MetaABC – an integrated Metagenomics platform for data Adjustment, Binning, and Clustering

Metagenomics is the study of the genetic materials sampled from uncultured microbial communities. To classify the sequences generated from metagenomes into functional groups and taxa, BLAST-based analyses are typically used. Software (such as Megan, Phymm, etc) and algorithms (such as lowest common ancestor (LCA), interpolated Markov models (IMMs), etc) have recently improved their performances. However, sampling biases, artificial duplicates, and improper binning may still affect the accuracy in estimating microbial compositions. In addition, there is no single platform that integrates analyzing tools and generates print-ready figures. Here, we present MetaABC, a platform that integrates advanced binning tools coupled with data filters and normalizations for advanced metagenomics analysis. MetaABC incorporates two methods for removing artifacts, seven methods for taxonomic binning, a method to re-analyze unassigned reads using conserved gene adjacency, and can control sampling biases via genome-length normalization. It can handle both Sanger and 454-pyrosequencing data, and the user-friendly interface allows easy adjustment of analyzing routes for an optimal outcome. Further, MetaABC embeds a hierarchical clustering program for metagenome comparative analysis, and different visualized output formats (including tables, bar charts, and pie charts of abundance profiles) can be generated per request. To facilitate the computing process for larger datasets, a stand-alone version of MetaABC is also available. In conclusion, MetaABC contains a web server and a software package that furnish various analyzing routes in a single platform. The platform not only filter artifact and correct bias to improve accuracy, its combinatorial analyzing pipelines and figure-ready
outcome visualization also facilitate analyzing efficiency and optimize data interpretation. MetaABC is available at http://bits2.iis.sinica.edu.tw/MetaABC/.