MICROBIOME INSIGHTS

How to Plan and Conduct a Microbiome Study

A GUIDE FROM MICROBIOME INSIGHTS

Introduction

Who is this guide for

Whether you are new to microbiome research or have previous experience, this guide was created for anyone interested in studying microbial communities. From scientists to business leaders, we have prepared a complete guide to walk the reader through planning and conducting a meaningful microbiome study.

What it covers

This comprehensive document is based on the combined multi-decade experience of our team in microbiome testing, having worked on almost 500 microbiome studies. It covers all the key steps from study design and execution through to analysis and interpretation, including many useful resources and supporting documents.

Why we created it

Our goal as an organization is to support our clients and advance microbiome research from basic science to commercial research and development. That's why we've done a deep dive into the process of conducting microbiome research to make it accessible and straightforward.

Authors

Ana Yun, MBA

Ana is a purpose-driven marketing leader with expertise in branding, marketing strategies, and product management. Her keen interest in innovation, combined with passion for enabling better health, have led her to the life sciences sector. Ana joined MBI as an MBA intern in 2018 before moving into a full time role as a product manager overseeing the development and commercialization of new testing services.

Bill Mohn, PhD

Dr. William (Bill) Mohn is a Professor in the Department of Microbiology & Immunology in the Life Sciences Institute at the University of British Columbia. Bill earned his B.A. from Colgate University and his Ph.D. from Michigan State. He was a Research Associate at the National Research Council of Canada in Ottawa prior to moving to UBC as an Assistant professor in 1993.

Karl Moran

Karl Moran joined the MBI team in late 2017 and is the Director of Sales and Business Development. Karl has a B.Sc. in Pharmacology from University College Dublin (Ireland). Prior to joining the team Karl worked as a Project Manager for Icon Clinical Research, where he managed the laboratory testing and operations for clinical trials.

Pedro Dimitriu, PhD

Pedro Dimitriu is MBI's senior scientist and Senior Director of Bioinformatics. He is in charge of integrating sequencing data, ecological statistics, and machine learning methods to study host–microbe interactions. Before joining the company, Pedro was a postdoctoral fellow at UBC's Department of Microbiology and Immunology.

Table of Contents

Before you start

Getting started

If you are reading this guide, you have taken a great first step. Seeking an expert opinion on how to design a study that will help you meet your objectives is the right place to start. Microbiome studies can be complex and there are many important steps that can significantly impact the overall quality of the resulting data. With this in mind, Microbiome Insights has put together this resource for our current and potential customers, independent researchers and anybody who is interested in knowing what a typical study involves. As complex as any study can be, we have broken it into simple-to-follow steps to set you up for success. Microbiome Insights is here to help you achieve your goals and can step in at any point in the process to provide the scientific expertise, technical capabilities and professional, friendly support.

Study Process

- Sampling
- Storage and shipment
- DNA extraction
- Profiling, including DNA sequencing and/or metabolomics
- Bioinformatics, data visualization and reporting

Depending on your internal capabilities, expertise and experience in study planning, you may need help with any of the steps involved, but regardless, it is beneficial to be aware of the whole process and the respective requirements and considerations of each subsequent step. For example, in order to collect samples, you need to consider things like sampling devices, storage temperature, shipping options and possibly contract research organizations (CRO) to help you with any of these steps.

At Microbiome Insights, we have customers who outsource the whole study to us and others who only require some steps, such as sequencing or bioinformatics analysis.

Main considerations for a study

Whether you are on your own or decide to partner with the reputable CRO organization, here are the main considerations for planning a study:

- **Question or hypothesis:** before you start, it is important to determine a hypothesis that you would like to support or reject. That will help you set the right objectives and the right expectations. Ask yourself what you would need to see in the data you receive in order to support or reject your hypothesis, with the latter scenario being just as important.

- **Objectives:** now that you have envisioned possible outcomes for your study, what are your primary and secondary objectives? Perhaps, your secondary objective is to find evidence for forming a follow-up hypothesis, or it may be to gain more detailed mechanistic understanding of the research topic.

> 🔗 Resource: Research questions, hypotheses and objectives

- **Novelty of question:** it is essential to know the past and ongoing research on the subject you are studying. Be sure to evaluate previously published literature.

- **Type of study:** another very important consideration that will help you better design your study. Is this an observational study, which would bring up a set of new factors to consider, such as stratification by potential confounders (some examples include age, sex, diet, lifestyle factors, medications)? By contrast, interventional studies evaluate the effect of medications, probiotics or other interventions on the microbiome.

- **Type of model:** a key consideration is whether you are working with an animal model. If so, how will the findings translate to human's microbiome?

- **Ethics:** there are multiple regulatory rules that need to be considered. For example, if it's a human study, how will the personal data be stored and is there an infrastructure in place to protect patient privacy. Institutional Review Board's (I.R.B) approvals are a common step in study planning as well. If it is an animal study, other regulations such as American Association for Laboratory Animal Science (IACUC) approval may be required. If you are working with soil or plant samples, field experiments will often have applicable regulations.

- **Sample type:** there is a broad range of samples that can be collected: fecal, skin, oral, urine, vaginal, soil, plant, water, and others. Depending on your particular study, you may or may not know what the best sample type is in your case. It often depends on the disease or condition you are studying. For example, one of the most common sample types, stool, is useful in studying gastrointestinal conditions and liver disease. It is also a tractable sample type for longitudinal studies, as it can be collected at multiple time points from the same individuals with minimal invasiveness, as compared to more invasive sample types, such as tissue obtained through colonoscopy. On the other hand, stool samples will be poorly representative of the upper gastrointestinal tract. Ultimately, sampling should be informed by the hypothesis being tested. From the testing standpoint, sequencing type and data depth may vary by sample type. Learn more about sample types and their applications in section 2 of this guide.

- **Study size:** is influenced by a number of factors that were addressed earlier in this guide. For example, if your study objective is to support association or a causative link between an intervention and changes in the microbiome, you may require a large number of samples. Where

treatments or subject groups are compared, the number of replicates is critical, as you must have enough statistical power to detect differences. Related previous studies will often suggest an appropriate study size. However, in many cases, you cannot know a priori how many samples will be necessary, and the decision will have to be based on the best estimate and availability of resources.

- **Other considerations:** these include control for experimental variation, particularly for multi-site or longitudinal studies, and sequencing technology (what type of technology is most appropriate to yield the data that will satisfy your study objectives), which can be important factors to consider. Also, for human studies, registration at clinicaltrials.gov should be established. Many journals now prefer to have clinical studies registered in advance of study execution for manuscripts to be considered for publication.

- **Setting realistic expectations for the final report:** this step is critical as you design your study. Some of the considerations you may want to think about are: who will I be presenting the study results to, what type of things will my target audience be looking for in order to inform future decisions, how can I help them better understand the data (or study in general), what are publication standards in the relevant field, what is my expertise with data visualization and is it worthwhile getting expert help with that. Particularly, if you are partnering with a CRO, those organizations have data scientists and bioinformaticians to help you with that and setting those expectations early on in the process will help your partners better serve your needs and save you time in the end.

- **Key success metrics:** subsequent to the previous step, determine specific metrics you are looking for. The best advice here is to remember that study integrity, objectivity and accuracy are often more important than getting desired outcomes. For example, your study may not yield the results you were hoping for, but that doesn't mean the study failed. It may simply mean that a more powerful study is needed, or a new hypothesis is required.

Now that you are clear on your objectives, considerations and expectations, you are all set to get started. This guide will focus on each of the critical steps in detail.

Sample Types & Collection Methods

In this section, we will cover some of the sample types which are commonly used in microbiome research along with some other key considerations for sample collection, storage, and shipment. For detailed collection instructions for each of the sample types you can follow the links to the dedicated sections on our website.

Sample types

I. Fecal samples

Fecal samples are the most commonly used non-invasive method of capturing a snapshot of the gut microbiome. Fresh fecal samples contain rich microbial communities — a representation of the microbial community existing in the gut — but proper conditions are required to maintain the integrity of this microbiome until it reaches the lab. For all samples, -80°C is recommended for long-term storage, and freeze-thaw cycles should be avoided (or minimized). A fecal swab is the easiest method for human stool sampling, but the sample volume may be insufficient for some analyses. Collecting whole stool will yield larger volumes that may be required for certain analyses (e.g., metabolomics) or for multiple analyses. Fecal samples may be collected whole, using any sealed container that can withstand -80°C or using a kit with a stabilization buffer that enables short-term storage and transport at ambient temperature.

🔗 Instructions for collection: Fecal Samples

II. Skin samples

Skin is the human body's largest organ and is teeming with microbes: bacteria, archaea, fungi, and viruses. But compared with the gut, skin does not support a high-biomass microbiome. Thus, microbiome sampling and sequencing for skin require special knowledge and handling.

Our internal validations have shown that skin microbiome recovery is optimized using pre-moistened swabs as compared to dry swabs or skin tapes. This optimization results in overall higher microbial DNA content while minimizing host DNA content. We recommend using the Becton-Dickinson, BBL Cultureswab EZ II (or similar), which includes a double swab encased in a non-breathable transport tube.

Because many skin microorganisms are firmly attached, the most representative samples are obtained with repeated and vigorous swabbing. Skin is inevitably also recovered - with substantial

host (human or animal) DNA. Host DNA can interfere with some downstream microbiome analyses.

More representative skin samples can be obtained with skin biopsies; however, this method is generally considered too invasive for human studies. It may be an option for animal studies.

Skin washing and use of some skin products can transiently but drastically impact the composition of the skin microbiome. It is recommended that the study protocol limits washing and use of products on the area(s) to be sampled for a defined period prior to sampling (eg, 12 hours).

Learn more about planning a skin microbiome study:
🔗 How to Design a Skin Microbiome Study, Part I: Sampling
🔗 How to Design a Skin Microbiome Study, Part II: Amplicon Sequencing
🔗 Instructions for collection: Skin Samples

III. Oral samples

The oral microbiome is both dense and highly diverse in bacteria, as well as other microbes such as viruses, fungi, and archaea. The oral microbiome interacts with the host, while also having many influences from the surrounding environment. Since this microbiome is transient in nature, to capture the oral microbiome it is imperative to stabilize the sample until it arrives in the lab (e.g. using a stabilization buffer or shipping on dry ice).

The mouth has many distinct microbiomes in various locations (periodontal, pallet, tongue, inner cheek, etc.). Therefore, the sampling location must be planned to target the microbiome under investigation (i.e., consistent with the research hypothesis).

🔗 Instructions for collection: Oral Samples

IV. Urine samples

Many assume urine is free of bacteria and other microbes in individuals without infections. However, certain conditions can result in higher levels of detectable microbes in the urinary system. This type of sample is considered low biomass and is also typically low in diversity. In common with other high-volume, low microbial biomass samples such as air particulates or seawater, microbes within a urine sample are typically concentrated through filtration and/or centrifugation prior to DNA extraction.

🔗 Instructions for collection: Urine Samples

V. Vaginal samples

The vaginal microbiome has been correlated with health issues such as bacterial vaginosis, sexually transmitted infections (STIs), and urinary tract infections (UTIs). The patient giving the sample may have sensitivities in the vagina and surrounding area so sample collection must be done by a professional in a clinic, or at home using a validated self-collection kit.

🔗 Instructions for collection: Vaginal Samples

VI. Soil samples

Soil microbial communities are both numerous (10^{10} to 10^{11} individual/ gram soil) and diverse (10^{4} bacterial species/gram soil). Soils can also present high spatial variability in the form of microenvironments, thus replication is very important when designing soil microbiome studies.

The most studied agricultural microbiome is the plant's rhizosphere: the microorganisms (including mycorrhizal fungi) living at the interface of the plant's root system and the surrounding soil. To the plant they provide fixed nitrogen, access to micronutrients, and protection from pathogens in the soil; in turn, the plant roots release nutrients (for example, sugars and amino acids) that sustain the microbes. Soil microbes are an important component of many biogeochemical cycles including Carbon, Nitrogen, Oxygen, Sulfur cycles.

🔗 Instructions for collection: Soil Samples

VII. Water samples

Water samples often contain a high diversity of microbes, including bacteria or archaea. Check out this link to a protocol prepared by Dr. Steven Hallam's lab that outlines small volume filtration of water samples. An important step is to provide at least 100mL of sample to follow this protocol and obtain enough DNA for sequencing.

🔗 Instructions for collection: Water Samples

Considerations

Find below the important considerations for sample collection, storage and sample shipment.

I. Sampling device/strategy is determined by factors like the facility used for sample collection, for example, is it a clinical facility? Is the sample collection going to be done by a research staff or study participants? If so, are the participants going to be asked to mail their samples or drop them off at the site? Is there a storage facility on-site? There are also multiple collection device types, including swabs, tubes, hats and others. In choosing your device also consider what needs to be extracted in addition to DNA, for example RNA, SCFA (short chain fatty acid) or other metabolites. Most importantly, does the sample type match your scientific question? For example, if you study hair microbiome, establishing chronic disease incidence is unrealistic. Learn more on choosing the right collection device for your study here:

🔗 Considerations when choosing a collection device for your fecal samples

II. **Storage conditions** are important for preserving the samples until further processing. What is the storage temperature requirement and how long will the samples need to be stored at the facility? Although depending on the sample type, some sample collection kits allow for the samples to be stored at room temperature, the gold standard for fecal samples without preservation is to snap freeze the sample at -80°C as fast as you can. If you are shipping your samples to the lab, check with your lab whether it has a freezer to store your samples. If you know what sample collection device you are using, you will be able to find storage requirements validated by the studies in the scientific literature. Sample collection kit providers will have both instructions and validation study papers posted on their websites or available by request.

III. **Some of available sampling and storage solutions are**:
- DNA Genotek
- Norgen Biotek
- RNAlater
- Zymo Research

IV. **Sample quantity:** it varies by sample type. For example, stool samples require very little quantity, but stool consistency can significantly impact downstream microbiome analyses. Bristol scale is

often used to record stool consistency at the time of collection. Learn more about Bristol scale. Also, consider if the body site is low or high in microbial biomass.

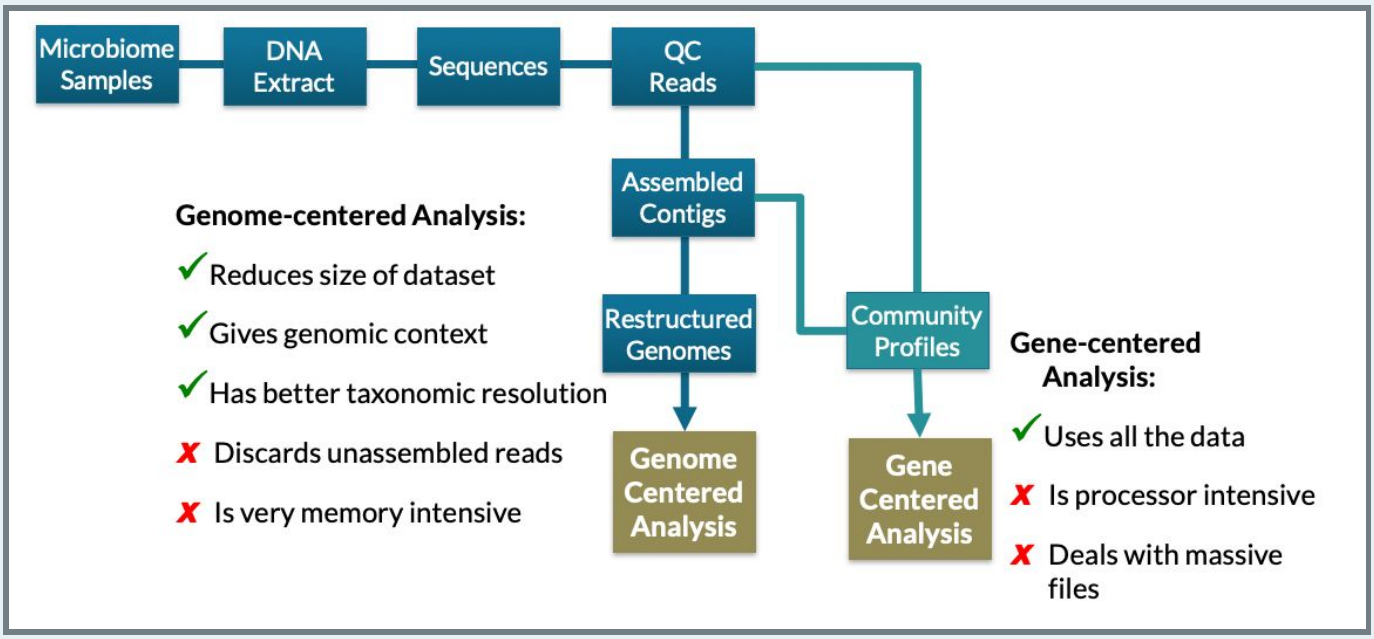
V. **Shipping temperature:** Samples that are being stored frozen should be shipped on dry ice with enough dry ice to last for the duration of transport. Samples that are collected using a stabilization buffer for storage at ambient temperature need not be kept in the freezer unless you will be storing your samples longer than the kit's stability timeframe as outlined by the collection kit manufacturer. Likewise, these samples can be transported to the lab facility without dry ice unless you are removing them from the freezer to ship. In this case, samples must be shipped on dry ice with enough dry ice to last for the duration of transport.

VI. **Shipping guidelines:** these depend on the country of destination, i.e. where you are shipping your samples to. It is critical to understand applicable regulations and laws, in order to avoid the loss of valuable samples. Permits are required to import certain sample types, such as soil. Generally, a waybill and commercial invoice are required for the samples to cross an international border.

VII. **Shipping preparation:** this will largely depend on the laboratory or CRO organization you are working with. Generally, you are required to submit the order request form or sample manifest before you ship your samples. Ask your CRO provider to send you a sample form, as well as specific requirements for filling out the form. It is important to ensure that all samples are labelled according to the identification numbers provided on the sample manifest. Samples should also be packaged in the same order as listed in the sample manifest.

DNA extraction

This step can be performed either in-house or by outsourcing it to another lab or CRO organization. Even if you decided to outsource this step to your trusted CRO partner, it is still helpful to know the common principles and the general flow that takes place in the laboratory from the moment the sample arrives to the point when raw data has been generated. The flow also depends on the type of analysis performed. For metagenomics, the common process is the following:



Main considerations if you are doing DNA extraction in your own lab:

It is important to note that DNA extraction is *very susceptible to bias* and as such it influences the analysis results.

Protocol choice:

- Application
- Sample type
- Required yield - volume and concentration, purity

Contaminant removal:

- Host DNA
- rRNAs
- Humic acids
- Etc.

Automation - robot vs. manual handling. Automation in certain cases may actually decrease DNA quality and throughput. Time and cost are the main considerations.

Profiling Methods

Before DNA samples can be sequenced by next-generation sequencing, they must be fragmented, end-repaired, and collected into adapter-ligated libraries. Library preparation protocols can influence the results generated by your NGS experiments.

Library preparation considerations:

Contamination:

- Well to well
- External
- Index hopping

Worth noting: contamination is increasingly problematic with decreasing biomass

Best practices for controlling contamination:

- Sequencing positive control not expected in real samples
- Including at least 2 types of negative controls
- Using mock communities for assessing bias in DNA extraction
- Using dual barcodes to help control index hopping
- Isolating pre- and post- PCR work
- Avoiding liquid handling robot for DNA extraction

Best practices for library preparations:

- It's best to collect all samples, randomize, extract and sequence samples all at once if possible.
- Be aware of "repeat offenders" (contaminants, like staphylococcus and E.coli).
- Removing everything found in negative controls can be a mistake.
- Contamination increases with lower biomass.

Other sources of variation (PCR bias):

- Choice of primers
- Chimeras
- Misincorporation of nucleotides
- PCR drift (accumulation of random amplification)

- Choice of high fidelity polymerase and number of PCR cycles

Sequencing Options

There is a range of sequencing options available, each method having its pros and cons. Typically, these are the kinds of analysis you can expect from a lab specialized in microbiome-based studies:

I. rRNA Gene Sequencing of 16S, 18S, ITS2 and archaeal V4-V5

Amplicon sequencing, or rRNA gene sequencing, is performed to determine the relative abundance of taxa in a bacterial community and to compare between groups of interest. This level of analysis can help to address changes in the overall microbial profile over time, or between treatment groups. Amplicon sequencing will give you reliable identification to the genus level. Depending on the sample type, the following amplicons are commonly used:

- Prokaryotic communities (16S V4, V1-V3)
- Eukaryotic communities (18S)
- Fungal communities (ITS2)
- Archaeal communities (16S V4-V5)

II. Total Bacterial Quantification via qPCR (quantitative PCR) for Total Bacterial Content

Quantitative PCR (qPCR), also known as Real-Time PCR, is a method that measures the number of copies of a DNA region defined by a particular PCR primer(s). Using this method, we specifically amplify the 16S amplicon and quantify the total bacterial content in each sample to determine the total bacterial load. This technique can also be used to determine the feasibility of your microbiome study by providing an exact measure of bacterial abundance in your samples. It is highly recommended using this technique prior to sequencing samples suspected to have low biomass, to determine whether these samples will result in positive amplification.

III. Shotgun Metagenome Sequencing

Shotgun metagenome sequencing is performed for taxonomic profiling (diversity and abundance), as well as functional analysis. This technique allows for parallel sequencing of DNA from all organisms within the community, with high coverage for species and strain level detection. The data generation allows for more advanced reporting and genome assemblies.

IV. Shallow Shotgun Metagenome Sequencing (SSMS)

Involves sequencing samples at a shallower depth than is applied in full shotgun metagenome sequencing. By combining many more samples into a single sequencing run and using a modified protocol that uses a lower volume of reagents for sequencing library preparation, SSMS is an economical way to provide compositional and functional sequencing data similar to deep shotgun metagenome sequencing. For host-associated samples, host contamination can pose a problem.

SSMS is best suited for samples that are rich in microbes and contain low levels of host DNA contamination such as fecal samples. For example, skin and biopsies can have 30-90% host contamination, which can lead to insufficient usable data with a SSMS approach.

V. Short Chain Fatty Acid Analysis

Short chain fatty acids (SCFA) are the products of fermentation of insoluble fibre from diet (eg. cellulose, resistant starch) by the bacteria in the gut. These fatty acids have been shown to play an important role in regulating metabolism in the gut and are closely associated with gastrointestinal diseases such as irritable bowel syndrome and other conditions such as obesity. By quantifying SCFA in stool, one can monitor gut health and inflammation. Learn more about how to interpret your fecal short chain fatty acid data below.

🔗 Interpreting your fecal short-chain fatty acid data

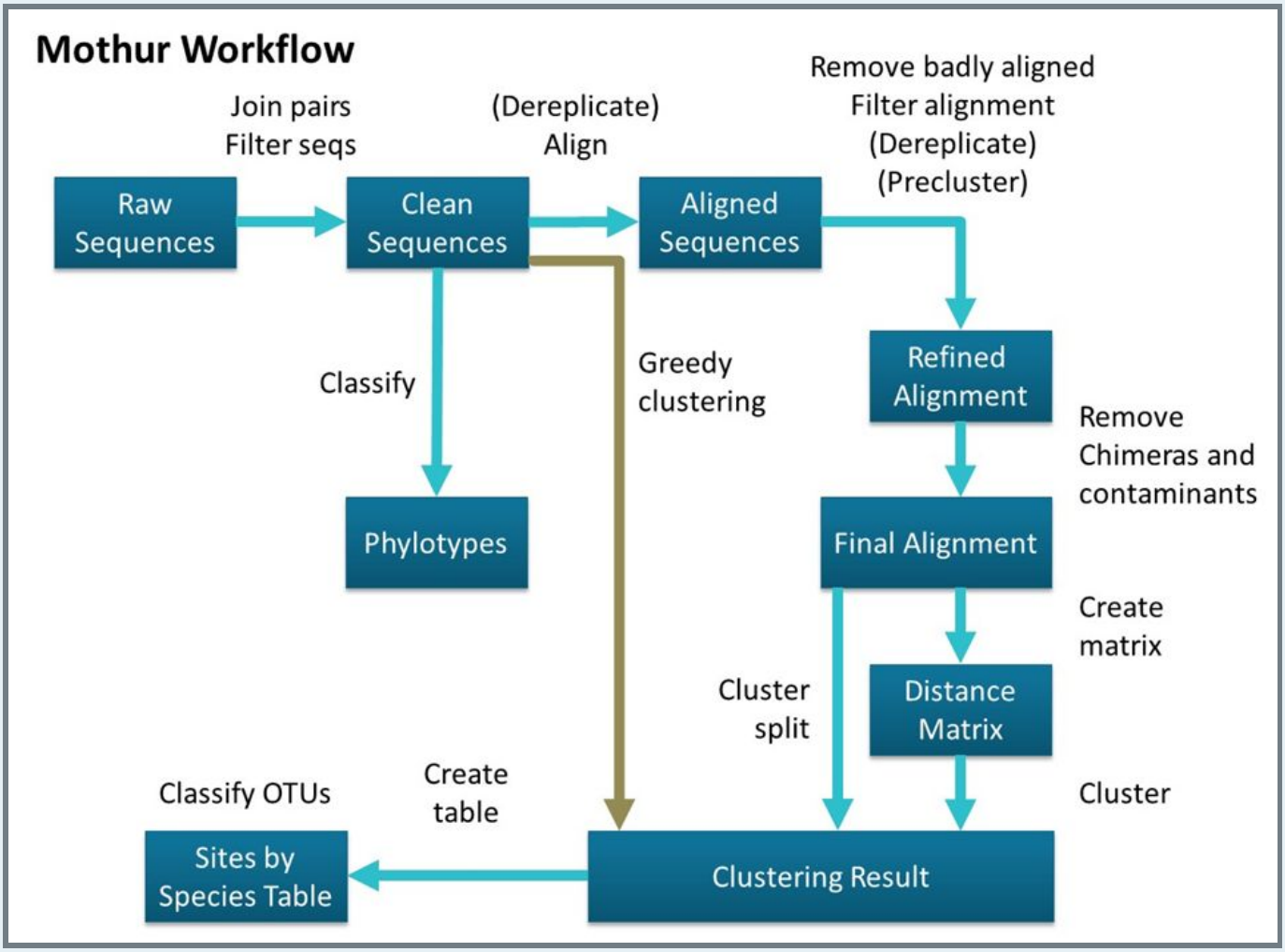
The volatility of SCFAs in stool samples, however, poses a challenge when it comes to research protocols. If you are interested in incorporating SCFA analysis into your study, it is important to consider your sample collection and stabilization options, as well as temperature control. Having received questions from our clients, our team put together a short internal investigation on how storage time, temperature, and stabilization reagent affect the quantities of SCFAs detected in a sample. Our white paper details the results of this investigation and is available below.

🔗 Factors that affect short-chain fatty acid preservation in human stool samples

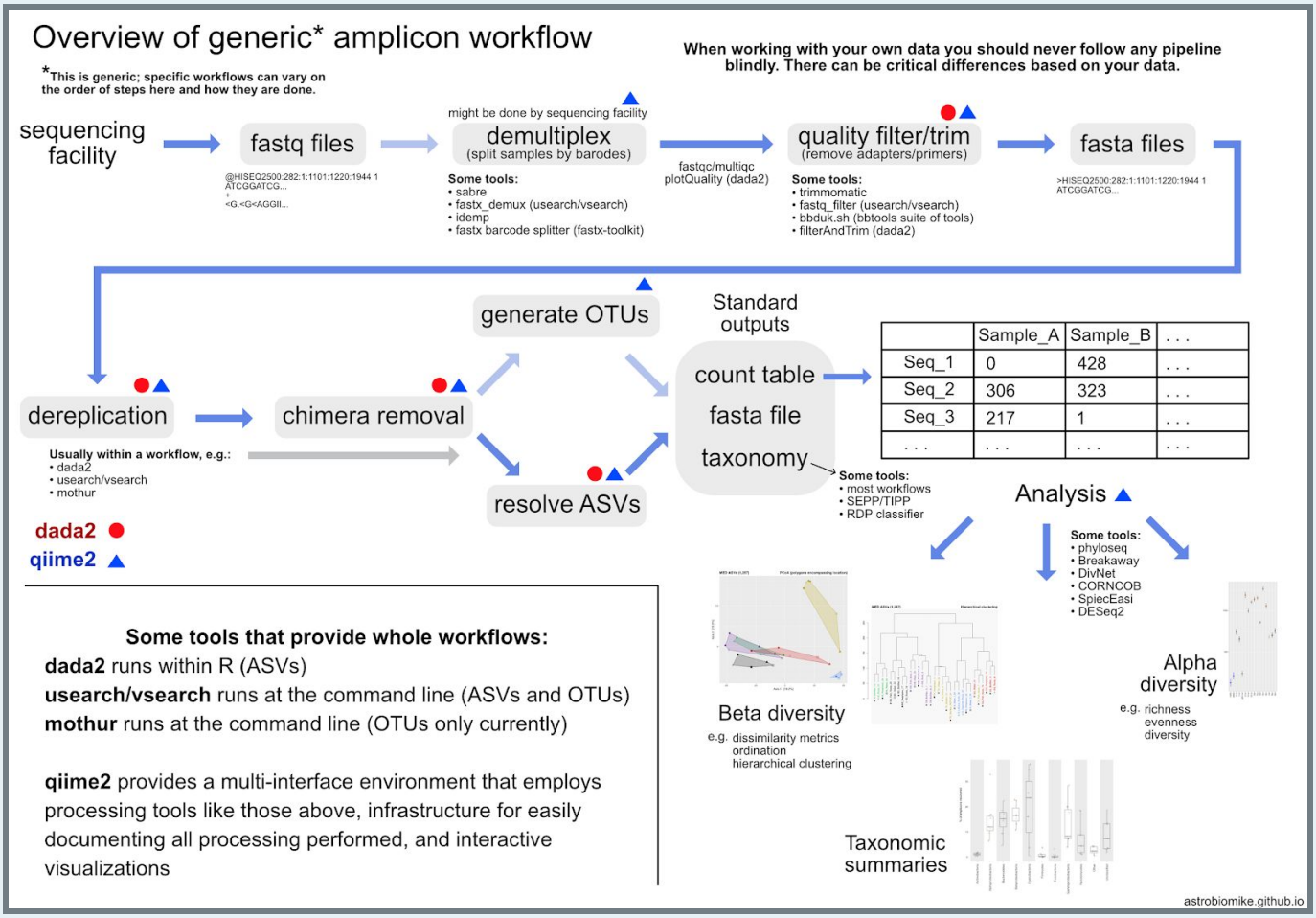
Bioinformatics, data visualization and reporting

Bioinformatics is an interdisciplinary field of science that combines biology, computer science, information engineering, mathematics and statistics to analyze and interpret the biological and clinical data.

Once your sequencing step has been completed, you should be left with the raw sequencing files in FASTQ format. FASTQ format is the standard for storing the output of high-throughput sequencing instruments such as the Illumina sequencers. From here there are several subsequent steps required to process the data before you are ready for data analysis and report generation. There are multiple methods of data processing. One of the widely used tools for amplicon data is Mothur. Below is a schematic of a Mothur workflow:



Here is an overview of a general amplicon workflow borrowed from https://astrobiomike.github.io/amplicon/



Summarizing the data

One way of summarizing data is to calculate diversity.

Diversity is a low dimensional representation of the entire community.

- Alpha-diversity is a summary of one community, for example, species richness, Simpson's index, inverse simpson, Shannon diversity, Shannon's richness. Different indices show different things. For example, species richness emphasizes total microbial taxa present in a sample and not taxonomic evenness. Shannon's index accounts for both taxonomic abundance and evenness.
- Beta diversity is a summary of multiple communities and how one community relates to another.

Limitations:
We just observe a sample from the community; however, real abundances or number of species are unknown. The classical approach is to assume the sample represents real diversity.

Diversity gives you compositional data, but change in diversity won't tell you which taxa have changed.

Multiple choices:

- Taxonomic level (strain, species, genus)
- Which diversity parameters
- Which estimate of diversity parameter

Data visualization

Data can help us tell the story. We need to keep the audience in mind. Context is important and so are the design and colors used.

Available report formats

- For amplicon studies standard bioinformatic reporting includes plots showing taxonomic composition, alpha diversity, beta diversity (ordination), differential abundance testing, and multivariate analysis (permutational analysis of variance). This is accompanied by raw bioinformatics outputs (OTU table, taxonomic classification of each OTU), differential abundance results, and relative abundances of taxa at various taxonomic levels.

  🔗 Download Amplicon Sequencing Sample Report

- For shotgun metagenome sequencing high-resolution taxonomic and functional profiles are analyzed observing the same principles that guide the analysis of amplicons. Our main metagenomics pipeline is based on tools developed to characterize human microbiomes -- specifically HUMAnN2, MetaPhlAn2, and StrainPhlAn form the Huttenhower lab -- but we can also accommodate approaches better suited to studying environmental microbiomes.

  🔗 Download Shotgun Sequencing Sample Report

- Advanced bioinformatic reporting may include:
  - Microbial epidemiology
  - Multiple regression
  - Partitioning sources of variation
  - Meta-analysis
  - Multi-omics integration analysis
  - Clustering algorithms
  - Machine-learning-based modelling
  - Biomarker discovery

Glossary

Finally, we will conclude the guide with a glossary of some technical terms used in microbiome studies.

Operational taxonomic unit
Operational taxonomic units (OTUs), are mathematical units used to classify groups of closely related sequences. OTUs are proxies for microbial 'species'. OTUs defined at 97% sequence similarity are loosely estimated as a species. The choice of algorithms and/or errors in sequencing strongly influence OTU calculations.

Diversity
The amount of variation of microbial community structure. Diversity within a sample, is referred to as alpha diversity, and diversity between samples referred to beta diversity. Measures of diversity are derived from tables of relative abundance and/or prevalence.

Alpha Diversity
Richness is simply the sum of unique OTUs found in each sample. Shannon diversity utilizes the richness of a sample along with the evenness (how evenly distributed the OTUs are) of the present OTUs to calculate a diversity index.

Beta Diversity
All profiles are inter-compared in a pair-wise fashion to determine a dissimilarity score and store it in a distance dissimilarity matrix. Distance functions produce low dissimilarity scores when comparing similar samples. Abundance-weighted sample pairwise differences are often calculated using Bray-Curtis dissimilarity. Bray-Curtis dissimilarity is calculated by the ratio of the summed absolute differences in counts to the sum of abundances in the two samples (Bray and Curtis 1957).

Ordination
Ordinations are dimensional-reduction techniques that are used to visualize complex relationships between communities. Using a dissimilarity matrix (see above) ordinations attempt to display sample-to-sample relationships in a 2-dimensional plot. Common methods include Principal Coordinates Analysis (PCoA) and Non-Metric Multidimensional Scaling (NMDS) analysis.

PERMANOVA
Permutational Analysis of Variance (PERMANOVA) tests for significant differences in the whole microbiome among discrete categorical or continuous variables. The samples are randomly reassigned to the various sample categories (Monte-Carlo Permutations), and the between-category differences are compared to the true between-category differences. PERMANOVA utilizes the sample-to-sample distance matrix (Bray-Curtis) directly to perform the calculation.

Relative abundance
Relative abundance refers to how common or rare an OTU is relative to other OTUs in the same community. In beta diversity analyses, relative abundance is useful for monitoring changes in a taxonomy's population between samples. The abundance scores are the counts of reads assigned to each OTU.

Dissimilarity matrix
A matrix of sample-to-sample dissimilarity values as calculated by the metric of choice. The Bray-Curtis dissimilarity matrix is the most conventionally used method to assess beta diversity.

Differential abundance testing
This is an estimate of variance-mean dependence in count data from high-throughput sequencing assays and tests for differential expression based on a model using the negative binomial distribution. See the likelihood-ratio test.

Negative binomial distribution
The negative binomial distribution is a discrete probability distribution of the number of successes in a sequence of independent and identically distributed Bernoulli trials before a specified (non-random) number of failures occurs. This is a useful distribution to model count-based data (such as microbiome), where biological replicates are low, and the data are not normally distributed. Statistical tests that utilize assumptions of normality cannot be used to accurately assess means and variances of microbiome data.

Likelihood-ratio test
The LRT examines two models for the counts, a full model with a certain number of terms and a reduced model, in which some of the terms of the full model are removed. The test determines if the increased likelihood of the data using the extra terms in the full model is more than expected if those extra terms are truly zero. The LRT is therefore useful for testing multiple terms at once, for example testing 3 or more levels of a factor at once, or all interactions between two variables. The LRT for count data is conceptually similar to an analysis of variance (ANOVA) calculation in linear regression, except that in the case of the Negative Binomial GLM, we use an analysis of deviance (ANODEV), where the deviance captures the difference in likelihood between a full and a reduced model.

Have a question for us?

Would you like to talk with one of our scientists? We're here to help.

Book a Meeting | Email Us | Call: +1 866-678-9544

Are you planning a study and require a sequencing provider?

🔗 Request a quote for your microbiome study

Visit our website:

microbiomeinsights.com