ZYMO RESEARCH
The Beauty of Science is to Make Things Simple

Considerations for Building an Accurate Shotgun Metagenomic Workflow for Gut Microbiome Profiling

Introduction

The human gut microbiome plays an important role in human health and various diseases, including many with no existing obvious correlations with the gut microbiome. Interest in using high-throughput DNA sequencing techniques to study the human gut microbiome has grown exponentially across many sectors, from academia to industry to the government.

As sequencing costs continue to drop, more researchers are considering using a shotgun metagenomic workflow rather than a targeted sequencing workflow (e.g., 16S rRNA gene sequencing) to obtain information around potential metabolic function, strain-level differences, or even the presence of antibiotic resistance genes. Interest in using RNA-seq to confirm predicted gene functions from metagenomic data is growing rapidly as well.

In this article, we discuss important considerations for building a good shotgun metagenomics workflow for microbiome profiling through real-world examples. Readers simply looking for the workflow that we recommend can refer to Table 1 for reagents and kits and Table 2 for an example workflow. Finally, Table 3 outlines key challenges and major considerations when building a shotgun metagenomic workflow for gut microbiome profiling.

Overview of Essential Challenges and Key Considerations

One common purpose of shotgun metagenomic sequencing is to determine the microbial composition, i.e., to identify the identity (often, up to species or strain-level) and abundance of microbes in a sample. Just like any other measurements, microbiome profiling can be assessed by accuracy, reproducibility, and sensitivity. From a practical point of view, simplicity, cost, throughput, automation, robustness, and biosafety must all be considered, especially when dealing with a large number of fecal samples. Shotgun metagenomic sequencing of gut microbiome samples is a long process consisting of six major steps. Each step has specific challenges (summarized in Table 2), which are outlined below.

(1) Microbiome Quality Control

As with any other scientific experiment, positive and negative controls should be used in all microbiome sequencing protocols, especially considering that the workflow is often complicated and vulnerable to improper handling. Without good microbiome controls, the performance of a workflow cannot be assessed. It is just like running a DNA electrophoresis gel without a ladder.

We previously discussed how lack of controls leads to poor data quality and poor data reproducibility across microbiome studies [1]. To address this issue, Zymo Research launched the Microbiome Standards and Controls Initiative (M-SCI) (https://www.zymoresearch.com/pages/m-sci), encouraging researchers to include reliable microbiome standards in their measurements.

(2) Sample Collection and Preservation

Microbes can grow or react incredibly quickly to changes in their environment (e.g., changes in nutrients, pH, temperature, oxygen level, etc.). The human gut is an anaerobic environment populated by obligate and facultative anaerobes. When feces leave the body, the change in environment, particularly oxygen levels, is drastic. Facultative anaerobes, such as *E. coli*, are happy to use oxygen as the perfect electron acceptor for energy production, while strict anaerobes can be killed within minutes if the sample is not preserved appropriately for maintaining anaerobe viability.

Because of these issues, many microbiome profiles are altered right at the start of the workflow. The American Gut Project [2], which asks participants to collect fecal samples on a sterile swab at home and then mail the sample back without any preservatives or temperature control, reported a significant increase in bacteria in the Proteobacteria phylum (specifically, Gammaproteobacteria). Microbiome profiles were artificially corrected by removing all sequences belonging to these "blooming" microbes at the risk of removing true observations. Similarly, we observed a bloom of Proteobacteria species in just one day when a fecal sample was left at room temperature without preservation (Figure 1).

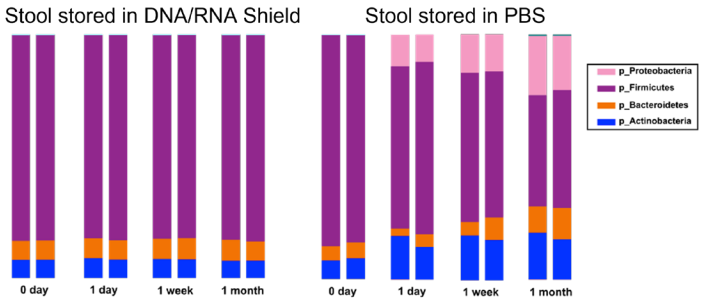

Figure 1. The phylum-level microbiome composition development of a fecal sample when left at ambient temperature with and without a preservative.

Freezing samples immediately and transporting the samples on dry ice is the gold standard, but is difficult in practice, particularly when asking study participants to collect samples at home. Even the most astute participants cannot control for automatic freeze-thaw cycles in home freezers, which can alter microbial communities considerably. In a case study, we found that the whole phyla of Bacteroidetes disappeared after five freeze-thaw cycles (Figure 2). It appeared that Gram-negative microbes are more vulnerable to the damage of free-thaw cycling.

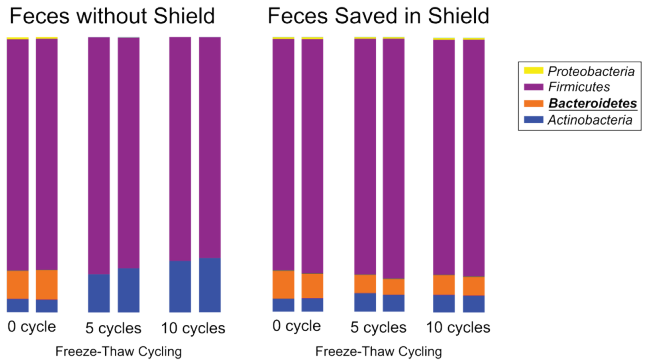

Figure 2. DNA/RNA Shield protects the microbiome profile of a fecal sample from freeze-thaw cycling damage.

Biosafety is another important concern when dealing with fecal samples and should be addressed during sample collection and transportation and DNA extraction. This is especially true in the middle of the SARS-CoV-2 pandemic as fecal shedding of the virus is frequently detected [3], and many studies transport biological samples across borders. An important step in DNA extraction is lysis, often done mechanically, which leads to almost unavoidable accidental leakages or spills. If fecal DNA cannot be extracted from fecal samples immediately after collection, the only way to eliminate biosafety concern is to inactivate microbes in feces as soon as they are collected.

(3) DNA Extraction

It is concerning to consider that poor quality data have made it almost impossible to compare microbiome studies coming out of different laboratories. The DNA extraction step appears to be the major source of variation causing this irreproducibility. Many research groups have reported dramatic variations in microbiome profiles when different DNA extraction methods are used [4-9]. This begs the question: what is the cause of variation during DNA extraction?

Mechanical lysis with bashing beads and bead beating devices is generally accepted as the gold standard for microbial lysis. But, commercially available DNA extraction kits come with different bashing beads in a wide variety of sizes and materials. Using the ZymoBIOMICS Microbial Community Standard and the Measurement Integrity Quotient (MIQ) score [10], we observed varying levels of performance among the different bead types and sizes (figure 3).

Additionally, there are many bead beating devices from different suppliers, each varying in power, capacity, and most importantly, the motion of the mechanism. We observed that the motion of the bead beater, not the speed, is the most important consideration. Generally, motions that create more random collisions between beads yield better results than linear motion; i.e., shaking back and forth or vertically and horizontally (Figure 4).

Unbiased microbial lysis is crucial because the human gut microbiome contains species with a wide range of recalcitrance to lysis. The most abundant microbes in the typical human microbiome come from four bacterial phyla: Bacteroidetes, Proteobacteria, Firmicutes, and Actinobacteria. Bacteroidetes and Proteobacteria are mainly Gram-negative microorganisms and are therefore easier to lyse; conversely, Firmicutes and Actinobacteria are mainly Gram-positive microorganisms and don't lyse as readily. This is an important consideration as there has been a large controversy in the field regarding the significance of the Bacteroidetes:Firmicutes ratio and its correlation with disease, especially metabolic diseases [11].

(4) Shotgun Library Preparation

Compared with 16S rRNA gene sequencing, which relies on PCR amplification of the target gene, shotgun metagenomic sequencing is generally less biased. Nevertheless, the library preparation process for shotgun metagenomics is more complicated and can introduce bias in downstream microbiome profiling. There are two popular kinds of library prep kits for shotgun sequencing: ligation-based kits and tagmentation-based kits. Both kits have a DNA fragmentation step, which is where some bias can be introduced, particularly when using enzymatic fragmentation as different enzymes may have different cut-site preferences. Bias can also be introduced in the final PCR step, due to inefficiencies of some polymerases for genomes with low GC content (Figure 5).

(5) Illumina Sequencing

Shotgun sequencing is generally performed at a greater depth compared with 16S rRNA gene sequencing. But how many reads do you really need? 1 million? 10 million? 20 million? Answering this question is critical because the target number of reads per sample determines which sequencing kit, sequencer, and multiplexing strategy are appropriate. And because more reads are more expensive, researchers often must balance sensitivity and cost. Assuming the major objective is to profile microbial composition, sequencing depth will affect only the detection limit of microbes. In a fecal sample where nearly 100% of the sequencing reads are derived from microbes, and 100 reads per genome is the minimum required to detect a species, 1 million

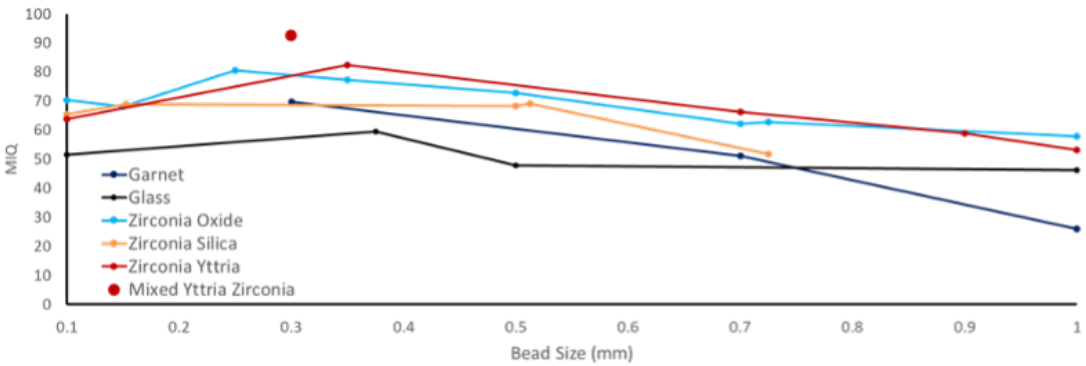

Figure 3. The MIQ score was used to evaluate the ability of different bead materials and sizes to effectively lyse the ZymoBIOMICS Microbial Community Standard.

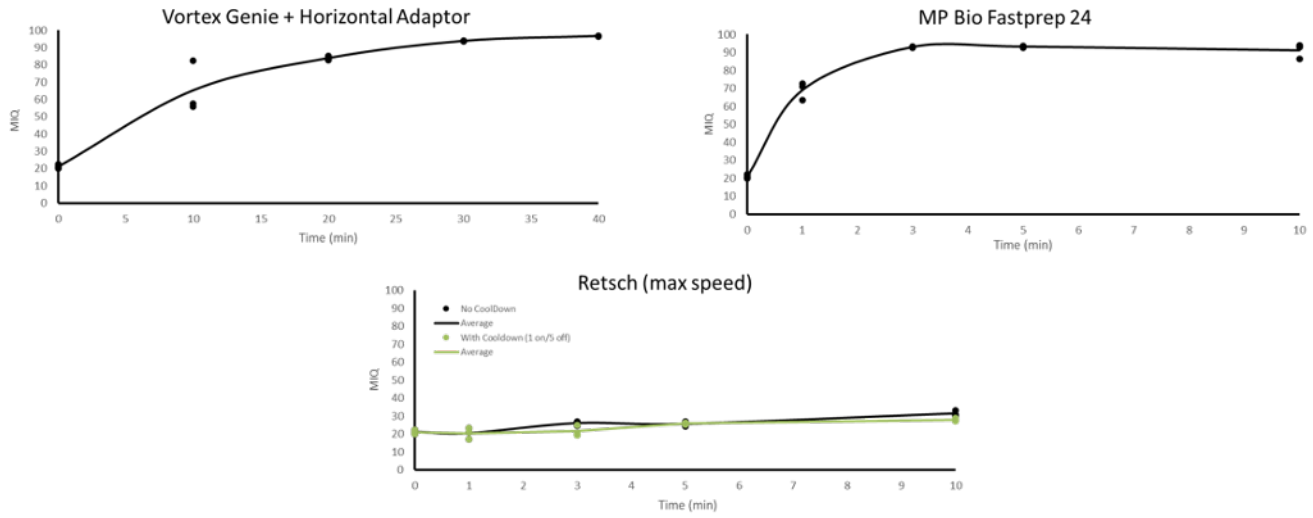

Figure 4. The MIQ score was used to evaluate the performance of several bead beaters using a mixture of 0.1 and 0.5 mm zirconia yttria beads and the ZymoBIOMICS Microbial Community Standard.

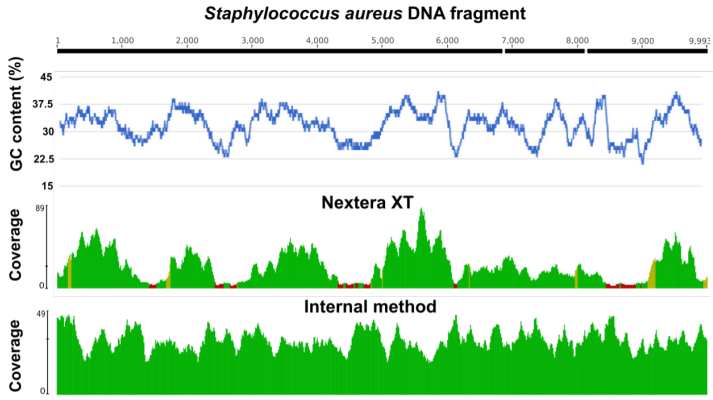

Figure 5. Shotgun sequencing of the ZymoBIOMICS Microbial Community DNA Standard. The coverage of a ~10k genomic region from the genome of *Staphylococcus aureus* was picked. The sequencing coverage distribution with Nextera XT (Illumina) synchronizes with the GC content variation graph, reducing coverage in regions with low GC density.

sequencing reads total is sufficient for detecting microbes comprising only 0.01% of the population (=100/1,000,000). Increasing sequencing depth ten-fold to 10 million reads per sample increases costs ten-fold but also increases sensitivity, enabling detection of very rare microbes comprising 0.001-0.01% of the population. Whether you need this level of sensitivity really depends on the goal of your research. Simple taxonomic profiling can get away with the popular shallow shotgun sequencing protocols; however, strain-level analyses or studies seeking to quantify antibiotic resistance markers will benefit from deeper sequencing protocols. A research group's chosen bioinformatics pipeline for analyzing sequencing results, discussed in more detail below, can also affect which sequencing depth is appropriate.

(6) Bioinformatics Analysis

Because researchers are continuously working to improve accuracy and reduce computational cost/time, there are many bioinformatics tools available. Unsurprisingly, these bioinformatics tools varied performance as reported in the primary literature. McIntyre et al. [12] compared the performance of 11 different shotgun metagenomic classifiers and found that no one is perfect; pairing tools with different classification strategies can be beneficial.

The basic principle of these pipelines is not complicated: assembly-free pipelines simply compare raw sequencing reads or read pairs directly with a reference database to assign taxonomy. Available tools generally fit into one of two categories based on the reference database they use. One category uses whole genomes as references (for example, Kraken [13] and Centrifuge [14]); the other uses a set of marker genes (i.e., a subset of the genome) as references (for example, mOTUs [15] and MetaPhlAn2 [16]).

Both types of pipelines have advantages and disadvantages. The advantage of using only a subset of marker genes is a dramatic reduction in computation cost and time. It is also easier to curate a subset of marker genes compared to whole genomes, and a better curated database equates to higher specificity. However, sensitivity suffers because only a small proportion of the reads that belong to chosen marker genes are used, and the majority of reads are simply wasted. Because a genome contains all genetic information that can define a microbial strain, shotgun metagenomic sequencing has the power to differentiate closely related microbial strains —but this power is lost when using marker gene-based databases for taxonomic classification. In contrast, using whole genomes solves this problem at the cost of an extremely high requirement on computation resources (hard drive storage, memory and

CPU cores). Even with sufficient resources, computation often takes a long time, requiring the use of cloud or supercomputer clusters.

Because most bacterial species have never been sequenced (even a large proportion of sequenced species are represented by unfinished genomes), false negatives are a common problem among metagenomic datasets. Even the human gut microbiome, comprised of more species that have been sequenced, suffers from limited coverage in taxonomy databases.

Even more concerning than false negatives are false positives, which are a common issue with pipelines using whole genome-based databases. Reads derived from species A can be mistakenly assigned to closely related species B simply because of mutations or chimeras introduced by PCR during library prep, or, more commonly, because of sequencing errors. Errors can also exist in the database itself; draft genomes can be contaminated by sequences from other species and is a common occurrence among public databases.

Recommended Workflow for Shotgun Metagenomic Sequencing of the Gut Microbiome

Given the multiple points of entry for variation described above, it can be overwhelming to know how to ensure your microbiome data are clean, standardized, and accurate. Below, we outline recommendations for addressing each of the challenges along the metagenomics pipeline.

(1) Microbiome Quality Control

In 2017, Zymo Research released the first commercial microbiome standard, the ZymoBIOMICS Microbial Community Standard (D6300), to address the urgent need for benchmarking controls in the microbiome field. Since then, we have been actively promoting the importance of validating microbiome workflows using microbiome standards. In 2020, in response to customers' requests for a mock microbial community mimicking the human gut microbiome, the ZymoBIOMICS Gut Microbiome Standard (D6331), consisting of 21 microbial strains mostly derived directly from the human gut, was released. This was followed by the 2021 release of a second control specific to the gut microbiome; the ZymoBIOMICS Fecal Reference with TRUMatrix Technology (D6323) was created from human feces to accurately represent the diversity of the human gut microbiome. Both can be used as defined inputs to assess and improve the accuracy and robustness of a given gut microbiome workflow. After workflow validation, these standards can also be used as routine quality controls (i.e., positive controls). Zymo Research offers DNA extraction and metagenomic sequencing services that always include a positive control (i.e., a microbiome standard) and a negative control (for detecting contamination) per 96-well plate.

Certain scenarios, such as clinical microbiome research, might require stricter quality control measures, such as a spike-in control for each individual sample. To address this need, the ZymoBIOMICS Spike-in Control I, consisting of two non-gut species, *Allobacillus halotolerans* and *Imtechella halotolerans*, was created. This control can be spiked into every fecal sample. Detecting these two microbes in a sample indicates the workflow is working as expected; if not detected in a sample, that sample may need to be re-sequenced or excluded from downstream analyses. And because the two microbial strains are never seen in the human gut microbiome, their sequences can be easily filtered out without impacting the rest of the study results.

(2) Sample Collection and Preservation

As explained above, a cost-effective method of microbiome sample inactivation and nucleic acid preservation is needed by microbiome researchers to avoid biases introduced by bacterial growth or nucleic acid degradation during sample transport. Zymo Research has developed a liquid preservative called DNA/RNA Shield, which can preserve a sample's microbiome profile (both DNA and RNA) at ambient temperature for over one month (Figure 1). The mixture of chemicals penetrates microbial cells, killing viable cells (and thus eliminating pathogens, including bacteria, fungi, and viruses) while inactivating DNases and RNases. DNA/RNA shield is available in pre-filled collection devices called the DNA/RNA Shield Fecal Collection Tube (R1101) and the DNA/RNA Shield Collection Tube w/Swab (R1107). These devices make sample collection both comfortable and simple for people to use in their own homes, and eliminate common challenges associated with shipping biological samples. In addition to its intended purpose, DNA/RNA Shield is also a widely used reagent for preserving nasal swabs, throat swabs, and saliva samples for SARS-CoV-2 testing. It is compatible with most commercially available DNA extraction kits. The Zymo Research team can provide any guidance needed for research groups using those extraction kits. If using Zymo Research DNA extraction kits, material can be taken from the DNA/RNA Shield tube and added directly into the lysis tube for the first step of DNA extraction.

For projects that won't process samples immediately or that will perform additional, future processing, samples in DNA/RNA Shield can be stored long-term at -80°C. Samples are additionally protected against freeze-thaw cycles that could occur due to power outages or equipment failure. To the best of our knowledge, this is the only commercially available product that can preserve both DNA and RNA (Figure 6). Even if RNASeq isn't part of the original research plan, long-term storage can facilitate the addition of metatranscriptomics analyses later if desired.

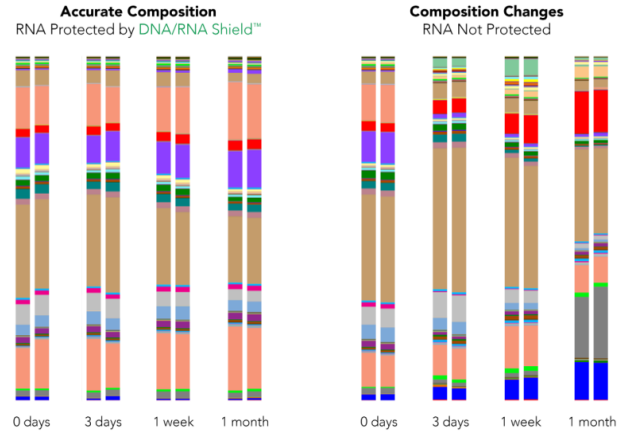

Figure 6. DNA/RNA shield preserves metatranscriptomic profile at ambient temperature for a month. The figure shows the taxonomy composition (genus level) derived from RNA-Seq data of a fecal sample saved at ambient temperature with (left) and without DNA/RNA Shield (right).

(3) DNA Extraction

As discussed above, most of the bias caused during DNA extraction is due to the lysis process. While mechanical lysis has been generally accepted as the gold standard method, not all protocols perform equally and not all mechanical bead beaters are qualified. Using the ZymoBIOMICS Microbial Community Standard for validation, we identified a list of mechanical bead beaters and working protocols that can minimize bias during microbial lysis (https://files.zymoresearch.com/documents/bead_beating_short_protocol_tables.pdf). Some of the high-speed homogenizers cost thousands of dollars, but for most users, a horizontal plate adaptor for a vortexer (such as the Vortex Genie 2) is an affordable and effective choice. Bead beating with the Vortex Genie 2 for 40 minutes is able to achieve similar extend of lysis with other high-speed bead beaters (Figure 4). To scale up, additional vortexers can simply be used in parallel. This is the approach used by Zymo Research when processing samples for customers, with up to 10 Vortex Genie 2 instruments operating simultaneously.

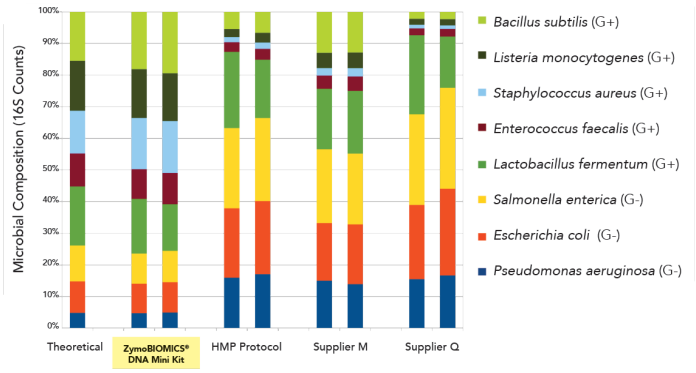

Figure 7. Comparing four different DNA extraction protocols using ZymoBIOMICS Microbial Community Standard (D6300). HMP protocol, Human Microbiome Project DNA extraction protocol. Supplier M, PowerSoil Kit (originally from MoBIO Inc). Supplier Q, QIAamp DNA Stool Mini Kit from Qiagen. The profile was determined by 16S rRNA gene sequencing targeted V3-V4 region.

In addition to the device used for mechanical lysis, the bead matrix used can significantly affect lysis. We previously reported that generally, small beads (0.1mm) are needed to lyse small bacteria and large beads (0.5mm) are required to lyse bigger microbes (such as yeasts). To avoid bias toward one type of microbe over another, ZymoBIOMICS extraction kits include a mixture of 0.1mm and 0.5mm beads. Beads made of high density, hard materials also maximize lysis efficiency (figure 3). Zirconia yttria, being the most effective material, is featured in the ZymoBIOMICS extraction kits. Comparing different DNA extraction kits using the ZymoBIOMICS Microbial Community Standard (D6300), the ZymoBIOMICS DNA Miniprep kit produced a microbial community profile equivalent to the defined composition of the standard, while the other two protocols (PowerSoil kit and HMP DNA extraction protocol) overestimated the abundance of easy-to-lyse Gram-negative bacteria (Figure 7). When the same comparison was performed with a real fecal sample, a similar phenomenon was observed, with the abundances of Gram-negative Bacteroidetes and Proteobacteria greatly overestimated (Figure 8). Wesolowska-Andersen et al [9] reported similar results, finding that compared with the MetaHIT DNA extraction protocol from MetaHIT, the HMP protocol recovers more Bacteroidetes [9].

After lysis, DNA is extracted using either silicon column-based manual DNA extraction or magnetic bead-based automatic DNA extraction. The ZymoBIOMICS DNA Miniprep kit (D4300), which has a very streamlined protocol enabling DNA extraction in as little as 20 minutes, is recommended for manual extraction. For automation or high-throughput users, the ZymoBIOMICS 96 Magbead DNA Kit (D4308) is recommended. This kit has a streamlined workflow consisting of only 5 steps: binding, pre-wash, wash (2x), and elute. Each step consists of simple operations and therefore can be easily programmed on different automatic liquid handlers (such as the Tecan DreamPrep NAP workstation and Hamilton Microlab Star) or automatic magnetic bead transfer devices (such as the KingFisher Flex). With KingFisher Flex and pre-aliquoted reagents, DNA can be extracted from fecal lysate from 96 samples in only 45 minutes.

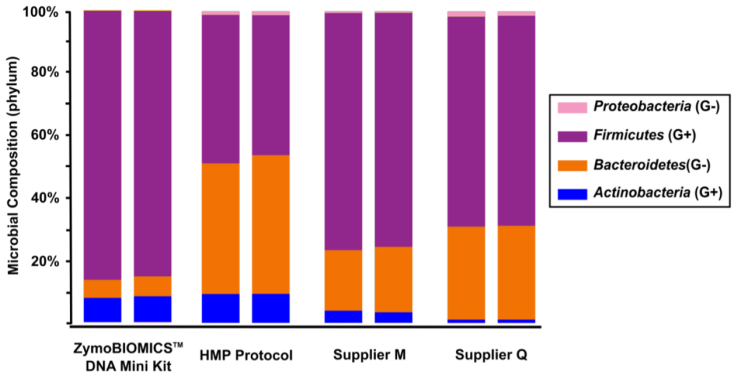

Figure 8. Comparing DNA extraction kits with a single fecal sample. The profile (at phylum level) was determined by 16S rRNA gene sequencing targeted V3-V4 region.

(4) Shotgun Library Preparation

For shotgun library preparation, the Illumina DNA Prep kit, formerly named the Nextera Flex, is recommended for its superior performance and minimal DNA input requirements. Unlike its predecessor, the Nextera XT kit, the Flex kit doesn't suffer dramatically from GC content related bias (Figure 5). Performance suffers only with regions with extremely high GC content; therefore, this kit is a great balance between performance and convenience.

(5) Illumina Sequencing

Given enough amount of samples to multiplex and process in one lane, NovaSeq is currently the most cost-effective choice for shotgun metagenomic sequencing. The sequencing reagent only costs ~ $17 for 10 million paired-end reads (i.e. 20 million reads in total). In contrast, it costs ~$1700 for 20 million paired-end reads when sequenced on MiniSeq. For applications that require deep sequencing like metagenomic sequencing, NovaSeq is the best choice. But, one inconvenience of using NovaSeq is that the libraries need to be barcoded with unique dual indexes to prevent index swapping (https://www.illumina.com/techniques/sequencing/ngs-library-prep/multiplexing/index-hopping.html).

(6) Bioinformatics Analysis

Zymo Research also offers a complete metagenomic data analysis package (MetaGeRM) through the Zymo Research Microbiome Sequencing Service. The underlying search/match engine is Centrifuge [14], which matches reads against whole genomes for the sensitivity and taxonomy resolution. Additionally, Centrifuge is rapid and has a low computation memory requirement. For functional profiling, MetaGeRM

uses Humann2, which includes reads from bacteria, archaea, eukaryotes, and viruses. MetaGeRM also profiles antibiotics resistance genes and virulence genes based on curated databases.

Similar to other assembly independent search engines, Centrifuge performs taxonomy search with each individual pair of paired-end reads. After that, it simply compiles these taxonomy search results together. The major issue with this is that the compiled result may produce false positives. For example, when you use shotgun sequencing to sequence a mock microbial community, such as the ZymoBIOMICS Microbial Community DNA Standard (D6306), which contains 10 microbes, you can end up with >100 species identified. Most of these false positives are species that are closely related to the 10 microbes contained. Even if you only count species that have reads that are uniquely assigned to them, there are still some closely relative false positives left. Including the exact genomes of the 10 microbes into the reference database can dramatically reduce false positives. But unfortunately for an unknown fecal sample, this is never possible. However, even if you include all 10 genomes in the reference database, you can still see some false positives, because sequencing errors can cause reads being assigned to closely related species. Actually, the presence of false positives is a common and difficult issue for all metagenomic taxonomy classifiers, especially those that use whole genomes as references (such as Kraken and Centrifuge).
But MetaGeRM uses proprietary methods to reduce false positives. In targeted amplicon sequencing (e.g. 16S rRNA gene sequencing), Dada2 [17] is a revolutionary tool that can effectively reduce false positives that are caused by sequencing errors and PCR chimera. Dada2 builds error correction models and is able to infer real amplicon sequences from the raw data. The principle of Dada2 is only applicable to amplicon sequencing data that have deep sequencing depth of targeted reagents. But fundamentally, it uses the fact that the reads from one sample are intrinsically related to each other. We found that by utilizing these relations, we can also significantly reduce false positives in shotgun metagenomic taxonomy classification, especially false positives that are caused by sequencing errors. In the case of a mock community, our MetaGeRM pipeline is able to achieve zero false positives while other existing pipelines cannot.

Discussion

The use of shotgun metagenomic sequencing to analyze the human microbiome is becoming increasing popular with many large companies, start-ups, hospitals, and universities beginning to apply the method for clinical and diagnostic applications. With such important applications, data accuracy is crucial and with such a complex workflow, there are many considerations to address. The choice of controls and methods for sample collection, DNA extraction, library prep, and bioinformatics are all important for minimizing bias and producing the highest quality microbiome data.

Table 1. Recommended reagents/kits for the workflow of shotgun metagenomic sequencing of gut microbiome.

Major Components	Recommended Kits/Products	Cost Estimation (per prep)	Key Considerations
Quality Control	ZymoBIOMICS Gut Microbiome Standard (D6331)	$0.4	Mock microbial community of 21 microbial strains from human gut. Containing strains from Bacteria, Archaea and Fungi. Ideal for assessing bias and consistency of the workflow.
	ZymoBIOMICS Fecal Reference with TruMatrix™ (D6323)	$0.3	A reference material created using real human fecal samples. A true representation of microbial diversity of human gut. Ideal for fecal workflow validation and assessing measurement consistency.
	ZymoBIOMICS spike-in Control I (D6320)	$2.0	A quality control to spike into every sample. Consists of two alien microbes never seen in human microbiome.
Sample Collection and Transportation	DNA/RNA Shield Fecal Collection Tube (R1101)	$9.0	Cold-free sample collection and transportation. Preserving both DNA and RNA profiles of a fecal sample at room temp. for >6 months. Inactivate potential pathogens including SARS-Cov-2. Different formats of collection devices for various application needs.
	DNA/RNA Shield Collection Tube w/ Swab (R1107-E)	$6.5	
DNA Extraction	ZymoBIOMICS DNA Miniprep Kit (D4300)	$5.5	Unbiased microbial DNA extraction with mechanical lysis. PCR-inhibitor free.
	ZymoBIOMICS 96 DNA Magbead Kit (D4308)	$4.9	96 samples from lysate to extracted DNA in 1hour. Extremely easy workflow to automate (binding, 3 washes and elution). Scripts ready for major automation platforms.
Library Preparation	Illumina DNA Prep	$41.1	Streamline workflow and automation friendly. Normalization free for most fecal samples. Little bias caused by GC content variations.
DNA Sequencing	Illumina NovaSeq 6000	$17.0*	Most cost-effective regarding cost per base
Bioinformatics	Zymo Research Shotgun Metagenomics Pipeline	$40.0	Comprehensive whole genome database of bacteria, archaea, eukaryote and virus. Sophisticate pipeline with low false positives. Functional metagenomics analysis using Humann2.

* cost per 10 million paired-end reads or 20 million reads

Table 2. A recommended workflow for shotgun metagenomic sequencing of gut microbiome

Major Steps	Brief Workflow Description	Equipment/Resources	Processing Time
Sample collection	1. DNA/RNA Shield Fecal Collection Tubes (R1101) were mailed to an end-user as requested.	USDA/UPS/FedEx	a few days to a couple of weeks
	2. The end-user collected fecal samples and mailed them back to a sequencing facility by portal or courier services (cold-free).		
DNA Extraction	3. Samples from the collection devices were transferred to ZR BashingBead Lysis Tubes (0.1 & 0.5 mm). Perform mechanical lysis using Vortex Genie 2. Microbiome standards/controls (such as D6331, D6323 and D6320) can be incorporated into the workflow from here.	Vortex Genie 2, horizontal vortex adapter from Scientific Industry	1 hour
	4a. Lysate was transferred to a deep well plate and the DNA was extracted using ZymoBIOMICS 96 Magbead DNA kit (D4308) using automation platforms	KingFisher Flex, Tecan, Hamilton liquid handlers	1.5 hour for 96 samples
	4b. Alternatively, the samples can be processed manually using ZymoBIOMICS DNA Miniprep Kit (D4300)	Microcentrifuge	1.5 hour for 24 samples
Shotgun library prep	5. The extracted DNA in 96-well PCR plate was then subjected to the library prep using the Illumina DNA Prep with automation platforms or prepared manually. The libraries were barcoded using unique-dual indexes and pooled into one or a few final libraries.	96-well thermocyclers, Tecan or Hamilton liquid handlers, Qubit DNA Quantification	3 hours with automation for 96
Sequencing	6. The final libraries were sequenced on NovaSeq 6000. using S4 300 cycle kit	NovaSeq 6000,	~2 days
Bioinformatics	7. The sequencing data were demultiplexed and uploaded onto AWS cloud and analyzed using Zymo Research MetaGeRM Shotgun Metagenomic Pipeline on AWS EC2 instances.	Amazon Cloud Computing (AWS)	several hours

Table 3. Key challenges and major considerations when building a shotgun metagenomic workflow for gut microbiome profiling

Major Components	Key challenges and Major Considerations
Quality Control	What quality controls to use?
	How to use a spike-in control?
Sample Collection	How to create a protocol of sample collection that is easy to follow by regular people?
	How to preserve the microbiome profile immediately after collection?
	Is it possible to preserve both DNA and RNA profiles?
	How to avoid bias caused by freeze-thaw cycling
Sample Transportation	How to safely transfer fecal sample during the pandemic of SARS-Cov-2?
	Shipping with dry ice is complicated and costly.
	Biosafety concerns when transporting feces across border.
DNA Extraction	How to avoid bias during microbial lysis?
	How to avoid PCR inhibition?
	Biosafety concerns when processing feces in the lab due to accidental spills or leakage.
	How to process hundreds of sample in a day
	Which library prep kits to use?
Library Preparation	How to avoid bias during library prep?
	How to process hundreds of sample in a day?
DNA Sequencing	Which sequencer is the right choice?
	What sequencing depth is needed?
Bioinformatics	Which pipeline and reference database to use?
	How to assess the accuracy of the taxonomy prediction?
	How to reduce false positives?
	How to have coverage beyond Bacteria?
	How many microbes do I miss?
	How to know if a detected taxon is not a false positive?

References

1. https://www.zymoresearch.com/blogs/blog/microbiomics-macro-bias
2. Amir, A., McDonald, D., Navas-Molina, J. A., Debelius, J., Morton, J. T., Hyde, E., ... & Knight, R. (2017). Correcting for microbial blooms in fecal samples during room-temperature shipping. MSystems, 2(2), e00199-16.
3. Jones, D. L., Baluja, M. Q., Graham, D. W., Corbishley, A., McDonald, J. E., Malham, S. K., ... & Farkas, K. (2020). Shedding of SARS-CoV-2 in feces and urine and its potential role in person-to-person transmission and the environment-based spread of COVID-19. Science of the Total Environment, 749, 141364.
4. Hsieh, Y. H., Peterson, C. M., Raggio, A., Keenan, M. J., Martin, R. J., Ravussin, E., & Marco, M. L. (2016). Impact of different fecal processing methods on assessments of bacterial diversity in the human intestine. Frontiers in microbiology, 7, 1643.
5. Vishnivetskaya, T. A., Layton, A. C., Lau, M. C., Chauhan, A., Cheng, K. R., Meyers, A. J., ... & Onstott, T. C. (2014). Commercial DNA extraction kits impact observed microbial community composition in permafrost samples. FEMS microbiology ecology, 87(1), 217-230.
6. Hart, M. L., Meyer, A., Johnson, P. J., & Ericsson, A. C. (2015). Comparative evaluation of DNA extraction methods from feces of multiple host species for downstream next-generation sequencing. PloS one, 10(11), e0143334.
7. Kennedy, N. A., Walker, A. W., Berry, S. H., Duncan, S. H., Farquarson, F. M., Louis, P., ... & Hold, G. L. (2014). The impact of different DNA extraction kits and laboratories upon the assessment of human gut microbiota composition by 16S rRNA gene sequencing. PloS one, 9(2), e88982.
8. Sohrabi, M., Nair, R. G., Samaranayake, L. P., Zhang, L., Zulfiker, A. H. M., Ahmetagic, A., ... & Wei, M. Q. (2016). The yield and quality of cellular and bacterial DNA extracts from human oral rinse samples are variably affected by the cell lysis methodology. Journal of microbiological methods, 122, 64-72.
9. Wesolowska-Andersen, A., Bahl, M. I., Carvalho, V., Kristiansen, K., Sicheritz-Pontén, T., Gupta, R., & Licht, T. R. (2014). Choice of bacterial DNA extraction method from fecal material influences community structure as evaluated by metagenomic analysis. Microbiome, 2(1), 1-11.
10. https://www.zymoresearch.com/blogs/blog/how-to-quantify-bias-with-mock-microbial-community-standards
11. Magne, F., Gotteland, M., Gauthier, L., Zazueta, A., Pesoa, S., Navarrete, P., & Balamurugan, R. (2020). The firmicutes/bacteroidetes ratio: a relevant marker of gut dysbiosis in obese patients?. Nutrients, 12(5), 1474.
12. McIntyre, A. B., Ounit, R., Afshinnekoo, E., Prill, R. J., Hénaff, E., Alexander, N., ... & Mason, C. E. (2017). Comprehensive benchmarking and ensemble approaches for metagenomic classifiers. Genome biology, 18(1), 1-19.
13. Wood, D. E., & Salzberg, S. L. (2014). Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome biology, 15(3), 1-12.
14. Kim, D., Song, L., Breitwieser, F. P., & Salzberg, S. L. (2016). Centrifuge: rapid and sensitive classification of metagenomic sequences. Genome research, 26(12), 1721-1729.
15. Milanese, A., Mende, D. R., Paoli, L., Salazar, G., Ruscheweyh, H. J., Cuenca, M., ... & Sunagawa, S. (2019). Microbial abundance, activity and population genomic profiling with mOTUs2. Nature communications, 10(1), 1-11.
16. Segata, N., Waldron, L., Ballarini, A., Narasimhan, V., Jousson, O., & Huttenhower, C. (2012). Metagenomic microbial community profiling using unique clade-specific marker genes. Nature methods, 9(8), 811-814.
17. Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: high-resolution sample inference from Illumina amplicon data. Nature methods, 13(7), 581-583.

ZYMO RESEARCH
The Beauty of Science is to Make Things Simple®

www.zymoresearch.com | info@zymoresearch.com | Toll Free: (888) 882-9682