[BIOINFO 2025]

Practical guidance for metagenomic binning derived from comprehensive benchmarking

Jungyeon Kim[1] and Insuk Lee[1,*]

[1] *Department of Biotechnology, College of Life Science & Biotechnology, Yonsei University, Seoul 03722, Republic of Korea*
Corresponding author: insuklee@yonsei.ac.kr

Metagenomic binning is a critical step for reconstructing microbial genomes from complex microbial communities, yet tool performance can vary substantially depending on data characteristics. Here, we present a systematic benchmarking of leading binning tools using CAMI-simulated, custom-simulated, and real metagenomic datasets, allowing a comprehensive evaluation of their strengths and limitations. Our analysis revealed several key factors that shape binning efficacy. Sequencing depth and taxonomic complexity had major impacts on performance, with CAMI datasets being considerably less complex than real human gut and environmental microbiomes. Substantial differences were observed in chimeric genome rates across tools, and multi-sample binning was most effective with approximately 20 samples, while both fewer and larger numbers reduced its benefits. Single-end sequencing further lowered genome recovery due to fragmented assemblies and reduced contig quality. Neural network–based methods consistently outperformed other approaches across both simulated and real datasets, particularly under conditions of realistic complexity. Moreover, by integrating and refining genome bins from the three top-performing tools, we achieved over a 30% increase in the recovery of high-quality genomes compared to existing strategies. Collectively, this study provides clear practical guidance for selecting and optimizing binning workflows, while also underscoring the importance of systematic evaluations for advancing genome-resolved metagenomics. These insights will help establish more reliable standards for future microbiome research and the development of next-generation binning approaches.