A Brief History of Microbial Study and Techniques for Exploring the Gastrointestinal Microbiome

Ashley M. Sidebottom, PhD[1]

[1] Duchossois Family Institute, University of Chicago, Chicago, Illinois

Clin Colon Rectal Surg 2023;36:98–104.

Address for correspondence Ashley M. Sidebottom, PhD, Duchossois Family Institute, University of Chicago, 900 East 57th Street, Chicago, IL 60637 (e-mail: asidebottom@medicine.bsd.uchicago.edu).

Abstract

Over the past 20 years, the study of microbial communities has benefited from simultaneous advancements across several fields resulting in a high-resolution view of human consortia. Although the first bacterium was described in the mid-1600s, the interest in community membership and function has not been a focus or feasible until recent decades. With strategies such as shotgun sequencing, microbes can be taxonomically profiled without culturing and their unique variants defined and compared across phenotypes. Approaches such as metatranscriptomics, metaproteomics, and metabolomics can define the current functional state of a population through the identification of bioactive compounds and significant pathways. Prior to sample collection in microbiome-based studies it is critical to evaluate the requirements of downstream analyses to ensure accurate processing and storage for generation of high data quality. A common pipeline for the analysis of human samples includes approval of collection protocols and method finalization, patient sample collection, sample processing, data analysis, and visualization. Human-based microbiome studies are inherently challenging but with the application of complementary multi-omic strategies there is an unbounded potential for discovery.

Keywords

- microbiome
- microbial communities
- multi-omics
- microbe

The History of Studying Bacteria

Microbial communities have been studied for millenia by civilizations across the world before they were described by modern science. In particular, beneficial techniques utilizing consortia such as food fermentation and the use of manure on crops date back to 6,000 BCE. Even though these early societies were not unaware of the molecular or mechanistic foundations generations of processes were built on, we now know these methods to be the result of complex, living ecosystems called microbiomes.

Before modern medicine, the gastrointestinal (GI) tract was frequently studied in the context of digestion complications or disease. Societies would use microbial community-derived products in the form of medicinal remedies and ancestral foods to "guard the stomach," treat ailments, and promote general well-being and health. There is evidence from the Romans that disease transmission through waste material was understood with the construction of aqueducts to introduce clean water and remove waste. However, we did not learn until the mid-1600s that healthy human GI tracts were home to microbes, and it was not until the 20th and 21st century we learned they were essential to human health.

In 1676, Antony van Leeuwenhoek reported the first visual observation of bacteria. With a single lens microscope of his own design, Leeuwenhoek studied the rod-shaped cells within a mouth scraping sample and he called the organisms little animals or "animalcules," a term he had coined a few years earlier. In his examination of these animalcules, Leeuwenhoek would go on to describe and draw their movement through liquid (i.e., cellular motility), detail his belief that some could survive without air, and he would accurately approximate their length (3 µm). From these descriptions and what is now known about the residents of the mouth microbiome, he was most likely describing a member of the

article published online January 25, 2023

Issue Theme Microbiome; Guest Editor: Benjamin D. Shogan, MD, FACS, FASCRS

Thieme Medical Publishers, Inc., 333 Seventh Avenue, 18th Floor, New York, NY 10001, USA

DOI https://doi.org/10.1055/s-0042-1760678.

ISSN 1531-0043.

genera, *Selenomonas*. Leeuwenhoek would also be the first to report a microbe, the parasite *Giardia*, from his own fecal material. While the discovery of unicellular organisms was fundamental to the fields of microscopy and microbiology, Leeuwenhoek would not share his microscopic methods and it would take approximately 200 years for scientists to validate his findings.

Discovering How to Culture and Define Bacteria

Following Louis Pasteur's report in 1860 on the first culture medium recipe that could support organism growth, scientists would quickly develop tools and methods for the selective isolation of bacteria from ecosystems. These critical studies included specific growth medias, development of solid agar media, and design of sterile growth containers (i.e., Petri dish). Additionally, microbes started to receive general classifications based on characteristics such as gas production on specific sugars, motility, and structure of their cell wall composition with the invention of the Gram stain by Hans Christian Gram in 1884. Despite the fact bacteria's discovery originated from an oral community and microbes were discovered within a fecal sample, the field would predominantly focus on the isolation and study of pure cultures for the next 100 years.

During this new era of bacterial study, the discovery of gut-derived microbes and anaerobic communities had begun with fundamental work by Pasteur in the 1860s. Pasteur discovered the first pathogenic anaerobe, *Clostridium septicum,* and in 1863 he coined the terms "aérobis" and "anaérobis" depending on the microbes' requirement for oxygen. During this time, additional methods for culturing gut-derived microbes were reported. For example, in 1905, Alfred MacConkey added bile salts to liquid culture to promote the growth of lactose fermenting bacteria from feces.[1] Pasteur and others also began experimenting with anaerobic growth conditions.[2,3] Studies at the end of the 19th century included the isolation of *Clostridium tetani* in pure culture with successful anti-toxin development against the anaerobe developed only 1 year later in 1890.[4]

From Studying Microbial Community Natural Product Production to Shotgun Sequencing

The interest in microbial community functions and interspecies interactions was ignited following the discovery of penicillin G from the fungus *Penicillium notatum* by Alexander Fleming in 1928. Fleming's foundational studies reminded the field that for millions of years microbes have interacted with each other and an understanding of this unknown world could provide significant breakthroughs for treatments and therapeutics. A significant natural product isolation effort from environmental source materials resulted in the discovery of several new chemical subclasses, many of which went on to be the building blocks of large scale, high-throughput screening studies by the pharmaceutical industry in the late 1900s (e.g., aminoglycosides, polyketides, cephalosporins, macrolides, and tetracyclines).

For the next several decades, investigators would continue to rely on experimental, in vitro evidence to determine the function of a microbe or a consortium. However, with the application of 16S rRNA-based sequencing and whole genome sequencing (WGS) in the 1970s and 1990s, the functional potential of a microbe and a population could be assessed without culturing. In 1995, back-to-back Science publications reported the first two WGS from bacteria: *Haemophilus influenzae* (1,830,137 bp) and *Mycoplasma genitalium* (580,070 bp).[5,6] Although WGS provides an organism's complete genome, the technique requires the growth and isolation of individual organisms which is a significant barrier for community-based studies. With the rapid progression of sequencing technologies and decreased run costs, 16S rRNA-based sequencing became a high-throughput technique that could be used to survey and define the taxa of consortia. From these data, it was clear that in vitro culturing techniques were insufficient for culturing all organisms in a sample, which further pushed the need of culture-independent methods for assessing microbial communities or isolates.

Jo Handelsman was the first to coin the term metagenomics.[7] From her work on assessing biosynthetic gene clusters and their resulting natural products, she was one of the first to suggest DNA analysis from an entire sample was useful for predicting if the community contained new bioactive small molecules. Two major questions remained: Can you and how do you reassemble every genome from a whole sample? In 2004, foundational work by Tyson and colleagues used deep shotgun sequencing with relaxed reassembly requirements to assemble and bin more contigs allowing for the near-complete reconstruction of two genomes and partial reconstruction for three other genomes from a single biofilm sample.[8] There are accepted best practices for sample processing,[9,10] however, there is not a standard method for data analysis pipelines although there has been some effort to create standard protocols despite the inherent challenges.[11–13] Essential work in genome curation[14] and data visualization strategies such as Anvi'o[15] are ongoing and because the field and its techniques are evolving, data analysis is not trivial and requires an investigator familiar with complex datasets to lead. Studies can now reconstruct individual strain genomes from consortia through shotgun sequencing to track biological function within communities (i.e., gene and bacterial fitness,[16] niche partitioning[17]), identify single nucleotide variants, identify single amino acid variants, and explore and compare genomes across all branches of life.

With the rapid increase in our ability to resolve whole genomes within consortia, the application of shotgun sequencing to human study has resulted in the generation of enormous datasets. In one publication by Pasolli and colleagues, 46 datasets were analyzed, and more than 150,000 genomes were reported from approximately 10,000 human metagenomes.[18] There have been recent cancer studies associating taxa with disease phenotypes such as *Flavonifractor plautii*, *Bacteroides vulgatus,* and *Parabacteroides* spp *CT06* in early-onset colorectal cancer[19] and *Streptococcus* spp, *Veillonella* spp, and *Actinomyces* spp in three independent cohorts of patients with pancreatic ductal carcinoma.[20]

Complementary to metagenomics, metatranscriptomics is the analysis of gene expression within a microbial community. Pathways are identified and can be mapped back to organisms through paired metagenomic data. RNA-based studies are inherently challenging due to the labile nature of the single-stranded biomolecule and processing requires specific preservation methods to prevent degradation.[21,22] One study that resulted from the Human Microbiome Project 2 (HMP2) reported species-specific transcriptional activity in their inflammatory bowel disease (IBD) patients compared with non-diseased controls. Two organisms, *Alistipes putredinis* and *Bacteroides vulgatus*, were correlated with disease severity and responsible for the expression of the methylerythritol phosphate pathway in IBD. *A. putredinis* was negatively correlated while *B. vulgatus* was positively correlated.[23] These data were of high interest because the mechanisms for gut dysbiosis and chronic diseases such as IBD remain unknown.

A New Era of Protein and Metabolite Analysis

Alongside the genomic revolution, metabolomics and proteomics have been experiencing a wave of advancement with the introduction of instrumentation into laboratories and more robust data analysis pipelines. Nuclear magnetic resonance (NMR) and a series of mass spectrometers such as the liquid chromatography-quadrupole time-of-flight-mass spectrometer (LC-QTOF-MS), liquid chromatography-orbitrap-mass spectrometer (LC-Orbitrap-MS), and electron impact/chemical ionization gas chromatography mass spectrometer (EI/CI-GC-MS) are most widely used in studies today. For one metabolomics project, several instruments can be used to detect a wide range of compounds with varying physicochemical properties. Although the technology for many instruments was developed decades before their widespread application to the multi-omics studies, systems have become more widely available to academic institutions since the early 2000s and these instruments have launched proteomics and metabolomics into a new era of discovery.

Metaproteomics is the study of expressed protein content within a microbial population. While genomic strategies can define the functional potential of a community, metaproteomics aim to characterize the active functional state of a population based on the detected peptide and protein profiles. Generally, proteins are extracted from samples and enzymatically cleaved by trypsin into peptides for analysis.[24] Because peptides are synthesized with known building blocks and trypsin specifically cleaves the C-terminal side of lysine and arginine, protein sequences can be reconstructed from peptide fragmentation spectra and mapped back to paired metagenomic data for the same sample. In a study by Tanca and coworkers, almost 30,000 microbial peptides detected from a human colonic luminal content cohort[25] were reported. The same group also reports distinguishing microbial peptides between three tumor clinicopathological features (294 distinguishing peptides for stage, 94 for grade, and 568 for tumor infiltrating lymphocytes, TILs) with distinguishing capabilities of 95% accuracy by stage, 100% accuracy by grade, and 100% accuracy based on the presence of TILs. In a separate study by Long and coworkers, 91,902 peptides and 30,062 protein groups from the fecal samples of colorectal cancer and healthy donor cohorts[26] are described. From this, 341 peptide groups were identified to be significantly altered in abundance between colorectal cancer and healthy donors. The groups were associated with functions such as iron uptake, oxidative stress, and DNA recombination, repair and replication. The clinical impact of these data will require follow-up, longitudinal studies, and paired in vivo work but the annotation of microbial community proteomes continues to grow as a powerful complement to other multi-omic strategies.

Metabolomics has long been a primary focus of microbial community studies because of the biosynthetic potential of consortia to produce therapeutically relevant compounds as previously described in this work. Additionally, microbe metabolite levels in the GI tract can impact host immune function, nutrient uptake, mental health disorders, and organ function. Metabolomics aims to detail the current functional state of the microbiome through the analysis of compounds approximately less than 2,000 Da in size. However, because microbes and humans produce many of the same compounds and the bioactivity of a single compound in the context of thousands is not well understood, interpreting metabolomics results remains a significant challenge. To address these issues, metabolic modeling studies and pipelines such as KBase have great promise to link assembled genomes to metabolic flux in a sample.[27] However, this analysis is currently limited to a few input genomes and not microbial communities in complex environments.

Biospecimens contain thousands of compounds within a single sample. Metabolites range in size, hydrophobicity, charge and other properties and a study will often require multiple instruments to describe several compound classes. Inherently, metabolomic studies have a significant bottleneck in compound identification and validation that is not experienced by genomics or proteomics. Metabolite structures are not composed of characterized, repeatable subunits like genes (nucleotides) or proteins (amino acids) and modifications (e.g., dehydration, decarboxylation, reduction, oxidation) can occur through known and unknown mechanisms by the host, other microbes, and the environment. Although *in silico* fragmentation modeling tools are becoming more advanced (SIRIUS[28]), feature validation requires authentic standard comparison on the same instrument as the sample was run to be considered the highest level of MS confirmation. Because of this, many studies begin with targeted methods to evaluate known compound classes such as bile acids, short chain fatty acids, amino acids, mono-/di-/tri-saccharides, fatty acids, tryptophan catabolites, indoles, and other small organic acids. Many of these compounds are routinely quantified and/or reported as normalized relative abundance by academic and commercial laboratories.

In addition to known features, biospecimens are composed of a substantial percentage of unknown features that are of significant interest due to unknown etiology of many diseases. NMR and MS^n techniques are frequently used to

evaluate the unknown metabolite space of a sample. In untargeted MS, compound fragmentation profiles can be compared with known compounds to putatively assign *m/z* values to a previously described subclass. The open-source pipeline Global Natural Products Social Molecular Networking (GNPS) compares fragmentation profiles to databases for identification of structural similarity to known subclasses.[29] In a study by Quinn and coworkers, GNPS uncovered new bile acid variants with amino acid modifications from murine material.[30] In humans, a recent application of the untargeted MS approach was applied by Gumpenberger and colleagues to a cohort of 88 colorectal cancer patients, 200 high-risk adenoma patients, and 200 low-risk adenoma patients.[31] They report 442 statistically significant molecular features from plasma discriminating between colorectal cancer and adenoma diagnosis. Similar to other strategies, metabolomics findings are often correlative and used as a starting point for studies and their findings require more rigorous experimentation to validate the significance of an altered metabolome.

The Foundational Microbiome Studies

The NIH HMP was a two phase, decade long, multi-institutional study that laid the groundwork for exploring human microbiomes and health outcomes.[32] In the first phase (HMP1), researchers collected and analyzed samples from 242 healthy adults across five major body sites (oral, skin, gut, airway, and vagina) at three time points for 16S rRNA-based sequencing and shotgun sequencing.[33] From their initial report of 5,177 16S rRNA sequencing profiles and 681 shotgun sequencing profiles, a major finding of this study was that the taxonomic profile of a subject did not always correlate with host phenotype. In the next phase of the project, the integrative HMP (iHMP or HMP2) aimed to expand sample collection and scope to address the findings from HMP1. HMP2 included additional sample types (blood, urine) from healthy, diseased (IBDs such as ulcerative colitis and Crohn's disease), pregnancy and preterm birth, and a prediabetic cohort. These cohorts were analyzed by several techniques including metagenomics, metatranscriptomics, metaproteomics, metabolomics, virome profiling, antibody profiling, host genome profiling, epigenetic profiling, and cytokine profiling. Many of these strategies such as antibody, cytokine, and epigenetic profiling were added to explore the impact of microbes on the host while strategies such as host genome profiling were added to elucidate if subjects had a predisposition to a particular disease state based on host genomic variation.

The HMP laid a roadmap for large-scale human microbiome studies. In addition to their biological findings, the study highlighted processes such as data repository development as key to building knowledge within the microbiome community (HMP data portal, https://www.hmpdacc.org/). Public databases for platforms continue to grow with the collective goal of increasing data availability for future data mining (e.g., MicrobiomeDB, Metabolomics Workbench, GNPS, MetaboLights, NCBI, INSDC).

Clinical Considerations

Sample Collection for Microbiome Analysis

To evaluate the role of host microbes, patient material such as intestinal content, urine, tissue, and blood (plasma, serum, blood spot cards) are analyzed by a growing platform of techniques. In all studies, there are critical steps throughout the pipeline that should be considered such as (1) Pre-sample collection planning, (2) Sample collection, (3) Sample processing, and (4) Sample submission and data analysis (►**Fig. 1**). Although this review cannot cover the best practices for acquisition, processing, and storage of all sample types across all platforms, there are several reviews and

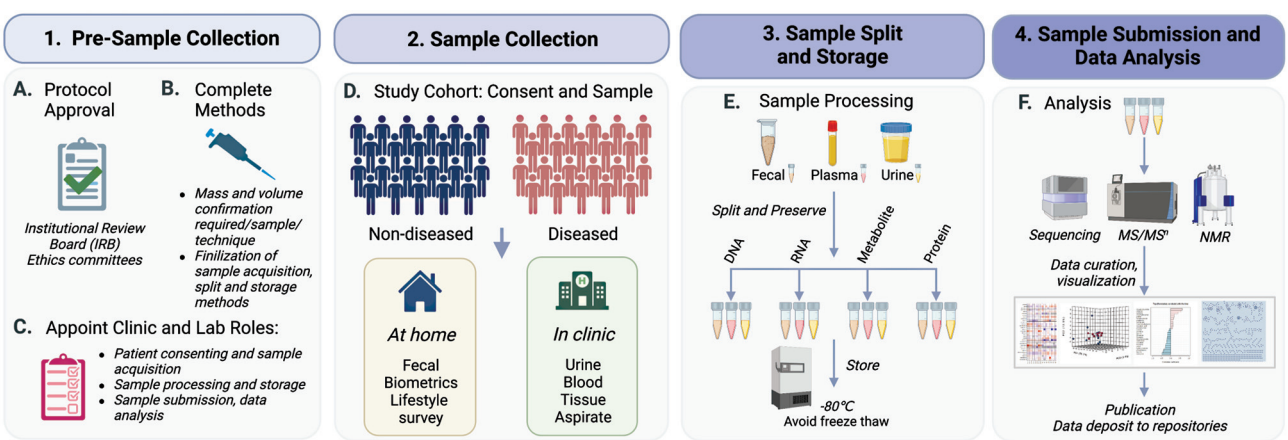

Fig. 1 *General pipeline for clinical human microbiome study.* Within a human microbiome study, pre-sample collection steps such as **(A)** protocol approval, **(B)** completion of project methods, and **(C)** appointment of necessary laboratory roles will need to be finalized prior to sample collection. **(D)** Sample collection will occur at home and in the clinic. **(E)** Once samples have been collected, they should be immediately split and preserved according to each downstream analyses. **(F)** Samples will be submitted for a variety of analysis across different platforms such as genome sequencing, mass spectrometry (MS), multistage fragmentation (MS^n), and nuclear magnetic resonance (NMR). Data will be curated and visualized prior to publication and deposit of data into data repositories (Created with BioRender.com).

works covering these topics.[14,21,34–36] Additionally, there is a body of work for analyzing the impact of the GI microbiome on the host that cannot be covered in this review such as murine studies, single cell RNAseq, cytokine profiling, antibody profiling, virome profiling, tissue histology, and organoid generation and testing.

Patient Consent, Sample Acquisition, and Data Analysis Considerations

At the start of a study, it is essential to plan for sample collection approval if no prior institutional protocol has been approved for the work. Initially, investigators are required to apply for Institutional Review Board or ethical committee approval for all protocols (►**Fig. 1A**). This process can take several months. General topics in the approval documentation include: a description of the specimens needed (i.e., fecal, urine, blood), description of the project in both lay and technical language, indication if these collections impose additional risk to the subjects, report how many subjects will be needed to complete study, report who will be recruited to your study and why, report what analyses will be completed with the samples, and how will data be stored and protected.

Prior to sample collection, the required downstream analyses for each sample type should be established to ensure adequate mass or volume acquisition and proper storage conditions are met (►**Fig. 1B**). To avoid poor data quality, investigators should collaborate and consult with the academic, clinical, or commercial laboratories that will be completing analyses on all sample types prior to sample acquisition. To oversee study processes, a clinical study coordinator, a sample processing lead, and a data analysis lead are recommended to be appointed as these roles are essential (►**Fig. 1C**). Microbiome experiments require sample splitting and analysis-dependent preservation methods, and it is critical to have a team member familiar with all study needs. Although many sites with routine collections have clinical coordinator teams that cover a project as needed, those without teams will need to collaborate with academic laboratories (if institution permits laboratory members to be trained for patient consent and sample collection in clinic) or they will need to hire specifically for the project.

Often, a sample will be analyzed by several methods, and it should be split and preserved for each method prior to storage (►**Fig. 1E**). For example, one fecal sample that requires 16S rRNA-based sequencing and shotgun sequencing from DNA, metatranscriptomics from RNA, metabolite quantitation, and cultivar isolation should be split and stored with four different methods, each with their own mass requirement. It is recommended that the collection and storage methods for each specimen are consistent. Additionally, a sample should remain in the optimal storage condition until analysis and freeze-thaw cycles should be avoided. If frozen samples must be shipped, they should be stored on dry ice and shipped overnight to the destination. Data curation and analysis are significant steps for all techniques (►**Fig. 1F**). For many data types, quality control evaluation, data interrogation, visualization, and presentation are not trivial and will require prior knowledge and training. Data analysis is routinely included or offered for an additional cost from academic and commercial laboratories. The sharing of curated data and subject metadata to public data repositories is encouraged following publication. Integrating common efforts such as the ability to mine previously acquired data for inter/intra-institutional benefit will increase the quality, number, and frequency of foundational and field progressing studies.

Outpatient Collection

Although many samples can be acquired while the subject is inpatient or onsite for an appointment, some samples such as stool and metadata such as lifestyle surveys and biometric information can be collected by the subject prior to the visit (►**Fig. 1D**). Lifestyle surveys can capture valuable information such as diet and exercise levels as well as biometric readouts (e.g., heart rate, sleeping patterns) from smart devices. For example, it has been shown that vegans have overall lower levels of primary and secondary bile acids.[37] For stool, at home collections the number of samples should be increased since not all subjects can or prefer not to deposit at the time of their appointment. There is not a standard kit for fecal collection, however, many institutions send home-made kits that include an instructional pamphlet, toilet collection vessel and scoop, and a sterilization product for cleaning hands after deposit. Depending on the analyses needed, samples can either be frozen in the subject's freezer (approximately −20°C), sent back to the institution by courier for highly sensitive material (i.e., RNA), or stored at room temperature with preservation solutions (ex. ethanol, RNA-later, Omni-met GUT or Omni-gene GUT kits).[38] If the subject stores their sample, they will bring it on the day of their visit and upon arrival the collection team will immediately process.

Institutional Sample Bank Protocols

In addition to individual project protocols, programs with potential future microbiome studies or those with multiple principal investigators consenting the same patient population, the creation of a sample bank might be of interest. In this way, the protocol can serve as an "umbrella" for the collection of medical records and biospecimens for the purpose of current and future research. In a gastroenterology clinic, for example, patients who are admitted for colonoscopies are consented and material such as luminal aspirate, colonic biopsies, fecal material, urine, and blood can be collected for the bank during their visit. Additionally, a bank of healthy or non-diseased donors is a challenging cohort to capture for many studies. Sites have found success in consenting and acquiring biospecimens from these non-diseased donors and making the material available to other investigators at their site. Researchers can apply to analyze banked samples based on their project goals. In this way, institutions can consolidate research efforts, normalize collection and storage methods and provide a greater number of patients to all studies, therefore increasing statistical significance and rate of study completion. Outside of institutions, there are foundations such as the Crohn's and Colitis

Foundation who have generated their own intra-institutional bank (IBD plexus) of samples that are available upon application approval.

The Compelling Potential of Clinical Microbiome Studies

In the era of robust data analysis pipelines, advanced instrumentation, and novel technologies for a high-resolution view of the microbiome, human-based projects remain largely descriptive, and most studies will require in vitro, in vivo, and follow-up studies to mechanistically define results. Because of this, projects require rigorous study design, collection plans, and data analysis pipelines to ensure high data quality. Despite these inherent yet surmountable hurdles, the future of population multi-omic analyses holds invaluable insight into host–microbe interactions for disease prevention, therapy, and treatment.

Conflict of Interest
None declared.

Acknowledgments
None.

References

1 Macconkey A. Lactose-fermenting bacteria in faeces. J Hyg (Lond) 1905;5(03):333–379
2 Hall IC. A constricted tube with mechanical seal for anaerobic fermentation tests. J Infect Dis 1921;29(04):317–320
3 Sebald M, Hauser D. Pasteur, oxygen and the anaerobes revisited. Anaerobe 1995;1(01):11–16
4 Behring E. Ueber das Zustandekommen der Diphtherie-Immunität und der Tetanus-Immunität bei Thieren. Dtsch Med Wochenschr 1890;16(49):1113–1114
5 Fleischmann RD, Adams MD, White O, et al. Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. Science 1995;269(5223):496–512
6 Fraser CM, Gocayne JD, White O, et al. The minimal gene complement of Mycoplasma genitalium. Science 1995;270 (5235):397–403
7 Handelsman J, Rondon MR, Brady SF, Clardy J, Goodman RM. Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. Chem Biol 1998;5 (10):R245–R249
8 Tyson GW, Chapman J, Hugenholtz P, et al. Community structure and metabolism through reconstruction of microbial genomes from the environment. Nature 2004;428(6978):37–43
9 Cardona S, Eck A, Cassellas M, et al. Storage conditions of intestinal microbiota matter in metagenomic analysis. BMC Microbiol 2012;12(01):158
10 Santiago A, Panda S, Mengels G, et al. Processing faecal samples: a step forward for standards in microbial community analysis. BMC Microbiol 2014;14:112
11 Quince C, Walker AW, Simpson JT, Loman NJ, Segata N. Shotgun metagenomics, from sampling to analysis. Nat Biotechnol 2017; 35(09):833–844
12 Wesolowska-Andersen A, Bahl MI, Carvalho V, et al. Choice of bacterial DNA extraction method from fecal material influences community structure as evaluated by metagenomic analysis. Microbiome 2014;2:19
13 Szóstak N, Szymanek A, Havránek J, et al. The standardisation of the approach to metagenomic human gut analysis: from sample collection to microbiome profiling. Sci Rep 2022;12 (01):8470
14 Chen LX, Anantharaman K, Shaiber A, Eren AM, Banfield JF. Accurate and complete genomes from metagenomes. Genome Res 2020;30(03):315–333
15 Eren AM, Kiefl E, Shaiber A, et al. Community-led, integrated, reproducible multi-omics with anvi'o. Nat Microbiol 2021;6(01): 3–6
16 Yaung SJ, Deng L, Li N, et al. Improving microbial fitness in the mammalian gut by in vivo temporal functional metagenomics. Mol Syst Biol 2015;11(03):788
17 Shaiber A, Willis AD, Delmont TO, et al. Functional and genetic markers of niche partitioning among enigmatic members of the human oral microbiome. Genome Biol 2020;21(01):292
18 Pasolli E, Asnicar F, Manara S, et al. Extensive unexplored human microbiome diversity revealed by over 150,000 genomes from metagenomes spanning age, geography, and lifestyle. Cell 2019; 176(03):649–662.e20
19 Kong C, Liang L, Liu G, et al. Integrated metagenomic and metabolomic analysis reveals distinct gut-microbiome-derived phenotypes in early-onset colorectal cancer. Gut 2022;(August):gutjnl-2022-327156
20 Nagata N, Nishijima S, Kojima Y, et al. Metagenomic identification of microbial signatures predicting pancreatic cancer from a multinational study. Gastroenterology 2022;163(01):222–238
21 Reck M, Tomasch J, Deng Z, Jarek M, Husemann P, Wagner-Döbler ICOMBACTE Consortium. Stool metatranscriptomics: a technical guideline for mRNA stabilisation and isolation. BMC Genomics 2015;16(01):494
22 Singh A, Vats S, Bhargava P. Advances and Challenges in Metatranscriptomic Analysis. Microbial Metatranscriptomics Belowground. 2021:453–469
23 Schirmer M, Franzosa EA, Lloyd-Price J, et al. Dynamics of metatranscription in the inflammatory bowel disease gut microbiome. Nat Microbiol 2018;3(03):337–346
24 Tanca A, Palomba A, Pisanu S, et al. A straightforward and efficient analytical pipeline for metaproteome characterization. Microbiome 2014;2(01):49
25 Tanca A, Abbondio M, Fiorito G, et al. Metaproteomic profile of the colonic luminal microbiota from patients with colon cancer. Front Microbiol 2022;13:869523
26 Long S, Yang Y, Shen C, et al. Metaproteomics characterizes human gut microbiome function in colorectal cancer. NPJ Biofilms Microbiomes 2020;6(01):14
27 Allen BH, Gupta N, Edirisinghe JN, Faria JP, Henry CS. Application of the metabolic modeling pipeline in KBase to categorize reactions, predict essential genes, and predict pathways in an isolate genome. Methods Mol Biol 2022;2349:291–320
28 Dührkop K, Fleischauer M, Ludwig M, et al. SIRIUS 4: a rapid tool for turning tandem mass spectra into metabolite structure information. Nat Methods 2019;16(04):299–302
29 Wang M, Carver JJ, Phelan VV, et al. Sharing and community curation of mass spectrometry data with global natural products social molecular networking. Nat Biotechnol 2016;34(08): 828–837
30 Quinn RA, Melnik AV, Vrbanac A, et al. Global chemical effects of the microbiome include new bile-acid conjugations. Nature 2020; 579(7797):123–129
31 Gumpenberger T, Brezina S, Keski-Rahkonen P, et al. Untargeted metabolomics reveals major differences in the plasma metabolome between colorectal cancer and colorectal adenomas. Metabolites 2021;11(02):119
32 Integrative HMP (iHMP) Research Network Consortium. The integrative human microbiome project. Nature 2019;569 (7758):641–648
33 Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. Nature 2012;486 (7402):207–214

34 Wang Z, Zolnik CP, Qiu Y, et al. Comparison of fecal collection methods for microbiome and metabolomics studies. Front Cell Infect Microbiol 2018;8:301
35 Costea PI, Zeller G, Sunagawa S, et al. Towards standards for human fecal sample processing in metagenomic studies. Nat Biotechnol 2017;35(11):1069–1076
36 Salvato F, Hettich RL, Kleiner M. Five key aspects of metaproteomics as a tool to understand functional interactions in host-associated microbiomes. PLoS Pathog 2021;17(02):e1009245
37 Trefflich I, Marschall HU, Giuseppe RD, et al. Associations between dietary patterns and bile acids-results from a cross-sectional study in vegans and omnivores. Nutrients 2019;12 (01):47
38 Ramamoorthy S, Levy S, Mohamed M, et al. An ambient-temperature storage and stabilization device performs comparably to flash-frozen collection for stool metabolomics in infants. BMC Microbiol 2021;21(01):59