

Shotgun Metagenomic Analysis, Assembly, and Annotation

Human reads within each sample were mapped to the HG38 reference index (GCF_000001405.40) using Minimap2¹. Paired, unmapped reads were pulled from the read set prior to microbial community analyses using and Samtools² fastq -f 13.

The composition of the microbial community for each metagenomic shotgun sample was profiled from the filtered Illumina read set using Kraken2³ via the Kraken2 Standard Database (Refseq: archaea, bacteria, viral, plasmid, and human; UniVec_Core). Profiled relative abundances were converted into Kronagrams using KronaTools⁴. A representative metagenome-assembled genome (MAG) was generated using MegaHit⁵. MAG assembly metrics were determined using Quast⁶.

The MAG assembly was subset into genomic bins using MetaBat2⁷. Individual bins were taxonomically assessed using Kraken2 and annotated using Bakta⁸. Binned assembly metrics were determined using Quast.

Generated output files:

- Metagenomic Assessment
 - *sample_taxonomy.tsv* – Filtered-read taxonomy reported as read counts and relative abundance
 - *sample_taxonomy.kreport* – Filtered-read taxonomy kraken2 source file
 - *sample_krona.html* – Interactive phylogeny of reported taxa and relative abundances
 - *sample_mag.fasta* – MAG assembly
 - *sample_assembly_metrics.tsv* – Captured metrics of the MAG assembly
- Metagenomic Reconstruction (by Bin)
 - *bin_taxonomy.tsv* – Identified taxonomy of bin reported as contig counts and relative abundance
 - *bin.fasta* – Binned subset from the MAG assembly
 - *binned_assembly_metrics.tsv* – Captured metrics of each binned assembly
 - *bin.gbk* – Annotation of the binned assembly

Tool	Version	Parameters
minimap2	2.26	default parameters; -ax sr
samtools	1.17	default parameters; sort, index, fastq -f 13
kraken2	2.1.3	default parameters; --memory-mapping; db=k2_plusfp_20240112
krona2	2.8.1	default parameters; ktlImportTaxonomy
megahit	1.2.9	default parameters
metabat2	2.15	default parameters
bakta	1.8.1	default parameters; db version 5.0
quast	5.2.0	default parameters

¹ Heng Li, Minimap2: pairwise alignment for nucleotide sequences, *Bioinformatics*, Volume 34, Issue 18, September 2018, Pages 3094–3100, <https://doi.org/10.1093/bioinformatics/bty191>

² Petr Danecek, James K Bonfield, Jennifer Liddle, John Marshall, Valeriu Ohan, Martin O Pollard, Andrew Whitwham, Thomas Keane, Shane A McCarthy, Robert M Davies, Heng Li. Twelve years of SAMtools and BCFtools. *GigaScience*. Volume 10, Issue 2. (2021).

³ Lu, J., Rincon, N., Wood, D.E. et al. Metagenome analysis using the Kraken software suite. *Nat Protoc* 17, 2815–2839 (2022).

⁴ Ondov BD, Bergman NH, Phillippy AM. Interactive metagenomic visualization in a Web browser. *BMC Bioinformatics*. 2011 Sep 30;12:385. doi: 10.1186/1471-2105-12-385. PMID: 21961884; PMCID: PMC3190407.

⁵ Dinghua Li and others, MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph, *Bioinformatics*, Volume 31, Issue 10, May 2015, Pages 1674–1676,

⁶ Alexey Gurevich, Vladislav Saveliev, Nikolay Vyahhi and Glenn Tesler. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29 (8): 1072-1075 (2013).

⁷ Kang DD, Li F, Kirtan E, et al. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ*. 7:e7359 (2019).

⁸ Schwengers O., Jelonek L., Dieckmann M. A., Beyvers S., Blom J., Goesmann A. (2021). Bakta: rapid and standardized annotation of bacterial genomes via alignment-free sequence identification. *Microbial Genomics*, 7(11).