Acta Pharmaceutica Sinica B 2025;15(2):1189−1192

Chinese Pharmaceutical Association
Institute of Materia Medica, Chinese Academy of Medical Sciences

Acta Pharmaceutica Sinica B

www.elsevier.com/locate/apsb
www.sciencedirect.com

HIGHLIGHT

Unlocking the healthy human microbiome: Redefining core microbial signatures

KEY WORDS

Gut health;
Guilds;
Biomarker;
Causal inference

Despite the enormous progress made in microbiome research, a comprehensive and universally accepted framework for identifying healthy microbiome biomarkers remains elusive. Defining what constitutes a "healthy" gut microbiome is not an easy task due to its complexity and dynamic nature. Historically, the quest to define a healthy microbiome has centered on identifying biomarkers like microbial diversity and the Firmicutes-to-Bacteroidetes (F/B) ratio, based on the assumption that these factors correlate with optimal health (Fig. 1A). While this notion is intuitively appealing, it has proven more complex in practice. These traditional approaches have been found unreliable as indicators of health, as they are often more closely associated with other traits like diet and lifestyle[2]. However, in the search for a healthy core microbiome, we have come to recognize its profound importance in human health, influencing everything from digestion and metabolism to immune function and even mental health.

Regardless of these complexities, there is compelling evidence that the microbiome plays a fundamental role in regulating human health. This suggests that underlying rules and mechanisms govern its influence on our physiological processes. Understanding these mechanisms could provide crucial insights into how the microbiome contributes to health and disease, reinforcing the idea that it is integral to our overall well-being.

1. The limits of traditional approaches: abundance, taxonomy, and cohort-specificity

The earliest attempts to define a healthy microbiome were based on taxonomic metrics, including microbial diversity indices and the abundance of specific taxa (Fig. 1A). A hallmark of these methods was the assumption that higher microbial diversity or specific taxonomic groups indicated a "healthy" microbiome. However, these approaches have consistently fallen short for several reasons.

Firstly, microbial diversity is highly influenced by dietary patterns. For example, individuals adhering to plant-based diets often exhibit higher levels of microbial diversity compared to those with omnivorous diets. Moreover, a high degree of microbial diversity does not necessarily guarantee that taxa producing harmful metabolites or contributing to systemic inflammation are outcompeted by the microbial community. Furthermore, traditional taxonomic profiling methods, such as 16S ribosomal RNA sequencing (16S rRNA sequencing), are often limited by their resolution to genus or species level. It lacks the resolution needed to distinguish microbial strains or identify SNP level variations, which are crucial for understanding the functional diversity of microbial communities (Fig. 1A). Large cohort studies such as the Human Microbiome Project (HMP) subsequently determined that there is no singular, universal microbiome composition that is considered "normal" across all individuals in terms of taxonomy[3]. As implied in this recent advance, health is likely associated with functional outcomes driven by microbial interactions rather than the mere presence or absence of specific species.

Furthermore, cohort-specific findings across studies have added another layer of complexity. The variability in microbiome composition, influenced by factors like lifestyle, genetic background, and geographic location, makes it difficult to generalize findings from one cohort to another (Fig. 1B). For instance, a microbiome profile associated with good health in a Western population may not resemble the profile in a rural, non-

Peer review under the responsibility of Chinese Pharmaceutical Association and Institute of Materia Medica, Chinese Academy of Medical Sciences.

https://doi.org/10.1016/j.apsb.2025.01.001

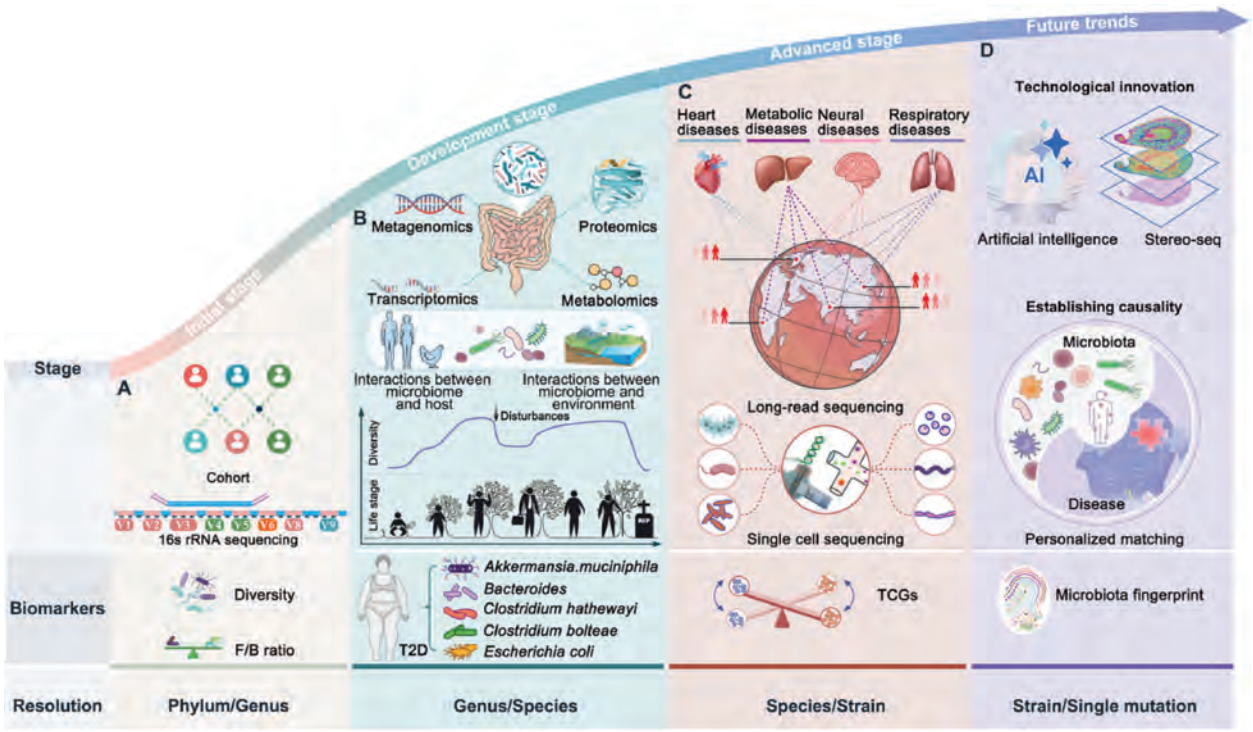

Figure 1 Evolution of microbiome research. (A) Initial stage: Focusing on a single cohort, this stage employs genomic methods such as 16S rRNA gene sequencing to pinpoint microbial markers linked to diseases (utilizing a single cohort and one technology). (B) Development stage: At this stage, data from various omics layers—metagenomics, transcriptomics, proteomics, and metabolomics—are integrated to gain a holistic view of microbial community structure and function. Interactions among different microbial components, including microbe-host and microbe-environment, are analyzed. Studies track the microbiome's dynamic changes over time to understand its health and disease implications (involves a single cohort, one disease, and multiple technologies). (C) Advanced stage: Cross-cohort validation strengthens the diagnostic precision and reliability of biomarkers. Cross-disease and population studies facilitate the identification of a universal core microbiome, encompassing conditions affecting the heart, metabolic systems, brain, and respiratory tract. High-precision technologies like long-read sequencing and single-cell sequencing are employed for in-depth species or strain-level research (engages multiple cohorts, various diseases, and a range of technologies). (D) Future trends: Technological advancements, including new sequencing technologies such as spatial transcriptomics and bioinformatics tools, will enhance microbiome data analysis, potentially incorporating AI. Future work aims to establish causal links between the microbiome and health. Personalized microbiome-host matching for clinical applications will be pursued (aims for personalized medicine).

industrialized cohort. This challenge has made it difficult to establish universal biomarkers that are valid across different environments and populations, further complicating efforts to define a "core" microbiome that universally signals health.

2. Technological and methodological advances: moving beyond taxonomy

In response to these challenges, the field has begun to shift from taxonomic classification to functional and strain-resolved analyses, providing a more precise understanding of microbial communities.

The concept of "core microbiomes" has evolved alongside these advancements. Rather than focusing solely on taxonomic annotations, researchers now emphasize the functional capacities of microbial communities—specifically, the core microbial functions essential for host health. Recent work by Wu et al.[1] introduced the "Two Competing Guilds" (TCGs) model: one responsible for beneficial functions such as fiber fermentation and butyrate production, and the other enriched in virulence factors and antibiotic resistance genes. This dualistic framework offers a promising perspective to universally conceptualize a healthy microbiome, where the balance between these two guilds could serve as a functional biomarker for health (Fig. 1C). Rather than reducing this balance to a single coefficient or index, Wu et al.[1] utilized random forest models trained on multi-cohort datasets to identify biomarkers representing a healthy "core microbiome". The random forest model utilizes the core microbiome signature as input features to classify subjects into case and control groups and to foresee outcomes of therapeutic interventions. Moreover, the use of high-quality metagenome-assembled genomes (HQMAGs)[4], which provide near-strain-level resolution, allows for more accurate and reproducible analyses, circumventing some of the limitations of earlier sequencing technologies.

3. The path forward: A comprehensive framework for defining a healthy microbiome

Given the complexity and variability of the microbiome, a single biomarker or simplistic abundance measure might be insufficient to define a healthy microbiome. Instead, a more sophisticated approach is required—one that integrates multiple layers of data and accounts for both microbial composition and function. Such a

framework must also prioritize uncovering the causal relationships between the microbiome and health, providing deeper insights into how microbial communities contribute to, or detract from, well-being.

3.1. Strain-resolved analyses

We believe it is crucial to analyze the microbiome at the strain-resolved or even higher resolution. The fecal microbiota transplantation (FMT) study revealed that strain-level variability, rather than species compositions, determined the colonization success in the patients. Therefore, understanding the strain-level variability within those species is critical for capturing the potential of microbial functionality. This new study[1] has further demonstrated the importance of higher resolution in understanding microbial interactions and their health implications. Also, the influence of human genetic variation on the gut microbiome is reported at the strain level, underscoring the necessity for strain-resolved analysis[5].

3.2. Multi-omics integration

A significant advancement in microbiome research, prominently showcased by the second phase of the Human Microbiome Project (HMP2), is the simultaneous profiling of both host and microbiome multi-omics data[6]. This dual-layer integration, which combines metagenomics, transcriptomics, metabolomics, and proteomics across both host and microbial profiles, offers unprecedented insights into host-microbiome interactions (Fig. 1B). By analyzing these layers together, HMP2 allows researchers to connect microbial activity directly with host biological responses, revealing how shifts in the microbiome influence health and disease at a molecular level.

This concurrent profiling is especially valuable in uncovering disease mechanisms, as it enables the identification of specific microbial metabolites and pathways that correspond to host immune responses and metabolic changes. For instance, multi-omic analyses of *Streptococcus* demonstrated its association with the host's WNT signaling and the NF-κB pathway in colorectal cancer (CRC) and macrophage inflammatory responses in irritable bowel syndrome (IBS)[7]. This indicates that the same microbes can impact various pathophysiological processes by associating with different host genes in each disease. This integrated approach provides a comprehensive framework for understanding complex host-microbiome dynamics, moving beyond isolated microbial or host data.

3.3. Cross-cohort validation and generalizability

Cross-cohort validation is essential for ensuring the robustness and generalizability of microbiome biomarkers. As we mentioned above, there are universally applied characteristics in the human microbiome, such as enterotype and TCGs[1,8]. The ability to reproduce findings across diverse populations, environments, and diseases is a critical step toward developing universally applicable biomarkers. Cohort diversity in microbiome studies is necessary to account for the complex interactions between host genetics, diet, lifestyle, and environmental factors (Fig. 1C). The public availability of large datasets, such as HMP and those from disease cohorts, will facilitate the validation, allowing for identifying generally applicable biomarkers.

3.4. Artificial intelligence (AI)-based causal inference methods

Determining causal relationships between specific microbes and diseases is straightforward in well-designed experiments based on Koch's postulates, such as *Helicobacter pylori* and gastric ulcers, *Human papillomavirus* (HPV) and cervical cancer, *Enterobacter cloacae* B29 and obesity. Multi-omics analysis techniques, combined with animal or clinical experiments, helped to confirm specific target microorganisms from the imbalanced microbiota and conduct causal relationship verification adhering to similar principles. For example, *Bacteroides thetaiotaomicron* and *Megamonas rupellensis* protected against or resulted in obesity in mice models, respectively. Fecal microbiota transplantation, which restores microbial imbalance, has been widely applied for treating *Clostridioides difficile*—induced diarrhea. Meanwhile, diseases, such as respiratory infections, could also lead to disruptions in the gut microbiota. However, it remains a great challenge to infer causal links between variations in the complex microbiomes and the general healthy status of the hosts. Traditional approaches, such as Mendelian randomization and mediation analysis, have been applied successfully in targeted studies but often fail when confronted with the complexity and scale of larger microbiome datasets.

AI-based causal inference methods have the potential to transform our understanding of the microbiome. For instance, the approach combining machine learning with causal inference has uncovered that alterations in gut-bacteria-associated bile acid metabolites influence neonatal jaundice by impacting the levels of total bilirubin[9]. Utilizing a high-quality, uniformly preprocessed dataset, advanced machine learning algorithms can elucidate complex, non-linear associations between microbial traits and health outcomes. These tools are particularly useful for discerning causality—something that traditional methods have struggled to achieve. Due to its powerful ability to analyze big data, recognize complex patterns, continuously learn and self-optimize, integrate multimodal data, and improve medical efficiency, AI can help predict disease outcomes, identify potential therapeutic targets, and optimize precision medicine approaches by integrating individual microbiome profiles with clinical data (Fig. 1D).

4. Challenges and future directions

Current approaches also come with challenges. First, the microbiome's dynamic nature means that even cross-sectional studies may not fully reflect its role in health or disease. Although intensively sampled longitudinal studies have shown the core stability and individual dynamics of the microbiome from all body sites[10], additional longitudinal studies are needed to generalize such trends in the context of inter-individual variability. Additionally, environmental factors and host genetic variations continue to create noise in identifying truly universal biomarkers. Finally, as the field moves toward clinical applications involving microbiome manipulation, it is essential to rigorously consider ethical implications. For example, there have been cases where six individuals developed infections caused by enteropathogenic *Escherichia coli* (EPEC) and Shiga toxin-producing *E. coli* (STEC) after receiving fecal microbiota transplantation (FMT) for *C. difficile* infection that did not respond to standard therapy. Therefore, it is imperative to implement more rigorous screening

protocols and adhere to stringent ethical guidelines when selecting donors and administering fecal microbiota transplantation (FMT) to ensure the safety and well-being of recipients.

5. Conclusion: A path toward precision medicine

The concept of a healthy microbiome biomarker is no longer a distant aspiration, it is an achievable goal. By integrating strain-resolved analytics, multi-omics data, cross-cohort validation, and AI-based causal inference, we can begin to define health in terms of microbial communities rather than individual species. The TCGs model and other similar frameworks represent a significant step forward in this effort, but much work remains. Future research should focus on longitudinal studies, multi-omics integration, and cross-population validation to further refine our understanding of the core microbiome. With continued technological advances and interdisciplinary approaches, we are poised to harness the power of the microbiome for personalized health and disease management.

Acknowledgments

The authors express their gratitude to all study participants for their invaluable contributions and cooperation. This work was supported by the National Natural Science Foundation of China (Nos. 82300345, 3217003, 32370099, and 81901287), the National High-Level Hospital Clinical Research Funding (2023-GSP-QN-23, China), the Provincial-level Talent Program for National Center of Technology Innovation for Biopharmaceuticals (NCTIB2024JS0101, China) and the Suzhou Top-Notch Talent Groups (ZXD2022003, China).

Author contributions

Shuting Xia and Diya Jiang drafted the manuscript and figure. Qianyi Zhou, Hairong Lyu, and Anita Y Voigt revised the manuscript. Yuan Huang, Zhemin Zhou and Xin Zhou designed the review framework and revised the manuscript. All of the authors have read and approved the final manuscript.

Conflicts of interest

The authors have no conflicts of interest to declare.

References

1. Wu GJ, Xu T, Zhao NS, Lam YY, Ding XY, Wei DQ, et al. A core microbiome signature as an indicator of health. *Cell* 2024;**187**:6550-65.e11.
2. Van Hul M, Cani PD, Petitfils C, De Vos WM, Tilg H, El-Omar EM. What defines a healthy gut microbiome?. *Gut* 2024;**73**:1893−908.
3. Consortium THMP. Structure, function and diversity of the healthy human microbiome. *Nature* 2012;**486**:207−14.
4. Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy TBK, et al. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat Biotechnol* 2017;**35**:725−31.
5. Zhernakova DV, Wang DM, Liu L, Andreu-Sánchez S, Zhang Y, Ruiz-Moreno AJ, et al. Host genetic regulation of human gut microbial structural variation. *Nature* 2024;**625**:813−21.
6. Consortium TIHiRN. The integrative human microbiome project. *Nature* 2019;**569**:641−8.
7. Priya S, Burns MB, Ward T, Mars Rat, Adamowicz B, Lock EF, et al. Identification of shared and disease-specific host gene-microbiome associations across human diseases using multi-omic integration. *Nat Microbiol* 2022;**7**:780−95.
8. Costea PI, Hildebrand F, Arumugam M, Bäckhed F, Blaser MJ, Bushman FD, et al. Enterotypes in the landscape of gut microbial community composition. *Nat Microbiol* 2018;**3**:8−16.
9. Chen WL, Zhang P, Zhang XL, Xiao TT, Zeng JH, Guo KP, et al. Machine learning-causal inference based on multi-omics data reveals the association of altered gut bacteria and bile acid metabolism with neonatal jaundice. *Gut Microbes* 2024;**16**:2388805.
10. Zhou X, Shen XT, Johnson JS, Spakowicz DJ, Agnello M, Zhou WY, et al. Longitudinal profiling of the microbiome at four body sites reveals core stability and individualized dynamics during health and disease. *Cell Host Microbe* 2024;**32**:506-26.e9.

Shuting Xia[b], Diya Jiang[b], Qianyi Zhou[c], Hairong Lyu[d], Anita Y. Voigt[e], Xin Zhou[f,*], Zhemin Zhou[b,g,h,*], Yuan Huang[a,*]

[a]*State Key Laboratory of Cardiovascular Disease, Fuwai Hospital, National Center for Cardiovascular Diseases, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100037, China*

[b]*Key Laboratory of Alkene-carbon Fibres-based Technology & Application for Detection of Major Infectious Diseases, MOE Key Laboratory of Geriatric Diseases and Immunology, Cancer Institute, Suzhou Medical College, Soochow University, Suzhou 215127, China*

[c]*The First Central Clinical College, Tianjin Medical University, Tianjin 300192, China*

[d]*Department of Hematology, Tianjin First Central Hospital, Tianjin 300192, China*

[e]*Department of Dermatology, Duke University School of Medicine, Durham, NC 27710, USA*

[f]*Department of Genetics, Stanford University School of Medicine, Stanford, CA 94305, USA*

[g]*Suzhou Key Laboratory of Pathogen Bioscience and Anti-infective Medicine, Jiangsu Province Engineering Research Center of Precision Diagnostics and Therapeutics Development, Suzhou Medical College, Soochow University, Suzhou 215127, China*

[h]*National Center of Technology Innovation for Biopharmaceuticals, Suzhou Biomedical Industry Innovation Center, Suzhou 215127, China*

*Corresponding authors.
E-mail addresses: xzhou7@stanford.edu (Xin Zhou), zmzhou@suda.edu.cn (Zhemin Zhou), clevelandhuangyuan@163.com (Yuan Huang)

Received 19 November 2024
Received in revised form 22 December 2024
Accepted 28 December 2024