5-2024

Current Trends and Challenges of Microbiome Research in Prostate Cancer

Shaun Trecarten

Bernard Fongang
University of Texas Health Science Center at San Antonio

Michael Liss

Recommended Citation

Trecarten, S., Fongang, B., & Liss, M. (2024). Current Trends and Challenges of Microbiome Research in Prostate Cancer. Current oncology reports, 26(5), 477–487. https://doi.org/10.1007/s11912-024-01520-x

Current Oncology Reports (2024) 26:477–487
https://doi.org/10.1007/s11912-024-01520-x

REVIEW

Current Trends and Challenges of Microbiome Research in Prostate Cancer

Shaun Trecarten[1] · Bernard Fongang[2,3,4] · Michael Liss[1]

Accepted: 18 March 2024 / Published online: 4 April 2024

Abstract

Purpose of Review The role of the gut microbiome in prostate cancer is an emerging area of research interest. However, no single causative organism has yet been identified. The goal of this paper is to examine the role of the microbiome in prostate cancer and summarize the challenges relating to methodology in specimen collection, sequencing technology, and interpretation of results.

Recent Findings Significant heterogeneity still exists in methodology for stool sampling/storage, preservative options, DNA extraction, and sequencing database selection/in silico processing. Debate persists over primer choice in amplicon sequencing as well as optimal methods for data normalization. Statistical methods for longitudinal microbiome analysis continue to undergo refinement.

Summary While standardization of methodology may help yield more consistent results for organism identification in prostate cancer, this is a difficult task due to considerable procedural variation at each step in the process. Further reproducibility and methodology research is required.

Keywords Gut microbiome · Prostate cancer · Methods · Challenges · Sample collection · Sequencing

Introduction

Cancer and the Microbiome

Cancer is often the result of multifactorial processes, including genetic predisposition and environmental/physiological factors. Recently, there has been increasing interest and research in the role of the human microbiome in cancer development. The human microbiome is composed of microorganisms, including bacteria, viruses, fungi, and protozoa, which are harbored externally (e.g., on our skin) and internally (e.g., oral cavity, genitourinary tract, and gastrointestinal tract) [1••]. When the physiologic composition of the microbiome is dysregulated and thought to contribute to disease processes, it is termed dysbiosis.

While there are standard organisms that colonize particular anatomical locations, there are some differences based on genetic predisposition, dietary, environmental exposures, and other individual factors [2•]. One of the leading hypotheses linking the microbiome to pathology is the concept of direct damage to an organ via inflammation or toxin exposures, and there is now a long-standing precedent for the direct involvement of microbes in cancer development (e.g., *Helicobacter pylori*, human papiloma virus (HPV). A more recent example from the gut microbiome is the potential role of genotoxins such as colibactin produced by bacteria harboring the polyketone synthase (pks) gene that can cause genetic insults in addition to the traditional reactive oxygen species generated from inflammation [3]. However, attributing causation continues to be elusive, especially with respect to general dysbiosis of the gut. General dysbiosis can lead to inflammation in the gut where wall integrity is impaired, increasing permeability to potentially damaging metabolites such as bacterial lipopolysaccharide or short-chain fatty acids, leading to systemic indirect effects [2•, 4•]. Indirect mechanisms have also been suggested with

✉ Michael Liss
liss@uthscsa.edu

1 Department of Urology, UT Health San Antonio, 7703 Floyd Curl Dr, San Antonio, TX 78229, USA

2 Glenn Biggs Institute for Alzheimer's & Neurodegenerative Diseases, UT Health San Antonio, San Antonio, TX, USA

3 Department of Biochemistry and Structural Biology, UT Health San Antonio, San Antonio, TX, USA

4 Department of Population Health Sciences, UT Health San Antonio, San Antonio, TX, USA

modulation of the immune system, and some specific bacteria have even been shown to slow tumor growth [5]. Furthermore, the gut microbiome has also been shown to have a significant impact on PD-1 targeted immunotherapy for cancer [4•].

Prostate Tissue Microbiome

The prostate microflora itself has been implicated as a direct mechanism, with certain pathogens being associated with prostate cancer. However, there is debate as to whether a prostate microflora truly exists, as some studies have shown that normal prostate tissue is unlikely to have commensal organisms, and elements of prostatic fluid (e.g., zinc and toll-like receptor 4) prevent colonization of pathogens [6, 7].

Nonetheless, *Propionobacterium acnes*, now renamed *Cutanibacterium acnes*, was demonstrated as the predominant pathogen in prostate specimens post radical prostatectomy in one study [8]. However, *Cutanibacterium acnes* has also been implicated in both benign pathology and as a contaminant on sequencing [9]. Another study ($N=30$) examined prostatectomy specimens sent for both bacterial culture and amplicon sequencing, revealing that while no organisms were grown, 83 distinct microorganisms were identified on sequencing [10]. Several studies have associated prostate cancer with some bacteria directly within the prostate, including but not limited to, *Escherichia coli* [11], *Staphylococci* [12], *H. pylori* [13], and *Mycoplasma genitalium* [14]. Some studies, however, have not shown any demonstratable difference between benign and malignant prostate microflora [13, 15].

Viruses have also been implicated in the development of prostate cancer, including HPV 16 and 18 as well as cytomegalovirus [13]. Despite heterogeneity in the literature, a recent meta-analysis of 27 case–control studies between 1991 and 2022 examined the link between prostate cancer and HPV, including 1607 patients with prostate cancer and 1515 control samples (317 normal tissues, 1198 benign prostatic hyperplasia (BPH)). The study reported significantly increased odds of developing prostate cancer with HPV infection (OR 3.07, 95% CI 1.80–5.21) compared to normal tissue [16•]. Furthermore, when the control group was patients with BPH, there was still an increased odds of developing prostate cancer (OR 1.94, 95% CI 1.43–2.63) [16•].

To date, no individual organism has consistently been shown to be the culprit link to prostate cancer, and studies examining the prostate flora are often limited due to contamination [1••].

Genitourinary Microbiome

The genitourinary microbiome, usually obtained from urine given its proximity to the prostate, has been examined as a possible contributor to prostate cancer, again with no definitively causative organisms identified. Traditionally thought to be sterile [17], microbial diversity has been demonstrated in the urinary tract, though with similar organisms to adjacent anatomical locations, including skin, vagina, and gastrointestinal tract [1••]. While most phyla identified include *Firmicutes*, *Bacteroidetes*, *Actinobacteria*, *Fusobacteria*, and *Proteobacteria*, there is still substantial variability at the genus level between individuals and sexes, and conflicting results seen in studies examining the role of genitourinary infection (urethritis, cystitis, and prostatitis) in prostate cancer [1••]. Another study examined the components of seminal fluid/urine and showed no significant differences in patients with prostate cancer compared to controls [4•]. To date, the role of the genitourinary microbiome in prostate cancer is still largely unknown, and ongoing research efforts are required to clarify if any meaningful association exists.

Gut Microbiome and the "Gut-Prostate" Axis

Special attention, however, is paid to the interaction between the gut microbiome and prostate cancer in a relationship termed the gut-prostate axis. With the highest microbe counts in the body, the gut microbiome has been implicated in regulating the tumor microenvironment, the intestinal-epithelial barrier, and the activity of lymphoid organs [18].

Differences in gut microbiome have been suggested in patients with prostate cancer. Studies have implicated numerous pathogens in the development of prostate cancer, including *Bacteroides massiliensis* (high risk prostate cancer patients vs controls), *Akkermansia muciniphila* and *Ruminococcaceae* (in patients on ADT), and *Streptococcus* and *Bacteroides* (in men with prostate cancer versus controls) [19–21]. *Rumminococcus* has also been associated as a predominant genus in patients with castrate-resistant prostate cancer [22, 23].

This gut-prostate microenvironment may further be altered by lifestyle habits, which then may predispose to prostate cancer. For example, high-fat diets (HFD) can damage gut wall integrity and induce subsequent systemic inflammation, which has been shown to have effects on the gut microbiome, circulating immune cells, and prostate cancer progression [4•].

While investigation of the gut microbiome is a developing research focus in prostate cancer, there are ongoing challenges with undertaking research in this field. With significant individual differences in the microbiome and heterogeneity between studies, understanding and optimizing the methodology is crucial in ensuring reproducibility and accurate results. This paper will focus on existing challenges with specimen collection and DNA extraction, sequencing techniques, and interpretation of results (Fig. 1).

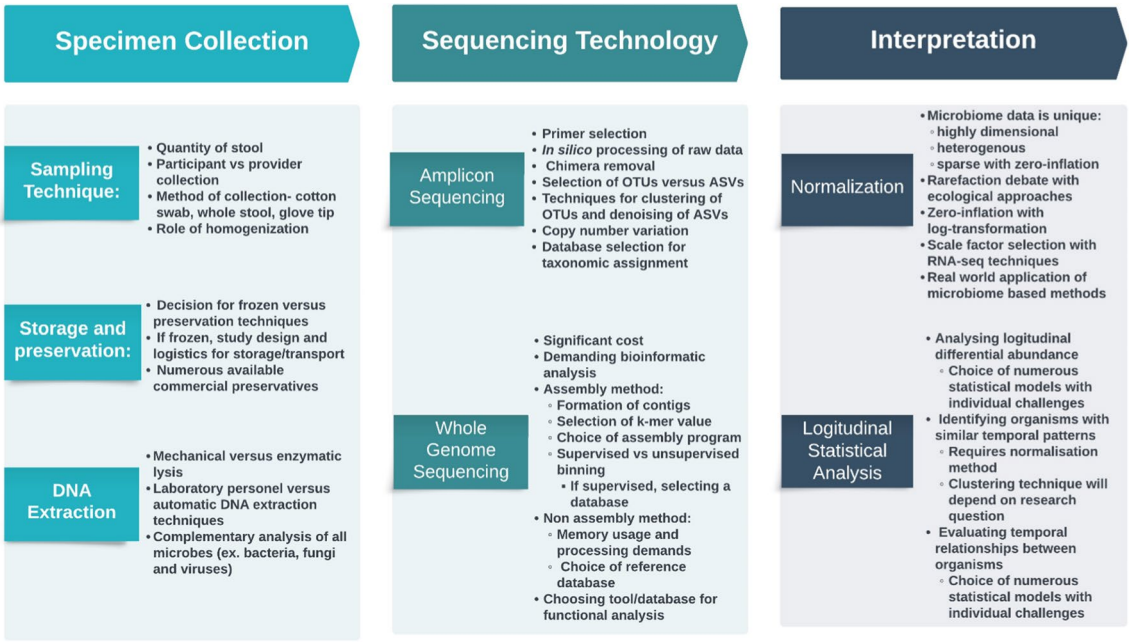

Fig. 1 Methodological challenges and considerations in gut microbiome research pertaining to specimen collection, sequencing technology, and interpretation of results

Specimen Collection

Challenges exist in collecting and processing samples due to considerable variation in methodology, including sampling technique, storage/preservation, and DNA extraction.

Sampling Technique

Fecal Sampling For fecal analysis, the technique for stool collection is often not specified, and if it is reported, there is no standardized approach. The quantity of sampled stool has also not been standardized, though some studies have described using 0.2–0.25 g of stool as optimal [24, 25]. Other collection methods include participant defecation into plastic toilet liner collection bags, with patients themselves then subsampling the specimen into two vials [26]. Efforts have been undertaken to improve collection strategies to improve efficiency of longitudinal studies requiring multiple specimens. Some have suggested sufficient biomass can be obtained by using a cotton swab on a used piece of toilet paper. However, the amount of material provided limits the study to only a few reactions and is normally satisfactory for 16S ribosomal gene-sequencing studies [27]. If further functional studies are required, then generally, larger pieces of stool are necessary and must be frozen immediately [27]. A recent study compared microbiome sampling via rectal swab, glove tip after digital rectal examination (DRE), and participant-collected stool samples in 22 men, demonstrating no difference in microbiological beta-diversity ($p > 0.05$) between glove tip and rectal swab specimens [28•]. The glove tip collection method was also generally similar to the home-based stool collection [28•].

There are also potential differences in microbiological composition within the stool since the surface rather than the core of a piece of stool is in close contact with the intestinal lumen and may be more susceptible to environmental influences (e.g., low oxygenation levels) [29]. To counteract differences in the distribution of bacteria based on location, many studies homogenize the sample [30]. One study compared microbiological composition at different locations within a flash-frozen stool specimen versus a homogenized sample and demonstrated minimal variability in microbiological abundance and diversity [31•]. However, there were differences in the metabolic profiles across sampling regions [31•]. No consensus exists currently regarding the necessity of homogenization, and within-sample variation may occur in aliquoting methods [30].

Urine Sampling Urine has been an intriguing biospecimen for prostate cancer detection. Traditionally, urine collected after a DRE could increase the contribution of prostatic fluid into the specimen. For microbiome research however, there remains the question of contamination from the urethra, meatal skin, and foreskin (if present) [32]. Some investigators try to overcome this with catheterization, yet the technique still suffers from skin and urethral contamination and is invasive, and is less likely to be translatable to clinical care [32]. Another consideration is that the

microbial biological load is low in the urine [32]. In historical biobanks, there tends to be some urine processing for pelleting, washing, and smaller volume storage. Preservatives in this space can widely vary if used, making the use of previously collected specimens for other uses a challenge [32]. Therefore, our group has used a dedicated microbiome processing protocol prospectively using 30-mL target volumes.

Tissue Sampling Significant advances in sequencing techniques allow for the investigation of the microbiome directly in the study tissue. A major concern with biopsy tissue and archival prostatectomy tissue is that these often occurred after a transrectal prostate biopsy, likely causing contamination from biopsy needle traversing the rectal wall carrying rectal flora into the prostate [33]. Moreover, any antibiotic used prior to either the prostate biopsy or prostatectomy will change the abundance of microbes. The difference between tumor and normal tissue should be compared in these scenarios; however, antibiotics may still impact composition of the overall microbiomes.

Storage and Preservation

Along with different stool collection methods, there are also numerous techniques to try and preserve the quality of the microbiome for testing, namely, immediate freezing or using preservatives [34]. The techniques used should be tailored to the specific question, especially if metabolomic evaluation is being investigated. For fresh stool samples, general principles include avoiding freeze–thaw cycles, and temperature fluctuations, and minimizing transport time [30]. As soon as feasible, collected stool should be transported to the laboratory within 4 h and frozen at − 20 °C (adequate for a few months) to − 80 °C (ideal for long-term storage) [30]. During transportation, stool should be placed in 4 °C cold storage, during which a 24–48-h window exists before arrival to the laboratory before meaningful changes in microbiological composition [30].

While immediate freezing is the gold standard, it is not always a viable solution based on study design (e.g., field research), logistical issues with transport, and associated costs. Consequently, many preservative solutions have been tested [35]. Numerous commercial preservation solutions have been tested including RNAlater™, preservation buffer (PB), 70% ethanol, 95% ethanol, fecal occult blood test (FOBT), fecal immunochemical test (FIT), dimethyl sulfoxide-ethylenediaminetetraacetic acid solution (DETA), DETA-NaCl, ethylenediaminetetraacetic acid (EDTA), PSP (Invitek) buffer™, DNA/RNA shield™, and OMNIgene™ among others [36••]. A recent meta-analysis on gut microbiome methodology examining preservation techniques demonstrated that out of 30 selected studies, only two had consistent sample preservation methods [36••]. While larger sample sizes are required, the article suggested acceptable outcomes with RNAlater™ (storage for 1 month only), preservation buffer, OMNIgene-Gut™, and FOBT cards stored at room temperature [36••].

DNA Extraction

While standardization in microbiome studies is lacking, DNA extraction has been identified as contributing to the most variability of results by the Microbiome Quality Control Project (MBQC) [37] and the International Human Microbiome Standards (IHMS) group [38]. Potential reasons for this variability include possible reagent contamination, mechanical versus enzymatic lysis techniques, differences between laboratory personnel, and automation of DNA extraction [39]. While different protocols for DNA extraction from fecal samples exist, with publications by groups including the Human Microbiome Project (HMP) [40], MetaHIT [41], and the Earth Microbiome Project [42], standardization is not an easy undertaking. Challenges with standardization exist as some protocols utilize automation of DNA extraction for larger sample sizes and lack of complete understanding of the true microbial diversity within a fecal sample [39].

Most protocols focus on DNA extraction techniques for subsequent bacterial sequencing, but a further challenge exists for DNA extraction of non-bacterial microbes. Protocols for fungal or viral DNA extraction methods isolate the non-bacterial microbe by removing contamination of human/bacterial cells [43, 44]. Consequently, complementary analysis of all microbes in a fecal specimen is difficult.

Sequencing Technologies

Currently, most microbiome research involves the measurement of microbiological ecology using next-generation sequencing techniques, namely, metatranscriptomics, via amplicon sequencing, and metagenomics through whole-genome/shotgun sequencing (Table 1). Both sequencing approaches also have "in silico" components, which refer to procedural steps conducted by computer processing/modeling.

Amplicon Sequencing

Amplicon sequencing generally uses variable regions of the bacterial 16S rRNA, but can occasionally use the 18S or internal transcribed spacer (ITS) component [45]. There are nine variable regions (V1–V9) within the bacterial 16S rRNA, which are next to highly conserved genes targeted for amplification with primers [45]. Using primers, these variable regions can

Table 1 Comparison of amplicon sequencing versus whole-genome/shotgun sequencing

	Amplicon sequencing	Whole-genome sequencing
Sequencing target	Region of ITS or subunit of 16S or 18S ribosome	DNA of both host and microbiome
Taxonomic detail	Phylum → genus	Species → strains
Cost	+	+ + +
Specificity	+ + +—improved due to plethora of reference databases	+—increased risk of host contamination
Functional analysis possible	Indirectly—prediction tools are available	Yes
Raw data organization	OTUs and ASVs	Assembly (via binning of contigs) and non-assembly approaches

be targeted for subsequent polymerase chain reaction (PCR). However, differences in which variable region is targeted can lead to alterations in sequencing and, consequently, different taxonomic outcomes [46].

Primer Selection

Primer selection remains a controversial topic in amplicon sequencing. With improvements in second-generation sequencers, sequencing up to approximately 600 base pairs was made possible, allowing for targeting one to three variable regions at a time [47]. Most commonly V1-V2/V3, V3-V4/V5, and V4 are used, though there are different results in taxonomic classification depending on which variable region is used [47]. Initially, V1-V2 sequencing was utilized, though as the Illumina protocol transitioned to V3-V4, analysis of this region became more common and is also felt to be the most cost-effective variable region [48]. Recent studies, however, have suggested that while the use of V3-4 primers is associated with amplification artifact, they are better suited for gut microbiome analysis due to improved detection of *Bifidobacteriales* [49, 50]. However, a contemporary modification to a V1 primer demonstrated improved *Bifidobacterium* detection compared to V3-4 [51]. Furthermore, a recent study using a modified V1-2 primer also produced more desirable analysis compared to V3-V4 data when comparing similarity to actual gut bacteria abundance [52•].

In Silico Processing of Raw Data

Sequencing produces raw data as multiple reads, which are processed to form a consensus output with an associated error rate [53]. There is a tradeoff between read length, which improves microbiological classification, and error rate, which often requires sophisticated software to remove artefacts [48]. To improve the quality of sequencing data, denoising algorithms and sequence curation strategies need to be applied in silico, which initially include removal of ambiguous bases and homopolymers larger than 8 nucleotides [54, 55]. Reads with anomalous lengths and below certain technology-specific quality control metrics are also removed [53, 54]. A further source of error is chimeras, which are sequences composed of two or more parents. This can lead to misinterpretations, such as the description of non-existent bacteria or confusing ecological diversity measurements. Chimeras likely occur from errors in PCR, where amplicons, which prematurely terminate, fuse to another homologous template. Detecting and eliminating chimeras is a challenging undertaking, and multiple algorithms exist for this purpose [48].

OTUs Versus ASVs

Raw sequencing data, prior to statistical analysis, has to be assigned to a particular organism, which can be accomplished via two approaches: operational taxonomic units (OTUs) and amplicon sequence variants (ASVs) [56]. OTU-based methods organize sequences together based on similarity (distance matrix among produced sequences) without initially relying on reference databases. OTUs with sequence similarities of > 97%, 95%, and 80% are typical definitions for species, genus, and phylum, respectively [56]. ASVs alternatively use exact nucleotide sequences and work on the basis that more frequently observed sequences are likely the result of true biological results rather than error [57]. Consequently, interpreting ASVs must be done for an entire sample rather than individual reads. Using denoising algorithms, erroneous sequences with sometimes as little as one nucleotide error are removed, and the remaining sequences are then compared to a reference for taxonomic assignment [58]. Since exact sequences are used, ASVs can be better compared across studies, though they are less optimal if the study design requires genomic heterogeneity.

Once the sequences are clustered (for OTUs) or denoised (for ASVs), they are provided with a taxonomic assignment, often with comparison to a database, examples of which include Greengenes [59], SILVA [60], Ribosomal Database Project (RDB) [61], and National Centre for Biotechnological Information (NCBI) BLAST Database [62]. Some authors recommend utilizing SILVA and RDB, since Greengenes has not been updated since May 2013 [63].

These databases are often incorporated into common pipelines for amplicon analysis including QIIME2, mothur, and RDP Classifier [64]. Despite substantial efforts to assign taxonomies effectively, the misclassification rate overall is quoted at approximately 16–20% [65].

Copy Number Variation

Another consideration for amplicon sequencing is a phenomenon known as copy number variation (CNV), where some species have more than one (sometimes up to 15) copy of the 16S rRNA [66]. This can lead to misinterpretation of abundance levels after taxonomic assignment and miscalculations in diversity assessments. To counteract this, read counts can be weighted based on gene numbers if they are known for a specific organism [48]. Otherwise, potential solutions include using reference databases or values of a closely related organism, which is not optimal for studying rarer organisms [67].

Whole Gene Sequencing

Whole genome sequencing involves analyses of all of the DNA data within the microbiome, also known as "shotgun" sequencing as no individual area is targeted. Rather than using primers, DNA is randomly fragmented after extraction, with barcodes and adapters ligated to the end of each segment to facilitate identification and subsequent sequencing [56]. With the vast amount of data collected, subspecies strain level resolution can be provided [68]. While it can also provide information regarding other non-bacterial organisms, limitations include high cost and more demanding bioinformatic analysis [69, 70]. Furthermore, since the entire genome is analyzed, a more robust functional analysis can be obtained compared to amplicon sequencing, where it can only be predicted indirectly with help from reference databases [68].

Assembly and Binning

After the raw reads are cleaned (similarly to amplicon sequencing), sequences can be assembled to form "contigs," which are longer (contiguous) sequences [64]. This process of forming contigs occurs in silico and is performed either de novo or using a reference genome. The assembly process is often based on dividing reads into a certain length (k) of nucleotides (k-mers) and reasoning the final sequence based on overlapping sequences [71]. The optimal value of k can vary based on estimated genome size and rate of heterozygosity, and, fortunately, numerous tools exist that can help the researcher with k-mer selection. Once the k-mer value is selected, a further challenge is deciding which assembly program to use, as some, for example, are more attractive if diversity capture is prioritized over contig length with limited computational resources (MEGAHIT), or if priority is to obtain large diversity regardless of complexity (metaSPAdes) [64].

Once contigs are assembled, they are organized into groups (bins) and classified taxonomically, in a process known as binning. This can be performed using a supervised (i.e., with a reference database) or unsupervised method, also known as taxonomy-dependent and taxonomy-independent, respectively. Using a database has limitations such as dependency on finite number of previously sequenced genomes and long computing times [64]. Alternatively, unsupervised methods are not dependent on a database and often combine analysis of nucleotide sequences and relative abundance to optimize binning [72]. With successful binning, each bin can be further analyzed and reassembled to form longer contigs [73].

After binning, if a functional analysis is required, the next step involves identifying genes and regulatory elements, a process known as annotation or gene-calling. As with other steps in WGS, there are a plethora of tools, some better depending on sequence length [74], computational requirements [75], and error rate [76]. Often, a combination of tools for this step is used [77]. With the genes identified, functional annotation can often be carried out utilizing a database, again challenged by a wide array of choices [64]. Depending on the research question, there may be a specific database, for example, when analyzing antibiotic resistance genes (e.g., CARD) [78]. With different tools required at each step of the process of WGS, efforts have been made to streamline analysis using pipelines, which can alleviate some of the complexity at the expense of potential oversimplification of answers to specific research questions [64]. Taxonomic profiling can be performed using many tools, including the MAGy pipeline [79] or DESMAN pipeline [80] if strain level analysis is required.

Non-Assembly Approach

Another approach to WGS involves avoiding assembly altogether, with analysis of raw data for taxonomic and functional assignment. Taxonomic classification can be accomplished using a reference database at the expense of slow processing, though processing times and memory usage requirements have been improved recently [64]. Functional analysis of raw data can also be performed using a host of different tools, for example, Carnelian [81] for comparative functional assessment and the Shot-MAP pipeline if user flexibility in analysis is desired [82].

Interpretation

Normalization

Microbiome data is unique to other forms of data in that it is heterogenous, highly dimensional, sparse with zero-inflation (excess zeros in particular fields), and grouped into taxonomic classifications [83]. As such, data requires normalization prior to statistical analysis, which can be grouped into four categories: ecology-based methods, traditional methods, RNA-sequencing–based methods, and microbiome-based methods. Each category has associated pros and cons (Table 2).

Data generated via amplicon sequencing or WGS is often presented in a taxa abundance table, which records each taxonomic unit in each sample [84]. An essential part of this table is the "sequencing depth," which is the computed column sum of the sequence reads. While one would expect sequence depth to be similar throughout individual samples, there is often significant variation in efficiency of the sequencing process and loading concentrations/volumes [85, 86]. Ecology-based methods involve a process known as rarefaction, a term derived from physics where it refers to reducing density. Many programs accomplish this by randomly subsampling the derived data to a common depth in order to improve comparisons to other individual samples [87]. Challenges include deciding the appropriate sequencing depth to balance the dataset without losing valuable information gained from the analysis [88]. Furthermore, rarefication as a method of normalizing remains a topic of debate, with some experts suggesting the practice is statistically "inadmissible," as it omits valid data [89].

Traditional methods, on the other hand, use proportions of gene abundances (also called total sum scaling) to normalize data, rather than subsampling [90]. While appropriate for community-level ecological differences, limitations include its inability to adequately account for outliers and compositionality, precluding satisfactory differential abundance detection [91, 92]. Log transformations have also been used to correct for skewness of data and heteroscedasticity but are challenged when dealing with zero inflation and when standard deviations are considerably large [83].

RNA-sequencing–based normalization methods also scale counts using different scale factors and assume taxa are not differentially abundant [83]. Many scale factors may be used, including quantiles, median values, and log upper quartiles, and an extensive review on the topic has been written by Xia et al. in 2023 [83]. RNA seq-based methods are thought to outperform traditional and rarefaction methods of normalization [89, 91].

Microbiome-based approaches can incorporate a hybrid approach of methods and can mitigate effects of compositionality, zero inflation, and over-dispersion [83]. While one study in machine learning showed a hybrid approach can improve performance in classification, it may change the dataset enough to the point where real-world application becomes challenging [93]. Another review found a combination of compositionality and zero inflation methods demonstrated superior performance compared to RNA sequencing–based methods [94•].

Statistical Analysis

While taxonomic or metabolomic data can be observed at one cross-sectional point, a specific challenge in microbiome research is that data is often collected longitudinally [95]. A thorough review of statistical models for longitudinal analysis and associated challenges was published by Kodikara et al. in 2022 and divided challenges into three categories: (1) analyzing longitudinal differential abundance, (2) identifying organisms with similar temporal patterns, and (3) identifying temporal relationships between organisms [96••].

While numerous models exist for longitudinal differential abundance analysis, each has its own limitations, with some models providing only univariate analysis (e.g., zero-inflated beta regression [ZIBR] [97] and Gaussian

Table 2 Pros and cons of methods for microbiome data normalization

	Pros	Cons
Ecology-based	• Rarefaction can provide measures of species richness • Methods available in most computational packages	• Variable subsampling depth to balance information loss and dataset balance • Rarefaction omits potentially valid data
Traditional	• Avoids subsampling by using gene abundances • Log transformations can account for heteroscedasticity	• Challenged when dealing with outliers or compositionality • Log transformations challenged with especially large standard deviations and zero inflation
RNA-sequencing based	• Outperforms ecology-based and traditional approaches	• Wide selection of scale factors
Microbiome-based	• Incorporates a hybrid approach, mitigating compositionality, zero inflation, and over dispersion • Can outperform RNA-sequencing based methods	• With degree of dataset processing required, may hinder real world application

Mixed Model [ZIBMM] [98]), unable to handle missing data (e.g., ZIBR[97]) or zero inflation (e.g., negative binomial mixed model [NBMM] [96••]), or failing to account for compositionality (e.g., SplinectomeR). Bayesian Sparse Multivariate regression, on the other hand, performs multivariable analysis and can account for sparse data [99].

Clustering models can be used to identify microorganisms with similar temporal patterns [96••]. Examples include dynamic time warping (DTW) distances, partitioning around medoids (PAM) and agglomerative clustering, and clustering using principal component analysis (PCA) or sparse principal component analysis (sPCA). An essential aspect of these methods is data normalization, and the exact clustering method applied may differ depending on the research question [96••].

A few methods can be used to understand temporal relationships between taxa, though each has associated challenges. For example, the two-stage dynamic Bayesian network (TS-DBN) is limited to only two-time intervals, may not perform well for rare taxa, and may result in over-fitting when combining clinical information with small sample sizes [100]. Other methods, including Granger Lasso Causality and Microbial Time-series Prior Lasso (MTP Lasso) do not take clinical or demographic variables into account [96••]. Furthermore, MTP Lasso requires using biological information from existing literature or previous datasets for regression analysis [96••]. Statistical longitudinal analysis of the microbiome is a challenging task, and methods are continuously being refined.

Conclusion

The role of the gut-prostate axis in the development of prostate cancer is an exciting area of research, with significant hurdles to overcome. These challenges apply to any form of microbiome research, but for prostate cancer, it means that consistent identification of a causative organism has not yet been achieved. Standardization of methodology remains difficult, with heterogeneity at each step, from sample collection, DNA extraction, and sequencing to in silico processing and interpretation of results. Extensive knowledge and expertise are required to balance the clinical, ecological, and bioinformatic demands, and a multi-disciplinary approach is essential.

Author Contributions ST wrote the initial draft of the manuscript and created the graphs/tables. Drs B.F. and M.L. were also integral in guiding the direction of the review, and contributing to the manuscript itself and making necessary edits. All authors reviewed the manuscript and approved its final version.

Data Availability No datasets were generated or analysed during the current study.

Declarations

Conflict of Interest Drs. Shaun Trecarten and Bernard Fongang declare no conflict of interest. Dr. Michael Liss has founded a microbiome startup Oncobiomix, which did not provide any funding for this manuscript. There is no reference to Oncoboiomix, or its products, in the text.

Human and Animal Rights and Informed Consent This article contains two references [21, 33] to previous studies performed by the authors.

References

Papers of particular interest, published recently, have been highlighted as:
- • Of importance
- •• Of major importance

1.•• Kustrimovic N, Bombelli R, Baci D, Mortara L. Microbiome and prostate cancer: a novel target for prevention and treatment. Int J Mol Sci. 2023;24:1511. https://doi.org/10.3390/ijms24021511. **This paper provides a thorough overview of the role of the intraprostatic, urinary, and gut microbiome in the development/progression of prostate cancer.**

2.• Fujita K, Matsushita M, De Velasco MA, Hatano K, Minami T, Nonomura N, et al. The gut-prostate axis: a new perspective of prostate cancer biology through the gut microbiome. Cancers. 2023;15:1375. https://doi.org/10.3390/cancers15051375. **This review describes the gut-prostate axis with role of dietary influences on gut microbiome.**

3.• Hsiao Y-C, Liu C-W, Yang Y, Feng J, Zhao H, Lu K. DNA damage and the gut microbiome: from mechanisms to disease outcomes. DNA. 2023;3:13–32. https://doi.org/10.3390/dna3010002.

4. Xia B, Wang J, Zhang D, Hu X. The human microbiome links to prostate cancer risk and treatment (Review). Oncol Rep. 2023;49:1–12. https://doi.org/10.3892/or.2023.8560. **This study discusses role of gut microbiome in prostate cancer and elaborates on direct and indirect mechanisms for cancer development.**

5. Ma J, Gnanasekar A, Lee A, Li WT, Haas M, Wang-Rodriguez J, et al. Influence of intratumor microbiome on clinical outcome and immune processes in prostate cancer. Cancers. 2020;12:2524. https://doi.org/10.3390/cancers12092524.

6. Stamey TA, Fair WR, Timothy MM, Chung HK. Antibacterial nature of prostatic fluid. Nature. 1968;218:444–7. https://doi.org/10.1038/218444a0.

7. Gatti G, Quintar AA, Andreani V, Nicola JP, Maldonado CA, Masini-Repiso AM, et al. Expression of Toll-like receptor 4 in the prostate gland and its association with the severity of prostate cancer. Prostate. 2009;69:1387–97. https://doi.org/10.1002/pros.20984.

8. Cohen RJ, Shannon BA, McNEAL JE, Shannon T, Garrett KL. Propionibacterium acnes associated with inflammation in radical prostatectomy specimens: a possible link to cancer evolution? J Urol. 2005;173:1969–74. https://doi.org/10.1097/01.ju.0000158161.15277.78.

9. Eisenhofer R, Minich JJ, Marotz C, Cooper A, Knight R, Weyrich LS. Contamination in low microbial biomass microbiome

studies: issues and recommendations. Trends Microbiol. 2019;27:105–17. https://doi.org/10.1016/j.tim.2018.11.003.
10. Sfanos KS, Sauvageot J, Fedor HL, Dick JD, De Marzo AM, Isaacs WB. A molecular analysis of prokaryotic and viral DNA sequences in prostate tissue from patients with prostate cancer indicates the presence of multiple and diverse microorganisms. Prostate. 2008;68:306–20. https://doi.org/10.1002/pros.20680.
11. Yow MA, Tabrizi SN, Severi G, Bolton DM, Pedersen J, Giles GG, et al. Characterisation of microbial communities within aggressive prostate cancer tissues. Infect Agent Cancer. 2017;12:4. https://doi.org/10.1186/s13027-016-0112-7.
12. Cavarretta I, Ferrarese R, Cazzaniga W, Saita D, Lucianò R, Ceresola ER, et al. The microbiome of the prostate tumor microenvironment. Eur Urol. 2017;72:625–31. https://doi.org/10.1016/j.eururo.2017.03.029.
13. Banerjee S, Alwine JC, Wei Z, Tian T, Shih N, Sperling C, et al. Microbiome signatures in prostate cancer. Carcinogenesis. 2019;40:749–64. https://doi.org/10.1093/carcin/bgz008.
14. Miyake M, Ohnishi K, Hori S, Nakano A, Nakano R, Yano H, et al. Mycoplasma genitalium infection and chronic inflammation in human prostate cancer: detection using prostatectomy and needle biopsy specimens. Cells. 2019;8:212. https://doi.org/10.3390/cells8030212.
15. Feng Y, Ramnarine VR, Bell R, Volik S, Davicioni E, Hayes VM, et al. Metagenomic and metatranscriptomic analysis of human prostate microbiota from patients with prostate cancer. BMC Genomics. 2019;20:146. https://doi.org/10.1186/s12864-019-5457-z.
16.• Tsydenova IA, Ibragimova MK, Tsyganov MM, Litviakov NV. Human papillomavirus and prostate cancer: systematic review and meta-analysis. Sci Rep. 2023;13:16597. https://doi.org/10.1038/s41598-023-43767-7. **Meta-analysis of 27 case-control trials describing higher risk of prostate cancer compared to normal tissue or controls with BPH.**
17. Aragón IM, Herrera-Imbroda B, Queipo-Ortuño MI, Castillo E, Del Moral JS-G, Gómez-Millán J, et al. The urinary tract microbiome in health and disease. Eur Urol Focus. 2018;4:128–38. https://doi.org/10.1016/j.euf.2016.11.001.
18. de Vos WM, Tilg H, Hul MV, Cani PD. Gut microbiome and health: mechanistic insights. Gut. 2022;71:1020–32. https://doi.org/10.1136/gutjnl-2021-326789.
19. Golombos DM, Ayangbesan A, O'Malley P, Lewicki P, Barlow L, Barbieri CE, et al. The role of gut microbiome in the pathogenesis of prostate cancer: a prospective, pilot study. Urology. 2018;111:122–8. https://doi.org/10.1016/j.urology.2017.08.039.
20. Sfanos KS, Markowski MC, Peiffer LB, Ernst SE, White JR, Pienta KJ, et al. Compositional differences in gastrointestinal microbiota in prostate cancer patients treated with androgen axis-targeted therapies. Prostate Cancer Prostatic Dis. 2018;21:539–48. https://doi.org/10.1038/s41391-018-0061-x.
21. Liss MA, White JR, Goros M, Gelfond J, Leach R, Johnson-Pais T, et al. Metabolic biosynthesis pathways identified from fecal microbiome associated with prostate cancer. Eur Urol. 2018;74:575–82. https://doi.org/10.1016/j.eururo.2018.06.033.
22. Liu Y, Yang C, Zhang Z, Jiang H. Gut microbiota dysbiosis accelerates prostate cancer progression through increased LPCAT1 expression and enhanced DNA repair pathways. Front Oncol. 2021;11:679712. https://doi.org/10.3389/fonc.2021.679712.
23. Pernigoni N, Zagato E, Calcinotto A, Troiani M, Mestre RP, Calì B, et al. Commensal bacteria promote endocrine resistance in prostate cancer through androgen biosynthesis. Science. 2021;374:216–24. https://doi.org/10.1126/science.abf8403.
24. Mathay C, Hamot G, Henry E, Georges L, Bellora C, Lebrun L, et al. Method optimization for fecal sample collection and fecal DNA extraction. Biopreservation Biobanking. 2015;13:79–93. https://doi.org/10.1089/bio.2014.0031.
25. Carroll IM, Ringel-Kulka T, Siddle JP, Klaenhammer TR, Ringel Y. Characterization of the fecal microbiota using high-throughput sequencing reveals a stable microbial community during storage. PLoS ONE. 2012;7:e46953. https://doi.org/10.1371/journal.pone.0046953.
26. Watson E-J, Giles J, Scherer BL, Blatchford P. Human faecal collection methods demonstrate a bias in microbiome composition by cell wall structure. Sci Rep. 2019;9:16831. https://doi.org/10.1038/s41598-019-53183-5.
27. Baranzini SE. Insights into microbiome research 2: experimental design, sample collection, and shipment. Mult Scler Houndmills Basingstoke Engl. 2018;24:1419–20. https://doi.org/10.1177/1352458518788962.
28.• Short MI, Hudson R, Besasie BD, Reveles KR, Shah DP, Nicholson S, et al. Comparison of rectal swab, glove tip, and participant-collected stool techniques for gut microbiome sampling. BMC Microbiol. 2021;21:26. https://doi.org/10.1186/s12866-020-02080-3. **Microbiome methods study comparing stool collection techniques including validation of the glove tip collection technique.**
29. Ahmed S, Macfarlane GT, Fite A, McBain AJ, Gilbert P, Macfarlane S. Mucosa-associated bacterial diversity in relation to human terminal ileum and colonic biopsy samples. Appl Environ Microbiol. 2007;73:7435–42. https://doi.org/10.1128/AEM.01143-07.
30. Wu W-K, Chen C-C, Panyod S, Chen R-A, Wu M-S, Sheen L-Y, et al. Optimization of fecal sample processing for microbiome study — the journey from bathroom to bench. J Formos Med Assoc. 2019;118:545–55. https://doi.org/10.1016/j.jfma.2018.02.005.
31.• Liang Y, Dong T, Chen M, He L, Wang T, Liu X, et al. Systematic analysis of impact of sampling regions and storage methods on fecal gut microbiome and metabolome profiles. MSphere. 2020;5:e00763-19. https://doi.org/10.1128/mSphere.00763-19. **Study examining effects of stool sampling, homogenization and storage conditions.**
32. Perez-Carrasco V, Soriano-Lerma A, Soriano M, Gutiérrez-Fernández J, Garcia-Salcedo JA. Urinary microbiome: yin and yang of the urinary tract. Front Cell Infect Microbiol. 2021; 11. https://doi.org/10.3389/fcimb.2021.617002
33. Wheeler KM, Liss MA. The microbiome and prostate cancer risk. Curr Urol Rep. 2019;20:66. https://doi.org/10.1007/s11934-019-0922-4.
34. Sidebottom AM. A brief history of microbial study and techniques for exploring the gastrointestinal microbiome. Clin Colon Rectal Surg. 2023;36:98–104. https://doi.org/10.1055/s-0042-1760678.
35. Kool J, Tymchenko L, Shetty SA, Fuentes S. Reducing bias in microbiome research: comparing methods from sample collection to sequencing. Front Microbiol. 2023;14:1094800. https://doi.org/10.3389/fmicb.2023.1094800.
36.•• Li X, Shi X, Yao Y, Shen Y, Wu X, Cai T, et al. Effects of stool sample preservation methods on gut microbiota biodiversity: new original data and systematic review with meta-analysis. Microbiol Spectr. 2023;11:e04297-22. https://doi.org/10.1128/spectrum.04297-22. **Review and meta-analysis focusing on the impact of stool preservation methods.**
37. Sinha R, Abu-Ali G, Vogtmann E, Fodor AA, Ren B, Amir A, et al. Assessment of variation in microbial community amplicon sequencing by the Microbiome Quality Control (MBQC) project consortium. Nat Biotechnol. 2017;35:1077–86. https://doi.org/10.1038/nbt.3981.
38. Costea PI, Zeller G, Sunagawa S, Pelletier E, Alberti A, Levenez F, et al. Towards standards for human fecal sample processing in

metagenomic studies. Nat Biotechnol. 2017;35:1069–76. https://doi.org/10.1038/nbt.3960.
39. Greathouse KL, Sinha R, Vogtmann E. DNA extraction for human microbiome studies: the issue of standardization. Genome Biol. 2019;20:212. https://doi.org/10.1186/s13059-019-1843-8.
40. Structure, function and diversity of the healthy human microbiome. Nature. 2012;486: 207–214. https://doi.org/10.1038/nature11234
41. Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, et al. A human gut microbial gene catalog established by metagenomic sequencing. Nature. 2010;464:59–65. https://doi.org/10.1038/nature08821.
42. Marotz C, Amir A, Humphrey G, Gaffney J, Gogul G, Knight R. DNA extraction for streamlined metagenomics of diverse environmental samples. Biotechniques. 2017;62:290–3. https://doi.org/10.2144/000114559.
43. Huseyin CE, Rubio RC, O'Sullivan O, Cotter PD, Scanlan PD. The fungal frontier: a comparative analysis of methods used in the study of the human gut mycobiome. Front Microbiol. 2017;8:1432. https://doi.org/10.3389/fmicb.2017.01432.
44. Shkoporov AN, Ryan FJ, Draper LA, Forde A, Stockdale SR, Daly KM, et al. Reproducible protocols for metagenomic analysis of human faecal phageomes. Microbiome. 2018;6:68. https://doi.org/10.1186/s40168-018-0446-z.
45. Kwa WT, Sundarajoo S, Toh KY, Lee J. Application of emerging technologies for gut microbiome research. Singapore Med J. 2023;64:45–52. https://doi.org/10.4103/singaporemedj.SMJ-2021-432.
46. Fischer MA, Güllert S, Neulinger SC, Streit WR, Schmitz RA. Evaluation of 16S rRNA gene primer pairs for monitoring microbial community structures showed high reproducibility within and low comparability between datasets generated with multiple archaeal and bacterial primer pairs. Front Microbiol. 2016;7:1297. https://doi.org/10.3389/fmicb.2016.01297.
47. Abellan-Schneyder I, Matchado MS, Reitmeier S, Sommer A, Sewald Z, Baumbach J, et al. Primer, pipelines, parameters: issues in 16S rRNA gene sequencing. mSphere. 2021;6:e01202-20. https://doi.org/10.1128/mSphere.01202-20.
48. de la Cuesta-Zuluaga J, Escobar JS. Considerations for optimizing microbiome analysis using a marker gene. Front Nutr. 2016;3:26. https://doi.org/10.3389/fnut.2016.00026.
49. Claesson MJ, Wang Q, O'Sullivan O, Greene-Diniz R, Cole JR, Ross RP, et al. Comparison of two next-generation sequencing technologies for resolving highly complex microbiota composition using tandem variable 16S rRNA gene regions. Nucleic Acids Res. 2010;38:e200. https://doi.org/10.1093/nar/gkq873.
50. Chen Z, Hui PC, Hui M, Yeoh YK, Wong PY, Chan MCW, et al. Impact of preservation method and 16S rRNA hypervariable region on gut microbiota profiling. mSystems. 2019;4:e00271-18. https://doi.org/10.1128/mSystems.00271-18.
51. Kim S-W, Suda W, Kim S, Oshima K, Fukuda S, Ohno H, et al. Robustness of gut microbiota of healthy adults in response to probiotic intervention revealed by high-throughput pyrosequencing. DNA Res Int J Rapid Publ Rep Genes Genomes. 2013;20:241–53. https://doi.org/10.1093/dnares/dst006.
52.• Kameoka S, Motooka D, Watanabe S, Kubo R, Jung N, Midorikawa Y, et al. Benchmark of 16S rRNA gene amplicon sequencing using Japanese gut microbiome data from the V1–V2 and V3–V4 primer sets. BMC Genomics. 2021;22:527. https://doi.org/10.1186/s12864-021-07746-4. **Comparison of V1-2 versus V3-4 in the analysis of gut microbiome in 192 volunteers.**
53. Schloss PD, Gevers D, Westcott SL. Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studieS. PLoS ONE. 2011;6:e27310. https://doi.org/10.1371/journal.pone.0027310.
54. Huse SM, Huber JA, Morrison HG, Sogin ML, Welch DM. Accuracy and quality of massively parallel DNA pyrosequencing. Genome Biol. 2007;8:R143. https://doi.org/10.1186/gb-2007-8-7-r143.
55. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, et al. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J. 2012;6:1621–4. https://doi.org/10.1038/ismej.2012.8.
56. Wensel CR, Pluznick JL, Salzberg SL, Sears CL. Next-generation sequencing: insights to advance clinical investigations of the microbiome. J Clin Invest. 2022;132:e154944. https://doi.org/10.1172/JCI154944.
57. Callahan BJ, McMurdie PJ, Holmes SP. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. ISME J. 2017;11:2639–43. https://doi.org/10.1038/ismej.2017.119.
58. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: high resolution sample inference from Illumina amplicon data. Nat Methods. 2016;13:581–3. https://doi.org/10.1038/nmeth.3869.
59. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, et al. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol. 2006;72:5069–72. https://doi.org/10.1128/AEM.03006-05.
60. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 2013;41:D590–6. https://doi.org/10.1093/nar/gks1219.
61. Wang Q, Garrity GM, Tiedje JM, Cole JR. Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl Environ Microbiol. 2007;73:5261–7. https://doi.org/10.1128/AEM.00062-07.
62. Nucleotide BLAST: search nucleotide databases using a nucleotide query. [cited 11 Sep 2023]. Available: https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=BlastSearch&BLAST_SPEC=MicrobialGenomes
63. Almeida A, Mitchell AL, Tarkowska A, Finn RD. Benchmarking taxonomic assignments based on 16S rRNA gene profiling of the microbiota from commonly sampled environments. GigaScience. 2018;7:giy054. https://doi.org/10.1093/gigascience/giy054.
64. Pérez-Cobas AE, Gomez-Valero L, Buchrieser C. Metagenomic approaches in microbial ecology: an update on whole-genome and marker gene sequencing analyses. Microb Genomics. 2020;6:mgen000409. https://doi.org/10.1099/mgen.0.000409.
65. Vinje H, Liland KH, Almøy T, Snipen L. Comparing K-mer based methods for improved classification of 16S sequences. BMC Bioinformatics. 2015;16:205. https://doi.org/10.1186/s12859-015-0647-4.
66. Klappenbach JA, Dunbar JM, Schmidt TM. rRNA operon copy number reflects ecological strategies of bacteria. Appl Environ Microbiol. 2000;66:1328–33.
67. Větrovský T, Baldrian P. The variability of the 16S rRNA gene in bacterial genomes and its consequences for bacterial community analyses. PLoS ONE. 2013;8:e57923. https://doi.org/10.1371/journal.pone.0057923.
68. Ranjan R, Rani A, Metwally A, McGee HS, Perkins DL. Analysis of the microbiome: advantages of whole genome shotgun versus 16S amplicon sequencing. Biochem Biophys Res Commun. 2016;469:967–77. https://doi.org/10.1016/j.bbrc.2015.12.083.
69. Sims D, Sudbery I, Ilott NE, Heger A, Ponting CP. Sequencing depth and coverage: key considerations in genomic analyses. Nat Rev Genet. 2014;15:121–32. https://doi.org/10.1038/nrg3642.
70. 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. 2011;13:47–58. https://doi.org/10.1038/nrg3129.
71. Miller JR, Koren S, Sutton G. Assembly algorithms for next-generation sequencing data. Genomics. 2010;95:315–27. https://doi.org/10.1016/j.ygeno.2010.03.001.

72. Sedlar K, Kupkova K, Provaznik I. Bioinformatics strategies for taxonomy independent binning and visualization of sequences in shotgun metagenomics. Comput Struct Biotechnol J. 2016;15:48–55. https://doi.org/10.1016/j.csbj.2016.11.005.
73. Breitwieser FP, Lu J, Salzberg SL. A review of methods and databases for metagenomic classification and assembly. Brief Bioinform. 2017;20:1125–36. https://doi.org/10.1093/bib/bbx120.
74. Yok NG, Rosen GL. Combining gene prediction methods to improve metagenomic gene annotation. BMC Bioinformatics. 2011;12:20. https://doi.org/10.1186/1471-2105-12-20.
75. Kelley DR, Liu B, Delcher AL, Pop M, Salzberg SL. Gene prediction with Glimmer for metagenomic sequences augmented by classification and clustering. Nucleic Acids Res. 2012;40:e9. https://doi.org/10.1093/nar/gkr1067.
76. Trimble WL, Keegan KP, D'Souza M, Wilke A, Wilkening J, Gilbert J, et al. Short-read reading-frame predictors are not created equal: sequence error causes loss of signal. BMC Bioinformatics. 2012;13:183. https://doi.org/10.1186/1471-2105-13-183.
77. Huntemann M, Ivanova NN, Mavromatis K, Tripp HJ, Paez-Espino D, Tennessen K, et al. The standard operating procedure of the DOE-JGI Metagenome Annotation Pipeline (MAP vol 4). Stand Genomic Sci. 2016;11:17. https://doi.org/10.1186/s40793-016-0138-x.
78. Alcock BP, Raphenya AR, Lau TTY, Tsang KK, Bouchard M, Edalatmand A, et al. CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. Nucleic Acids Res. 2020;48:D517–25. https://doi.org/10.1093/nar/gkz935.
79. Stewart RD, Auffret MD, Snelling TJ, Roehe R, Watson M. MAGpy: a reproducible pipeline for the downstream analysis of metagenome-assembled genomes (MAGs). Bioinformatics. 2019;35:2150–2. https://doi.org/10.1093/bioinformatics/bty905.
80. Quince C, Delmont TO, Raguideau S, Alneberg J, Darling AE, Collins G, et al. DESMAN: a new tool for de novo extraction of strains from metagenomes. Genome Biol. 2017;18:181. https://doi.org/10.1186/s13059-017-1309-9.
81. Nazeen S, Yu YW, Berger B. Carnelian uncovers hidden functional patterns across diverse study populations from whole metagenome sequencing reads. Genome Biol. 2020;21:47. https://doi.org/10.1186/s13059-020-1933-7.
82. Nayfach S, Bradley PH, Wyman SK, Laurent TJ, Williams A, Eisen JA, et al. Automated and accurate estimation of gene family abundance from shotgun metagenomes. PLoS Comput Biol. 2015;11:e1004573. https://doi.org/10.1371/journal.pcbi.1004573.
83. Xia Y. Statistical normalization methods in microbiome data with application to microbiome cancer research. Gut Microbes. 2023;15:2244139. https://doi.org/10.1080/19490976.2023.2244139.
84. Hong J, Karaoz U, de Valpine P, Fithian W. To rarefy or not to rarefy: robustness and efficiency trade-offs of rarefying microbiome data. Bioinformatics. 2022;38:2389–96. https://doi.org/10.1093/bioinformatics/btac127.
85. Robin JD, Ludlow AT, LaRanger R, Wright WE, Shay JW. Comparison of DNA quantification methods for next generation sequencing. Sci Rep. 2016;6:24067. https://doi.org/10.1038/srep24067.
86. Wu WW, Phue J-N, Lee C-T, Lin C, Xu L, Wang R, et al. Robust sub-nanomolar library preparation for high throughput next generation sequencing. BMC Genomics. 2018;19:326. https://doi.org/10.1186/s12864-018-4677-y.
87. Lozupone C, Lladser ME, Knights D, Stombaugh J, Knight R. UniFrac: an effective distance metric for microbial community comparison. ISME J. 2011;5:169–72. https://doi.org/10.1038/ismej.2010.133.
88. de Aguirre Cárcer D, Denman SE, McSweeney C, Morrison M. Evaluation of subsampling-based normalization strategies for tagged high-throughput sequencing data sets from gut microbiomes. Appl Environ Microbiol. 2011;77:8795–8. https://doi.org/10.1128/AEM.05491-11.
89. McMurdie PJ, Holmes S. Waste not, want not: why rarefying microbiome data is inadmissible. PLoS Comput Biol. 2014;10:e1003531. https://doi.org/10.1371/journal.pcbi.1003531.
90. Weiss S, Xu ZZ, Peddada S, Amir A, Bittinger K, Gonzalez A, et al. Normalization and microbial differential abundance strategies depend upon data characteristics. Microbiome. 2017;5:27. https://doi.org/10.1186/s40168-017-0237-y.
91. McKnight DT, Huerlimann R, Bower DS, Schwarzkopf L, Alford RA, Zenger KR. Methods for normalizing microbiome data: an ecological perspective. Methods Ecol Evol. 2019;10:389–400. https://doi.org/10.1111/2041-210X.13115.
92. Chen L, Reeve J, Zhang L, Huang S, Wang X, Chen J. GMPR: A robust normalization method for zero-inflated count data with application to microbiome sequencing data. PeerJ. 2018;6:e4600. https://doi.org/10.7717/peerj.4600.
93. Mulenga M, Kareem SA, Sabri AQ, Seera M, Govind S, Samudi C, et al. Feature extension of gut microbiome data for deep neural network-based colorectal cancer classification. IEEE Access. 2021;9:23565–78. https://doi.org/10.1109/ACCESS.2021.3050838.
94.• Swift D, Cresswell K, Johnson R, Stilianoudakis S, Wei X. A review of normalization and differential abundance methods for microbiome counts data. WIREs Comput Stat. 2023;15:e1586. https://doi.org/10.1002/wics.1586. **Comprehensive review and comparison of methods for differential abundance and normalization.**
95. Park S-Y, Ufondu A, Lee K, Jayaraman A. Emerging computational tools and models for studying gut microbiota composition and function. Curr Opin Biotechnol. 2020;66:301–11. https://doi.org/10.1016/j.copbio.2020.10.005.
96.•• Kodikara S, Ellul S, Lê Cao K-A. Statistical challenges in longitudinal microbiome data analysis. Brief Bioinform. 2022;23:bbac273. https://doi.org/10.1093/bib/bbac273. **Review of the challenges associated with statistical analysis of logitudinal microbiome data.**
97. Liu L, Shih Y-CT, Strawderman RL, Zhang D, Johnson BA, Chai H. Statistical analysis of zero-inflated nonnegative continuous data: a review. Stat Sci. 2019;34:253–79. https://doi.org/10.1214/18-STS681.
98. Zhang X, Guo B, Yi N. Zero-Inflated gaussian mixed models for analyzing longitudinal microbiome data. PLoS ONE. 2020;15:e0242073. https://doi.org/10.1371/journal.pone.0242073.
99. Lee J, Sison-Mangus M. A Bayesian semiparametric regression model for joint analysis of microbiome data. Front Microbiol. 2018; 9. https://doi.org/10.3389/fmicb.2018.00522
100. McGeachie MJ, Chang H-H, Weiss ST. CGBayesNets: conditional Gaussian Bayesian network learning and inference with mixed discrete and continuous data. PLOS Comput Biol. 2014;10:e1003676. https://doi.org/10.1371/journal.pcbi.1003676.