Databases, Web-servers and Prediction Tools

The management of experimental data is organized in a way as to make them available as soon as possible and over a long period of time for molecular network modeling and, thus, for generations of scientific hypotheses and the next steps of model-based experimental design. Therefore, a data warehouse has been developed and used for service-oriented data and knowledge management: data collection, storage, pre-processing and standardized primary analysis of experimental data including genome, transcriptome and proteome data as well as other biochemical, microbiological, clinical data.

VirMiner 147.8.185.62/VirMiner is a software tool which provides comparatively comprehensive phage information in metagenomic data: 1) identifies phage contigs using reliable pre-trained model; 2) gets full functional annotation for these phage contigs; 3) predicts possible phage-host relationships using existing tools; 4) if users upload two groups of metagenomic samples, downstream analysis for comparison among different groups would be done.

COMAN (**CO**mprehensive **M**etatranscriptome **AN**alysis web-server, sbb.hku.hk/COMAN) is an integrated web server dedicated to comprehensive functional analysis of metatranscriptomic data, translating massive amount of reads to data tables and high-standard figures. It is expected to facilitate the researchers with less expertise in bioinformatics in answering microbiota-related biological questions and to increase the accessibility and interpretation of microbiota RNA-Seq data.

MESSI (**M**etabolic **E**ngineering Target **S**election and Best **S**train **I**dentification Tool, sbb.hku.hk/MESSI is a webserver for predicting efficient chassis and regulatory components for yeast bio-based production. The server provides an integrative platform for users to analyze ready-to-use public high-throughput metabolomic data, which are transformed to metabolic pathway activities for identifying the most efficient S. cerevisiae strain for the production of a compound of interest.

NutriChem, available at 147.8.185.62/services/NutriChem-2.0, is a database generated by text mining of 21 million MEDLINE abstracts for information that links plant-based foods with their small molecule components and human disease phenotypes. NutriChem contains text-mined data for 18478 pairs of 1772 plant-based foods and 7898 phytochemicals, and 6242 pairs of 1066 plant-based foods and 751 diseases. In addition, it includes predicted associations for 548 phytochemicals and 252 diseases.

FungiFun: Characteristic patterns in candidate lists are often identified by enrichment analysis. FungiFun is a user-friendly web tool for functional enrichment analysis of fungal genes and proteins. The new tool, FungiFun2, uses a completely revised data management system, enabling analyses for 298 fungal strains currently published in common databases. FungiFun2 offers a modern web interface and generates interactive tables, diagrams and figures that users can

directly customise according to their needs. FungiFun2 as well as examples and instructions are publicly available at https://elbe.hki-jena.de/fungifun/fungifun.php.

DAnIEL (**D**escribing, **An**alyzing and **I**ntegrating fungal **E**cology to effectively study the systems of **L**ife): DAnIEL (https://sbi.hki-jena.de/daniel) is a web server to analyze fungal communities using the genome marker ITS (Internal transcribed spacer). One can upload own nucleotide sequences and supplement them with already existing projects. Afterwards, plots displaying the fungal abundance and diversity of the samples are automatically generated and can be interactively customized. Correlation networks and Machine Learning complete the analysis. Taxa found significant are annotated with a manually curated database to support the biological interpretation.