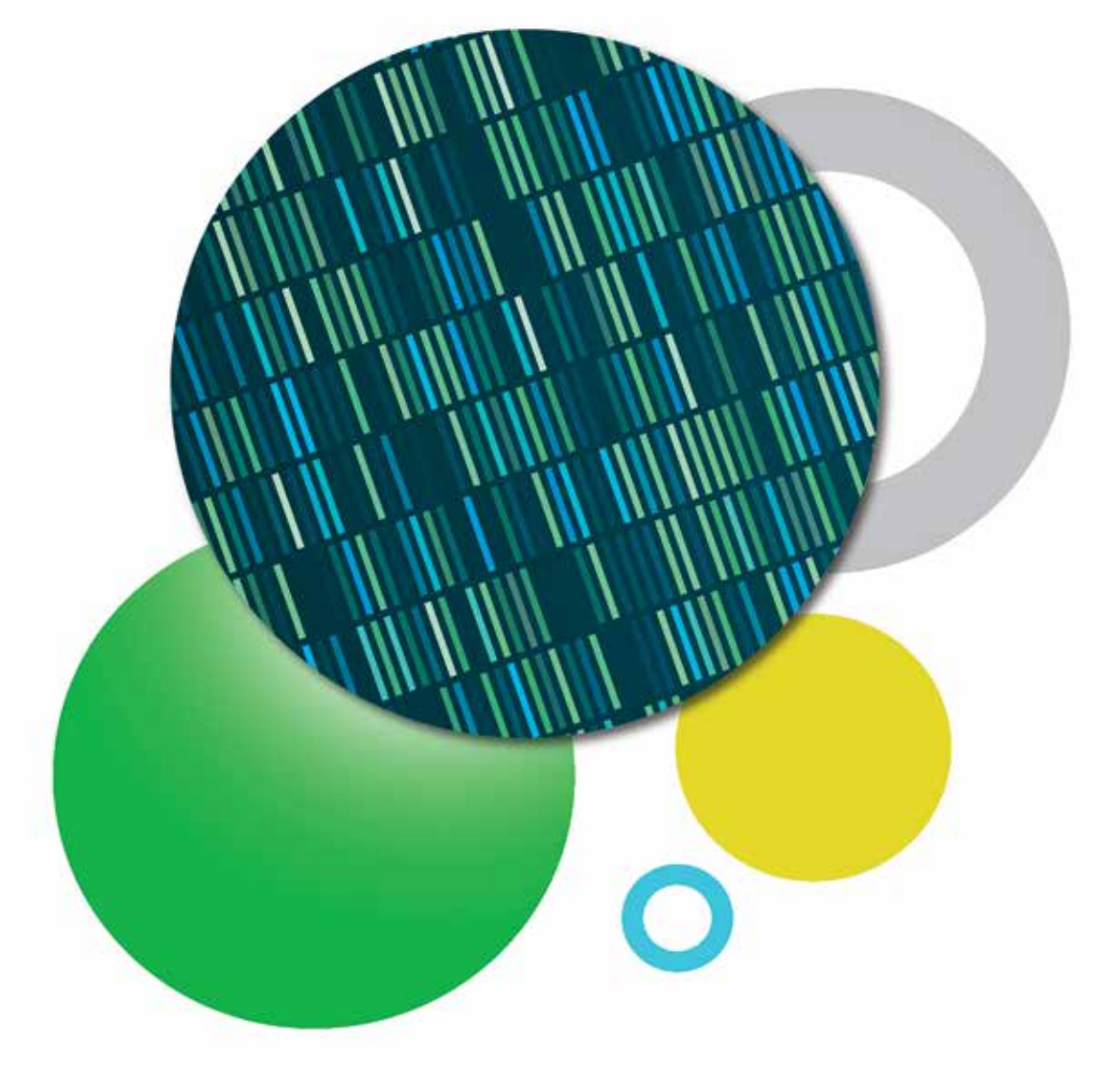
ZYMO RESEARCH

The Beauty of Science is to Make Things Simple®

Optimizing Sample Preparation for Metagenomic Assembly using Long-Read Sequencing



Introduction

Next Generation Sequencing (NGS) is widely used for microbiome analysis due to its high-throughput and cost-effectiveness. NGS utilizes short read lengths, which allow for more efficient and accurate sequencing of complex genomes and metagenomes, where high-coverage and rare species identification is required. Illumina sequencing represents the most commonly used form of NGS, and is capable of read lengths of 100-300 bp. This technology enables researchers to use paired-end sequencing of 2x150bp for metagenomic sequencing, and 2x250bp or 2x300bp for amplicon-based sequencing of the 16S/ITS regions. This provides sufficient read length to cover large, commonly sequenced amplicons, such as the V1-V3 and V3-V4 hypervariable regions of the 16S rRNA gene, which are approximately 490bp and 460bp, respectively.

Short-read sequencing with NGS is sufficient for many common applications of microbiome research, when paired with proper bioinformatics analysis. However, because short reads have limited overlap, complete genome assembly is challenging, particularly in complex regions.

De novo assembly is required to identify unknown and newly discovered microbes with no reliable existing reference genome. However, the process of assembly is computationally intensive and time consuming. The resulting contigs are often highly fragmented, making it challenging to determine their origin. Despite the existence of bioinformatics tools (e.g. MetaVelvet[6], metaSPAdes[7], and Ray Meta[8]) and specialized library preparation methods (e.g. Mate-Pair Sequencing and Metagenomic Hi-C [9]), metagenomic assembly of short reads remains a challenging task, typically requiring significant time and effort. The emergence of long-read sequencing (e.g. PacBio or Oxford Nanopore sequencing) offers new opportunities to solve current challenges in metagenomic assembly. Longer reads provide greater sequence overlap, simplifying the assembly process and improving assembly quality (e.g. N50). Long-read sequencing also enables direct assembly of complete or circular bacterial chromosomes from metagenomes.

In a recent collaboration between Zymo Research and PacBio, complete circularized genomes were assembled directly from a standardized fecal reference material, the ZymoBIOMICS Fecal Reference with TruMatrix™ Technology (D6323), using PacBio HiFi sequencing. A challenge unique to long-read sequencing is the requirement of high molecular weight (HMW) DNA, which can be difficult to obtain from complex microbial samples. Below, we breakdown how to obtain HMW DNA from fecal samples for accurate microbiome profiling and assembly. These methods are not unique to fecal DNA extraction and can potentially be used for any complex microbial sample type.

Sample Collection and Preservation

Sample collection and preservation are often overlooked when designing a microbiome workflow, but proper consideration is essential for project success. Without adequate sample preservation, an entire project can be compromised, with serious consequences in downstream analysis. The American Gut project, for example, used dry swabs without any preservation media to collect fecal samples, resulting in the overgrowth of certain bacteria, depletion of others, and the need for additional analysis filters to minimize the impact[10].

A reliable sample preservation reagent is a simple solution to this problem. DNA/RNA Shield, preserves microbial profiles at ambient temperature for at least a month (Figure 1).

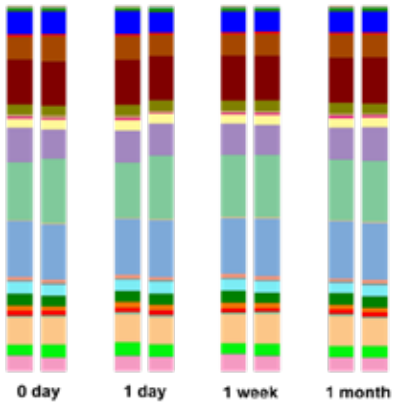


Figure 1. Shield preserved microbiome profile at ambient temperature. One fecal sample was preserved in DNA/RNA Shield and stored at room temperature. Liquids were withdrawn from the sample at different time points. DNA were extracted using ZymoBIOMICS DNA miniprep kit and profiling using 16S rRNA gene sequencing by Illumina MiSeq targeting the v3-v4 variable region.

DNA integrity is paramount when attempting to isolate HMW DNA from samples. Fortunately, DNA/RNA Shield has also been demonstrated to preserve DNA fragment size (Figure 2),

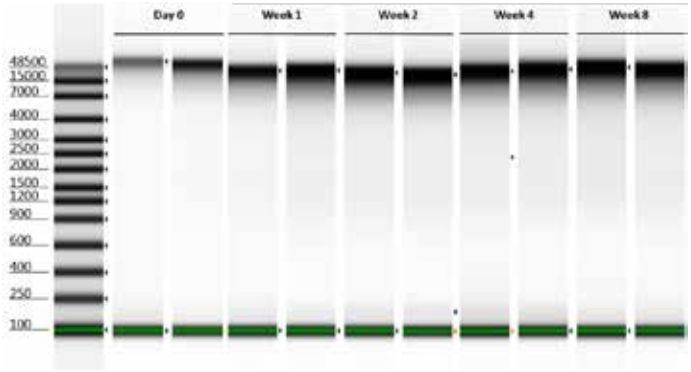


Figure 2. DNA/RNA Shield preserves DNA fragment size in fecal samples. Feces were preserved in DNA/RNA Shield and sampled at different time point. DNA was extracted using Quick-DNA HMW MagBead Kit with enzymatic lysis with lyzozyme. DNA fragment size was analyzed with TapeStation.

maintaining optimal input quality for long-read sequencing. DNA/RNA Shield is offered as a standalone reagent, as well as pre-filled in a variety of sample collection devices to simplify the sample collection process. (https://www.zymoresearch.com/collections/dna-rna-shield).

DNA Extraction

Most DNA extraction methods utilized for short-read sequencing rely on mechanical lysis, for example, bead beating to break open microbial cells. This method is easy-to-use, provides consistent results, and can handle a variety of sample types. With proper conditions, bead beating can result in unbiased lysis of microorganisms of different recalcitrance, such as bacteria, archaea, fungi, and viruses. Unbiased lysis is crucial for accurate composition microbiome profiling. However, it is also well-known that mechanical lysis can lead to DNA shearing and fragmentation. With this in mind, can unbiased DNA extraction methods that utilize mechanical lysis be applied to long-read sequencing applications?

The short answer is it depends.

The requirements for DNA fragment size can vary widely between different long-read sequencing protocols. For example, Oxford Nanopore's Rapid Barcoding Kit calls for DNA fragment sizes greater than 30kb. This kit utilizes a tagmentation-based approach for library preparation and can generate reads in excess of 100kb if the input DNA allows, as seen in Figure 3.

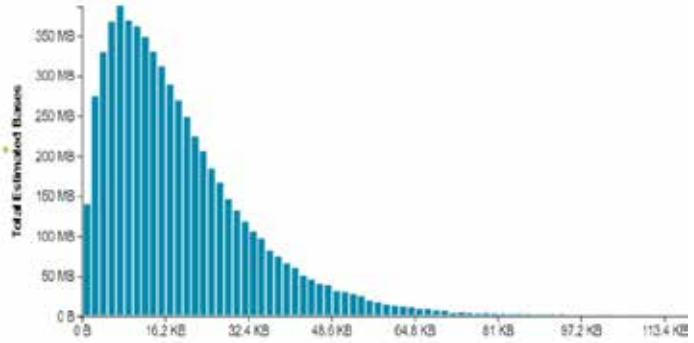


Figure 3. Read length distribution of Oxford Nanopore sequencing with DNA extracted from ZymoBIOMICS Microbial Community Standard. The cells were pelleted by centrifugation and the pellet was washed with PBS and lyzed with Lyzozyme and Zymolyase. HMW DNA was extracted using Quick-DNA HMW MagBead Kit. Sequencing was performed on Oxford Nanopore MinION™ device using an R9.4 flow cell. Library was prepared using the RBK-004 kit (Oxford Nanopore Technologies, UK) per manufacturer's protocol.

Oxford Nanopore's Ligation Sequencing Kit is more flexible with DNA input requirements, but this frequently comes at the cost of read length. Examining other long-read technologies, PacBio HiFi sequencing requires input DNA within a range of 10-20kb. This is because PacBio HiFi reads are produced with circular consensus sequencing (CCS). Limiting the DNA fragment size to 10-20kb allows for multiple readings of the same DNA sequences, resulting in highly accurate consensus reads for each. Therefore, additional fragmentation is necessary when the input DNA is above 20kb.

Although mechanical lysis is prone to significantly fragmented DNA, proper optimization of lysis conditions can still yield relatively large DNA fragments suitable for long-read sequencing. Not all methods of mechanical lysis shear DNA equally. The use of high-speed homogenization instruments like MP Bio's Fastprep-24 tend to shear DNA more than low-speed alternatives like the Vortex Genie 2 (Figure 4). Using the Vortex Genie 2 for 40 minutes can produce DNA fragments of 8-15 kb from fecal samples, which is ideal for PacBio HiFi sequencing, removing the necessity for additional DNA fragmentation.

The type of DNA binding matrix used during purification will also contribute to DNA fragment size. Most commercial DNA extraction kits use either magnetic beads or centrifugation-based spin columns as the DNA binding matrix. Magnetic beads tend to preserve fragment length better than columns, especially when working with very long DNA fragments (Figure 4).

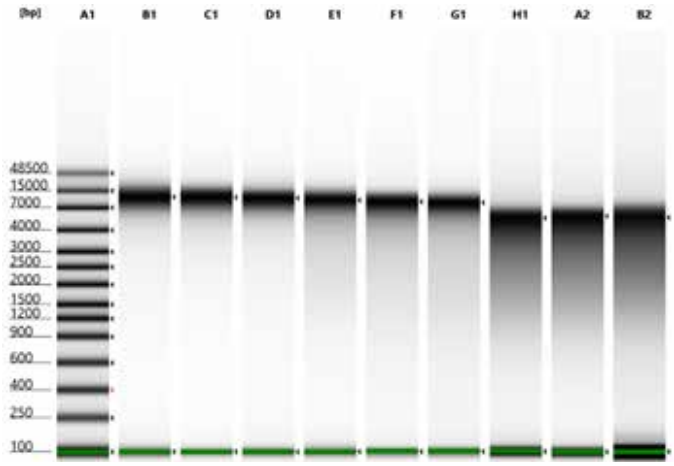


Figure 4. Fragment size distribution of fecal DNA extracted with three DNA extraction protocols using mechanical lysis. (1) B1-C1, bead beating with vortex genie 2 for 40 minutes and DNA extracted using ZymoBIOMICS-96 MagBead kit (magnetic bead based). (2) E1-G1, bead beating with vortex genie 2 for 40 minutes and DNA extracted using ZymoBIOMICS DNA miniprep kit (spin column based). (3) H1-B2, bead beating with FastPrep-24 for 5 minutes and DNA extracted using ZymoBIOMICS DNA miniprep kit (spin column based).

The nature of the raw sample itself will also play a role. For example, solid particles in soil samples can cause additional shearing during mechanical lysis. Other components of soil can also cause DNA degradation and make extraction of HMW DNA more difficult. Recommended extraction methods for long-read metagenomic sequencing are provided below in Table 1.

Application	Extraction Kit	Microbial lysis	DNA Binding Matrix	DNA Fragment Size	Features
PacBio HiFi Sequencing	ZymoBIOMICS DNA MiniPrep Kit	Vortex Genie 2	Spin-Column	8-15kb	Manual, low throughput
PacBio HiFi Sequencing	ZymoBIOMICS-96 MagBead DNA	Vortex Genie 2	Magnetic beads	8-15kb	Automatic, high throughput
Nanopore Se-quencing	Quick-DNA HMW Mag-Bead Kit	Lyzozyme or meta-polyzyme	Magnetic beads	>30 kb	Manual, low throughput
Nanopore Se-quencing	ZymoBIOMICS-96 MagBead DNA	Lyzozyme or meta-polyzyme	Magnetic beads	>30 kb	Automatic, high throughput

Table 1. Microbial DNA extraction methods for long-read sequencing

All DNA fragment sizes shown are based on testing with human fecal samples.

Alternative lysis methods are available, but additional factors must be considered when applying these methods to microbiome research. Enzymatic lysis using enzymes such as lyzozyme or meta-polyzyme, and chemical lysis using agents such as SDS, are generally considered milder methods that can allow retrieval of DNA fragments greater than 50kb. However, these methods also have limitations. Many microbes, especially fungi and gram-positive bacteria, are resistant to chemical and enzymatic lysis to varying degrees, which introduce potential bias. Enzymatic lysis efficiency may also be limited by denaturants and enzyme inhibitors present in raw samples and may require certain pretreatment of samples to be effective. By contrast, mechanical lysis methods offer a one-size-fits-all approach that is generally unbiased against the vast array of different microbes.

Quality Controls

Sequencing-based microbiome workflows can be complex, making it crucial to benchmark them with a reliable standard. Zymo Research offers a variety of microbiome standards and quality controls. For the application of long-read sequencing, the ZymoBIOMICS Microbial Community Standard (D6300) can be used to benchmark DNA extraction (Figure 5),

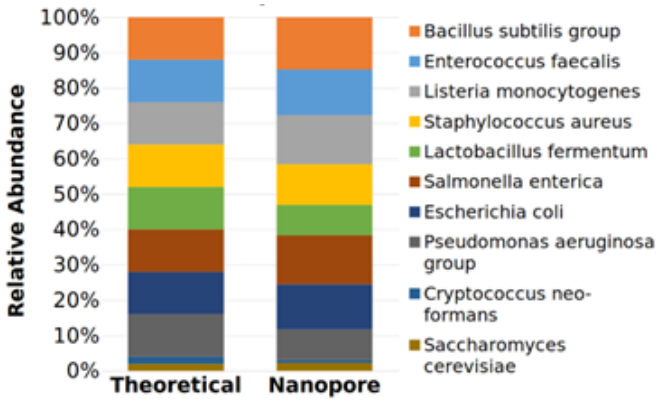


Figure 5. Benchmark a workflow of Oxford Nanopore sequencing using ZymoBIOMICS Microbial Community Standard (D6300). DNA was extracted using Quick-DNA HMW MagBead Kit with enzymatic lysis using Lyzozyme and Zymolysase. Sequencing was performed on Oxford Nanopore MinION™ device using an R9.4 flow cell. Library was prepared using the RBK-004 kit (Oxford Nanopore Technologies, UK) per manufacturer's protocol.

and the ZymoBIOMICS HMW DNA Standard (D6322) can be used to evaluate library preparation. These standards are widely used and cited by microbiome researchers to benchmark and validate both short and long-read sequencing techniques. In addition, because they are composed of intact microbial cells, whole cell mock communities can be used as positive extraction controls for HMW DNA.

Long-read sequencing provides several advantages over traditional NGS methods, including improved accuracy, and simplified genome assembly. One major challenge in long-read sequencing is to avoid DNA shearing of the easy-to-lyse organisms while still achieving lysis of difficult-to-lyse organisms. The methods above have been optimized for long-read sequencing sample preparation, and can be used to preserve the length and integrity of nucleic acids. This achieves unbiased lysis, and incorporates microbiome standards and quality controls.

References:

1. **Blanco-Miguez, A., Beghini, F., Cumbo, F., McIver, L. J., Thompson, K. N., Zolfo, M., ... & Segata, N. (2022). Extending and improving metagenomic taxonomic profiling with uncharacterized species with MetaPhlAn 4. bioRxiv, 2022-08.**
2. **Truong, D. T., Tett, A., Pasolli, E., Huttenhower, C., & Segata, N. (2017). Microbial strain-level population structure and genetic diversity from metagenomes. Genome research, 27(4), 626-638.**
3. **Wood, D. E., Lu, J., & Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. Genome biology, 20, 1-13.**
4. **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.**
5. **Ruscheweyh, H. J., Milanese, A., Paoli, L., Sintsova, A., Mende, D. R., Zeller, G., & Sunagawa, S. (2021). mOTUs: profiling taxonomic composition, transcriptional activity and strain populations of microbial communities. Current Protocols, 1(8), e218.**
6. **Namiki, T., Hachiya, T., Tanaka, H., & Sakakibara, Y. (2011, August). MetaVelvet: an extension of Velvet assembler to de novo metagenome assembly from short sequence reads. In Proceedings of the 2nd ACM conference on bioinformatics, computational biology and biomedicine (pp. 116-124).**
7. **Nurk, S., Meleshko, D., Korobeynikov, A., & Pevzner, P. A. (2017). metaSPAdes: a new versatile metagenomic assembler. Genome research, 27(5), 824-834.**
8. **Boisvert, S., Raymond, F., Godzaridis, É., Laviolette, F., & Corbeil, J. (2012). Ray Meta: scalable de novo metagenome assembly and profiling. Genome biology, 13(12), 1-13.**
9. **DeMaere, M. Z., Liu, M. Y., Lin, E., Djordjevic, S. P., Charles, I. G., Worden, P., ... & Darling, A. E. (2020). Metagenomic Hi-C of a healthy human fecal microbiome transplant donor. Microbiology resource announcements, 9(6), e01523-19.**
10. **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.**