Bioinformatics Center
- Bioknowledge Systems -

http://kanehisa.kuicr.kyoto-u.ac.jp/

Prof
KANEHISA, Minoru
(D Sc)

Assoc Prof
GOTO, Susumu
(D Eng)

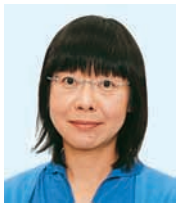
Assist Prof
HATTORI, Masahiro
(D Sc)

Program-Specific Assist Prof
TOKIMATSU, Toshiaki
(D Agr)

Program-Specific Assist Prof
KOTERA, Masaaki
(D Sc)

PD
RUIZ, Diego Deiz
(Ph D)

PD
HU, Qian-Nan
(Ph D)



PD
HIRAKAWA, Mika
(D Sc)

Researchers

MORIYA, Yuki
NAKAGAWA, Zenichi
SANO, Etsuko

Students

SAKIYAMA, Tadahiko (RF)
ONUKI, Ritsuko (D3)
HONDA, Wataru (D3)
MUTO, Ai (D3)
SUGA, Akitsugu (D3)
TANAKA, Michihiro (D3)
TAKARABE, Masataka (D3)
SHIMIZU, Yugo (D2)
DONG, Hong Ju (D2)
MIZUTANI, Sayaka (D1)
HIRANUKA, Kazushi (D1)
NISHIMURA, Yosuke (M2)
ERGUNER, Bekir (M2)
OKADA,Yuki (M1)
KONG, Xiang Shuo (M1)
JIN, Zhao (M1)
CHEN, Zhang (RS)

Scope of Research

DNA, RNA, and proteins are the basic molecular building blocks of life, but the living cell contains additional molecules, including water, ions, small chemical compounds, glycans, lipids, and other biochemical molecules, without which the cell would not function. Because the proteins responsible for biosynthesis, biodegradation, and transport of these additional molecules are encoded in the genome, one may assert that all cellular functions are specified by the genomic DNA sequence. In practice, however, it is not possible to infer higher-level systemic functions of the cell or the organism simply from the molecular sequence infomation alone. We are developing bioinformatics methods to integrate different types of data and knowledge on various aspects of the biological systems towards basic understanding of life as a molecular interaction/reaction system and also for practical applications in medical and pharmaceutical sciences.

Research Activities (Year 2009)

Grants

Kanehisa M, Backbone Database for Analysis of the Biological Systems and Environment, Grants-in-Aid for Scientific Research on Priority Areas, MEXT.

Kanehisa M, Deciphering Systemic Biological Functions by Integration of Genomic and Environmental Information, Bioinformatics Research and Development, JST.

Goto S, Hierarchical Structuring and Integration of Knowledge in Life Sciences, Integrated Database Project, MEXT.

E-zyme: Predicting Potential EC Numbers from the Putative Enzyme Reactions

The high-throughput screenings of biochemical compound libraries have been producing huge amounts of chemical data, and we are now confronted with the necessity to automate the processing and interpretation of such chemical data in order to derive biologically meaningful information. There are numerous enzyme reactions known to be present in various metabolic pathways but without any official EC (Enzyme Commission) numbers, most of which have no hope to be given ones because of the lack of the published articles on enzyme assays.

We have been developing a new method to predict an EC sub-subclass based on our original biochemical transformation pattern which we call an "RDM pattern", and develop a web-server called "E-zyme" which enables us to automatically assign the potential EC numbers to given pairs of substrates and products, or uncharacterized reactions. The original version of the E-zyme was established in 2004 (Kotera et al., *J. Am. Chem. Soc.*, 2004, **126(50)**: 16487-16498), and we published the latest version with improved coverage, recall and precision in 2009 (Yamanishi et al., *Bioinformatics*, 2009, **25(12)**: i179-i186). The E-zyme system can provide a link to the corresponding enzyme candidate genes. The next possible development involves specifying which genes are actually involved in the reaction of interest for a specific organism.

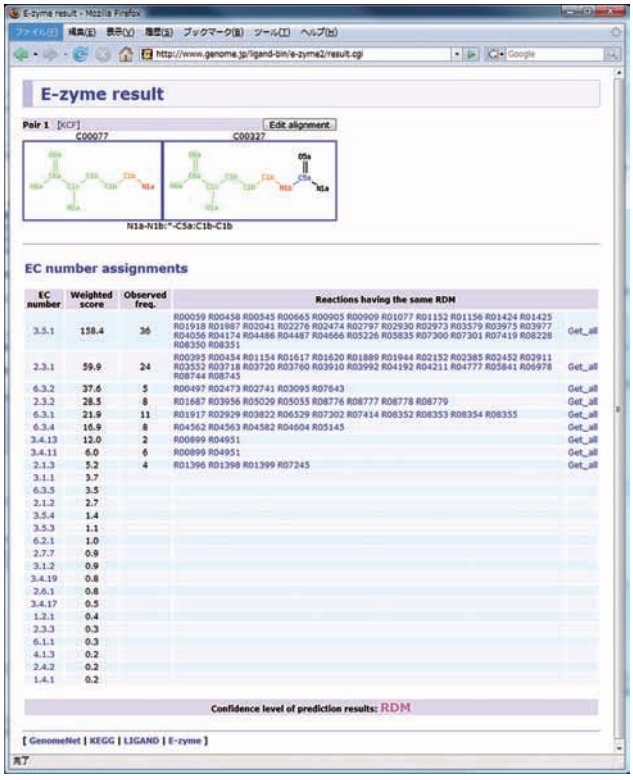


Figure 1. A screenshot of the E-zyme output page.

varDB: a Database for Studying Antigenic Variation

Antigenic variation plays a major role in immune evasion and establishment of persistent infections for many pathogens, including HIV (AIDS), *Plasmodium falciparum* (malaria), etc. Due to the inherent complexities associated with multi-gene families, antigenic variation studies are usually focused on single gene families and restricted to a small number of organisms. To lessen these limitations and promote cross-species and comparative genomic studies, we have developed varDB. VarDB is a public resource that collects genes and proteins from known antigenic variant gene families. The main goals of the varDB project are: I) to serve as a repository for antigenic variant gene families, II) to work as a platform for the analysis of antigenic variation between different organisms, and III) to be a community driven resource enabling synergistic cooperation from experts in different antigenic gene families. As of October 2009, the varDB project contains sequence data of 49 gene families, from 31 different pathogens that cause 22 diseases. Together, more than 68,000 sequences are available, including those obtained from clinical samples around the world. The database is expected to expand in the future as new sequences are being submitted to repositories like GenBank, and more antigenic variant gene families are identified. Many different tools for sequence analysis are integrated, providing a unique framework for cross-species analysis.

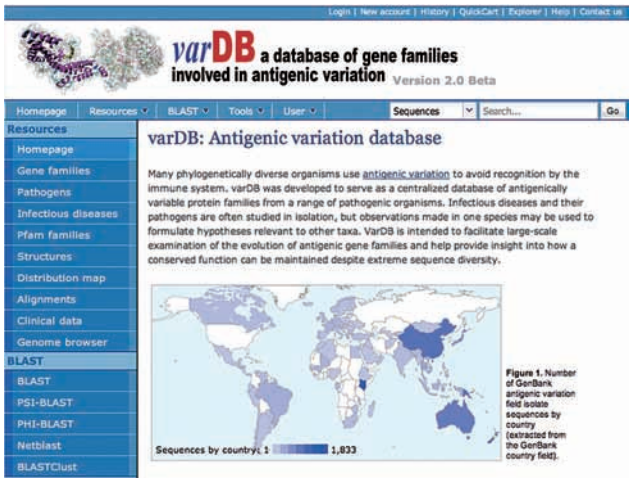


Figure 2. Home page of the varDB project, showing in the left panel the resources and part of the available tools. The map indicates the location and density of antigenic variant sequences collected in the database.