Y., and Chen, Z. (2025). The role of gut microbiota in the modulation of pulmonary immune response to viral infection through the gut-lung axis. *JIR* 18, 11755–11781. doi: 10.2147/JIR.S525880
- Chen, Y., Xiao, L., Zhou, M., and Zhang, H. (2024). The microbiota: a crucial mediator in gut homeostasis and colonization resistance. *Front. Microbiol.* 15:1417864. doi: 10.3389/fmicb.2024.1417864
- Côté, R., Reisinger, F., Martens, L., Barsnes, H., Vizcaino, J. A., and Hermjakob, H. (2010). The ontology lookup service: bigger and better. *Nucleic Acids Res.* 38, W155–W160. doi: 10.1093/nar/gkq331
- Dai, D., Zhu, J., Sun, C., Li, M., Liu, J., Wu, S., et al. (2022). GMrepo v2: a curated human gut microbiome database with special focus on disease markers and cross-dataset comparison. *Nucleic Acids Res.* 50, D777–D784. doi: 10.1093/nar/gkab1019
- Díez-Madueño, K., de la Cueva Dobao, P., Torres-Rojas, I., Fernández-Gosende, M., Hidalgo-Cantabrana, C., and Coto-Segura, P. (2025). Gut dysbiosis and adult atopic dermatitis: a systematic review. *J. Clin. Med.* 14:19. doi: 10.3390/jcm14010019
- Effendi, R. M. R. A., Anshory, M., Kalim, H., Dwiyana, R. F., Suwarsa, O., Pardo, L. M., et al. (2022). Akkermansia muciniphila and Faecalibacterium prausnitzii in immune-related diseases. *Microorganisms* 10:2382. doi: 10.3390/microorganisms10122382
- Field, D., Garrity, G., Gray, T., Morrison, N., Selengut, J., Sterk, P., et al. (2008). The minimum information about a genome sequence (MIGS) specification. *Nat. Biotechnol.* 26, 541–547. doi: 10.1038/nbt1360
- Franzosa, E. A., Morgan, X. C., Segata, N., Waldron, L., Reyes, J., Earl, A. M., et al. (2014). Relating the metatranscriptome and metagenome of the human gut. *Proc. Natl. Acad. Sci. USA* 111, E2329–E2338. doi: 10.1073/pnas.1319284111
- Gomes, A. C., Hoffmann, C., and Mota, J. F. (2018). The human gut microbiota: metabolism and perspective in obesity. *Gut Microbes* 9, 308–325. doi: 10.1080/19490976.2018.1465157
- Gonzalez, A., Navas-Molina, J. A., Kosciolke, T., McDonald, D., Vázquez-Baeza, Y., Ackermann, G., et al. (2018). Qiita: rapid, web-enabled microbiome meta-analysis. *Nat. Methods* 15, 796–798. doi: 10.1038/s41592-018-0141-9
- Goudman, L., Demuyser, T., Pilitis, J. G., Billot, M., Roulaud, M., Rigoard, P., et al. (2024). Gut dysbiosis in patients with chronic pain: a systematic review and meta-analysis. *Front. Immunol.* 15:1342833. doi: 10.3389/fimmu.2024.1342833
- Hamjane, N., Mechita, M. B., Nourouti, N. G., and Barakat, A. (2024). Gut microbiota dysbiosis-associated obesity and its involvement in cardiovascular diseases and type 2 diabetes. A systematic review. *Microvasc. Res.* 151:104601. doi: 10.1016/j.mvr.2023.104601
- Hou, K., Wu, Z.-X., Chen, X.-Y., Wang, J.-Q., Zhang, D., Xiao, C., et al. (2022). Microbiota in health and diseases. *Signal Transduct. Target. Ther.* 7, 1–28. doi: 10.1038/s41392-022-00974-4
- Hug, L. A., Hatzenpichler, R., Moraru, C., Soares, A. R., Meyer, F., Heyder, A., et al. (2025). A roadmap for equitable reuse of public microbiome data. *Nat. Microbiol.* 10, 2384–2395. doi: 10.1038/s41564-025-02116-2
- Huntemann, M., Ivanova, N. N., Mavromatis, K., Tripp, H. J., Paez-Espino, D., Tennessen, K., et al. (2016). The standard operating procedure of the DOE-JGI metagenome annotation pipeline (MAP v.4). *Stand. Genomic Sci.* 11:17. doi: 10.1186/s40793-016-0138-x

- Iebba, V., Totino, V., Gagliardi, A., Santangelo, F., Cacciotti, F., Trancassini, M., et al. (2016). Eubiosis and dysbiosis: the two sides of the microbiota. *New Microbiol.* 39, 1–12.
- Integrative HMP (iHMP) Research Network Consortium (2019). The integrative human microbiome project. *Nature* 569, 641–648. doi: 10.1038/s41586-019-1238-8
- Janssens, Y., Nielandt, J., Bronselaer, A., Debonne, N., Verbeke, F., Wynendaele, E., et al. (2018). Disbiome database: linking the microbiome to disease. *BMC Microbiol.* 18:50. doi: 10.1186/s12866-018-1197-5
- Katz, K., Shutov, O., Lapoint, R., Kimelman, M., Brister, J. R., and O'Sullivan, C. (2022). The sequence read archive: a decade more of explosive growth. *Nucleic Acids Res.* 50, D387–D390. doi: 10.1093/nar/gkab1053
- Keegan, K. P., Glass, E. M., and Meyer, F. (2016). MG-RAST, a metagenomics Service for Analysis of microbial community structure and function. *Methods Mol. Biol.* 1399, 207–233. doi: 10.1007/978-1-4939-3369-3_13
- Khalil, M., Di Ciaula, A., Mahdi, L., Jaber, N., Di Palo, D. M., Graziani, A., et al. (2024). Unraveling the role of the human gut microbiome in health and diseases. *Microorganisms* 12:2333. doi: 10.3390/microorganisms12112333
- Kreouzi, M., Theodorakis, N., Nikolou, M., Feretzakis, G., Anastasiou, A., Kalodanis, K., et al. (2025). Skin microbiota: mediator of interactions between metabolic disorders and cutaneous health and disease. *Microorganisms* 13:161. doi: 10.3390/microorganisms13010161
- Li, S., Liu, J., Zheng, X., Ren, L., Yang, Y., Li, W., et al. (2021). Tumorigenic bacteria in colorectal cancer: mechanisms and treatments. *Cancer Biol. Med.* 19, 147–162. doi: 10.20892/j.issn.2095-3941.2020.0651
- Liu, Y., Yan, D., Chen, R., Zhang, Y., Wang, C., and Qian, G. (2025). Recent insights and advances in gut microbiota's influence on host antiviral immunity. *Front. Microbiol.* 16:1536778. doi: 10.3389/fmicb.2025.1536778
- Ma, L., Terwilliger, A., and Maresso, A. W. (2015). Iron and zinc exploitation during bacterial pathogenesis. *Metallomics* 7, 1541–1554. doi: 10.1039/c5mt00170f
- Martin, C. R., Osadchij, V., Kalani, A., and Mayer, E. A. (2018). The brain-gut-microbiome Axis. *Cell. Mol. Gastroenterol. Hepatol.* 6, 133–148. doi: 10.1016/j.jcmgh.2018.04.003
- Mehta, I., Juneja, K., Nimmakayala, T., Bansal, L., Pulekar, S., Duggineni, D., et al. (2025). Gut microbiota and mental health: a comprehensive review of gut-brain interactions in mood disorders. *Cureus* 17:e81447. doi: 10.7759/cureus.81447
- Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E., Kubal, M., et al. (2008). The metagenomics RAST server – a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* 9:386. doi: 10.1186/1471-2105-9-386
- Mithradas, N., Sudhakar, U., Shanmugapriya, K., Jeddy, N., and Ram, S. (2024). The oral-lung microbiome dysbiosis: unravelling its role in implications for chronic obstructive pulmonary disease (COPD) pathogenesis. *J. Oral. Maxillofac. Pathol.* 28, 619–625. doi: 10.4103/jomfp.jomfp_277_24
- Morais, L. H., Schreiber, H. L., and Mazmanian, S. K. (2021). The gut microbiota–brain axis in behaviour and brain disorders. *Nat. Rev. Microbiol.* 19, 241–255. doi: 10.1038/s41579-020-00460-0
- Mukherjee, S., Stamatis, D., Li, C. T., Ovchinnikova, G., Bertsch, J., Sundaramurthi, J. C., et al. (2023). Twenty-five years of genomes OnLine database (GOLD): data updates and new features in v.9. *Nucleic Acids Res.* 51, D957–D963. doi: 10.1093/nar/gkac974
- Mukherjee, S., Stamatis, D., Li, C. T., Ovchinnikova, G., Kandimalla, M., Handke, V., et al. (2025). Genomes OnLine database (GOLD) v.10: new features and updates. *Nucleic Acids Res.* 53, D989–D997. doi: 10.1093/nar/gkae1000
- Noy, N. F., Shah, N. H., Whetzel, P. L., Dai, B., Dorf, M., Griffith, N., et al. (2009). BioPortal: ontologies and integrated data resources at the click of a mouse. *Nucleic Acids Res.* 37, W170–W173. doi: 10.1093/nar/gkp440
- O'Cathail, C., Ahamed, A., Burgin, J., Cummins, C., Devaraj, R., Gueye, K., et al. (2025). The European nucleotide archive in 2024. *Nucleic Acids Res.* 53, D49–D55. doi: 10.1093/nar/gkae975
- Origiuela, V., and Lopez-Zaplana, A. (2025). Gut microbiota: an immersion in dysbiosis, associated pathologies, and probiotics. *Microorganisms* 13:1084. doi: 10.20944/preprints202503.1398.v1
- Pasolli, E., Schiffer, L., Manghi, P., Renson, A., Obenchain, V., Truong, D. T., et al. (2017). Accessible, curated metagenomic data through ExperimentHub. *Nat. Methods* 14, 1023–1024. doi: 10.1038/nmeth.4468
- Qi, C., Cai, Y., Qian, K., Li, X., Ren, J., Wang, P., et al. (2022). gutMDisorder v2.0: a comprehensive database for dysbiosis of gut microbiota in phenotypes and interventions. *Nucleic Acids Res.* 51, D717–D722. doi: 10.1093/nar/gkac871
- Qiu, P., Ishimoto, T., Fu, L., Zhang, J., Zhang, Z., and Liu, Y. (2022). The gut microbiota in inflammatory bowel disease. *Front. Cell. Infect. Microbiol.* 12:733992. doi: 10.3389/fcimb.2022.733992
- Richardson, L., Allen, B., Baldi, G., Beracochea, M., Bileschi, M. L., Burdett, T., et al. (2023). MGnify: the microbiome sequence data analysis resource in 2023. *Nucleic Acids Res.* 51, D753–D759. doi: 10.1093/nar/gkac1080
- Rinninella, E., Raoul, P., Cintoni, M., Franceschi, F., Miggiano, G. A. D., Gasbarrini, A., et al. (2019). What is the healthy gut microbiota composition? A changing ecosystem across age, environment, diet, and diseases. *Microorganisms* 7:14. doi: 10.3390/microorganisms7010014
- Shi, W., Qi, H., Sun, Q., Fan, G., Liu, S., Wang, J., et al. (2019). gcMeta: a global catalogue of metagenomics platform to support the archiving, standardization and analysis of microbiome data. *Nucleic Acids Res.* 47, D637–D648. doi: 10.1093/nar/gky1008
- Speir, M. L., Teh, W. K., Perry, M. D., Schwartz, R., Nejad, P., Harris, T., et al. (2025). Making genomic data FAIR through effective data portals. *Sci Data* 12:1872. doi: 10.1038/s41597-025-06142-x
- Tang, J., Wu, X., Mou, M., Wang, C., Wang, L., Li, F., et al. (2021). GIMICA: host genetic and immune factors shaping human microbiota. *Nucleic Acids Res.* 49, D715–D722. doi: 10.1093/nar/gkaa851
- Turnbaugh, P. J., Ley, R. E., Hamady, M., Fraser-Liggett, C. M., Knight, R., and Gordon, J. I. (2007). The human microbiome project. *Nature* 449, 804–810. doi: 10.1038/nature06244
- van Ommen, B., Bouwman, J., Dragsted, L. O., Drevon, C. A., Elliott, R., de Groot, P., et al. (2010). Challenges of molecular nutrition research 6: the nutritional phenotype database to store, share and evaluate nutritional systems biology studies. *Genes Nutr.* 5, 189–203. doi: 10.1007/s12263-010-0167-9
- Vitali, F., Lombardo, R., Rivero, D., Mattivi, F., Franceschi, P., Bordoni, A., et al. (2018). ONS: an ontology for a standardized description of interventions and observational studies in nutrition. *Genes Nutr.* 13:12. doi: 10.1186/s12263-018-0601-y
- Wilke, A., Harrison, T., Wilkening, J., Field, D., Glass, E. M., Kyrpides, N., et al. (2012). The M5nr: a novel non-redundant database containing protein sequences and annotations from multiple sources and associated tools. *BMC Bioinformatics* 13:141. doi: 10.1186/1471-2105-13-141
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J. J., Appleton, G., Axton, M., Baak, A., et al. (2016). The FAIR guiding principles for scientific data management and stewardship. *Sci Data* 3:160018. doi: 10.1038/sdata.2016.18
- Yamashita, Y., and Takeshita, T. (2017). The oral microbiome and human health. *J. Oral Sci.* 59, 201–206. doi: 10.2334/josnusd.16-0856
- Yassin, L. K., Nakhla, M. M., Alderei, A., Almehairbi, A., Mydeen, A. B., Akour, A., et al. (2025). Exploring the microbiota-gut-brain axis: impact on brain structure and function. *Front. Neuroanat.* 19:1504065. doi: 10.3389/fnana.2025.1504065
- Yilmaz, P., Kottmann, R., Field, D., Knight, R., Cole, J. R., Amaral-Zettler, L., et al. (2011). Minimum information about a marker gene sequence (MIMARKS) and minimum information about any (x) sequence (MIxS) specifications. *Nat. Biotechnol.* 29, 415–420. doi: 10.1038/nbt.1823
- Yu, L., Zhao, G., Wang, L., Zhou, X., Sun, J., Li, X., et al. (2022). A systematic review of microbial markers for risk prediction of colorectal neoplasia. *British J. Cancer* 126, 1318–1328. doi: 10.1038/s41416-022-01740-7
- Zhao, M., Chu, J., Feng, S., Guo, C., Xue, B., He, K., et al. (2023). Immunological mechanisms of inflammatory diseases caused by gut microbiota dysbiosis: a review. *Biomed. Pharmacother.* 164:114985. doi: 10.1016/j.biopha.2023.114985
- Zhou, M., Liu, Y., Yin, X., Gong, J., and Li, J. (2025). The role of oral microbiota in lung carcinogenesis through the oral-lung axis: a comprehensive review of mechanisms and therapeutic potential. *Discover Oncol.* 16:1651. doi: 10.1007/s12672-025-03440-z