

Integrated GENES Database in KEGG

Kazushige Sato ¹

kazs@scl.kyoto-u.ac.jp

Toyoko Katsurada

toyoko@scl.kyoto-u.ac.jp

Yoshiko Kimura

yoshiko@scl.kyoto-u.ac.jp

Minoru Kanehisa

kanehisa@kuicr.kyoto-u.ac.jp

Institute for Chemical Research, Kyoto University

Uji, Kyoto 611, Japan

Abstract

KEGG provides the linkage between the catalog of molecular components and the network of molecular interactions in living cells and organisms. While the latter has been organized as the PATHWAY database in KEGG, we have relied on the existing databases for the former except the LIGAND database of chemical compounds. Because genes and gene products are the major molecular components, and because of the conflicts that we identify with the original authors and with the existing databases, we have decided to maintain our own gene catalogs, eventually for all organisms, but starting with the organisms whose genomes are completely sequenced. We report here the implementation of the GENES database in KEGG.

1 Introduction

KEGG (Kyoto Encyclopedia of Genes and Genomes) [1] is our attempt to describe, utilize, predict, and possibly design biological systems based on the genomic information. The existing sequence databases are organized in such a way that the sequence is the core information with everything else is an attribute or an annotation to the sequence. We take a view at a higher level where a gene or a molecule is the basic element that interacts with each other to form a network of living systems. Thus, the amino acid sequence and the nucleotide sequence are considered just two of the many attributes of this element.

The KEGG/GENES database is being organized from this viewpoint. At the moment, most of its contents are automatically generated from a number of sources, we plan to incorporate information on interacting partners that are manually identified in the KEGG project.

2 System Overview

KEGG provides a functional hierarchy of genes based on the classification of metabolic and regulatory pathways, as well as other hierarchies. The GENES database corresponds to the KEGG version of the hierarchy. That is, any gene name in the KEGG gene catalog has a link to this database.

Each GENES entry has ENTRY, NAME, CLASS, POSITION, CODON_USAGE, AASEQ, and NTSEQ items (if any). ENTRY and NAME are the gene accession and the description, which are exactly the same as the KEGG gene catalog. AASEQ and NTSEQ are amino acid and nucleotide sequences, respectively. The sequence information is taken mostly from the SWISS-PROT and GenBank databases. CODON_USAGE is a codon usage table for this gene. A summary of codon usage in all genes is also compiled by KEGG for each organism, and the statistics can be viewed graphically.

¹Permanent address: Nihon SiliconGraphics Cray K.K., Umeda, Kita-ku, Osaka 530, Japan

CLASS is a classification according to the KEGG pathways. This item has corresponding map numbers that link to the KEGG pathway maps. By clicking on the link, the pathway map is displayed with this gene marked in red, thus showing the functional location on the pathway. It is easy to switch from one organism to another on the pathway map to see if homologs exist in other organisms. Note that the pathway component is marked green if the corresponding gene exists in the given organism.

POSITION represents the physical position in the chromosome. It links to the genome map viewer written in Java. The viewer shows physical position of the gene graphically and provides an enlarged window showing the neighboring genes on the chromosome.

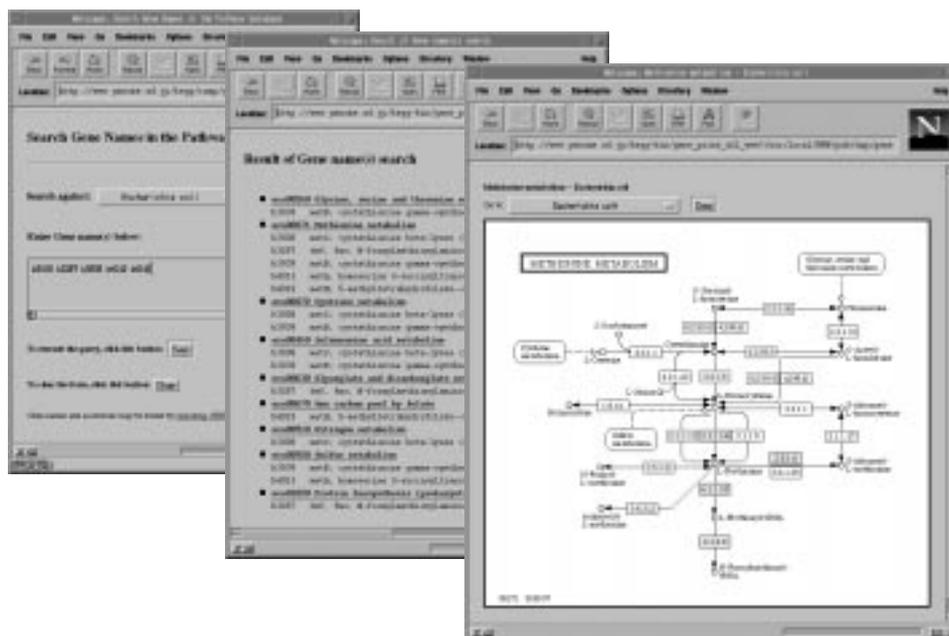


Figure 1: Genes marking interface

While the KEGG databases provide numerous links that can be used to browse a number of different aspects of genes and genomes, KEGG can also be used as a search tool. For example, genes can be searched on the known pathways in KEGG by EC numbers, by gene names, or sequence similarity. Figure 1 shows an example of the search by gene name.

Acknowledgments

This work was supported in part by a Grant-in-Aid for Scientific Research on Priority Areas, ‘Genome Science’, from the Ministry of Education, Science, Sports and Culture of Japan. The computation time was provided by the Supercomputer Laboratory, Institute for Chemical Research, Kyoto University.

References

- [1] Goto, S., Bono, H., Ogata, H., Fujibuchi, W., Nishioka, T., Sato, K., Kanehisa, M., “Organizing and computing metabolic pathway data in terms of binary relations,” *Pacific Symposium on Biocomputing '97*, pp. 175–186 (1997).
- [2] Sato, K., Katsurada, T., Kamiya, T., Egoshi, M., Kanehisa, M., “Java applets for handling genome maps in KEGG,” *Proc. Genome Informatics Workshop 1996*, pp. 254–255 (1996).