



INTRODUCTION

The use of metagenome-assembled genomes (MAGs) in microbial surveillance has increased in popularity. Metagenome assembly can allow for fine-scale resolution of epidemiological tracing and improve recovery of antimicrobial resistance (AMR) genes. Highly accurate PacBio HiFi reads can dramatically improve metagenome assembly, particularly when used with assemblers like hifiasm-meta¹, and can produce complete genomes (Fig. 1). We present HiFi-MAG-Pipeline, a comprehensive workflow that identifies and extracts high-quality MAGs from long-read metagenomic assemblies.

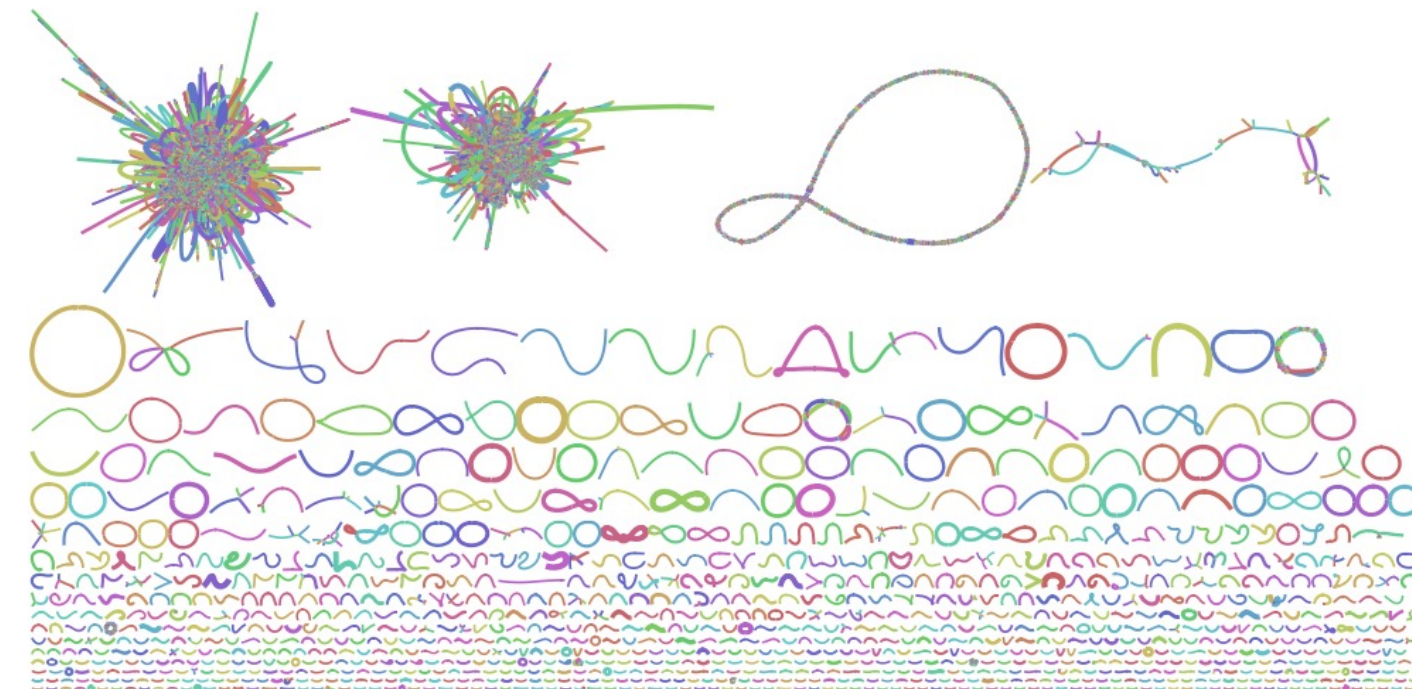


Figure 1. A hifiasm-meta assembly graph for a human gut microbiome dataset. The graph reveals many large (>1 Mb) circular contigs produced directly from assembly. These represent complete MAGs and do not require binning methods to be discovered. However, many large linear contigs are also produced in the assembly. These represent fragmented genomes and postprocessing is required to recover these additional high-quality MAGs.

METHODS

HiFi-MAG-Pipeline and other workflows are freely available on github:

[PacificBiosciences / pb-metagenomics-tools](https://github.com/PacificBiosciences/pb-metagenomics-tools)

Completeness-aware binning strategy (Fig. 2)

- Standard binning assumes genomes occur as fragmented contigs.
- This can cause unexpected behavior for HiFi assembly: complete contigs can be mis-binned with additional contigs, inflating contamination scores and causing removal during filtering.
- The completeness-aware strategy first extracts long, complete contigs, and then uses a multi-binning strategy on all remaining contigs.
- The bin sets are de-replicated, combined with the long contig bins, and filtered to extract all high-quality MAGs and associated metadata.

Comparative analysis and benchmarking

- We assembled 12 publicly available HiFi metagenomic datasets with hifiasm-meta and then performed binning analyses.
- We compared a standard binning pipeline (MetaBat2) to HiFi-MAG-Pipeline v2.0.

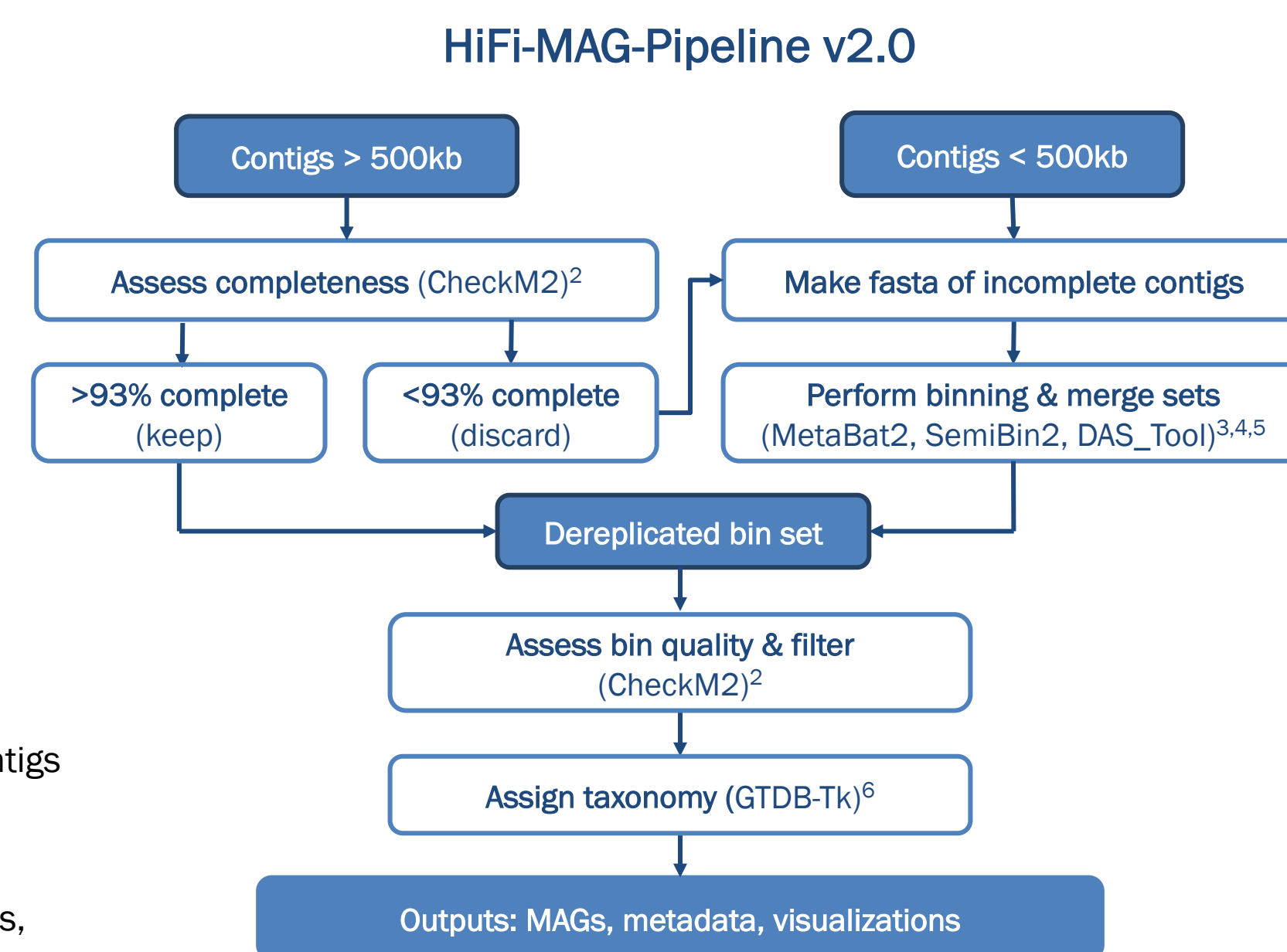


Figure 2. Overview of the completeness-aware binning strategy implemented in HiFi-MAG-Pipeline v2.0.

RESULTS

HiFi assemblies produce many high-quality MAGs

- Recovered 60–325 MAGs per sample
- Found 33–193 MAGs (up to 65%) are single-contig (Fig. 3)

HiFi-MAG-Pipeline yields more total MAGs than other methods

- Found 14–67% increase in total MAGs (Fig. 3)
- Gain of 12–120 total MAGs per sample

Completeness-aware binning rescues single-contig, complete MAGs

- Found 10–142% increase in single-contig, complete MAGs
- Incomplete circular contigs are successfully binned

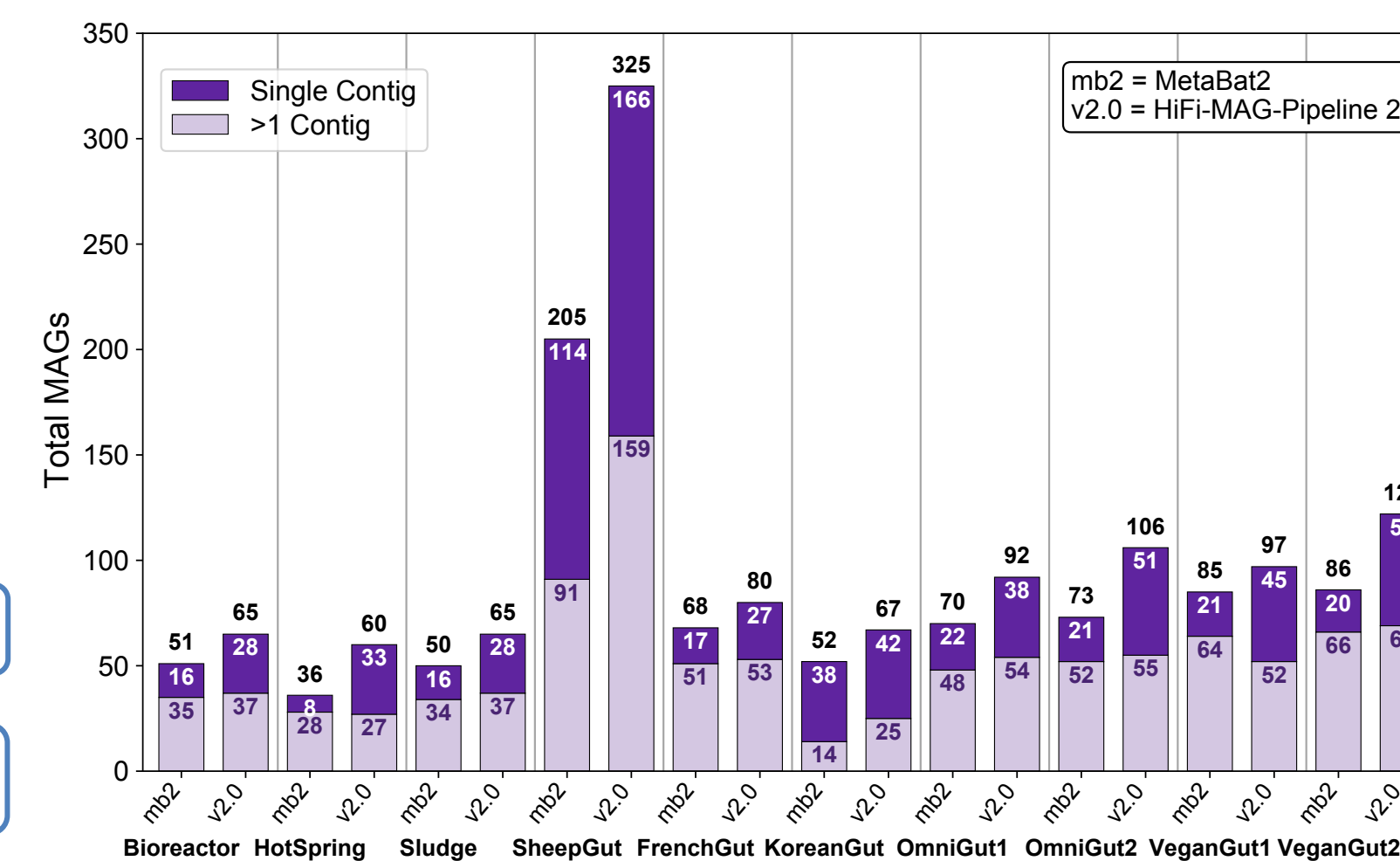


Figure 3. MAG yields from standard binning with MetaBat2 (mb2) vs. HiFi-MAG-Pipeline (v2.0). Dark purple represents single-contig circular MAGs and light purple represents MAGs containing >1 contigs. Numbers in the stacked bars represent each category, and numbers above represent total MAGs.

Pipeline Outputs

HiFi-MAG-Pipeline produces several informative figures displaying quality characteristics for MAGs recovered (Fig. 4), and provides metadata from CheckM2 and GTDB-Tk. All MAG sequences are provided as individual fasta files for downstream analysis.

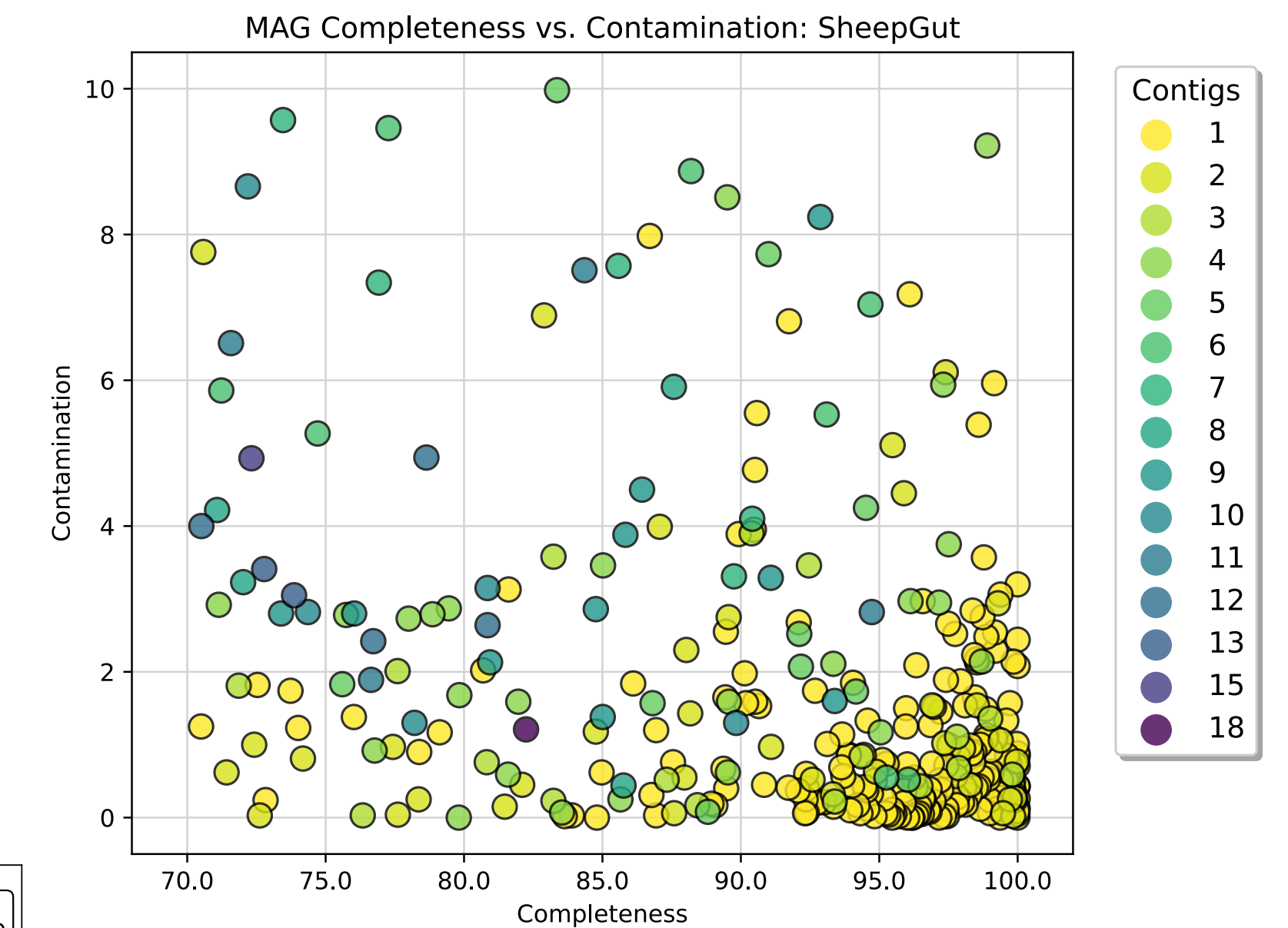


Figure 4. Completeness versus contamination scores for 325 high-quality MAGs found by HiFi-MAG-Pipeline for the sheep gut assembly. Each dot represents a MAG, and colors indicate the number of contigs contained in the MAG. We found 156 MAGs (48%) were exceptionally high quality and displayed >95% completeness (pink outline), with 126 being MAGs composed of a single contig.

CONCLUSION

- PacBio HiFi sequencing offers major advantages for metagenome assembly.
- Complete, single-contig MAGs can be routinely assembled from HiFi reads (33-62% of total MAGs).
- The HiFi-MAG-Pipeline automates all key steps required to obtain high-quality MAGs from long-read metagenome assemblies.
- Completeness-aware binning recovers substantially more MAGs than other methods (67% increase in total MAGs, 142% increase in single-contig MAGs).
- HiFi sequencing is an effective strategy for obtaining large numbers of high-quality MAGs, which can be used to improve microbial surveillance efforts.

For More Information

To see the PacBio github site containing publicly available metagenomics tools and resources, please scan the QR code. To access more information, please visit:

- <https://www.pacb.com/wp-content/uploads/Application-Brief-Metagenomic-Sequencing-with-HiFi-Reads-Best-Practices.pdf>
- <https://www.pacb.com/wp-content/uploads/metagenomics-solutions-brochure.pdf>
- <https://www.pacb.com/products-and-services/applications/complex-populations/microbial/>



Sources

- Feng, X., et al. 2022. Metagenome assembly of high-fidelity long reads with hifiasm-meta. *Nature Methods*, 19: 671–674.
- Chklovski et al. 2023. CheckM2: a rapid, scalable and accurate tool for assessing microbial genome quality using machine learning. *bioRxiv*, <https://doi.org/10.1101/2022.07.11.499243>
- Kang, D.D., et al. 2019. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ*, 7: e7359.
- Pan et al. 2023. SemiBin2: self-supervised contrastive learning leads to better MAGs for short- and long-read sequencing. *bioRxiv*, <https://doi.org/10.1101/2023.01.09.523201>
- Sieber, C.M.K., et al. 2018. Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. *Nature Microbiology*, 3: 836–843.
- Chaumeil, P.-A., et al. 2019. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics*, 35: 1925–1927.

Research use only. Not for use in diagnostic procedures. © 2023 Pacific Biosciences of California, Inc. ("PacBio"). All rights reserved. Information in this document is subject to change without notice. PacBio assumes no responsibility for any errors or omissions in this document. Certain notices, terms, conditions and/or use restrictions may pertain to your use of PacBio products and/or third-party products. Refer to the applicable PacBio terms and conditions of sale and to the applicable license terms at pacbio.com/license. Pacific Biosciences, the PacBio logo, PacBio, Circulomics, Omniome, SMRT, SMRTbell, Iso-Seq, Sequel, Nanobind, SBB, Revo, and Onso are trademarks of PacBio.