SAÚDE E PESQUISA
e-ISSN 2176-9206

ORIGINAL ARTICLE

https://doi.org/10.17765/2176-9206.2025v18e13398

BIOINFORMATICS TOOLS IN HUMAN GUT MICROBIOME RESEARCH: A SCIENTOMETRIC REVIEW

FERRAMENTAS DE BIOINFORMÁTICA UTILIZADAS NAS PESQUISAS EM MICROBIOMA INTESTINAL HUMANO: UMA REVISÃO CIENTOMÉTRICA

Arthur Lima da Silva[1*], Anna Beatriz Machado Lima[2], João Gonçalves da Costa Neto[2], Lara Luisa Lopes Chrisóstomo[2], Márcio Fabrício Falcão de Paula Filho[2], Isaac Farias Cansanção[3]

[1] PhD Student in Ecology at the Postgraduate Program in Ecology, Federal University of Bahia, Salvador (BA), Brazil; Professor in the Natural Sciences Department. Federal University of Sao Francisco Valley. São Raimundo Nonato (PI), Brazil,[2] Medical Student. Federal University of Sao Francisco Valley. Paulo Afonso (BA), Brazil,[3] PhD in Biotechnology by Postgraduate Program in Biotechnology RENORBIO, Federal University of Ceará, Fortaleza (CE), Brazil. Professor at Medical Department. Federal University of Sao Francisco Valley, Paulo Afonso (BA), Brazil.

***Corresponding author**: Arthur Lima da Silva – **Email:** arthur.limas@univasf.edu.br

Received: 30 Dec. 2024
Accepted: 16 Feb. 2025

ABSTRACT: The human intestinal microbiota plays important roles in the metabolism and immunology of the host, associated with various metabolic, psychiatric and immunological diseases. The impossibility of analyzing microbiomes in vitro has led to the application of bioinformatics in the knowledge of this booming area. We analyzed the temporal/spatial dynamics of studies on MIH with bioinformatics through a scientometric review, using systematic review and bibliometric techniques. Articles and meta-analyses published between January 2018 and August 2023 were selected, using the descriptors "*gastrointestinal microbiome*", "*gut microbiome*", "*gut microbiota*", "*human gut*", "*metagenomics*", "*meta-analysis*" and "*bioinformatics*". The results revealed the growth of publications, their impact and relevance and the articulation of partnerships between countries, institutions and authors. The panorama of studies on MIH with the aid of bioinformatics techniques indicates the potential for growth in the area and the importance of future studies for better understanding and clinical and preventive health applications.

KEYWORDS: Bioinfo. Gastrointestinal. Scientometrics. Science mapping. Systematic review.

RESUMO: A microbiota intestinal humana desempenha funções importantes no metabolismo e imunologia do hospedeiro associada a diversas doenças metabólicas, psiquiátricas e imunológicas. A impossibilidade de análise *in vitro* dos microbiomas, levaram à aplicação da bioinformática no conhecimento dessa área em plena expansão. Analisamos a dinâmica temporal/espacial dos estudos sobre MIH com a bioinformática através de uma revisão cientométrica, utilizando técnicas de revisão sistemática e bibliometria. Foram selecionados artigos e meta-análises publicados entre janeiro de 2018 e agosto de 2023, usando os descritores *"gastrointestinal microbiome", "gut microbiome", "gut microbiota", "human gut", "metagenomics", "meta-analysis" and "bioinformatics"*. Os resultados revelaram o crescimento das publicações, seu impacto e relevância e a articulação de parcerias entre países, instituições e autores. O panorama dos estudos sobre MIH com auxílio das técnicas de bioinformática indicam potencial de crescimento da área e a importância de estudos futuros para o melhor entendimento e aplicações clínicas e de prevenção à saúde.

PALAVRAS-CHAVE: Bioinfo. Cientometria. Gastrointestinal. *Science mapping*. Revisão sistemática.

INTRODUCTION

The human gut microbiota (HGM) is a complex and highly diverse ecosystem that remains little understood. It is estimated that the gut microbiota consists of approximately 100 trillion microorganisms, including bacteria, fungi, viruses, and protozoa. Of these, bacteria represent the predominant group, comprising 1,000 to 1,150 different species[1,2,3]. The predominant bacterial phyla are *Bacteroidetes*, *Firmicutes*, *Proteobacteria*, and *Actinobacteria*, collectively representing 97% of the microbiota, with approximately 70% residing in the large intestine. These microorganisms engage in a mutually beneficial relationship with gut epithelial cells, influencing various aspects of human health[4,5,6].

The composition of the microbiota varies across different regions of the body, life stages, and in response to external and internal factors, including stress, diet, genetics, and diseases[4,6,7]. An imbalance in the microbiota, or dysbiosis, can affect the regulation of the host's gastrointestinal functions and brings implications to many health and disease processes[10,11,12,13,14,15,16]. By enhancing metabolic processes and nutrient absorption, a balanced microbiome plays a pivotal role in prevention of chronic diseases, including obesity and type 2 diabetes. Various studies have demonstrated correlations between HGM and physiological processes, including immune system activation, gut permeability, digestive functions, and neurological pathways. This has led to the hypothesis that alterations in the HGM contributes to the development of various health conditions, including irritable bowel disease, obesity, diabetes, and mental health disorders[2,3,4,6,19,20,21,22,23]. Research indicates that the microbiome exerts a significant influence on mental health through the gut-brain axis, affecting mood and behavior. Consequently, the importance of maintaining a healthy microbiome by adhering to a diet abundant in fiber, prebiotics, and probiotics, this is crucial for enhancing well-being and preventing diseases[24,25,26].

The use of new tools, such as bioinformatics, helps to fill gaps in the relationship between HGM and other diseases, as well as contributing to the development of treatments and preventive strategies[10]. The recent advancement of sequencing technologies has enabled an exhaustive examination of microbial communities within the human gut[8,9.] This has in turn led to a more comprehensive understanding of their role in human health and disease. The use of bioinformatic tools is of great importance in the identification of microbial species, the assessment of their functional potential, and the study of microbe-host interactions within the gut microbiome[27]. Thus, to gain a deeper insight into the use of bioinformatic tools in studies on HGM, a bibliometric analysis was performed. This allowed us to identify key statistical and structural insights, while observing the production and evolution of scientific knowledge, including its historical progress and/or decline, major trends, techniques, and research opportunities[39]. By examining the scientific literature and applying bibliometric techniques, the aim was to identify trends, research topics, key players, and knowledge gaps revealed by bioinformatics in the study of HGM[39].

The objective of this study was to provide a comprehensive overview and analysis of the evolution of bioinformatic tools utilization in HGM research through bibliometric analysis[28,29]. This analysis will contribute to defining the current state of research in this field and will also inform future directions of research for younger researchers.

METHODOLOGY

This is a scientometric review that uses systematic review and bibliometric techniques. It was conducted in 3 phases: 1). bibliographic survey; 2). systematic review and 3). scientific mapping (Figure

1). In the first phase, keywords were extracted from MeSH using descriptors related to the topic (Table 1). The set of keywords showing potential was subjected to specificity and sensitivity analysis through exploratory searches in databases.

A comprehensive search of the literature was conducted in databases including *PubMed*, *Scopus*, and *Web of Science* (Table 2). The inclusion criteria adopted were: Articles that were published in accepted journals and use of the following descriptors: *gastrointestinal microbiome*, *gut microbiome*, *gut microbiota*, *human gut*, *metagenomics*, *meta-analysis*, and *bioinformatics*. Reviews were excluded; as well as descriptors such as: *mice*, *animals*, *mouse*, *rats*, *murine*, *rodent*, *fishes*, *piglets*, *dogs*, *canine*, *porcine*, *broilers*, *equine*, *rabbit*, and *birds*; and others that did not address the topic. Duplicate papers were excluded, and the remaining articles were selected in accordance with the previously defined eligibility, exclusion, and inclusion criteria (Table 3).

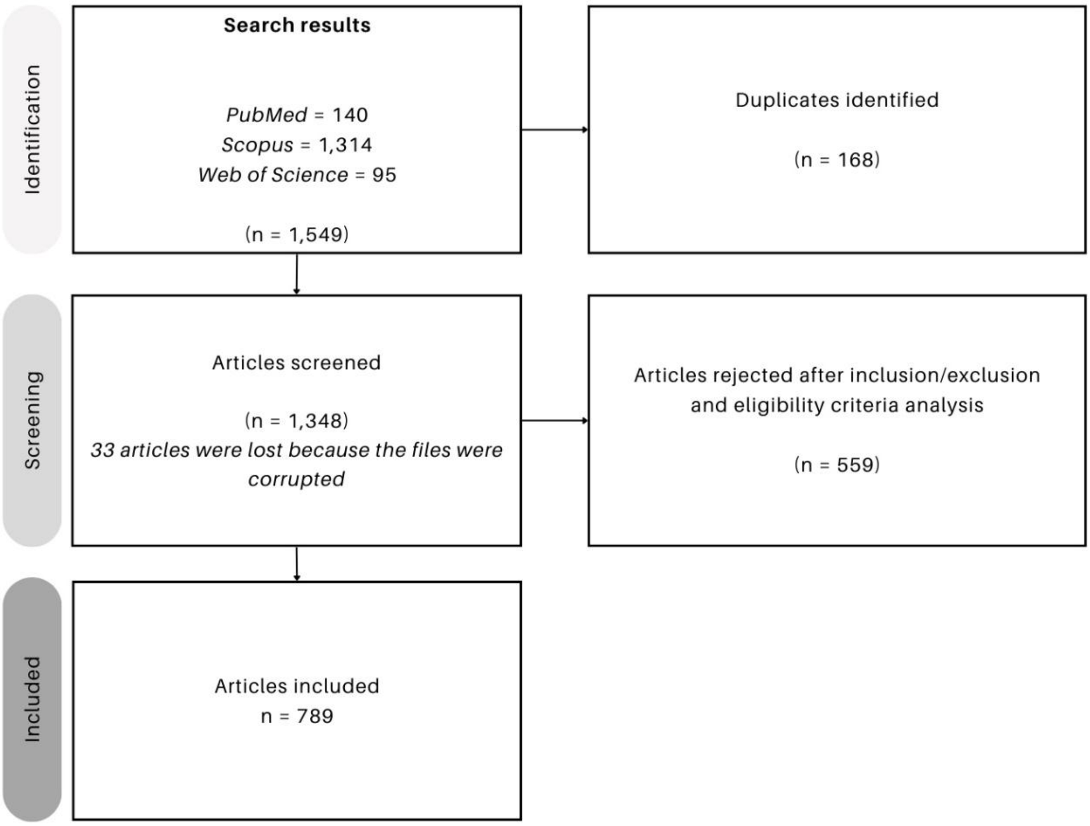

Figure 1. Flow chart of the process of carrying out and analysis of the results of the systematic search. During the exporting of raw archives from *PubMed* to *Rayyan*, 32 papers were lost. Due to this, the total analyzed archives were 1345, not 1377 as previously obtained (1549 - 172 = 1377 papers).
Source: the authors, 2024.

Table 1. Stages and steps of the methodology.

Stages	Steps
1. Systematic search strategy	1.1. Keywords were extracted from MesH 1.2. Sensitivity analysis test was conducted 1.3. The Systematic Search Strategy was defined
2. Systematic Search	2.1. Articles retrieved (*Scopus*, *WoS* and *PubMed*) 2.2. Duplicates were identified
3. Science Mapping	3.1. Included or excluded articles were selected 3.2. Data analysis with bibliometrix package – R software was conducted

Source: the authors, 2024.

For the temporal scope, analysis was confined to publications from the past five years (2018-2023). This time frame was chosen because research on this topic increased greatly during this period. In addition, during the COVID-19 pandemic, there was a boom in publications on the subject, with some topics related to change. The results were exported in *ris format from the *Scopus* and *Web of Science* databases and in *txt (plaintext) from *PubMed*. The results were imported in text format (*txt, or plain text) (*PubMed* and *Web of Science*) and in comma-separated values format (*csv) (*Scopus*)[27]. Relevance and adequacy of the results obtained were evaluated, and a systematic research strategy was then employed. The research strategy, data collection, and data analysis were conducted using reference management software.

Table 2. Research strategy and quantity of obtained papers in each database.

Database	Strings	Results
Scopus	(TITLE-ABS-KEY (“gastrointestinal microbiome” OR “gut microbiome”) AND TITLE-ABS-KEY (“computational biology” OR bioinformatics) AND TITLE-ABS-KEY (human)) AND PUBYEAR > 2017 AND PUBYEAR < 2024 AND (LIMIT-TO (DOCTYPE “ar”)) AND (LIMIT-TO (SRCTYPE, “j”))	1314
PubMed	((“gastrointestinal microbiome”[Title/Abstract] OR “gut microbiome” [Title/Abstract]) AND (“computational biology” [Title/Abstract] OR “bioinformatics”[Title/Abstract]) AND “human[Title/Abstract]”) AND (2018/1/1:2023/8/31[pdat]))	140
Web of Science	**“gastrointestinal microbiome” OR “gut microbiome”** (Topic) and **“computational biology” OR bioinformatics** (Topic) and **human** (Topic) and **2018** or **2019** or **2020** or **2021** or **2022** or **2023** (Publication Years) and **Article** (Document Types)	95

Source: the authors, 2024.

Table 3. Terms of reference for inclusion/exclusion criteria and eligibility used in paper selection.

Step	Analyzed metadata	Selected terms
Inclusion/ exclusion	Title, abstract and Keywords	Inclusion: *bioinformatics, gastrointestinal microbiome*, *gut microbiome*, *gut microbiota*, *human gut*, *meta-analysis* e *metagenomics* Exclusion: *review*, *mice*, *animals*, *birds*, *broilers*, *canine*, *dogs*, *equine*, *fishs*, *mouse*, *murine*, *piglets*, *porcine*, *rabbit*, *rats* e *rodent*
Eligibility	Title, abstract, keywords, introduction, materials and methods	i) Research that used bioinformatics tools in HGM research ii) Research conducted in human groups iii) Research realized with data from meta-analyzes

Source: the authors, 2024.

During the 2nd phase, *Rayyan* and R package *bibliometrix* software were employed. The papers were selected based on identification of specific terms after a thorough examination of the title, abstract, and keywords of each paper. Papers that satisfied both criteria were labeled as *included*, while those that satisfied only one criterion were classified as *maybe*. Papers that did not meet any of the established criteria were excluded. Concurrently, papers were selected in accordance with the established eligibility criteria.

The *bibliometrix* functions were only able to identify 168 duplicated papers It was then necessary to undertake a manual check and exclude duplicate papers. Subsequently, as a way of comparing and checking, papers that did not meet the established inclusion and exclusion criteria were excluded, resulting in the same quantity of included papers being systematically organized using the *Rayyan* software. All information and data presented were generated using the *bibliometrix* package and the *biblioshiny* application[28] from the RStudio 2023.06.1 programming kit, which is part of the R software (version 4.2.2)[30]. The raw data as well as complementary graphics and sheets is available at https://github.com/arthur-82/biblioanalysis_bioinfo_hgm_2024.git. Additional information can be requested via email to the corresponding author.

RESULTS

The absolute number of publications followed an upward trend until 2021, followed by a decline in the following years (Figure 2A). A total of 123 articles were published through August 2023. The annual mean for citations decreased, ranging from 8.42 in 2018 to 1.28 in 2022. The downward trend was confirmed for the year 2023, with an average of 0.29 citations through August 2023 (Figure 2B). Roughly 322 journals have published articles on this topic, and 5 of them with more than 20 articles each. The periodical *Frontiers in Cellular and Infection Microbiology* published the greatest number of articles on the theme (71), followed by *Gut Microbes* (62), *Frontiers in Microbiology* (41), *Microbiome* (36), *Frontiers in Immunology* (26), *Nutrients* (19), *Scientific Reports* (16), *International Journal of Molecular Sciences* (15), *PLOS One* (15) and *Nature Communications* (14) (Figure 2C). The ten most relevant journals, as determined by the h-index for impact measurement (Figure 2D), were *Microbiome* (h=17), *Frontiers in Cellular and Infection Microbiology* (h=16) and *Gut Microbes* (h=15). The list is completed by *Nature Communications* (h=13), *Frontiers in Immunology* (h=12), *Frontiers in Microbiology* (h=12), *Nutrients* (h=10), *International Journal of Molecular Sciences* (h=9), *Plos ONE* (h=8), and *Scientific Reports* (h=8).

AUTHOR PRODUCTIVITY

A total of 4,534 researchers contributed to knowledge in this field of study during the period under review (Table S7). The ten researchers who published the most were *Wang*, *Y*. (40), *Li, Y*. (37), *Liu, Y*. (31), *Zhang, Y*. (26), *Li, X*. (24), *Zhang, X*. (23), *Li, J*. (22), *Zhang, J*. (22), *Chen, Y*. (22), and *Wang, H*. (21) (Figure 2E).

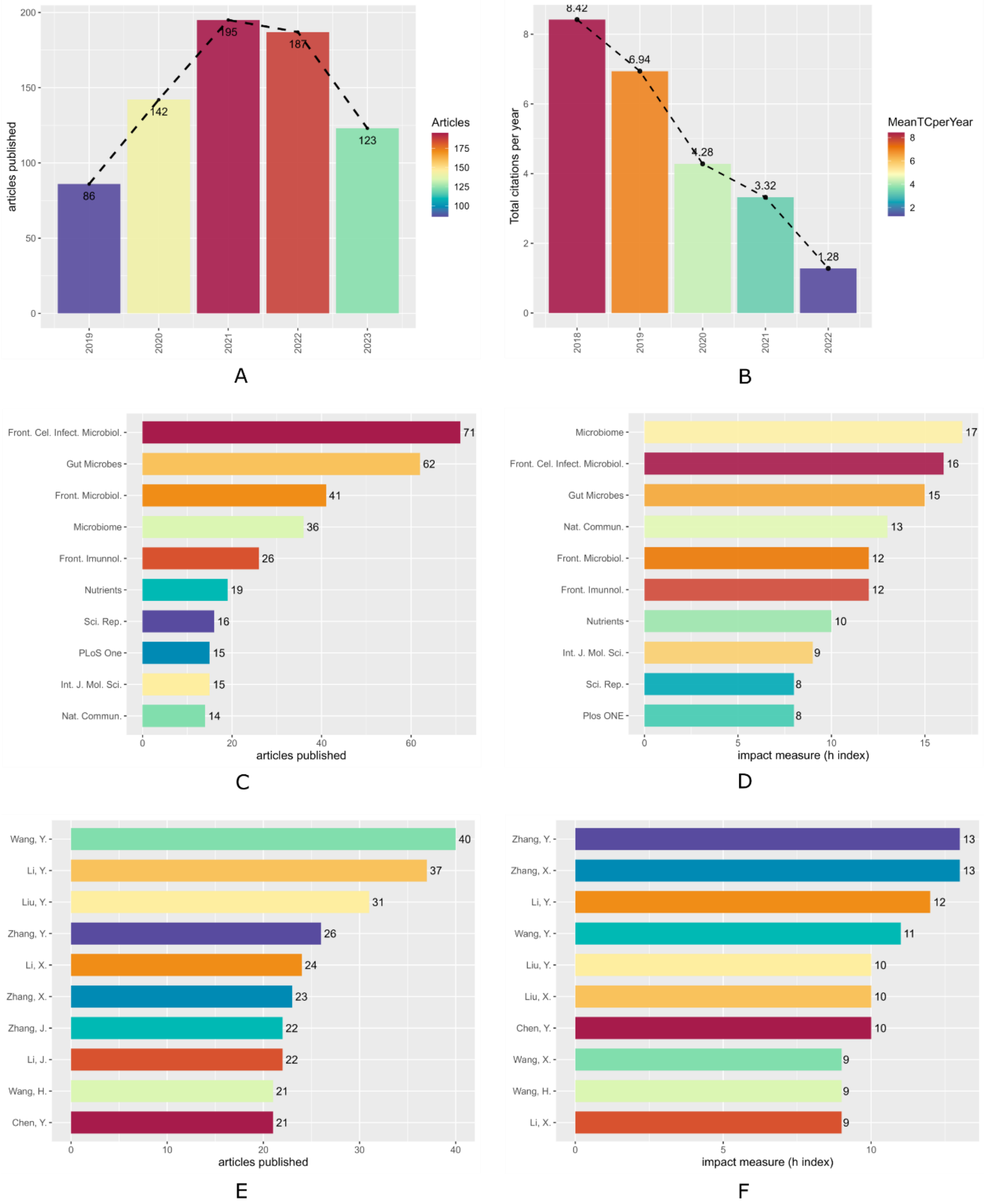

Figure 2. (A) Annual scientific publishing in absolute numbers. (B) Annual average of citations. (C) The 10 periodicals with more publications. (D) The 10 most relevant periodicals, ranked by impact measuring (h index). (E) Number of published papers for each author. (G) Measurement of impact of each author (*h index*).
Source: the authors, 2024.

The ten researchers with the greatest impact, as measured by the h-index, are shown in Figure 2F. Note that the authors *Zhang, X*. and *Zhang, Y*. presented comparable values for the h-, g- and m indices. The lists of the ten most prolific authors and the ten most impactful researchers exhibit a high degree of overlap. It is notable that all ten of the most influential and impactful researchers are of Chinese nationality, which highlights China's significant contribution to research on this topic. The collaborative author network (Figure S23) depicts four clusters. The size of each circle presents the number of published papers for each author, while the width of each line is proportional to the frequency of collaboration between researchers. The colors represent clusters of authors who collaborated the most with one another.

INSTITUTIONAL PRODUCTIVITY

Authorship analysis shows that the authors of the analyzed articles were associated with 1,365 research institutions. Only ten institutions were responsible for more than half of the total number of publications in the period (Tables S11, S12). The ten most productive universities were as follows: *Capital Medical University* (94), China; *Southern Medical University* (91), China; *University of Bologna* (74), Italy; *University of Ottawa* (71), Canada; *Fudan University* (70), China; *Huazhong University of Science and Technology* (65), China; *Shanghai Jiao Tong University* (57), China; *Sichuan University* (56), China; *University College Cork* (54), Ireland; and *Central South University* (53), China (Figure 3A). The collaborative network between the analyzed research institutions highlights significant insights such as partnerships and interrelations between institutions, represented by color clusters, and the extent of cooperation between them, represented by line width (Figure S24). It was observed that some universities, such as *Jiangnan University* and *Zhejiang University School of Medicine*, demonstrated a proclivity towards the development of research and papers in collaboration with other institutions. In contrast, other institutions, such as *Imperial College London* and *Central South University*, exhibited a tendency to conduct research with minimal institutional collaboration or even in isolation from other institutions.

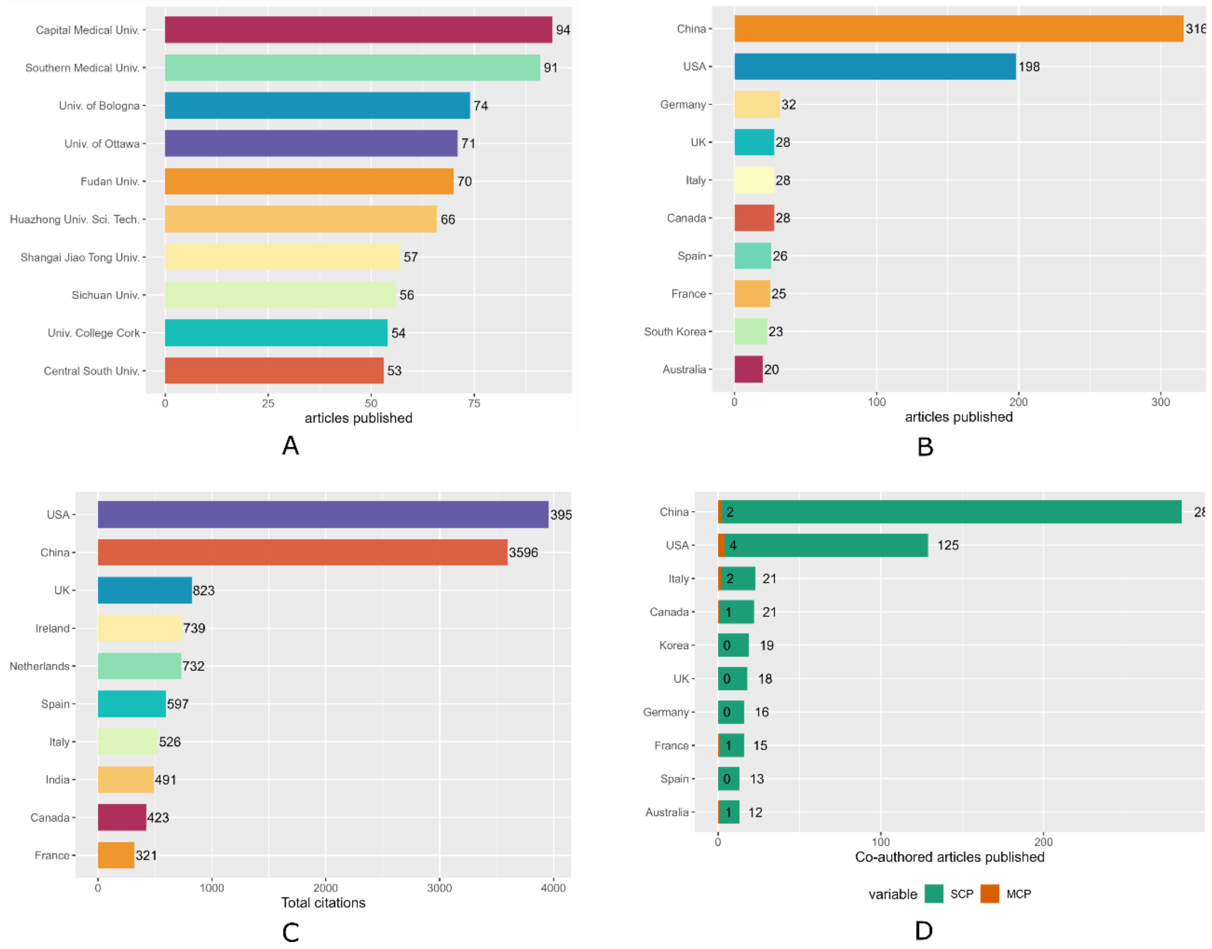

Figure 3. (A) Number of publications of the 10 most productive research institutions. (B) Number of publications per country. (C) Number of times papers from each country were cited. (D) Proportion of papers in co-authorship for the 10 most productive countries. *SCP = Single Country Publication*. *MCP = Multiple Countries Publication*.
Source: the authors, 2024.

PRODUCTIVITY BY COUNTRY

A total of 55 countries were represented by the researchers who published papers on the theme (Tables S14, S15). *China* demonstrated the greatest output of published papers (316), followed by the *USA* (198), *Germany* (32), the *UK* (28) and *Italy* (28) in descending order (Figure 3B). Figure 3C provides a summary of the results for the most cited countries. *USA* was the most cited country with 3,959 citations, followed by *China* (3,596), *UK* (823), *Ireland* (739), and the *Netherlands* (732). A total of seven countries had papers that had been cited on more than 500 occasions. In terms of the mean number of citations per paper, the Northern European countries of *Norway* (95.00 citations), *Netherlands* (73.20 citations per paper), *Ireland* (61.60), and *UK* (45.70 citations per paper) presented the highest number of citations. This was followed by *India* (49.10 citations per paper), *Spain* (45.90 citations per paper), *Pakistan* (36.00 citations per paper), *Ecuador* (36.00 citations per paper), the *USA* (30.70 citations per paper), and *Portugal* (30.00 citations per paper) which comprise the list of the ten most frequently cited countries (Table S16).

China, the country with the highest number of citations, occupied the 28th position in the list of cited papers, with an average of 12.60 citations per paper. This is due to the high quantity of published papers. Even though papers from *Ireland* were cited on 739 occasions in other published papers, there were only 9 published papers from this country, which resulted in the higher average and gives *Ireland* the first position on the list. The role of international collaboration in relation to the countries under analysis was not a significant factor, exhibiting limited variability. Figure 3D illustrates the extent of international collaboration in the publications of the ten countries that published the most. It demonstrates that a significant proportion of these publications involved authors within the same country. Of the ten countries that published the most, only four (*China, USA, Italy,* and *Canada*) published papers with international collaboration: the *USA* with four papers of a sample of 129 (16.3%), *China* and *Italy* with two papers each from a sample of 285 (36.10%) and 23 (2.9%) respectively, and Canada with one paper from a sample of 22 (2.8%) (Table S13).

Relations between countries are contingent upon a multitude of factors. A review of the data revealed a mere 14 instances of inter-country collaboration out of a total of 789 papers, which serves to illustrate the dearth of international collaboration. An examination of the collaborative network between countries reveals that, of the 55 countries that published on the theme, only 17 engaged in international co-authorship (Figure S25). *China* and *USA* demonstrated the greatest degree of collaboration, publishing papers in conjunction with eight countries (Table S23). This resulted in a total of 12 collaborations between *China* and 8 other countries, and 10 collaborations between the *USA* and 8 other countries. A sample of 789 papers revealed that only 50 (6.33%) were published in international collaboration, indicating a low level of collaboration overall. *China*, the country with the highest number of publications, authors, and impactful papers, published seven papers in collaboration with other countries (*Canada*, *France*, *Iran*, *Spain*, *Sweden* and *Turkey*), being one paper with each country. Collaborations between *China* and the *USA*, the most prolific in terms of published papers, resulted in only five papers (0.006%).

MOST FREQUENT KEYWORDS

The most frequently used authors' keywords among the papers analyzed were *gut microbiota* (172) and *gut microbiome* (119), followed by *microbiome* (84) and *microbiota* (57). Each of these terms is related to the intestinal microbiome (Figure 4A). Keywords related to bioinformatics were also analyzed. The most frequently used of these were *bioinformatics* (40), *16S rRNA* (25), and *metagenomics* (25).

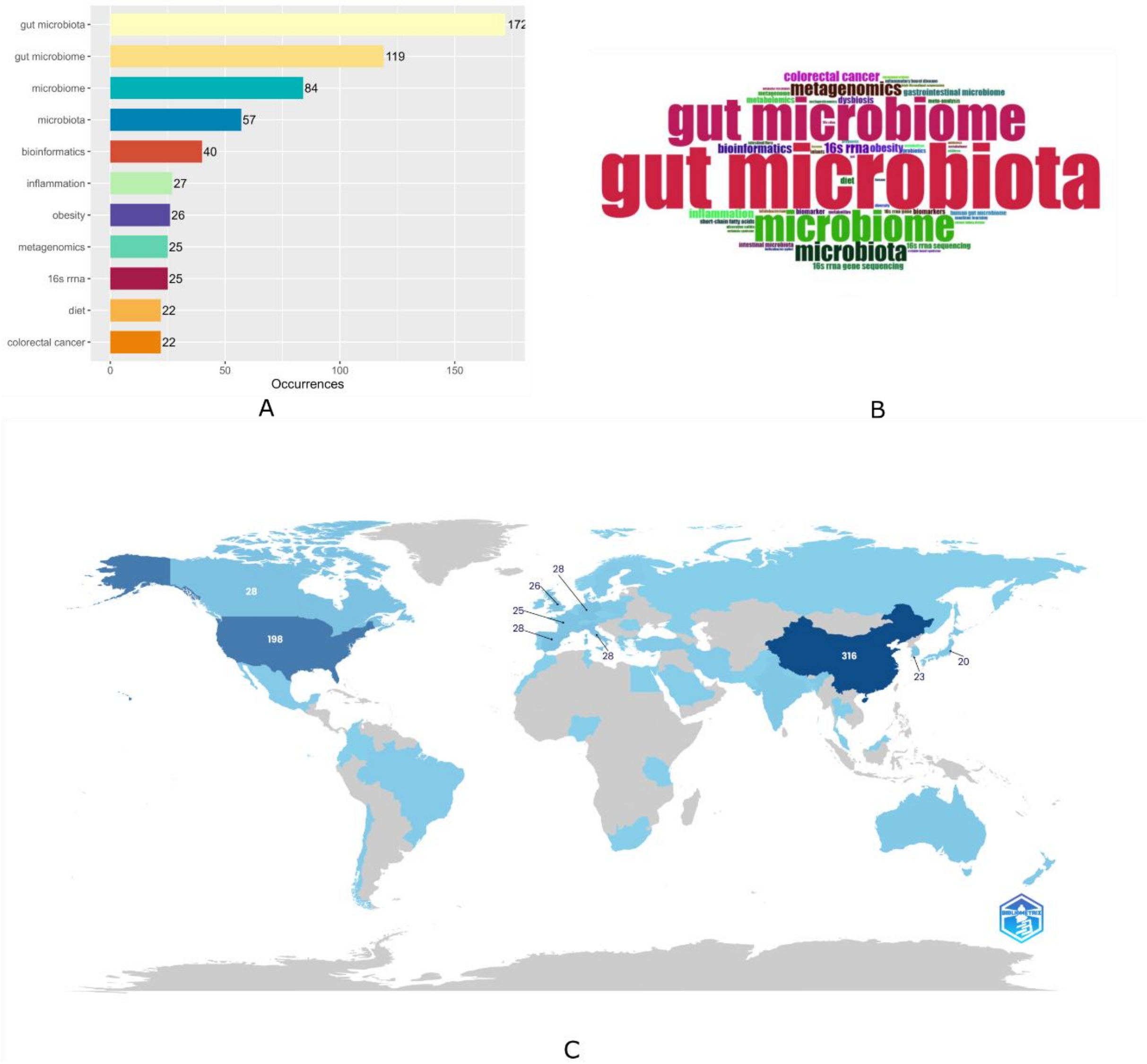

Figure 4. (A) Most frequently used keywords. (B) Word Cloud Map. (C) Map of countries productivity (top ten). Source: the authors, 2024.

The Word Cloud map (Figure 4B) illustrates the most pertinent keywords employed by the authors in the sample analyzed. The keywords exhibited *trending topics* over the period, with the years 2020, 2021, and 2022 presenting the greatest diversity of keywords (Figure S13). In 2019, the most frequent keywords were *microbial diversity* and *rheumatoid arthritis*. In 2021, the most studied subjects were associated with the term's *gut microbiome* (119), *microbiome* (84), and *bioinformatics* (40). In 2023, the final year of the period under analysis, the term *mycobiota* appeared five times, indicating an interest in another group of microorganisms that play an essential role in regulating gastrointestinal functions.

QUALITATIVE ASPECTS

The concept of bibliographic linkage is a fundamental tenet of bibliometrics, denoting the act of citing the same reference by two or more academic papers. If two articles, designated as A and B, make simultaneous citations to a third article, these papers are said to be bibliographically linked[25]. This linkage indicates the degree of correlation between the two works, considering the references in common that are cited by both. This approach allows the identification of not only frequently cited articles by authors within a research area, but also of cited references that are bibliographically linked. Further, it is possible to differentiate between clusters of articles based on the keywords employed by the authors. Bibliometrics assess the impact of each cluster, quantifying the number of global citations that contain the specified keywords and determining the cluster's centrality, which is defined as the recurrence of keywords within the analyzed articles (Figure S14-S19). The keywords *microbiome*, *microbiota,* and *bioinformatics* were used with greater frequency in articles with a high number of citations, whereas the keywords *gut microbiome* and *gut microbiota* were employed in a greater quantity of articles (Figure S14). The group comprising the terms *gut microbiota*, *inflammation*, and *16S rRNA* is distinguished from other groups by its simultaneous high impact and high centrality. Figure S14 also demonstrates a robust correlation between the terms *obesity* and *inflammation* in papers of moderate or high impact, which reinforces a potential relationship between the most studied topics in this field.

CONCEPTUAL STRUCTURE

The co-occurrence network demonstrates the extent of correspondence between specific keywords within the specified period (Figure S20). The network demonstrates that the most frequently used terms form a multitude of combinations both within and between their respective clusters, resulting in associations such as *gut microbiota* and *Helicobacter pylori*. The approach allows for the measurement of the relevance (centrality) and development (density) of different themes within a study, as indicated by the keywords employed in the investigation. Figure S21 uses two axes to separate the keywords into 4 thematic groups: *niche themes*, *motor themes*, *basic themes*, and *emergent/declining themes*. *Niche themes*, defined by high density and centrality, are well-developed but with few external linkages, and can be exemplified by terms related to microorganisms, such as *Enterococcus*, *Lactobacillus*, and *Bacteroides*; *Motor themes* are those with high centrality and high density, which characterizes them as important and well developed, being exemplified by the keywords *obesity*, *intestinal flora*, and *high-throughput sequencing*; *Basic themes* possess high density and low centrality, being poorly developed but frequently used in publications; and *Emergent/declining themes* simultaneously possess low density and low centrality, explained by the uncertainty of these themes due to fewer studies with these themes – indicating potential for future study – or due to a lack of interest in the theme(s).

INTELLECTUAL STRUCTURE

In the context of bibliometrics, co-citation refers to the phenomenon of two papers or authors (A and B) being cited by the same document or author (C). The recurrence of co-citation indicates a high level of association between the two papers in question, as well as their relevance as references among authors in their respective fields of knowledge. A particularly useful tool for elucidating the way disparate authors can contribute to the advancement of this field of study is the co-citation network (Figure S23). The co-citation network illustrates mutual citations between groups of authors, elucidating interconnections between the subjects under investigation.

DISCUSSION

To advance our understanding of microbial communities in organs and systems of the human body, it is important to study the genetic characteristics and interactions of the HGM with the environment (microbiome). The potential of microbiota studies to inform the development of new therapeutic approaches and in the clinical detection of diseases is immense. Since its inception, microbiome study and research has undergone rapid and profound changes driven by advances in sequencing tools and techniques. The difficulty in laboratory cultivation of microorganisms from the human microbiome, coupled with the loss of information even when cultivation was possible, has led researchers and companies to perfect devices and techniques for sequencing the genetic data of these microorganisms in situ, i.e. directly in their natural environment[27]. Our results can help researchers and students evaluate their research trajectories, choose potential topics and gaps, plan studies and network their research through future collaborations.

Advancements in sequencing technology have enabled researchers to obtain vast amounts of data for processing and extraction of genetic information. Developing the requisite equipment, computational tools, and methods for extraction and analysis genetic data, researchers employ statistical techniques that give rise to the field of bioinformatics[8,27]. The advent of New Generation Sequencing (NGS) technologies in the early 2000s enabled detection, identification, and analysis of microorganisms, particularly those belonging to the HGM. The combination of these technologies with bioinformatic tools has facilitated research while integrating bioinformatics and NGS in HGM investigations [31].

Due to such technological advances, the study data show a growing and cumulative scientific production on topics related to HGM between 2018 and 2021. This is probably due to the greater accessibility of genetic sequencing technologies and the progressive improvement of bioinformatics software and tools[31]. Despite this, there was a persistent and gradual decline in scientific output on the topic after 2022. The comparison of Figures 2A and 2B reveals a notable decline in the number of citations per year over the period. It might be that the decline in citations was a contributing factor to the observed reduction in scientific output on the topic after the year 2022. However, it is important that other factors may also be at play, as there was no clear causal relationship between the decline in citations and the observed reduction in scientific output in the earlier periods under analysis. Another reason for the decline in scientific production in this area is the complexity of analyzing and interpreting the data obtained after sequencing. This, coupled with the need for adequate computer resources and information technologies to store and process data poses significant challenges for studies in the field[31].

Of the five journals with the highest number of publications on the subject, the majority were also the most relevant according to their h and g indexes (Figures 2C and 2D). Despite having a lower impact factor than *Microbiome*, which had 36 publications and the first-highest impact (h=17) among the analyzed journals, the periodical *Frontiers in Microbiology* presented 71 publications on the theme. However in general, the impact of each journal was found to be related to the theme.

Most of the authors included (being also those who contributed the most) were of Chinese nationality, including the ten most relevant and prolific authors who presented similar h-index values. For example, Li, Y. and Liu, Y. have clearly established the most central scientific relationships and networks with other authors. Through a comparative analysis of the most important authors, institutions and countries in terms of production and impact, a pattern could be identified in which China, North America, and Europe consistently ranked among the top five in every category. Consequently, China, Canada, and Italy were identified as the most prolific countries, which was directly correlated with the institutions situated in these countries (those that were most productive). Another reason (apart from the economic output) which is remarkable for China is the growing body of research examining the role of gut microbiota in complementary and alternative medical practices. This is collectively known as “traditional Chinese medicine” (TCM), and has a longstanding and pervasive presence in the country[32].

Scientific consortia such as the Human Microbiome Project (HMP), the European MetaHIT, and the Integrative Human Microbiome Project (iHMP) are interested in metagenomic studies on the HGM. These initiatives are in line with research being conducted in countries such as China, Canada and Italy[33]. An exception to this is the United States (USA), which occupies second place in academic production, and yet is also a member of these research consortia, despite not hosting any of the ten most productive institutions.

Concentration on a single institution, which plays a central role in promoting research in this area, could explain this result. Among Chinese universities, *Jiangnan University* and *Zhejiang University School of Medicine* were identified as the most collaborative institutions, although only when compared with other Chinese universities. However, though China produced the greatest number of published papers, it did not engage in significant collaborative efforts with other countries. This reveals a scenario of scientific isolation that attributable to political and economic factors that shape the way Chinese intellectuals engage with international scholars[34]. Interestingly, the USA's most important scientific collaborators are Brazil, England, and India.

With respect to academic productivity, China has been identified as the leading nation, having published over 200 articles more than the other countries included in the analysis. Moreover, China is the nation with the second highest number of citations. This is the result of two factors: its high production of publications and the focus on collaboration with local universities. Notwithstanding, the highest positions in the ranking for number of citations are held by *UK*, *Ireland*, *Netherlands,* and *Spain*, in that order (Figure 3B).

A review of the most frequently used keywords revealed that *gut microbiota*, *gut microbiome*, *metagenomics*, and *colorectal cancer* exhibited higher correlations to gut microbiota and bioinformatics (Figures 4A and 4B). This represents one of the lines of study in the field that addresses the importance of the relationship between HGM and colorectal cancer, and depends not only on the presence of microorganisms, but the fecal microbiota as well, which was first scientifically noted at the beginning of the 20th century[35].

Co-occurrence analysis demonstrated that despite the absence of direct collaboration between authors, their publications are similar in terms of keywords and themes. This approach allows for an examination of the influence of specific keywords within the context of the analyzed theme. Thus, the keyword *gut microbiota* is among the most frequently utilized, with associations extending to other themes such as *inflammation* and *16s rRNA*, which is linked to technological advancements in sequencing and shaped by the HMP. Figure S14 reveals that *microbiome* is closely related to *bioinformatics*. This is due to studies investigating the association between diseases and new therapeutic approaches for conditions such as cancer[37].

Also, according to the mapping, the relevance and development of these themes, as well as related studies in bacterium of the gut microbiota, are increasing, but they are no longer considered to be as relevant as other themes. Since the inception of the HMP in 2007, there has been a notable surge in research activity, with a particular focus on identifying the existing species and functions of the gut microbiota[35]. The relationship between obesity and microbiota, as well as studies on next-generation sequencing and the microbiota, remain pertinent. This, given the ongoing discovery of new phenomena involving the brain-gut axis.

The co-citation network demonstrates that certain authors are particularly prominent within the context of the themes analyzed (Figure S17). These include Zhang, J., Zhang, X, and Li, Y., who were the most frequently cited authors. Figure S23 depicts four nodes, with Wang, Y. and Li, Y., represented by the blue node, pertaining to computational biology in microbiomes. Liu, Y. is the most frequently cited author in the field of bioinformatics applied to the study of gut microbiota (red node). The topic is also represented by the purple node, which is led by Turroni, S. The final node, depicted in green, represents researchers who have published in the field of genomics incorporating computational elements. The network demonstrates that most collaborations are between scholars from universities in China.

Analysis of the data reveals that the number of studies on HGM has decreased. However, research on the relationship between HGM and other diseases continues, either directly or indirectly. In addition, previously abandoned topics such as the fecal microbiota have resurfaced, helping to understand HGM, and how to modify it to improve quality of life. Finally, this study contributes to understanding the behavior of scientific research in this field, analyzing potentially new and expanding lines of research, and as the field of research evolves, reflecting on and promoting international cooperation.

CONCLUSION

Bibliometric studies combined with systematic literature searches are an effective way to analyze scientific production. This is a fundamental way to analyze and anticipate findings in a field of study and contribute to a better understanding of HGM in comparison to other diseases. No scientific mapping study of the subject has been published before, which gives this work its originality, but prevents us from comparing the results obtained with other studies. The data from 789 published papers were analyzed with consideration of several factors, including publication in periodicals, impact (h-index), productivity and the collaboration of institutions and countries, impact and collaborations between authors, most frequently used keywords, and correspondence within the knowledge area. The period from 2018 to 2023 evidenced a notable increase in the quantity of publications. China demonstrated a notable presence across all analyzed categories, with North American and Europe (despite a slight decline in absolute publication and citation numbers) also exhibiting robust scientific output. Papers incorporating bioinformatics as a supplementary field in diverse research endeavors have played a pivotal role in advancing scientific understanding, with most of the papers analyzed relating pathophysiology and disease diagnosis as associated with gut dysbiosis.

The sample analyzed, despite a considerable number of authors and a relatively low number of papers authored by a single individual exhibited a paucity of international collaboration. This suggests a scenario of elevated internal collaboration yet diminished external collaboration. Some of the papers analyzed presented keywords indicating utilization of bioinformatic tools and techniques, including *metagenomics*, *shotgun*, and *16S rRNA*. However, comprehensive analysis of the specific techniques and tools employed was not feasible, and therefore coming research should investigate further the use of bioinformatic techniques, examining impacts and objectives to gain insight into this growing phenomenon.

To gain a more comprehensive and robust understanding of scientific production, studies should consider a longer time frame and include all available publications in the databases. Although this study employed the three most utilized databases for reviews and bibliometric research, it was determined that the *Scopus* database was inadequate for the purposes of this investigation. However, to provide a more realistic representation of the bioinformatics applications in HGM studies, future research should include other databases within the health knowledge domain.

REFERENCES

1. Vyas U, Ranganathan N. Probiotics, Prebiotics, and Synbiotics: Gut and beyond. Gastroenterol Res Pract [Internet]. 2012 [cited 2025 Feb 1];2012:1–16. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3459241/. doi: https://doi.org/10.1155/2012/872716.
2. McGuinness AJ, Davis JA, Dawson SL, Loughman A, Collier F, O'Hely M, et al. A Systematic Review of Gut Microbiota Composition in Observational Studies of Major Depressive disorder, Bipolar Disorder and Schizophrenia. Mol Psychiatry [Internet]. 2022 Apr 1 [cited 2025 Feb 7];27(4):1920–35. Available from: https://www.nature.com/articles/s41380-022-01456-3. doi: https://doi.org/https://doi.org/10.1038/s41380-022-01456-3.
3. Góralczyk-Bińkowska A, Szmajda-Krygier D, Kozłowska E. The Microbiota–Gut–Brain Axis in Psychiatric Disorders. Int J Mol Sci [Internet]. 2022 Sep 24 [cited 2025 Feb 7];23(19):11245. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9570195/. doi: https://doi.org/10.3390/ijms231911245.
4. Illiano P, Brambilla R, Parolini C. The Mutual Interplay of Gut microbiota, Diet and Human Disease. FEBS J [Internet]. 2020 Mar 1 [cited 2025 Feb 7];287(5):833–55. Available from: https://pubmed.ncbi.nlm.nih.gov/31955527/. doi: https://doi.org/10.1111/febs.15217
5. Ruan W, Engevik MA, Spinler JK, Versalovic J. Healthy Human Gastrointestinal Microbiome: Composition and Function after a Decade of Exploration. Dig Dis Sci [Internet]. 2020 Feb 18 [cited 2025 Feb 1];65(3):695–705. Available from: https://link.springer.com/article/10.1007/s10620-020-06118-4. doi: https://doi.org/10.1007/s10620-020-06118-4
6. Quaglio AEV, Grillo TG, Oliveira ECSD, Stasi LCD, Sassaki LY. Gut microbiota, Inflammatory Bowel Disease and Colorectal Cancer. World J Gastroenterol [Internet]. 2022 Aug 14 [cited 2025 Feb 7];28(30):4053–60. Available from: https://www.wjgnet.com/1007-9327/full/v28/i30/4053.htm. doi: https://doi.org/ 10.3748/wjg.v28.i30.4053
7. Manichanh C, Borruel N, Casellas F, Guarner F. The Gut Microbiota in IBD. Nat Rev Gastroenterol Hepatol [Internet]. 2012 Aug 21 [cited 2025 Feb 1];9(10):599–608. Available from: https://www.nature.com/articles/nrgastro.2012.152. doi: https://doi.org/10.1038/nrgastro.2012.152
8. Lepage P, Leclerc MC, Joossens M, Mondot S, Blottière HM, Raes J, et al. A Metagenomic Insight into Our gut's Microbiome. Gut [Internet]. 2012 Apr 23 [cited 2025 Feb 1];62(1):146–58. Available from: https://gut.bmj.com/content/62/1/146. doi: https://doi.org/10.1136/gutjnl-2011-301805
9. 9. Aguiar-Pulido V, Huang W, Suarez-Ulloa V, Cickovski T, Mathee K, Narasimhan G. Metagenomics, Metatranscriptomics, and Metabolomics Approaches for Microbiome Analysis. Evol Bioinf [Internet]. 2016 Jan [cited 2025 Feb 1];12(Suppl 1):EBO.S36436. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4869604/. doi: https://doi.org/10.4137/EBO.S36436
10. Chen Y, Zhou J, Wang L. Role and Mechanism of Gut Microbiota in Human Disease. Front Cell Infect Microbiol [Internet]. 2021 Mar 17 [cited 2025 Feb 7];11. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8010197/#:~:text=Hypertension%20is%20associated%20with%20altered%20gut%20function%2C%20altered. doi: https://doi.org/10.3389/fcimb.2021.625913
11. Adak A, Khan MR. An Insight into Gut Microbiota and Its Functionalities. Cell Mol Life Sci [Internet]. 2018 Oct 13 [cited 2025 Feb 1];76(3):473–93. Available from: https://link.springer.com/article/10.1007/s00018-018-2943-4. doi: https://doi.org/10.1007/s00018-018-2943-4

12. Zhu L, Liu W, Alkhouri R, Baker RD, Bard JE, Quigley EM, et al. Structural Changes in the Gut Microbiome of Constipated Patients. Physiol Genomics [Internet]. 2014 Sep 15 [cited 2025 Feb 1];46(18):679–86. Available from: https://journals.physiology.org/doi/full/10.1152/physiolgenomics.00082.2014. doi: https://doi.org/10.1152/physiolgenomics.00082.2014
13. Jandhyala SM. Role of the Normal Gut Microbiota. World J Gastroenterol [Internet]. 2015 Aug 7 [cited 2025 Feb 1];21(29):8787. Available from: https://www.wjgnet.com/1007-9327/full/v21/i29/8787.htm. doi: https://doi.org/10.3748/wjg.v21.i29.8787
14. Chang C, Lin H. Dysbiosis in Gastrointestinal Disorders. Best Pract Res Clin Gastroenterol [Internet]. 2016 Feb 1 [cited 2025 Feb 1];30(1):3–15. Available from: https://www.sciencedirect.com/science/article/abs/pii/S1521691816000056. doi: https://doi.org/10.1016/j.bpg.2016.02.001
15. Wang B, Yao M, Lv L, Ling Z, Li L. The Human Microbiota in Health and Disease. Engineering [Internet]. 2017 Feb [cited 2025 Feb 1];3(1):71–82. Available from: https://www.sciencedirect.com/science/article/pii/S2095809917301492. doi: https://doi.org/10.1016/J.ENG.2017.01.008
16. Weiss GA, Hennet T. Mechanisms and Consequences of Intestinal Dysbiosis. Cell Mol Life Sci [Internet]. 2017 [cited 2025 Feb 1];74(16):2959–77. Available from: https://www.ncbi.nlm.nih.gov/pubmed/28352996. doi: https://doi.org/10.1007/s00018-017-2509-x
17. Foster J, Zhou L (Linda). Psychobiotics and the Gut–brain axis: in the Pursuit of Happiness. Neuropsychiatr Dis Treat [Internet]. 2015 Mar [cited 2025 Feb 1];11:715. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4370913/. doi: https://doi.org/10.2147/NDT.S61997
18. 18. Tomasello G, Mazzola M, Leone A, Sinagra E, Zummo G, Farina F, et al. Nutrition, Oxidative Stress and Intestinal dysbiosis: Influence of Diet on Gut Microbiota in Inflammatory Bowel Diseases. Biomed Pap Med Fac Univ Palacky Olomouc Czech Repub [Internet]. 2016 [cited 2025 Feb 1];160(4):461–6. Available from: https://www.ncbi.nlm.nih.gov/pubmed/27812084. doi: https://doi.org/10.5507/bp.2016.052
19. Perler BK, Friedman ES, Wu GD. The Role of the Gut Microbiota in the Relationship between Diet and Human Health. Ann Rev Physiol [Internet]. 2022 Nov 14 [cited 2025 Feb 7];85(1). Available from: https://www.annualreviews.org/content/journals/10.1146/annurev-physiol-031522-092054. doi: https://doi.org/10.1146/annurev-physiol-031522-092054
20. Carroll IM, Ringel-Kulka T, Siddle JP, Ringel Y. Alterations in Composition and Diversity of the Intestinal Microbiota in Patients with diarrhea-predominant Irritable Bowel Syndrome. Neurogastroenterol Motil [Internet]. 2012 Feb 20 [cited 2025 Feb 1];24(6):521-e248. Available from: https://onlinelibrary.wiley.com/doi/10.1111/j.1365-2982.2012.01891.x. doi: https://doi.org/10.1111/j.1365-2982.2012.01891.x
21. Yang G, Wei J, Liu P, Zhang Q, Tian Y, Hou G, et al. Role of the Gut Microbiota in Type 2 Diabetes and Related Diseases. Metabolism [Internet]. 2021 Apr [cited 2025 Feb 7];117:154712. Available from: https://www.metabolismjournal.com/article/S0026-0495(21)00012-3/abstract. doi: https://doi.org/10.1016/j.metabol.2021.154712.
22. Qiu P, Ishimoto T, Fu L, Zhang J, Zhang Z, Liu Y. The Gut Microbiota in Inflammatory Bowel Disease. Front Cell Infect Microbiol [Internet]. 2022 Feb 22 [cited 2025 Feb 7];12. Available from: https://www.frontiersin.org/journals/cellular-and-infection-microbiology/articles/10.3389/fcimb.2022.733992/full. doi: https://doi.org/10.3389/fcimb.2022.733992.
23. Tap J, Derrien M, Törnblom H, Brazeilles R, Cools-Portier S, Doré J, et al. Identification of an Intestinal Microbiota Signature Associated with Severity of Irritable Bowel Syndrome. Gastroenterology [Internet]. 2017 Jan [cited 2025 Feb 1];152(1):111-123.e8. Available from:

https://www.gastrojournal.org/article/S0016-5085(16)35174-5/fulltext. doi: https://doi.org/10.1053/j.gastro.2016.09.049

24. Afzaal M, Saeed F, Shah YA, Hussain M, Rabail R, Socol CT, et al. Human Gut Microbiota in Health and disease: Unveiling the Relationship. Frontiers in Microbiology [Internet]. 2022 Sep 26 [cited 2025 Feb 16];13. Available from: https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2022.999001/full
25. Khalil M, Agostino DC, Mahdi L, Jaber N, Maria D, Graziani A, et al. Unraveling the Role of the Human Gut Microbiome in Health and Diseases. Microorganisms [Internet]. 2024 Nov [cited 2025 Feb 16];12(11):2333. Available from: https://www.mdpi.com/2076-2607/12/11/2333
26. Lee JY, Bays DJ, Savage HP, Bäumler AJ. The Human Gut Microbiome in Health and disease: Time for a New chapter? Infection and Immunity [Internet]. 2024 Sep 30 [cited 2025 Feb 16];92(11). Available from: https://pubmed.ncbi.nlm.nih.gov/39347570/
27. Kuczynski J, Lauber CL, Walters WA, Parfrey LW, Clemente JC, Gevers D, et al. Experimental and Analytical Tools for Studying the Human Microbiome. Nat Rev Genet. 2012 Jan;13(1):47–58. doi: https://doi.org/10.1038/nrg3129
28. Aria M, Cuccurullo C. Bibliometrix : an R-tool for Comprehensive Science Mapping Analysis. J Informetr. [Internet]. 2017 Nov [cited 2025 Feb 1];11(4):959–75. Available from: https://www.sciencedirect.com/science/article/abs/pii/S1751157717300500?via%3Dihub. doi: https://doi.org/10.1016/j.joi.2017.08.007
29. D'Aniello L, Spano M, Cuccurullo C, Aria M. Academic Health Centers' configurations, Scientific productivity, and impact: Insights from the Italian Setting. Health Policy [Internet]. 2022 Sep [cited 2022 Oct 18];126(12). Available from: https://www.sciencedirect.com/science/article/pii/S0168851022002512?via%3Dihub. doi: https://doi.org/10.1016/j.healthpol.2022.09.007
30. R Core Team. R: a Language and Environment for Statistical Computing [Internet]. R Foundation for Statistical Computing. 2024 [cited 2025 Feb 1]. Available from: https://www.r-project.org/
31. Maljkovic Berry I, Melendrez MC, Bishop-Lilly KA, Rutvisuttinunt W, Pollett S, Talundzic E, et al. Next Generation Sequencing and Bioinformatics Methodologies for Infectious Disease Research and Public Health: Approaches, Applications, and Considerations for Development of Laboratory Capacity. J Infect Dis [Internet]. 2020 Mar 28 [cited 2025 Feb 1];221(Supplement_3):S292–307. Available from: https://academic.oup.com/jid/article/221/Supplement_3/S292/5586940. doi: https://doi.org/10.1093/infdis/jiz286
32. Feng W, Ao H, Peng C, Yan D. Gut microbiota, a New Frontier to Understand Traditional Chinese Medicines. Pharmacol Res [Internet]. 2019 Apr [cited 2025 Feb 1];142:176–91. Available from: https://www.sciencedirect.com/science/article/abs/pii/S1043661818318346?via%3Dihub. doi: https://doi.org/10.1016/j.phrs.2019.02.024
33. Galloway-Peña J, Hanson B. Tools for Analysis of the Microbiome. Dig Dis Sci [Internet]. 2020 Jan 31 [cited 2025 Feb 1];65(3):674–85. Available from: https://link.springer.com/article/10.1007/s10620-020-06091-y. doi: https://doi.org/10.1007/s10620-020-06091-y
34. Leite ACC, Máximo JCR. Através dos Olhos do dragão: Um Estudo Sobre o Campo das Relações Internacionais na China. Contexto Int [Internet]. 2013 Jun [cited 2025 Feb 1];35(1):231–59. Available from: https://www.scielo.br/j/cint/a/GgNy5mvbGYxwxcPTxxyMqtk/?lang=pt. doi: https://doi.org/10.1590/S0102-85292013000100008
35. Lewandowska-Pietruszka Z, Figlerowicz M, Mazur-Melewska K. The History of the Intestinal Microbiota and the Gut-Brain Axis. Pathogens [Internet]. 2022 Dec 15 [cited 2025 Feb 1];11(12):1540. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9786924/. doi: https://doi.org/10.3390/pathogens11121540
36. Mendes N, Innocentini L. Colaboração Científica intraorganizacional: Análise de Redes por Co-ocorrência de palavras-chave. Em Questão [Internet]. 2019 Jan 1 [cited 2024 Mar 14];25(1):87–

110. Available from: https://seer.ufrgs.br/EmQuestao/article/view/79906. doi: http://dx.doi.org/10.19132/1808-5245251.87-110

37. Wang J, Maniruzzaman M. A Global Bibliometric and Visualized Analysis of Bacteria-mediated Cancer Therapy. Drug Discov Today [Internet]. 2022 May [cited 2022 Jun 27];27(10). Available from: https://www.sciencedirect.com/science/article/abs/pii/S1359644622002148?via%3Dihub. doi: https://doi.org/10.1016/j.drudis.2022.05.023
38. Chang L, Wei Y, Hashimoto K. Brain–gut–microbiota Axis in Depression: A Historical Overview and Future Directions. Brain Res Bull [Internet]. 2022 May [cited 2025 Feb 1];182:44–56. Available from: https://www.sciencedirect.com/science/article/pii/S0361923022000375?via%3Dihub. doi: https://doi.org/10.1016/j.brainresbull.2022.02.004v
39. Nunes M, Pereira E, Luiz A. Bibliometrics: What Is it? What Is It Used for? and How to Do it? Cuad Ed Desar [Internet]. 2024 Feb 9 [cited 2025 Feb 7];16(2):e3042–2. Available from: https://ojs.cuadernoseducacion.com/ojs/index.php/ced/article/view/3042. doi: https://doi.org/10.55905/cuadv16n2-021