

## Supplemental References

- Huang W, Li L, Myers JR, and Marth GT. 2012. ART: A next-generation sequencing read simulator. *Bioinformatics*. **28**: 593–594.
- Lu J, Breitwieser FP, Thielen P, and Salzberg SL. 2017. Bracken: estimating species abundance in metagenomics data. *PeerJ Computer Science*. **3**: e104.
- Marçais G and Kingsford C. 2011. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics*. **27**: 764–770.
- Meyer F, Bremges A, Belmann P, Janssen S, McHardy AC, and Koslicki D. 2019. Assessing taxonomic metagenome profilers with OPAL. *Genome Biology*. **20**: 51.
- Meyer F, Fritz A, Deng ZL, Koslicki D, Lesker TR, Gurevich A, Robertson G, Alser M, Antipov D, Beghini F, et al. 2022. Critical Assessment of Metagenome Interpretation: the second round of challenges. *Nature Methods*. **19**: 429–440.
- Ondov BD, Treangen TJ, Melsted P, Mallonee AB, Bergman NH, Koren S, and Phillippy AM. 2016. Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biology*. **17**: 132.
- Unit R, Wanamaker S, Close TJ, and Lonardi S. 2015. CLARK: fast and accurate classification of metagenomic and genomic sequences using discriminative k-mers. *BMC Genomics*. **16**: 236.
- Rachtman E, Balaban M, Bafna V, and Mirarab S. 2020. The impact of contaminants on the accuracy of genome skimming and the effectiveness of exclusion read filters. *Molecular Ecology Resources*. **20**: 1755–0998.13135.
- Şapcı AOB, Rachtman E, and Mirarab S. 2024. CONSULT-II: accurate taxonomic identification and profiling using locality-sensitive hashing. *Bioinformatics*. **40**: btae150.
- Wood DE, Lu J, and Langmead B. 2019. Improved metagenomic analysis with Kraken 2. *Genome Biology*. **20**: 257.