

EDITORIAL

Open Access



A blueprint for contemporary studies of microbiomes

Laure B. Bindels^{1,2*}, Joy E. M. Watts³, Kevin R. Theis⁴, Víctor J. Carrion^{5,6,7,8}, Adam Ossowicki^{5,6}, Jana Seifert⁹, Julia Oh¹⁰, Yongqi Shao¹¹, Markus Hilty¹², Purnima Kumar¹³, Falk Hildebrand^{14,15}, Connie Lovejoy¹⁶, Paul Wigley¹⁷, Ke Yu¹⁸, Meiling Zhang¹⁹, Tong Zhang^{20,21}, Jens Walter²², Mahesh S. Desai²³, Sharon Ann Huws²⁴, Lynn M. Schriml²⁵, Jacques Ravel^{26,27}, W. Florian Fricke^{28,29}, Emiley A. Eloë-Fadrosh³⁰, Charles K. Lee^{31,32} and Thomas Clavel³³

Abstract

This editorial piece co-authored by the Senior Editors at *Microbiome* aims to highlight current challenges in the field of environmental and host-associated microbiome research. We also take the opportunity to clarify our expectations for the articles submitted to the journal. At *Microbiome*, we are seeking studies that provide either new mechanistic insights into the role of microbiomes in health and environmental systems or substantial conceptual or technical advances. Manuscripts need to meet high standards of language accuracy, quality of microbiome analyses, and data and protocol availability, including detailed reporting of wet-lab and in silico protocols, all of which can critically enhance transparency and reproducibility. We think that such efforts are essential to push the boundaries of our knowledge on microbiomes in a concerted, international effort.

On the importance of accurate terminology

The misuse of terms such as microbiome, microbiota, abundance, 16S ribosomal RNA (rRNA) gene, and metagenomics, among others, has contributed to the misunderstanding of many studies by the scientific community and the lay public [1–3]. The microbiota includes the communities of microorganisms (bacteria, archaea, fungi, viruses, and protists, including micro-algae) that inhabit an ecosystem, although most studies investigate only the bacterial component of the microbiota [3]. The term “microbiota” should be differentiated from the term “microbiome”, the latter encompassing the entire microbial ecosystem, namely the community of microorganisms together with their theater of activity (including structural elements (gene, expressed gene and proteins),

metabolites/signal molecules, and their surrounding environmental conditions) [3]. The term “microflora”, often wrongly used as a synonym for “microbiota”, refers to “microscopic plants or the plants or flora of a micro-habitat” [2]. An additional emerging semantic issue is the use of the term “pseudo-germfree mice” when referring to antibiotic-treated mice. This term is impossible to define, biologically implausible and misleading and should therefore not be used.

Confusion also exists around the denomination of methodologies used to assess the composition and function of microbiota. The sequencing of PCR amplicons of hypervariable regions of the gene encoding for 16S or 18S rRNA should be described as “16S rRNA gene amplicon sequencing”, or “18S rRNA gene amplicon sequencing”; the use of truncated versions of this term (e.g. 18S or 16S sequencing) or referring to rDNA is inaccurate. Whilst this technique provides information on the diversity and taxonomic composition of prokaryotic members of the microbiota, it cannot be described as “metagenomics”.

*Correspondence:

Laure B. Bindels

laure.bindels@uclouvain.be

Full list of author information is available at the end of the article



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Metagenomics refers to the random sequencing of all DNA within a given sample [2] and provides information on the functional potential of the microbial communities under study.

Another important semantic problem leading to technical misconception is the widespread use of the term “abundance” when reporting proportional results of amplicon sequencing or metagenomic studies [1]. It is misleading to refer to “abundance” and the term “relative abundance” must be used. General terms such as “occurrence” or “presence” may be used whenever appropriate if they improve the clarity of the text, keeping in mind that sequencing approaches are usually restricted to dominant populations and do not enable the detection of all members in a complex microbial community. Quantitative microbiome analysis approaches exist, such as cultivation or sequencing profiles corrected by quantitative measurement of bacterial load by flow cytometry or quantitative PCR [4–6]. In such cases, the data are in “absolute abundances”, but these approaches are more complex to implement and always need to be benchmarked using complex mock communities.

Misleading language can also include the way microbes are associated with supposed beneficial or negative effects on their host or environment. For example, host-associated lactobacilli and bifidobacteria are often wrongly referred to as probiotics [1]. Exogenous microbes that are administered to a host and hold potential health benefits can be referred to as probiotic candidates; in contrast, endogenous members of microbial communities are commensals, not probiotics. In general, care should be taken not to label entire microbial taxa, especially families or genera, as probiotic, beneficial, pathogenic, detrimental, or pathobionts, because the observed effects can be highly species- or even strain-specific [7, 8].

Nomenclature rules and taxonomic accuracy also matter. There are rules dictating how organisms must be named. For cultured bacteria these are set by The International Code of Nomenclature for Prokaryotes (ICNP), maintained by the International Committee on Systematics of Prokaryotes (ICSP) [9, 10]. In 2021, phylum was added to the ranks covered by the rules of the Code, which led to the update of widely used names of numerous phyla, such as *Bacillota* (formerly Firmicutes) or *Bacteroidota* (formerly Bacteroidetes) [11]. The List of Prokaryotic names with Standing in Nomenclature (LPSN) is a good reference for authors to check the validity of bacterial names [12, 13]. Importantly, not all databases agree on the taxonomic placement and naming of bacteria, and they are not all systematically updated, highlighting the importance of reporting the name and version of resources used to name microbes during

sequencing-based analysis of microbiomes. One of the most comprehensive databases is the Genome Taxonomy Database (GTDB) [14]. It provides a phylogenetically consistent and rank-normalized genome-based taxonomy for prokaryotic genomes. However, it does not strictly follow naming rules according to the ICNP and also contains names validated under the SeqCode [15]. At *Microbiome*, we require that validly published bacterial names are written according to the guidelines of the *American Society for Microbiology* and the *Journal of Bacteriology* [16, 17]. Essentially, validly published Latin names at any taxonomic level (kingdom, phyla, class, order, family, genus, species, and subspecies) should be italicized. In contrast, anglicized versions of Latin bacterial names, such as bifidobacteria, lactobacilli, clostridia, or enterococci are written without a capital letter and are not italicized. Strain designations or numbers should also not be italicized. For taxa belonging to the category ‘*Candidatus*’, the word ‘*Candidatus*’, but not the genus name and/or the vernacular epithet, should be printed in italics [18, 19]. For recent valid names of reclassified taxa or in the case of synonyms, authors may add in brackets a reference to the previous, or other name to facilitate understanding by the readers. Similarly, there are international frameworks for the taxonomy of fungi, protists and algae. We recommend that researchers working with these groups become familiar with these conventions and clearly state the source of names and classification systems used in publications. For viruses, the International Committee on Taxonomy of Viruses (ICTV) is the authority for developing, refining, and maintaining a universal virus taxonomy [20].

Avoiding pitfalls in taxonomic and functional analyses of microbiomes

Biases, such as the over/underestimation of individual microbiota members as well as contamination/depletion issues that result in false positive/negative taxon identification, can be introduced at every step of microbiota analysis: sample collection and preservation, DNA extraction, library construction, sequencing, bioinformatics, biostatistics, and data visualization. The recent controversy over the existence of the placental or pre-natal human microbiome, now widely considered to be the result of contaminations and misinterpretation, is an example of how reports of microbiomes can have a lasting impact on scientific discourse [21, 22]. This risk should be minimized by following general and current technical recommendations, including controls and adequate replication, as detailed elsewhere [23–25]. Importantly, manuscripts should always contain a detailed description of these technical aspects, including all the experimental details required to reproduce the work.

Examples of details that are often omitted include (non-exhaustive list): method of sample collection, duration and conditions of sample storage, method of lysis during DNA extraction, primers used and number of cycles (for amplicon studies), wet-lab and in silico strategies to prevent the analysis of spurious taxa. As the field continues to evolve, and contamination, experimental challenges, and technical biases can never be completely avoided, their detailed, comprehensive, and transparent assessment, characterization and discussion in manuscripts remains a critical requirement for studies published in *Microbiome*.

Appropriate controls

Adequate experimental controls, reagent-negative controls (also called “blanks”) and mock communities should be included in the entire extraction, sequencing and analysis process, and their exact nature adapted to the specific study objectives. Negative controls, introduced at each step of the process (e.g., DNA stabilization solutions without sample, controls for collection devices, extraction controls, PCR controls) are essential when working with low biomass samples, like oral, respiratory, or environmental microbiota, to better control for the inclusion of artefacts from reagents, and other sources of contamination [26, 27]. Data from negative controls should be released alongside the sample data, and the results of their analysis should be compared to those of the samples under study, with an interpretation that will depend on the project [22, 27]. Similarly, biological mock communities (i.e., known mixtures of microorganisms or their DNA that, if possible, reflect the diversity and taxonomic composition of the microbial communities under investigation) should be included to assess potential bias in taxonomic analyses. The mock community composition and sequence results should also be made publicly available with other data from the study; such composition should be compared to the theoretical composition [28, 29]. In addition to generic controls, there is a need to develop habitat-specific mock communities. For complex environments, mock communities with higher diversity (e.g. more than 10 taxa) should be considered. Non-biological mock communities (i.e., samples that contain lab-made variable regions that do not exist in nature) are also relevant to assess cross-sample contamination and tag switching, and to parametrize bioinformatics pipelines [30].

DNA extraction

DNA extraction methods have a large effect on the outcome of microbiome analysis [31]. Nucleic acid extraction should be optimized to obtain accurate representation of

the microbial community present in the sample/environmental matrix [32]. For example, including a bead-beating in the extraction protocol is highly recommended for fecal and soil samples to avoid losing specific taxa [33, 34]. However, biases seem to be inherently related to extraction; there is not a “one size fits all” protocol that would allow to accurately capture the genomes of all strains present in a given sample [35]. The detection of all taxa in complex communities is also hampered by the nature of current sequencing approaches, which are de facto limited to dominant populations. DNA extraction of complex samples such as plant tissues can lead to contamination with mitochondria and chloroplast rRNA genes, which represent a significant source of bias [36]. In these cases, we recommend the use of specific blockers and/or discriminating PCR primers [37].

Sequencing

Several sequencing-based approaches and techniques for the analyses of microbiomes exist, with their own strengths and weaknesses [38]. In general, we advise the use of unique dual sequencing indices to reduce the risk of misassigned reads during the demultiplexing step [39]. Sequencing amplicons of the 16S rRNA gene, 18S rRNA gene, and the fungal internal transcribed spacer (ITS) region allow researchers to assess the diversity and composition of bacterial, eukaryotic, and fungal communities, in most cases at the family or genus level depending on the ecosystem and the taxonomic group studied. The implementation of protocols that enable sequencing of the entire gene or even ribosomal RNA operon will improve resolution [40–43]. Amplicon-based approaches can be supplemented or replaced by shotgun sequencing for metagenomic analysis. Although more expensive, computationally demanding, and difficult to perform using low-biomass samples, metagenomics facilitates strain-level resolution and gain of insight on functional potential, pending sufficient sequencing depth [44]. Shallow metagenomics has the advantage of avoiding PCR bias compared to amplicon sequencing but does not enable the level of taxonomic accuracy and functionality provided by deep shotgun sequencing. Prediction of functional microbiota profiles based on 18S or 16S rRNA gene-derived taxonomic information is not reliable, mainly due to the large fraction of unknown microbes, microbes without appropriate reference genomes [45, 46] and substantial gene content differences even between strains of the same species [47, 48]. We therefore do not recommend marker gene-based in silico functional predictions, unless supported by additional analyses or to generate hypotheses that are then tested experimentally.

Sequencing data analysis

Preprocessing and downstream analyses of sequencing data can be done using multiple tools. The goal here is not to provide recommendations but to clarify important general aspects.

Firstly, at *Microbiome*, we strongly encourage the use of open access tools, which are essential to ensure reproducibility of the analyses and assessment of potential biases related to this step; this aspect is further developed below.

Secondly, recent versions of taxonomic databases should be used to ensure adequate taxonomic analysis. Such databases are evolving quickly due to the ongoing effort to characterize unknown taxa in microbial ecosystems [12, 49–51]. As mentioned above, inconsistencies in the nomenclature across databases exist and databases can suffer from shortcomings in their annotation. The authors are thus required to state the version number or date of accession to ensure that the results reported are traceable. Metagenomics faces similar challenges as amplicon sequencing in terms of annotation accuracy and maintaining high-quality reference databases. Improvements in genome-resolved metagenomics have enabled the creation of comprehensive and ecosystem-specific catalogs of population genomes or taxonomically annotated genes [52–57]. These resources, however, are being generated so quickly that the maintenance and harmonization across systems are major challenges [58–60] and de novo generation of sample-specific gene catalogues or genomes may be preferable [61].

Thirdly, *Microbiome* encourages the submission of thorough and comprehensive studies that aim at benchmarking new or even established metagenomic software. Guidelines for appropriate benchmarking are emerging, supported by the Critical Assessment of Metagenome Interpretation (CAMI) initiative [62, 63]. In future benchmarking studies, biological mock communities of higher complexity and specificity to different microbial habitats will play an important role, for example to assess bioinformatic workflows used to reconstruct metagenome-assembled genomes (MAGs), as they are widely used as references to analyze metagenomic data and to draw conclusions on strain-level differences in microbiomes.

Biostatistical analyses

A key aspect of biostatistical analyses is the use of methods that consider the compositional nature of microbial datasets generated by sequencing, are suited for the study design, and include false discovery rate (FDR) correction [64–67]. Rarefaction techniques are recommended to be used for specific analyses that are known to be sensitive to sequencing depth. Microbiome sequencing datasets

are typically sparse, meaning that they frequently contain zero-counts (0), due to the taxon or gene not being detectable or present. The use of null values in statistical analysis and data presentation has been a matter of debate, as a null value can only indicate that the given taxon was not detected, possibly because it is part of subdominant communities [68]. The choice of whether to include null values ultimately depends on the goal of the study. One solution is to compare the occurrence of taxa only in those samples that contain them and assess presence/absence separately by prevalence testing. The choice of the method used also impacts the results. Regarding the identification of differentially abundant microbes, methods such as ALDEx2 and ANCOM-II produce the most consistent results across studies and agree best with the intersect of results from different approaches [64]. In line with this study, we propose the use of a consensus approach based on multiple differential abundance methods to help ensure robust biological interpretations. All these aspects of the analysis should always be carefully considered and clearly reported.

Data interpretation

As mentioned above, the abundance of taxa or genes, expressed either at relative or absolute levels, can lead to different interpretations [4, 69]. Considering the potential bias that can arise from comparing the taxonomic composition of microbial communities that differ in their microbial load, we also recommend (when possible) to complement compositional analyses with absolute quantification of microbial taxa of interest for the study [1, 70–72]. A known example of misconception arising from differences in relative abundances is the so-called Firmicutes to Bacteroidetes ratio (*Bacillota* to *Bacteroidota* ratio as per the current nomenclature). The original studies proposing this ratio were underpowered and phyla encompass a broad variety of bacterial strains, hence with no functional insight provided by such ratio [35]. Therefore, studies submitted to *Microbiome* are discouraged to draw conclusions based on this ratio without providing thorough information on why doing so and further demonstrating the validity of the findings.

Another main challenge in microbiome studies is the interpretation of alpha-diversity measures. Diversity is a general term often used (and sometimes misused) in microbiome studies and it became of major interest due to “low diversity” gut microbiomes proposed to be associated with multiple diseases. As diversity analysis in microbial ecology has many facets and different opinions exist, we only report here our own view on a limited number of elements. Alpha-diversity describes the bacterial diversity within a given sample, with metrics focusing on the richness, reflecting the observed

number of different taxa in a sample, or the evenness, evaluating the uniformity of distribution of these taxa, or both parameters at once [73, 74]. In general, when reporting alpha-diversity, we recommend working with data that have undergone rarefaction to account for differences in sequencing depth [75]. Importantly, many of the approaches used to assess ecosystem diversity were developed for landscape-based diversity estimation and their applicability to microbiome data should be questioned before use. For instance, the Chao1 and ACE estimators are commonly used, but their calculation is based on the occurrence of singletons in a dataset. Given that microbiome data are sparse, that the occurrence of singletons varies with sequencing depth, and that bioinformatic data processing frequently involves the removal of singletons, the Chao1 and ACE indices should not be used in most cases, especially in amplicon sequence variant (ASV)-based analysis approaches [76]. The Shannon index is also commonly used, but just as diameter is an index of the volume of a sphere, the Shannon index is only one parameter associated with diversity. There is no direct proportionality between the index and diversity (a doubling of the Shannon index does not mean a doubling of diversity), and the meaning of “statistically significant” changes in Shannon diversity for an ecosystem is difficult to comprehend and can be misleading. An alternative is to calculate the effective number of species based on the Shannon index, which provides an easily interpretable number of species that integrates their relative proportion within the sample [77]. Lastly, alpha-diversity metrics from different studies cannot be directly compared due to the normalization and filtering steps that are inherently different between studies. New concepts, such as effective microbial richness [28], are emerging to address this issue and their evaluation and use is encouraged.

Data visualization

We recommend the usage of box plots or violin plots instead of stacked bar plots for several reasons. Firstly, stacked bar plots do not allow the visualization of data distributions and their standard deviations. Secondly, low abundance taxa, which can be numerous, are not visible when included in stacked bar plots. Thirdly, plotting many taxa prevents straightforward visual recognition, as too many similar colors are required. If it is unavoidable for the study to include stacked bar plots to represent marked change, they must show the taxonomic composition in individual samples (presenting biological replicates individually). In addition, strategies for presenting low abundance taxa, such as relative abundance or prevalence filters, should be used to enhance visualization.

We also recommend the use of color vision deficiency (CVD)-friendly color palettes [78].

Reporting and sharing microbiome studies

Detailed reporting of microbiome studies is of utmost importance for reproducibility of the research, which has always been a priority at *Microbiome* [79]. We are endorsing the Strengthening The Organization and Reporting of Microbiome Studies (STORMS) guidelines, which describe reporting elements for laboratory, bioinformatic, and statistical analyses tailored to human microbiome studies [80]. The Standards for Technical Reporting in Environmental and host-Associated Microbiome Studies (STREAMS) Guidelines are being written based on the STORMS reporting checklist, which would allow to expand these guidelines to environmental, non-human host-associated, and synthetic microbiome studies [81].

In line with the importance of reproducibility, *Microbiome* follows a strict data release policy and expects datasets to comply with the Findability, Accessibility, Interoperability, and Reusability (FAIR) principles [82]. The FAIR Principles put specific emphasis on enhancing the ability of machines to automatically find and use the data, in addition to supporting its reuse by individuals. The principles apply not only to ‘data’ sensu stricto, but also to the metadata, algorithms, tools, and workflows that led to the presented findings. In line with these principles, we require that all datasets used in articles to generate findings and draw conclusions are available to the reviewers at the time of submission and publicly available at the time of publication, with raw data deposited in public repositories [83]. In addition to the metadata provided in international sequence repositories, metadata that are used in any analysis within a submitted manuscript need to be made available in its entirety in a recognized repository or as supporting files to the manuscript. We also suggest inclusion of supplementary table(s) that link(s) samples, sources, sequences and metadata. Metadata should be formatted according to the MIxS (Minimum Information about any (x) Sequence) standards developed by the Genomic Standards Consortium (GSC) [84]. The version of software used needs to be reported to facilitate reproducibility, with adequate citation to support the work of researchers that are making most of these tools freely accessible. We are also requesting that authors make the code/scripts used for their analysis available through public sources, with the associated license. This cultural change to make microbiology data Open and FAIR is promoted by initiatives and consortia like NFDI4MICROBIOTA [85] and NMDC [86]. These efforts are designed to promote transparency and complete reproducibility of microbiome analyses on a long-term basis.

Moving the microbiome field towards a mechanistic understanding

At *Microbiome*, we are looking for manuscripts of general interest, which provide significant new insights into the mechanistic role of microbiomes for health and the environment, or that lead to substantial conceptual or technical advances in the field. *Microbiome* is especially interested in studies that go beyond descriptive omics surveys and include experimental or theoretical approaches that mechanistically support proposed microbiome functions, and establish, if possible, cause-and-effect relationships. Associative studies, while valuable for generating hypotheses, often lead to overstatements and misconceptions about causal relationships, resulting in public skepticism and hesitation from the scientific community to embrace the importance of microbiomes, and should therefore be self-critically and conservatively interpreted. This position is also motivated by the risk of spurious correlations existing in large multivariate datasets [87–89]. These bioinformatically inferred associations are useful for reducing the number of potential hypotheses to be tested, but do not preclude the ultimate necessity for experimental validation [88]. For microbiome research to foster meaningful advancements, it is essential to prioritize studies that go beyond correlations and provide robust mechanistic insights.

Methods for establishing causality in the field of microbiomes have been widely discussed [1, 90, 91]. The most reliable way of generating mechanistic insight is to have a carefully constructed hypothesis, a sampling or experimental design that robustly interrogates the hypothesis (i.e., including all necessary controls and measuring all relevant variables), and appropriate validated methodologies to generate data that allow researchers to corroborate or reject the hypothesis. Useful field-advancing hypotheses must address an important knowledge gap that is well-supported by literature review and outlined, with supporting citations, in the introduction of the manuscript. Exceptions always exist since useful mechanistic insight can be generated without an explicit hypothesis (e.g., using techniques like metabolic labeling or click chemistry approaches [92, 93]), and there are habitats and ecosystems for which so little knowledge exists that useful hypotheses cannot be constructed de novo.

This editorial was informed by papers submitted to *Microbiome* and discussions between the current Senior Editors at *Microbiome*. *Microbiome* will continue to strive toward high-quality mechanistic research in this ever-expanding field as we are convinced that following this path is essential to collectively push the boundaries of our knowledge on microbiomes.

Acknowledgements

LBB is a Collen-Francqui Research Professor and grateful for the support of the Francqui Fondation. TC received funding from the German Research Foundation (DFG), project no. 460129525 – NFDI4Microbiota. CL is supported by the Natural Research and Engineering Council of Canada.

Authors' contributions

This Editorial arises from a discussion initiated by Thomas Clavel. A plan for this Editorial and associated strategy was drafted by Laure B. Bindels and further discussed and amended during a Senior Editor meeting. The manuscript was written by Laure B. Bindels and Thomas Clavel, with contributions from Joy Watts, Charles Lee, Kevin Theis, Emiley Eloe-Fadrosh, Victor Carrión, Adam Ossowski, Lynn Schriml, W. Florian Fricke, Jana Seifert, Connie Lovejoy and Falk Hildebrand. All Senior Editors read and approved the final version of the manuscript.

Declarations

Competing interests

Jacques Ravel is Editor-in-Chief of *Microbiome*, and all other authors of this Editorial are Senior Editors of *Microbiome*. Mahesh S. Desai works as a consultant and an advisory board member at Theralution GmbH, Germany.

Author details

¹Metabolism and Nutrition Research Group, Louvain Drug Research Institute, UCLouvain, Université catholique de Louvain, Brussels, Belgium. ²Welbio Department, WEL Research Institute, Wavre, Belgium. ³School of Environment and Life Sciences, University of Portsmouth, Portsmouth, UK. ⁴Department of Biochemistry, Microbiology, and Immunology, Wayne State University School of Medicine, Detroit, MI, USA. ⁵Departamento de Microbiología, Facultad de Ciencias, Campus Universitario de Teatino s/n, Universidad de Málaga, Málaga, Spain. ⁶Departamento de Protección de Cultivos, Instituto de Hortofruticultura Subtropical y Mediterránea "La Mayora", Campus Universitario de Teatino, Universidad de Málaga-Consejo Superior de Investigaciones Científicas (IHSM-UMA-CSIC), Málaga, Spain. ⁷Institute of Biology, Leiden University, Leiden, The Netherlands. ⁸Department of Microbial Ecology, Netherlands Institute of Ecology (NIOO-KNAW), Wageningen, The Netherlands. ⁹HoLMIR – Hohenheim Center for Livestock Microbiome Research, Functional Microbiology of Livestock, University of Hohenheim, Stuttgart, Germany. ¹⁰Department of Dermatology, Duke University, Durham, NC, USA. ¹¹Faculty of Agriculture, Life and Environmental Sciences, Zhejiang University, Hangzhou, China. ¹²Institute for Infectious Diseases, University of Bern, Bern, Switzerland. ¹³Department of Periodontics and Oral Medicine, School of Dentistry, University of Michigan, Ann Arbor, MI, USA. ¹⁴Quadram Institute Bioscience, Norwich Research Park, Norwich, Norfolk NR4 7UQ, UK. ¹⁵Earlham Institute, Norwich Research Park, Norwich, Norfolk NR4 7UQ, UK. ¹⁶Département de Biologie and Institut de Biologie Intégrative Et Des Systèmes, Université Laval, Québec, QC, Canada. ¹⁷Bristol Veterinary School, Langford Campus, University of Bristol, Bristol, UK. ¹⁸Shenzhen Graduate School, Peking University, Shenzhen, China. ¹⁹School of Life Sciences, East China Normal University, Shanghai 200241, China. ²⁰Environmental Microbiome Engineering and Biotechnology Laboratory, Department of Civil Engineering, Faculty of Engineering, The University of Hong Kong, Hong Kong, China. ²¹School of Public Health, The University of Hong Kong, Pokfulam Road, Hong Kong, China. ²²APC Microbiome Ireland, School of Microbiology, and Department of Medicine, University College Cork, Cork, Ireland. ²³Nutrition, Microbiome and Immunity Group, Department of Infection and Immunity, Luxembourg Institute of Health, Esch-Sur-Alzette, Luxembourg. ²⁴School of Biological Sciences, Institute for Global Food Security, Queen's University, Belfast, UK. ²⁵Department of Epidemiology and Public Health, Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, MD, USA. ²⁶Center for Advanced Microbiome Research and Innovation, Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, MD, USA. ²⁷Department of Microbiology and Immunology, University of Maryland School of Medicine, Baltimore, MD, USA. ²⁸Department of Microbiome Research and Applied Bioinformatics, University of Hohenheim, Stuttgart, Germany. ²⁹Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, MD, USA. ³⁰DOE Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, CA, USA.

³¹Environmental Research Institute, University of Waikato, Hamilton, New Zealand. ³²School of Science, University of Waikato, Hamilton, New Zealand. ³³Functional Microbiome Research Group, Institute of Medical Microbiology, RWTH University Hospital, Aachen, Germany.

Received: 17 February 2025 Accepted: 7 March 2025

Published online: 08 April 2025

References

- Shanahan F, Hill C. Language, numeracy and logic in microbiome science. *Nat Rev Gastroenterol Hepatol*. 2019;16(7):387–8.
- Marchesi JR, Ravel J. The vocabulary of microbiome research: a proposal. *Microbiome*. 2015;3:31.
- Berg G, Rybakova D, Fischer D, Cernava T, Verges MC, Charles T, et al. Microbiome definition re-visited: old concepts and new challenges. *Microbiome*. 2020;8(1):103.
- Vandeputte D, Kathagen G, D'Hoe K, Vieira-Silva S, Valles-Colomer M, Sabino J, et al. Quantitative microbiome profiling links gut community variation to microbial load. *Nature*. 2017;551(7681):507–11.
- Jian C, Luukkainen P, Yki-Jarvinen H, Salonen A, Korpela K. Quantitative PCR provides a simple and accessible method for quantitative microbiota profiling. *PLoS ONE*. 2020;15(1):e0227285.
- Tettamanti Boshier FA, Srinivasan S, Lopez A, Hoffman NG, Proll S, Fredricks DN, et al. Complementing 16S rRNA gene amplicon sequencing with total bacterial load to infer absolute species concentrations in the vaginal microbiome. *mSystems*. 2020;5(2):e00777.
- Hill C, Guarner F, Reid G, Gibson GR, Merenstein DJ, Pot B, et al. Expert consensus document: the international scientific association for probiotics and prebiotics consensus statement on the scope and appropriate use of the term probiotic. *Nat Rev Gastroenterol Hepatol*. 2014;11(8):506–14.
- Jochum L, Stecher B. Label or concept - what is a pathobiont? *Trends Microbiol*. 2020;28(10):789–92.
- Oren A, Arahal DR, Goker M, Moore ERB, Rossello-Mora R, Sutcliffe IC. International code of nomenclature of prokaryotes. *Prokaryotic code* (2022 revision). *Int J Syst Evol Microbiol*. 2023;73(5a):005782.
- International Committee on Systematics of Prokaryotes (ICSP). <https://www.the-icsp.org/>. Accessed 22 Mar 2025.
- Oren A, Arahal DR, Rossello-Mora R, Sutcliffe IC, Moore ERB. Emendation of Rules 5b, 8, 15 and 22 of the International Code of Nomenclature of Prokaryotes to include the rank of phylum. *Int J Syst Evol Microbiol*. 2021;71(6):004851.
- Parte AC, Sarda Carbasse J, Meier-Kolthoff JP, Reimer LC, Goker M. List of Prokaryotic names with Standing in Nomenclature (LPSN) moves to the DSMZ. *Int J Syst Evol Microbiol*. 2020;70(11):5607–12.
- List of Prokaryotic names with Standing in Nomenclature (LPSN). <https://lpsn.dsmz.de/>. Accessed 22 Mar 2025.
- Parks DH, Chuvochina M, Rinke C, Mussig AJ, Chaumeil PA, Hugenholtz P. GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. *Nucleic Acids Res*. 2022;50(D1):D785–94.
- Hedlund BP, Chuvochina M, Hugenholtz P, Konstantinidis KT, Murray AE, Palmer M, et al. SeqCode: a nomenclatural code for prokaryotes described from sequence data. *Nat Microbiol*. 2022;7(10):1702–8.
- American Society for Microbiology. <https://journals.asm.org/writing-your-paper#nomenclature>. Accessed 22 Mar 2025.
- Thines M, Aoki T, Crous PW, Hyde KD, Lücking R, Malosso E, et al. Setting scientific names at all taxonomic ranks in italics facilitates their quick recognition in scientific papers. *IMA Fungus*. 2020;11(1):25.
- International Code of Nomenclature of Prokaryotes. *Int J Syst Evol Microbiol*. 2019;69(1A):S1–111.
- Arahal D, Bisgaard M, Christensen H, Clermont D, Dijkshoorn L, Duim B, et al. The best of both worlds: a proposal for further integration of Candidatus names into the International Code of Nomenclature of Prokaryotes. *Int J Syst Evol Microbiol*. 2024;74(1):006188.
- Lefkowitz EJ, Dempsey DM, Hendrickson RC, Orton RJ, Siddell SG, Smith DB. Virus taxonomy: the database of the International Committee on Taxonomy of Viruses (ICTV). *Nucleic Acids Res*. 2018;46(D1):D708–17.
- Fricke WF, Ravel J. Microbiome or no microbiome: are we looking at the prenatal environment through the right lens? *Microbiome*. 2021;9(1):9.
- Kennedy KM, de Goffau MC, Perez-Munoz ME, Arrieta MC, Backhed F, Bork P, et al. Questioning the fetal microbiome illustrates pitfalls of low-biomass microbial studies. *Nature*. 2023;613(7945):639–49.
- Nearing JT, Comeau AM, Langille MGI. Identifying biases and their potential solutions in human microbiome studies. *Microbiome*. 2021;9(1):113.
- Eisenhofer R, Minich JJ, Marotz C, Cooper A, Knight R, Weyrich LS. Contamination in Low Microbial Biomass Microbiome Studies: Issues and Recommendations. *Trends Microbiol*. 2019;27(2):105–17.
- Karstens L, Asquith M, Davin S, Fair D, Gregory WT, Wolfe AJ, et al. Controlling for contaminants in low-biomass 16S rRNA gene sequencing experiments. *mSystems*. 2019;4(4):10.
- Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, et al. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol*. 2014;12:87.
- Bedarf JR, Beraza N, Khazneh H, Ozkurt E, Baker D, Borger V, et al. Much ado about nothing? Off-target amplification can lead to false-positive bacterial brain microbiome detection in healthy and Parkinson's disease individuals. *Microbiome*. 2021;9(1):75.
- Reitmeier S, Hitch TCA, Treichel N, Fikas N, Hausmann B, Ramer-Tait AE, et al. Handling of spurious sequences affects the outcome of high-throughput 16S rRNA gene amplicon profiling. *ISME Commun*. 2021;1(1):31.
- Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl Environ Microbiol*. 2013;79(17):5112–20.
- Palmer JM, Jusino MA, Banik MT, Lindner DL. Non-biological synthetic spike-in controls and the AMPtk software pipeline improve mycobiome data. *PeerJ*. 2018;6:e4925.
- Costea PI, Zeller G, Sunagawa S, Pelletier E, Alberti A, Levenez F, et al. Towards standards for human fecal sample processing in metagenomic studies. *Nat Biotechnol*. 2017;35(11):1069–76.
- Miller DN, Bryant JE, Madsen EL, Ghiorse WC. Evaluation and optimization of DNA extraction and purification procedures for soil and sediment samples. *Appl Environ Microbiol*. 1999;65(11):4715–24.
- Walker AW, Martin JC, Scott P, Parkhill J, Flint HJ, Scott KP. 16S rRNA gene-based profiling of the human infant gut microbiota is strongly influenced by sample processing and PCR primer choice. *Microbiome*. 2015;3:26.
- Galla G, Praeg N, Rzehak T, Sprecher E, Colla F, Seeber J, et al. Comparison of DNA extraction methods on different sample matrices within the same terrestrial ecosystem. *Sci Rep*. 2024;14(1):8715.
- Walker AW, Hoyle L. Human microbiome myths and misconceptions. *Nat Microbiol*. 2023;8(8):1392–6.
- Giangacomo C, Mohseni M, Kovar L, Wallace JG. Comparing DNA Extraction and 16S rRNA Gene Amplification Methods for Plant-Associated Bacterial Communities. *Phytopharmaceuticals*. 2021;5(2):190–201.
- Fitzpatrick CR, Lu-Irving P, Copeland J, Guttman DS, Wang PW, Baltrus DA, et al. Chloroplast sequence variation and the efficacy of peptide nucleic acids for blocking host amplification in plant microbiome studies. *Microbiome*. 2018;6(1):144.
- Wensel CR, Pluznick JL, Salzberg SL, Sears CL. Next-generation sequencing: insights to advance clinical investigations of the microbiome. *J Clin Invest*. 2022;132(7):e154944.
- MacConaill LE, Burns RT, Nag A, Coleman HA, Slevin MK, Giorda K, et al. Unique, dual-indexed sequencing adapters with UMIs effectively eliminate index cross-talk and significantly improve sensitivity of massively parallel sequencing. *BMC Genomics*. 2018;19(1):30.
- Johnson JS, Spakowicz DJ, Hong BY, Petersen LM, Demkowicz P, Chen L, et al. Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. *Nat Commun*. 2019;10(1):5029.
- Curry KD, Wang Q, Nute MG, Tyshaieva A, Reeves E, Soriano S, et al. Emu: species-level microbial community profiling of full-length 16S rRNA Oxford Nanopore sequencing data. *Nat Methods*. 2022;19(7):845–53.
- Jamy M, Foster R, Barbera P, Czech L, Kozlov A, Stamatakis A, et al. Long-read metabarcoding of the eukaryotic rDNA operon to phylogenetically and taxonomically resolve environmental diversity. *Mol Ecol Resour*. 2020;20(2):429–43.
- Srinivas M, Walsh CJ, Crispie F, O'Sullivan O, Cotter PD, van Sinderen D, et al. Evaluating the efficiency of 16S-ITS-23S operon

- sequencing for species level resolution in microbial communities. *Sci Rep.* 2025;15(1):2822.
44. Pinto Y, Bhatt AS. Sequencing-based analysis of microbiomes. *Nat Rev Genet.* 2024;25(12):829–45.
 45. Sun S, Jones RB, Fodor AA. Inference-based accuracy of metagenome prediction tools varies across sample types and functional categories. *Microbiome.* 2020;8(1):46.
 46. Machado MS, Rühlemann M, Reitmeier S, Kacprowski T, Frost F, Haller D, Baumbach J, List M. On the limits of 16S rRNA gene-based metagenome prediction and functional profiling. *Microb Genom.* 2024;10(2):001203.
 47. Li F, Li X, Cheng CC, Bujdos D, Tollenaar S, Simpson DJ, et al. A phylogenomic analysis of *Limosilactobacillus reuteri* reveals ancient and stable evolutionary relationships with rodents and birds and zoonotic transmission to humans. *BMC Biol.* 2023;21(1):53.
 48. Maistrenko OM, Mende DR, Luetge M, Hildebrand F, Schmidt TSB, Li SS, et al. Disentangling the impact of environmental and phylogenetic constraints on prokaryotic within-species diversity. *ISME J.* 2020;14(5):1247–59.
 49. Hitch TCA, Masson JM, Pauvert C, Bosch J, Nüchtern S, Treichel N, et al. Broad diversity of human gut bacteria accessible via a traceable strain deposition system. *bioRxiv.* 2024:2024.06.20.599854. <https://www.biorxiv.org/content/10.1101/2024.06.20.599854v2>.
 50. Bai Y, Muller DB, Srinivas G, Garrido-Oter R, Potthoff E, Rott M, et al. Functional overlap of the *Arabidopsis* leaf and root microbiota. *Nature.* 2015;528(7582):364–9.
 51. Northen TR, Kleiner M, Torres M, Kovacs AT, Nicolaisen MH, Krzyżanowska DM, et al. Community standards and future opportunities for synthetic communities in plant-microbiota research. *Nat Microbiol.* 2024;9(11):2774–84.
 52. Stewart RD, Auffret MD, Warr A, Walker AW, Roehle R, Watson M. Compendium of 4,941 rumen metagenome-assembled genomes for rumen microbiome biology and enzyme discovery. *Nat Biotechnol.* 2019;37(8):953–61.
 53. Nayfach S, Roux S, Seshadri R, Udwy D, Varghese N, Schulz F, et al. A genomic catalog of Earth's microbiomes. *Nat Biotechnol.* 2021;39(4):499–509.
 54. Royo-Llonch M, Sanchez P, Ruiz-Gonzalez C, Salazar G, Pedros-Alio C, Sebastian M, et al. Compendium of 530 metagenome-assembled bacterial and archaeal genomes from the polar Arctic Ocean. *Nat Microbiol.* 2021;6(12):1561–74.
 55. Borton MA, McGivern BB, Willi KR, Woodcroft BJ, Mosier AC, Singleton DM, et al. A functional microbiome catalogue crowdsourced from North American rivers. *Nature.* 2025;637(8044):103–12.
 56. Almeida A, Mitchell AL, Boland M, Forster SC, Gloor GB, Tarkowska A, et al. A new genomic blueprint of the human gut microbiota. *Nature.* 2019;568(7753):499–504.
 57. Blanco-Miguez A, Beghini F, Cumbo F, McIver LJ, Thompson KN, Zolfo M, et al. Extending and improving metagenomic taxonomic profiling with uncharacterized species using MetaPhlAn 4. *Nat Biotechnol.* 2023;41(11):1633–44.
 58. Olson RD, Assaf R, Brettin T, Conrad N, Cucinell C, Davis JJ, et al. Introducing the Bacterial and Viral Bioinformatics Resource Center (BV-BRC): a resource combining PATRIC, IRD and ViPR. *Nucleic Acids Res.* 2023;51(D1):D678–89.
 59. Richardson L, Allen B, Baldi G, Beracochea M, Bileschi ML, Burdett T, et al. MGnify: the microbiome sequence data analysis resource in 2023. *Nucleic Acids Res.* 2023;51(D1):D753–9.
 60. Chen IA, Chu K, Palaniappan K, Ratner A, Huang J, Huntemann M, et al. The IMG/M data management and analysis system vol 7: content updates and new features. *Nucleic Acids Res.* 2023;51(D1):D723–32.
 61. Frioux C, Singh D, Korcsmaros T, Hildebrand F. From bag-of-genes to bag-of-genomes: metabolic modelling of communities in the era of metagenome-assembled genomes. *Comput Struct Biotechnol J.* 2020;18:1722–34.
 62. Meyer F, Lesker TR, Koslicki D, Fritz A, Gurevich A, Darling AE, et al. Tutorial: assessing metagenomics software with the CAMI benchmarking toolkit. *Nat Protoc.* 2021;16(4):1785–801.
 63. Meyer F, Fritz A, Deng ZL, Koslicki D, Lesker TR, Gurevich A, et al. Critical Assessment of Metagenome Interpretation: the second round of challenges. *Nat Methods.* 2022;19(4):429–40.
 64. Nearing JT, Douglas GM, Hayes MG, MacDonald J, Desai DK, Allward N, et al. Microbiome differential abundance methods produce different results across 38 datasets. *Nat Commun.* 2022;13(1):342.
 65. Gloor GB, Macklaim JM, Pawlowsky-Glahn V, Egozcue JJ. Microbiome Datasets Are Compositional: And This Is Not Optional. *Front Microbiol.* 2017;8:2224.
 66. Boshuizen HC, Te Beest DE. Pitfalls in the statistical analysis of microbiome amplicon sequencing data. *Mol Ecol Resour.* 2023;23(3):539–48.
 67. Wirbel J, Essex M, Forslund SK, Zeller G. A realistic benchmark for differential abundance testing and confounder adjustment in human microbiome studies. *Genome Biol.* 2024;25(1):247.
 68. Legendre P, Legendre L. Numerical ecology. 3rd Edition ed. 2012.
 69. Rao C, Coyte KZ, Bainter W, Geha RS, Martin CR, Rakoff-Nahoum S. Multi-kingdom ecological drivers of microbiota assembly in preterm infants. *Nature.* 2021;591(7851):633–8.
 70. Tkacz A, Hortala M, Poole PS. Absolute quantitation of microbiota abundance in environmental samples. *Microbiome.* 2018;6(1):110.
 71. Li F, Liu J, Maldonado-Gomez MX, Frese SA, Ganzle MG, Walter J. Highly accurate and sensitive absolute quantification of bacterial strains in human fecal samples. *Microbiome.* 2024;12(1):168.
 72. Munch MM, Strenk SM, Srinivasan S, Fiedler TL, Proll S, Fredricks DN. Gardnerella Species and Their Association With Bacterial Vaginosis. *J Infect Dis.* 2024;230(1):e171–81.
 73. Magurran AE. Measuring biological diversity. Malden: Blackwell Publishing; 2004.
 74. Quinn GP, Keough MJ. Experimental design and data analysis for biologists. Cambridge: Cambridge University Press; 2002.
 75. Schloss PD. Waste not, want not: revisiting the analysis that called into question the practice of rarefaction. *mSphere.* 2024;9(1):e0035523.
 76. Deng Y, Umbach AK, Neufeld JD. Nonparametric richness estimators Chao1 and ACE must not be used with amplicon sequence variant data. *ISME J.* 2024;18(1):wrae106.
 77. Jost L. Partitioning diversity into independent alpha and beta components. *Ecology.* 2007;88(10):2427–39.
 78. Dahl EM, Neer E, Bowie KR, Leung ET, Karstens L. microshades: An R Package for Improving Color Accessibility and Organization of Microbiome Data. *Microbiol Resour Announc.* 2022;11(11):e0079522.
 79. Langille MGI, Ravel J, Fricke WF. "Available upon request": not good enough for microbiome data! *Microbiome.* 2018;6(1):8.
 80. Mirzayi C, Renson A, Genomic Standards C, Massive A, Quality Control S, Zohra F, et al. Reporting guidelines for human microbiome research: the STORMS checklist. *Nat Med.* 2021;27(11):1885–92.
 81. STREAMS Microbiome Guidelines. <https://streamsmicrobiome.org/>. Accessed 22 Mar 2025.
 82. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data.* 2016;3:160018.
 83. Repository options in the Springer Nature Research data policy. <https://www.springernature.com/gp/authors/research-data-policy/repositories-mandates/19540364>. Accessed 22 Mar 2025.
 84. Yilmaz P, Kottmann R, Field D, Knight R, Cole JR, Amaral-Zettler L, et al. Minimum information about a marker gene sequence (MIMARKS) and minimum information about any (x) sequence (MIXS) specifications. *Nat Biotechnol.* 2011;29(5):415–20.
 85. NFDI4Microbiota. <https://nfdi4microbiota.de/>. Accessed 22 Mar 2025.
 86. National Microbiome Data Collaborative (NMDC). <https://microbiome.data.org/>. Accessed 22 Mar 2025.
 87. Weiss S, Van Treuren W, Lozupone C, Faust K, Friedman J, Deng Y, et al. Correlation detection strategies in microbial data sets vary widely in sensitivity and precision. *ISME J.* 2016;10(7):1669–81.
 88. Carr A, Diener C, Baliga NS, Gibbons SM. Use and abuse of correlation analyses in microbial ecology. *ISME J.* 2019;13(11):2647–55.
 89. Wang M, Tu Q. Effective data filtering is prerequisite for robust microbial association network construction. *Front Microbiol.* 2022;13:1016947.
 90. Fischbach MA. Microbiome: focus on causation and mechanism. *Cell.* 2018;174(4):785–90.
 91. Walter J, Armet AM, Finlay BB, Shanahan F. Establishing or exaggerating causality for the gut microbiome: lessons from human microbiota-associated rodents. *Cell.* 2020;180(2):221–32.

92. Hellwig P, Dittrich A, Heyer R, Reichl U, Benndorf D. Detection, isolation and characterization of phage-host complexes using BONCAT and click chemistry. *Front Microbiol.* 2024;15:1434301.
93. Kleiner M, Kouris A, Violette M, D'Angelo G, Liu Y, Korenek A, et al. Ultra-sensitive isotope probing to quantify activity and substrate assimilation in microbiomes. *Microbiome.* 2023;11(1):24.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.