Extension of the Integrated Database "HyperGenome" for Genome Maps and Sequence Information

Susumu Goto[1]
goto@kuicr.kyoto-u.ac.jp

Satoru Kuhara[2]
kuhara@grt.kyushu-u.ac.jp

Toshihisa Takagi[3]
takagi@ims.u-tokyo.ac.jp

Minoru Kanehisa[1]
kanehisa@kuicr.kyoto-u.ac.jp

[1] Institute for Chemical Research, Kyoto University
[2] Graduate School of Genetic Resources Technology, Kyushu University
[3] Human Genome Center, Institute of Medical Science, The University of Tokyo

1 Introduction

The Human Genome Project is rapidly accumulating various kinds of data on the genome. Databases were developed in order to organize data obtained during experiments (GenBank, EMBL, DDBJ, PIR, Swissprot, PDB, GDB, MIM, etc.). We developed a database system called "HyperGenome"[1, 2] which integrates data on genome mapping and DNA sequences. The HyperGenome system provides a user friendly integrated interface for a database of genome maps and DNA sequences. We assume that the average users of the HyperGenome will not enter their own data into the database, rather they retrieve information which is currently stored separately in genome mapping databases and/or in DNA sequence databases.

2 System Overview

Using the HyperGenome system, the user can retrieve the following information related to a locus by directly pointing to the locus on a chromosome displayed on a window.

1. Information of locus mapped in certain region on chromosome (from GDB)
2. Cell line information (from GDB)

[1]五斗 進，金久 實：京都大学化学研究所，〒611 京都府宇治市五ヶ庄
[2]久原 哲：九州大学農学研究科遺伝子資源工学，〒812 福岡市東区箱崎 6-10-1
[3]高木利久：東京大学医科学研究所ヒトゲノム解析センター，〒108 東京都港区白金台 4-6-1

3. DNA library information (from GDB)
4. Probe, PCR, aso information (from GDB)
5. Order of loci on a chromosome (from GDB)
6. DNA and amino acid sequences related to loci on a chromosome (from GenBank, GDB and PIR)
7. Relation between a chromosome map and a genetic map (from GDB and linkage data (NIH/CEPH Collaborating Mapping Group, 1992))
8. Information of genetic disorders (from MIM)

Data are organized around locus data from the genome mapping database (GDB). Graphical images of chromosomes (chromosome windows) are generated from the locus data. A user of the HyperGenome selects a locus in the chromosome window and retrieves information related to the locus. The HyperGenome is currently developed on SUN Sparc series workstations running SUN OS 4.1.X and OpenWindows Ver.2. There are two versions of the HyperGenome which use different data management methods for data on genome mapping. One version uses an internal database management routine while the other has a relational database interface and manages the genome mapping data in relational tables. One of the important problems is the update of data in the HyperGenome. Both GDB and GenBank are updated frequently. The database of the HyperGenome can be updated by importing data from GDB and GenBank.

3 Future Plan

We are planning to implement a client and server system by using internet. We are designing GDML (Genome Database Manipulation Language), which manipulates the query and/or formulea of SQL, based on HTML (Hyper Text Markup Language). We are also designing an extended HTTP (Hyper Text Transfer Protocol) in order to transfer GDML. The server has GDML generator, GDML-HTML trasnlator, GDML-SQL translator and some kinds of data. The server allows searches of sequence information from WebDBGET through GDML-HTML translator, and also searches of map information from GDB through GDML-SQL translator. Consequently, our system enables the user to continuous working with the GUI independently of updating data.

References

[1] T. Suzuki, T. Takagi, S. Kuhara and M. Kanehisa, Development of an integrated database for genome mapping and nucleotide sequence, *Proc. Twenty-Seventh Annual Hawaii International Conference of System Sciences*, pp. 68-76, 1994.

[2] T. Suzuki, S. Nakashima, T. Takagi, S. Kuhara and M. Kanehisa, Enhancement of the Integrated Database "HyperGenome" for Genome Maps and Sequence Information, *Proc. Genome Informatics Workshop IV*, pp. 362-369, 1994.