[L5] Secondary Metabolite-Activity Relationship Based on Metabolic Pathways: Pathway Prediction of Secondary Metabolites Based on Machine Learning

Shigehiko Kanaya, Naoaki Ono, Morita (Hirai) Aki, Ryohei Eguchi

Computational Systems Biology Lab., Graduate School of Science and Technology, Nara Institute of Science and Technology (NAIST)

Data in omics biology has explosively increased in areas including large-scale DNA sequencing, expression analyses, mass spectrometry of metabolites and phenotype studies including very diverged biological activities focused on facilitate a comprehensive understanding of the interactions between medicinal/edible plants and human healthcare and nutrition in particular medicine. Omics data make it possible to understand interactions between species based on metabolites in various ecosystems. To achieve this goal, it is necessary to develop DBs that include the relationships among these omics data. Currently, we introduce KNApSAcK family database based on accessibility of edible species from around the world. For healthcare and pharmacology applications, we have developed a total of eight DBs consisting of four DBs (Lunch Box, DietNavi, FoodProcessor and DietDish) related to the effects of popular Japanese foodstuff on health and four other DBs to systematize crude drugs: WorldMap (relationships between geographic zones and the usage of edible and medicinal plants), KAMPO (crude drugs in Japan), JAMU (crude drugs in Indonesia), and Tea Pot (relationships between herbal tea and health care).

Species-metabolite information is accumulated in the KNApSAcK Core DB that has been utilized extensively in omics-based science. Activity-species and activity-metabolite relations are accumulated in Biological Activity DB and Metabolite Activity DB, respectively. KNApSAcK Family DBs play important roles in data-intensive or data-driven biological discovery because a large open pool of data that covers the full breadth of the life sciences therefore facilitating comprehensive research. The KNApSAcK Family DB provides systematized information related to metabolites in various fields, such as omics sciences, particularly metabolomics, nutrigenomics, and foodomics.
Figure 1 shows the last version of main window in KNApSAcK family DB
(http://kanaya.naist.jp/KNApSAcK_Family/).

In the present study, we introduce TWINS database, which makes it possible to retrieve compounds by chemical similarities in 3D Graph Match algorithm (COMPLIG) developed by Prof. Shirai et al (J.Mol.Biol, 424, 379-390, 2012). Taking databases such as TWINS and the other information including KNApSAcK family DB, and Deep learning so called Graph Convolution technique applying to chemical compounds into consideration, we also examine biological activities associated with secondary metabolites and their pathways produced by organism including secondary metabolic pathways.

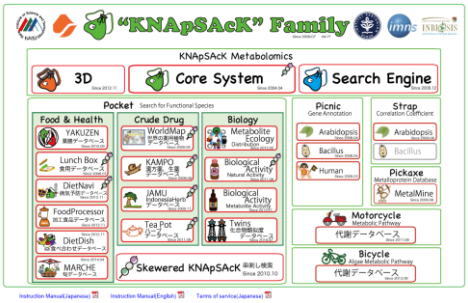

Figure 1. Main window in KNApSAcK family DB (http://kanaya.naist.jp/KNApSAcK_Family/).

Acknowledgement
This work was partially supported by the National Bioscience Database Center in Japan; the Ministry of Education, Culture, Sports, Science, and Technology of Japan (16K07223 and 17K00406); NAIST Big Data Project; the Platform Project for Supporting Drug Discovery and Life Science Research funded by Japan Agency for Medical Research and Development (AMED).