[BIOINFO 2024]

Metabuli App: A High-Performance, Accessible Platform for Metagenomic Taxonomic Profiling

SunJae Lee[1], Jaebeom Kim[1], and Martin Steinegger[1]

[1] *Department of Biological Sciences, Seoul National University*

Taxonomic profiling of metagenomes identifies microbial community compositions by analyzing environmental DNA. However, users without computational expertise may face difficulties with complex command-line tools and data visualization. Additionally, traditional taxonomic profiling tools often require costly server-level resources, creating a barrier for researchers lacking access to computing servers.

To address these issues, we present the Metabuli App, a user-friendly graphical interface that enables metagenomic taxonomic profiling on a standard computer and easy visualization of results with Sankey and Krona plots, all without requiring command-line knowledge. The app streamlines analysis by providing customizable Sankey plots, allowing users to filter taxa by read counts or proportions and adjust rank limits. It further supports downstream analysis of specific taxa by offering a focused subtree view, displaying only the selected taxon and its descendants while excluding unrelated taxa, along with the functionality to extract reads from the selected taxon.

At the core of the app is Metabuli, our taxonomic profiling tool that integrates the strengths of both DNA- and amino acid (AA)-based methods. DNA-based tools excel at distinguishing closely related taxa through DNA mutations, while AA-based tools are better at detecting under-studied clades due to AA conservation. Metabuli leverages novel k-mers that store AA 8-mers and their DNA codon usage to combine these strengths, consistently achieving top-level performance in benchmarks that favor either approach. Optimized to run on regular PCs with improved parallelization, reduced memory demands, and algorithmic enhancements, it can now classify 22 million paired-end reads from a human gut metagenome (6.6Gbp) against ~36K genomes from 8,465 species (141GB) in just 43 minutes on a MacBook Pro M2 (32GB RAM) and 59 minutes on a Windows Intel i9-9900 (32GB RAM).

This new user-friendly and efficient platform enables researchers to gain quick insights into microbial communities, making metagenomic profiling more accessible.