

TOWARDS THE ASIA-PACIFIC BIOINFORMATICS NETWORK

H. SUGAWARA and S. MIYAZAKI

*DNA Data Bank of Japan, Center for Informatics Biology,
National Institute of Genetics,
111 Yatai Mishimizuoka 41, Japan*

A vigorous effort has been undertaken to develop a robust and high speed network in the Asia-Pacific region which satisfies the needs of the scientific community. Consortium for Asia-Pacific Advanced Network (APAN) are established in some countries in this region. Our group, along with collaborators in Singapore and Australia, is moving towards the establishment of Asia-Pacific Bioinformatics Network (APBioNet) based on APAN. In this paper, we describe the concept of APBioNet and introduce the activity of DNA Data Bank of Japan (DDBJ) which will play an important role in APBioNet. Some salient features of APBioNet are also discussed.

1 Introduction

The Internet and the World Wide Web help capture, disseminate and share biological data, information and knowledge within the scientific community. We could expect that this technology diminishes any barriers against information flow worldwide. However, it is highly probable that this technology increases the gap between developing countries and developed countries from the viewpoint of information access contrary to the expectation. It is obvious that scientists who have no access to the Internet suffer from disadvantages relative to those who enjoy networking. It is not feasible for those who have limited facilities to surf the International Nucleotide Sequence Databases (INSDB), the amino acid sequence databases, the Protein Data Bank and many other biological sites.

Thus, there exists a strong demand for establishing broad-band networks to every corner of the world. A vigorous effort has been undertaken to develop a robust and high speed network in the Asia-Pacific region which satisfies the needs of the scientific community. Consortium for Asia-Pacific Advanced Network (APAN) are established in some countries in this region¹. Our group, along with collaborators in Singapore and Australia, is moving towards the establishment of Asia-Pacific Bioinformatics Network (APBioNet) based on APAN

2 Concept of APBioNet

In the Asia-Pacific region, there are many countries which will produce biological data (especially data on biodiversity) in a few years. It is very important to prepare the infrastructure for the sharing of information among developing countries and developed countries such as US. For example, the National Institute of Genetics in Japan provides a Web data submission tool for DNA Data Bank of Japan (DDBJ) which organizes the International Nucleotide Sequence Data Bank with GenBank of National Center for Biotechnology Information in Bethesda and EMBL of European Bioinformatics Institute in Hinxton, and the tool is used not only in Japan but also in other Asia-Pacific countries. However, it takes scientists in Asia and Australia more than one hour to register a piece of nucleotide sequence and its relevant biological information due to the slow network. Thus the tool is not efficient for those scientists.

Therefore APBioNet was proposed³ to facilitate smooth flow and sharing of biological data, information and knowledge free from boundaries caused by physical distance, social system and gaps in technology and culture. APBioNet will consist of:

- a fast and robust links for communication
- a network for the real-time integrated use of distributed databases and applications
- facilities for visualization of large scale and heterogeneous data
- an interactive simulation of life phenomena for the advanced study on life sciences and biotechnology
- distance learning systems based on multi-media server to foster bioinformatics research in the region

3 Existing elements of the future APBioNet

There are several activities of bioinformatics which may compose the core of APBioNet. They are DNA Data Bank of Japan (DDBJ)⁴ and GenomeNet⁵ in Japan, Australian National Genomic Information Service (ANGIS)⁶ and Bioinformatics Center in Singapore⁷.

DDBJ started its function in 1986 and has a special relationship of close international cooperation with INSDB, to which DDBJ contributes about 10% of the data contents. DDBJ exchanges updates with EMBL and GenBank daily which are the other two members of INSDB. Therefore, DDBJ, EMBL

and GenBank provide the same data contents to the scientific communities from three different places. The services of DDBJ are introduced in the next section and the introduction of the other two elements will be found in other literature.

4 Services of DDBJ

The primary task of DDBJ is the collection and dissemination of sequence data and DDBJ heavily relies on the World Wide Web server⁸ for this task. For example, 80% of the data submissions done through Sakura⁹ which is a data submission system for Web browsers. Data dissemination including usage of FASTA and BLAST through Web browsers has been increasing since February 1997 when DDBJ enabled Web interfaces to the chromosome search programs.

DDBJ provides many types of services in addition to data submission and chromosome searches. For example users can have E-mail accounts for their research based on sequence data. The Home Page menu of DDBJ (shown below) clearly introduces users to the wide variety of DDBJ's roles.

- What's new and others

- We released the window analysis program for synonymous and non-synonymous nucleotide substitutions, developed by Toshinori Endo and Takashi Gotohori.
- The Japan Escherichia coli genome project team newly determined the range of 12.6-70 mb and compiled the whole genome sequence of E. coli (as of 97/01/23) that is represented in an interactive Web with graphics.
- Latest News from DDBJ (Japanese Page Only)
- Introduction of DDBJ, DDBJ KIT (Japanese Page Only), DDBJ News Letter No. 17
- International Collaboration of DDBJ/EMBL/GenBank Nucleotide Sequence Databases

- Submission of Nucleotide Sequences

- DDBJ "DNA Data Submission System" SAKURA " ← Please try!
- Frequently asked questions on Data Submission

- Explanation of data items and description
 - **DDBJ Searches and Data Analysis Services**
 - Data Retrieval by accession numbers using getentry
 - Data Retrieval by key words using SFGate & WAIS
 - Homology Search using FASTA and BLAST
 - Multiple Alignment using nalign
 - Multiple Alignment using clustalw
 - Gopher/Anonymous FTP/E-mail Search server
 - Links to Other Databases
 - **Other Databases developed at NIG Mshina**
 - Codon Usage Database (CUG) [developed by Toshihide Ikemura's group]
 - Protein mutant DB [developed by Ken Nishikawa's group]
 - C. elegans EST database (YK Project) [developed by Yujiko Kohara's group]
 - Expression pattern map of the C. elegans genome [developed by Yujiko Kohara's group]
 - Bacillus subtilis Non-Redundant Database (NRSub) [developed by Guy Perrier and Takashi Gjobori]
 - Databases maintained at Genetic Resources Laboratory
 - HLA database [Molecular Classification Lab. with Osaka U. and RIKEN]
 - WDCM [VCC World Data Centre for Microorganisms by SUGAWARA Hiideaki's group]
 - Escherichia coli genome browser (by Molecular Classification Lab and et al.) (List of Clones)
 - **Others**
 - Introduction of the NIG Super Computer System
-

It is now obvious that DDBJ services cover

- communication
- data retrieval
- homology search
- sequence analysis
- value-added databases
- genome databases
- link to other bioinformatics sources.

These will also be fundamental to APBioNet.

5 Advanced features of APBioNet

Broad-band networks are required even for the most basic procedures such as data submission to INSDB. In the current state, Sakura is not a practical tool for data submission outside of Japan.

In addition, broad-band networks like APAN promise more than that. In the case of biodiversity study, a tremendous amount of samples are scattered in fields and laboratories in regions from the deep sea to the high mountains. The original data sources are distributed among a number of distant sites. However, large scale databases frequently updated such as INSDB and high-performance data analysis computing reside on a limited number of sites. Therefore virtual laboratories which facilitate remote integration of data sources and remote computing are required. They will only be feasible if broad-band networks are established.

Broad-band networks also allow us distant learning systems based on multi-media text on server. With this system students situated at home, school and university are able to study several aspects including the manipulation of samples and instruments through interactive multi-media sessions. The distant learning is very useful to foster local research in specific subjects and is indispensable to erase the gap between developing countries and developed countries.

References

1. <http://apan.imnet.ad.jp/>
2. R. J. Robbins, in *Bioinformatics: Essential infrastructure for global biology* (OECD Megascience Forum Biological Informatics Working Group Washington DC 1996).
3. H. Sugawara *et al.*, *APBoNet: Asia-Pacific Bioinformatics Network* (APAN Workshop, Tokyo 1997).
4. Y. Tateno and T. G. Jobori, *Nucleic Acids Research*, **25-1**, 14 (1997).
5. <http://genom.ad.jp/>
6. <http://www.angis.su.oz.au/>
7. <http://bi.c.nus.sg/>
8. <http://www.ddb.j.nig.ac.jp/>
9. H. Yamamoto *et al.*, *SAKURAI: A new data submission system of DDB to meet users' needs in the age of mass production of DNA sequences*, (Genom Informatics Workshop 96, Tokyo 1996).
10. T. Koike *et al.*, *Development of New DDB DNA Sequence Database with Data Annotation Tool Yantol II*, *Genom Informatics 1996*, (Proceedings of the Seventh Workshop on Genom Informatics, 157 1996).