HMB-SUNDAY-974

6/22/2025
Abstract 5062

Genome-resolved metagenome assembly of human oral microbiome using highly accurate long-read sequencing

Daniel M. Portik[1], Dominic O'Neil[2], Jeremy E. Wilkinson[1]
1. PacBio, 1305 O'Brien Drive, Menlo Park, CA 94025; 2. QIAGEN GmbH, Qiagen Strasse 1, 40724 Hilden, Germany

Introduction

There are many challenges associated with metagenome assembly:

- the presence of multiple species
- uneven and unknown species abundances
- conserved genomic regions shared across species
- strain-level variation within species

PacBio HiFi sequencing produces highly accurate long reads (>Q20, >99% accuracy) which provide major advantages for metagenome assembly. New metagenome assembly algorithms have been developed to assemble PacBio HiFi reads, including **hifiasm-meta**[1], **metaMDBG**[2], and **metaFlye**[3]. These methods can reconstruct full metagenome-assembled genomes (**MAGs**) for many higher abundance species.

Metagenome assembly of human oral microbiomes can be difficult using short reads due to strain-level diversity and high levels of host contamination (30% of reads for tongue scrape and 80% for saliva[4]). Here, we demonstrate that HiFi data obtained from the high-throughput PacBio Revio system is low in host contamination and sufficient to assemble high-quality MAGs from the oral microbiome.

Methods

PacBio HiFi sequencing

A tongue scraping was extracted using the Qiagen QIAamp DNA Microbiome Kit, then prepped and sequenced on a PacBio Revio system using one SMRT Cell. Sequencing results are shown in the table below.

Sequencing	HiFi reads (million)	Total data (gigabases)	Average read length (kb)	Median QV
Revio SMRT Cell	7.90 M	77.27 Gb	9.8 kb	Q44

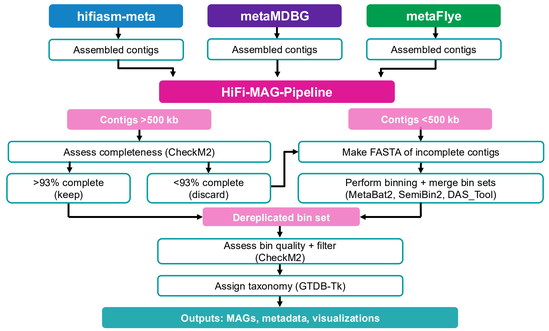
Metagenome assembly and postprocessing

The HiFi reads were assembled using **hifiasm-meta**, **metaMDBG**, and **metaFlye**. Each contig set was processed using the PacBio **HiFi-MAG-Pipeline**[5]. The complete analysis workflow is shown visually in Figure 1.

MAGs were categorized based on quality scores from CheckM2[6]:

Quality	Completeness	Contamination	Contig number
Medium (MQ)	>50%	<10%	≥ 1
High, multi-contig (mc-HQ)	>90%	<5%	≥ 2
High, single-contig (sc-HQ)	>90%	<5%	1

Figure 1. Metagenome assembly workflow.



Results

HiFi metagenomics routinely produces single-contig HQ-MAGs

The assemblers produced 55–99 sc-HQ-MAGs directly from the assembly step (Fig. 2), which can appear as circular in the assembly graphs (Fig. 3).

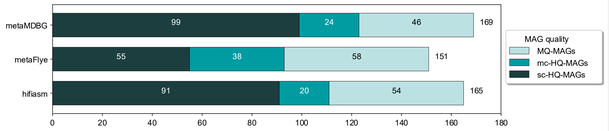


Figure 2. Counts of MAGs assigned to different quality categories.

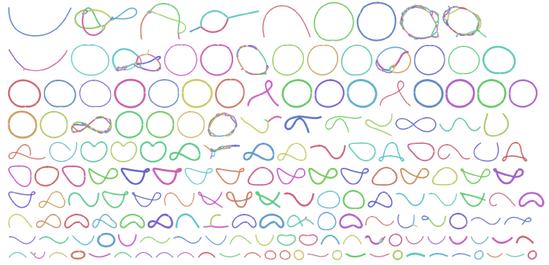


Figure 3. A partial view of the hifiasm-meta assembly graph for the oral microbiome dataset. The graph reveals many large circular contigs (0.5–4.2 Mb) produced directly from assembly.

HiFi metagenomics recovers genomes with diverse properties

- Size range across HQ-MAGs was 0.6–3.8Mb (Fig. 4)
- HQ-MAGs display a range of 28–68% GC content
- HQ-MAGs displayed 5X–2400X depth of coverage

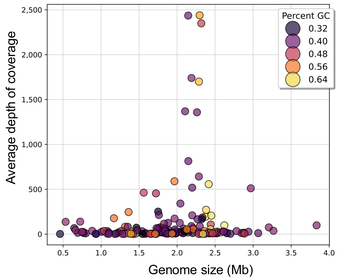
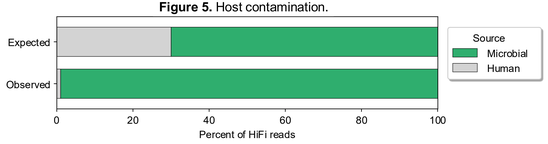


Figure 4. Characteristics of MAGs recovered by hifiasm-meta, including the genome size and average depth of coverage per genome. Each point represents an individual MAG, which are color-coded based on estimated percent GC content.

Host contamination is very low in long-read tongue scrapings

- <1% human HiFi reads vs. ~30% in short-read datasets (Fig. 5)

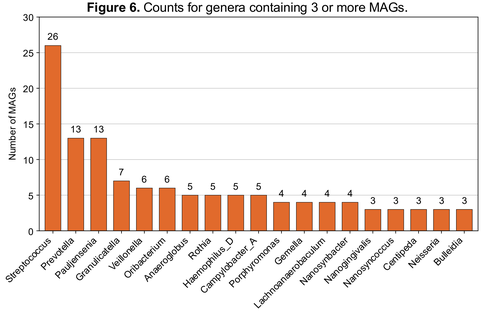
Figure 5. Host contamination.



Key oral microbiome genera are well represented in MAGs

- 26 *Streptococcus* MAGs from 18 species, 13 *Prevotella* MAGs from 8 species, and 13 *Pauljensenia* MAGs from 11 species (Fig. 6)

Figure 6. Counts for genera containing 3 or more MAGs.



Downsampling reveals most genomes are being recovered

- Oral microbiome is relatively low complexity, and we observe a saturation curve with 77Gb data from one Revio SMRT Cell (Fig. 7)
- Additional sequencing could recover more rare taxa, and potentially improve quality of MQ-MAGs

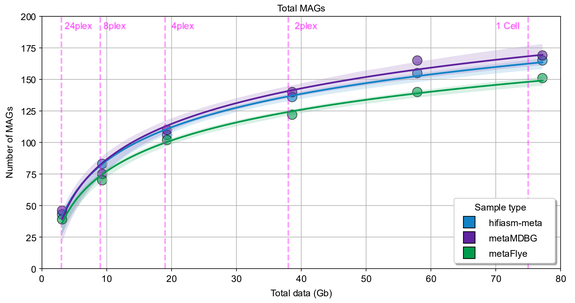


Figure 7. Effects of total data on MAG recovery. The full dataset was downsampled, then assembled and processed using identical methods. We assessed if total MAGs could be predicted by total data (Gb), either by a linear relationship (linear regression) or a non-linear relationship (log-transforming total data values before linear regression). The expected total data for different multiplex levels for one Revio SMRT Cell is shown by the magenta dotted lines.

Conclusions

- HiFi sequencing offers major advantages for metagenome assembly.
- Single-contig HQ-MAGs are routinely assembled with HiFi reads.
- One cell of PacBio Revio sequencing provides sufficient depth to assemble nearly all species, and with low host contamination.

All PacBio metagenomics workflows are open-source and publicly available on **Github**:

PacificBiosciences / pb-metagenomics-tools

References

1. Feng et al. 2022. Metagenome assembly of high-fidelity long reads with hifiasm-meta. *Nature Methods*, 19: 671–674.
2. Benoit et al. 2024. High-quality metagenome assembly from long accurate reads with metaMDBG. *Nature Biotechnology*, https://doi.org/10.1038/s41587-023-01983-6
3. Kolmogorov et al. 2020. metaFlye: scalable long-read metagenome assembly using repeat graphs. *Nature Methods*, 17: 1103–1110.
4. Liu et al. 2021, Metagenome-genome-wide association studies reveal human genetic impact on the oral microbiome. *Cell Discovery*, 7: 117.
5. Portik et al. 2024. Highly accurate metagenome-assembled genomes from human gut microbiota using long-read assembly, binning, and consolidation methods. *bioRxiv*, https://doi.org/10.1101/2024.05.10.593587
6. Chklovski et al. 2023. CheckM2: a rapid, scalable and accurate tool for assessing microbial genome quality using machine learning. *Nature Methods*, 20: 1203–1212.