[BIOINFO 2023]

Metabank for Metagenome Repository

Suvin Baek[1,2], Jooseong Oh[1], Eugene Rha[1], Kil-Koang Kwon[1], Dae-Hee Lee[1], Seung-Goo Lee[1]

and Haseong Kim[1,2]

[1] *Synthetic Biology Research Center, Korea Research Institute of Bioscience and Biotechnology (KRIBB), Daejeon, Korea*
[2] *Graduate School of Engineering Biology, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea*
**Corresponding author: haseong@kribb.re.kr*

Metagenomics is one of the largest microbial genetic resources that can be used to discover new industrial enzymes and microbial strains. However, most of the new enzyme discovery is based on sequence alignment with a reference sequence, and which is hard to find a novel enzyme that significantly deviates from the original enzyme sequence. In this study, we built Metabank which stores real metagenome DNA libraries with their analyzed information. We collected 20 real metagenomic DNA libraries sequenced with short read sequencing techniques. Four of them were sequenced with a long-read sequencing method for the comparison of assembly accuracy. We investigated the microbial community composition through metagenome taxonomic profiling. The assembled metagenomic sequences are stored along with the description of the samples, strain distribution, and ORF annotation. MongoDB and R-based shiny tools were used for the construction of Metabank. We expect that the real DNA samples with accurately analyzed properties can provide a rapid enzyme testing platform along with biofoundry and AI in the near future.