Supplementary Material

MetaBakery: a Singularity implementation of bioBakery tools as skeleton application for efficient HPC deconvolution of microbiome metagenomic sequencing data to machine learning ready information

Boštjan Murovec1, Leon Deutsch2,3, JADBio4, Damjan Osredkar5,6, Blaž Stres2,7,8,9\*

1 University of Ljubljana, Faculty of Electrical Engineering, Ljubljana, Slovenia

2 Department of Animal Science, Biotechnical faculty, University of Ljubljana, Ljubljana, Slovenia

3 The NU, The NU B.V., Leiden, The Netherlands

4Jadbio

4Damjan

4Damjan

7D13 Department of Catalysis and Chemical reaction Engineering, National Institute of Chemistry, Ljubljana, Slovenia

8Institute of Sanitary Engineering, Faculty of Civil and Geodetic Engineering, Ljubljana, Slovenia

9Department of Automation, Biocybernetics and Robotics, Jožef Stefan Institute, Ljubljana, Slovenia

A screenshot of a computer

Description automatically generated

**Supplementary Figure 1.** Start of MetaBakery execution.

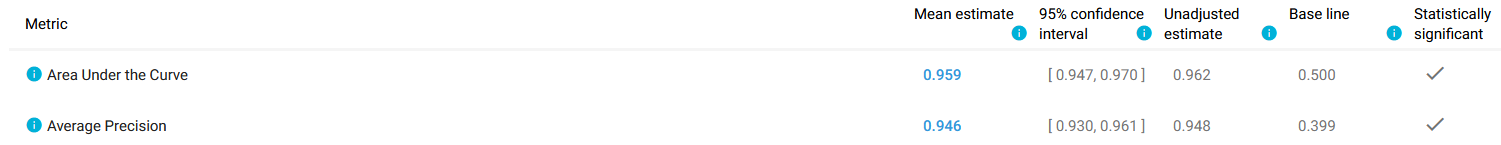
A graph with blue and orange lines

Description automatically generated

**Supplementary Figure 2.** BioBakery and MetaBakery performance comparison; MetaBakery is self tuned.

A screenshot of a computer

Description automatically generated

**Supplementary Figure 3.** An example of various descriptive metrics illustrating the performance of JADBio machine learning on GMHI taxonomy dataset.

A chart with colorful rectangular objects

Description automatically generated with medium confidence

**Supplementary Figure 4.** Shannon diversity across all disease in large dataset calculated with MetaBakery. Numerous diversity indices can be calculated using the standard approaches within the program Mothur that was inbuilt into the MetaBakery presented in this work.

A graph of a number of tax

Description automatically generated with medium confidence

**Supplementary Figure 5.** Performance of JADBio three different versions (MetaBakery2, Metabakery3, MetaBakery4) on taxonomy data on depression dataset. TPR – True Prediction Rate; FPR – False Prediction Rate.

A graph of a number of data

Description automatically generated with medium confidence

**Supplementary Figure 6.** Performance of JADBio machine learning on the separate matrices from the microbiome information layers produced with MetaBakery3. TPR – True Prediction Rate; FPR – False Prediction Rate.