W9

HyperGenomeの紹介

HyperGenome : An integrated system for mapping data

久原　哲[1]　高木利久[2]　金久　實[3]

[1]九州大学農学研究科　[2]東京大学医科学研究所　[3]京都大学化学研究所

Satoru KUHARA[1] Toshihisa TAKGI[2] Minoru KANEHISA[3]

1Graduate School of Genetic Resources and Technology, Kyushu University

2Institute of Medical Science, The University of Tokyo

3Institute for Chemical Research, Kyoto University

Abstract

Features of the HyperGenome System are described herein. This system was developed to integrate information from the human genomic map from Genome Data Base (GDB) to DNA sequences from GenBank. DNA sequence data were assigned to locus on genomic map by checking cross references. The Open Window has been adopted for display information.

1. Introduction

There are various facts of experimental information related to the human genome project. The top of reference information we focused on is the genomic map, and the bottom information we have focused on is DNA sequences. Information on the human genomic map was collected in Genome Data Base(GDB). The DNA sequence information was gathered in GenBank, EMBL, and DDBJ. And the amino acid sequence were gathered in Protein Identification Resources(PIR). The formats of databases are different from each other. In recent years, certain kinds of cross references which link two databases were created by some databases [1]. A researcher, however, needs access to numerous databases to acquire information of the genome related to a particular research. It would be difficult for a molecular biologist to handle a SQL system for retrieving information. Integrated database systems have been constructed, for example Chromosome Information System [2] and Encyclopedia of the Mouse Genome [3]. Almost of those systems involve use of a pointer device as a mouse and a window system to get the information that a researcher requires.

from Locus to sequence by pointing button or locus in a window.

Hypergenome has nine functions: Chromosome, Locus, Select, Full, Reference, Sequence, Probe, Keywords, and Superimpose. These functions are performed using a clicking button (Fig. 2). In the locus window (which is the basic window), the banding pattern and assigned gene symbol will be displayed and any locus can be selected by clicking on a gene symbol. Information on reference, probe, and sequence of a selected locus will be shown in each window by clinking on the button. Clicking on any chromosome in the chromosome window will make it appear in the locus window. The keywords are collected from the gene symbol, name, location, title of reference, probe name, and GenBank accession number. We can therefore retrieve a locus by keywords in a keyword window. The chromosome on which the selected locus assigned will be displayed and the gene symbol will be darkened in the locus window.

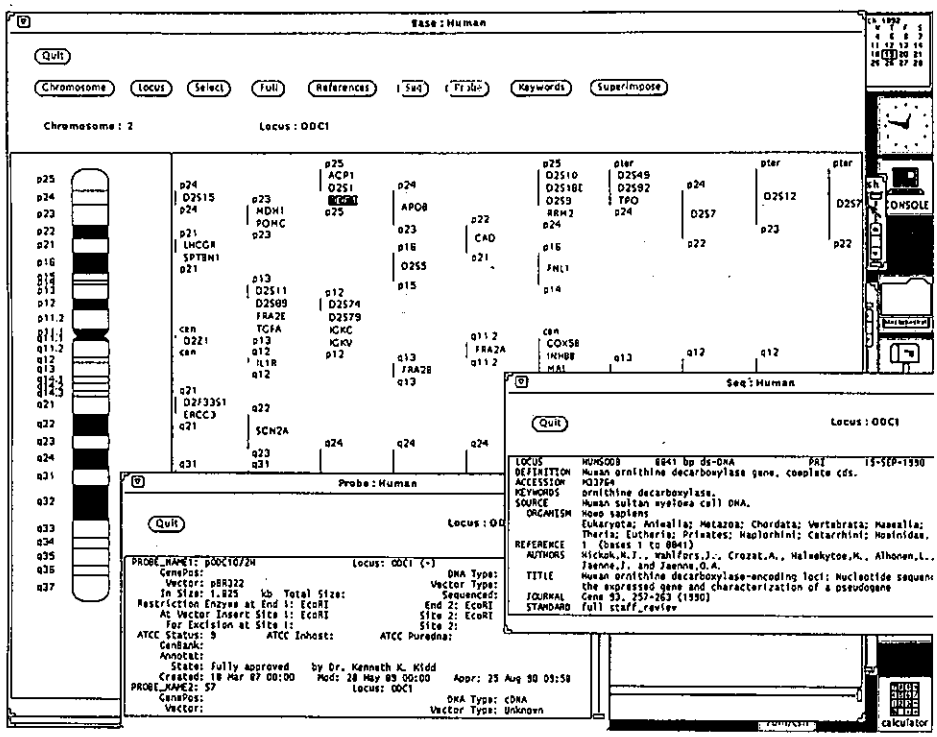

Fig. 2. Major types of windows in the HyperGenome System

The function of superimpose is to overlap a chromosome on the selected chromosome to determine the difference between chromosomes of different species.

In attempts to integrate information from the human genomic map and DNA sequences, we set up an integrated database Hypergenome for ready access to information that the user needs.

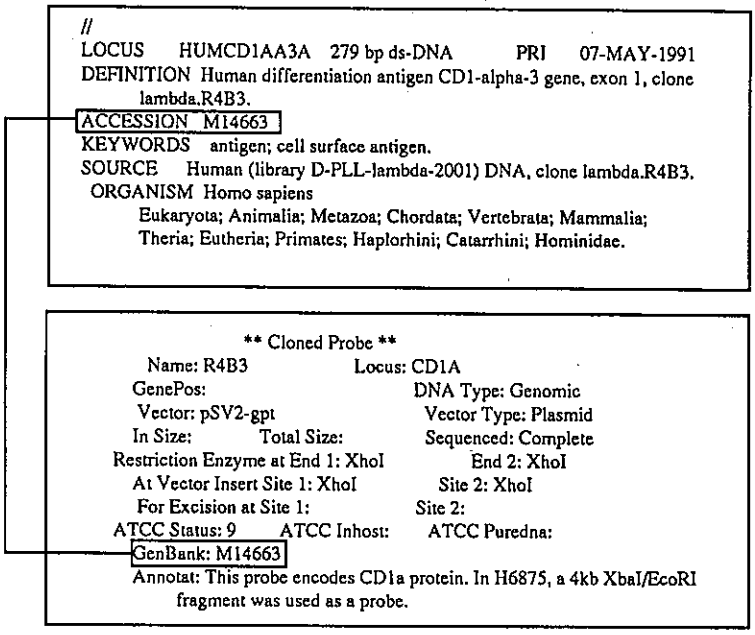

Fig. 1. Integration of GDB and GenBank

2. System configuration

The integrated database Hypergenome supports the parallel display of information with each appearing in a separate window. This system works in Open Window on SUN SPARCstation 2.

In the genomic map, we gathered information on the locus, probe and reference in GDB. In the DNA sequence, GenBank entry which corresponds to each locus in GDB was chosen from GenBnak. The map information was chosen according to assignment to each chromosome, and the input data were transformed to the Hypergenome. The selected entries from GenBank were collected according to assignment to each chromosome then were transformed into input data to the Hypergenome(Fig 1).

The assignment of entry of GenBank to locus in GDB was confirmed by agreement between the accession number described in probe of GDB and the map position mentioned in features of GneBank. The input errors were picked up in cross references. Appropriate correction for input error was made by inspection of the definition of the gene.

We applied the Open Window system and graphical user interface to acquire information

3. Conclusion

We have registered over 4000 gene symbols and DNA fragments from GDB and over 1500 DNA sequence entries have been registered. Almost all entries related to the human genome project are incorporated into Hypergenome. This system can provide ready retrieval by graphical user interface. Thus, molecular biologists with little computer experience, can search for any kind of information on the genomic map to the DNA sequence.

References

[1] Bilofsky, H. S. and Burks, C., The GenBank(R) Genetic Sequence Data Bank, Nucl. Acids Res, Vol. 16 1861-1864, 1988

[2] Johnston, W., Lewis, S., Markowitz, V., MaCarthy, J., Olken, F., and Zorn, M., Chromosome Information System, Lawrence Berkeley Laboratory Technical Report LBL 29675, 1991

[3] Encyclopedia of the mouse genome ver. 1.0, the Jackson Laboratory 600 Main Street Bar Harbor, ME, 1990