

DNA Information and Stock Center in Japan

Chikayoshi KITAMURA*

Abstract

Recently, studies in the genome researches of several agricultural species have been carried out actively. AFFRC/MAFF is developing the DNA Information and Stock Center (tentatively designated as "DISC") at NIAR to assist researches involved in DNA studies. DISC will provide tools to enable researchers to search DNA data and to analyze them under the co-ordination of DNA Data Bank of Japan (DDBJ). DISC will also provide genome maps in future as well as DNA clones such as RFLP markers and cDNA clones.

The development of the DNA information and clones center in the world and our plan to set up DISC will be introduced in this paper.

Introduction

The large genome projects aimed at decoding the totality of genetic information encoded in the double strands of DNA in the chromosomes were initiated from the mid-1980. Among these projects, the Human Genome Project is prominent in scale and international co-operation. Its early development occurred in USA under the support of DOE (Department of Energy) and NIH (National Institutes of Health).

Organisms listed in Table 1 with their genome size are targeted in the genome projects. The ultimate goal of these projects is to determine the whole nucleotide sequences of these genomes. The development of improved technologies for mapping and sequencing, and of informatics system to handle diverse data may facilitate the generation of many sequence data, mapping data and biological materials (e.g. cell lines, libraries and DNA fragments). Therefore, DNA banking becomes important.

Repository of DNA

In the repository services, collection, evaluation, amplification, storage and distribution of DNA, and management of updated information about the accessions are included. Recently some repositories have been set up which deal with DNA materials from each different organisms.

1 Human genome project

In USA, ATCC (American Type Culture Collection), a nonprofit organization, which is funded under contract with the National Institute of Child Health and Human Development (NICHD) and National Center for Research Resources, maintains the repositories of human and mouse DNA probes and libraries.

Cloned genes and DNA probes are maintained as plasmid-containing bacterial cultures, as phage lysates and purified DNA. Oligonucleotides for amplification of human DNA at a specific locus are also distributed. ATCC offers more than 300 human brain cDNA clones. These EST (expressed sequence tagged) clones described in *Science* 252: 1651-1656 (1991) have come from the laboratory of J. Craig Venter and they are also available. [Contact to the ATCC repository information: Donna Maglott, Fax +1-301-770-1541]

In Europe, DNA Probe Bank has been set up in the UK HGMP (Human Genome Mapping Project)

Presented at the 27th International Symposium on "Plant Genetic Resource Management in the Tropics", Tsukuba, Ibaraki, Japan, 25-26 August 1993, held by Tropical Agriculture Research Center (TARC).

*Department of Research Planning and Coordination, National Institute of Agrobiological Resources

Table 1 Organisms targeted in genome research and genome size (base pairs in haploid genome)

Species name	Genome size
Microorganisms	
<i>Escherichia coli</i>	4.7×10^6 bp
<i>Bacillus subtilis</i>	4.2×10^6 bp
<i>Saccharomyces cerevisiae</i>	1.5×10^7 bp
<i>Schizosaccharomyces pombe</i>	1.4×10^6 bp
Plants	
<i>Oryza sativa</i>	3×10^8 bp
<i>Arabidopsis thaliana</i>	7×10^7 bp
<i>Zea mays</i>	5.7×10^9 bp
<i>Glycine max</i>	1.8×10^9 bp
<i>Triticum aestivum</i>	6×10^9 bp
<i>Hordeum vulgare</i>	5×10^9 bp
<i>Decale cereale</i>	7×10^9 bp
<i>Pinus taeda</i>	1.1×10^{10} bp
Animals	
<i>Caenorhabditis elegans</i>	1×10^8 bp
<i>Drosophila melanogaster</i>	1.4×10^8 bp
<i>Mus musculus</i>	2.7×10^9 bp
<i>Homo sapiens</i>	3.2×10^9 bp

Resource Center funded by the MRC (Medical Research Council). It offers more than 800 probes to the UK research community freely and charges foreign researchers. [Contact: Fax +44-81-707-49527 or Fax +44-81-869-3807]. CEPH (Center d'Etude du Polymorphisme Humain) in France provides DNA samples to collaborating investigators [Contact: Fax +33-1-4206-1619].

In Japan, JCRB (Japanese Cancer Research Resources Bank) has been organized in the Foundation for Promotion of Cancer Research under the supervision of the Ministry of Health and Welfare of Japan in 1985. Its activity is supported by the Japan Ship-building Industry Foundation. JCRB has collected more than 1500 DNA clones and delivered more than 13,000 clones to about 3,000 researchers in this seven year period [Contact: K. Hahimoto, Fax +81-3-5285-1181]. RIKEN Gene Bank started RGB DNA Bank

Respository of DNA materials	
	American Type Culture Collection (USA)
	UK HGMP Resource Center (UK)
	Centre d'Etude du Polymorphisme Humain (France)
	Japanese Cancer Research Resource Bank (Japan)
	Riken Gene Bank DNA Bank (Japan)
Mapping Database	
	Genome Data Base
	integrated with Online Mendelian Inheritance in Man
Sequence Databases and Software Tools	
	National Center for Biotechnology Information (USA)
	integrated with Medline/NLM/NIH

Fig. 1 DNA Respository and Information Center in Human Genome Project

from 1991. RGB distributes DNA clones, RFLP markers and STS oligonucleotides. [Contact : E. Soeda, +81-298-36-9120]

2 Mouse genome project

In USA, ATCC maintains the collection of mouse DNA besides human DNA. Jackson Laboratory provides genomic DNA from mouse strains and mutant stocks maintained there. A Mouse DNA Resource catalogue is available [Contact : Fax +1-207-288-3398].

3 Arabidopsis genome project

European DNA Resource Center at Koeln is funded by EC BRIDGE to provide several DNA libraries [Fax +49-221-5062-613]. It collaborates with the Arabidopsis Stock Center at Nottingham University [Fax +44-602-424270].

In USA, Arabidopsis Biological Resource Center at Ohio University is funded by NSF (National Science Foundation) to provide seeds and DNA stocks [Contact : Randy Scholl, Fax +1-614-292-0603].

Mapping Database

Mapping databases are of special importance for genome research. The scientific community has begun to collect valuable genetic maps of man, of agriculturally significant plants and animals, and of model organisms, though often not in computer-readable forms.

GDB (Genome Data Base), the database of human chromosome mapping, is a Howard Hughes Medical Institute project in collaboration with The Johns Hopkins University School of Medicine and the William H. Welch Medical Library, and is financially supported by NIH and DOE (Department of Energy). GDB is implemented in the Sybase relational database management system, and is integrated with OMIM (Online Mendelian Inheritance in Man), the full-text database of inherited traits and diseases [Contact : Fax +1-301-955-0054 ; e-mail help@welch.jhu.edu].

The nematode database, ACEDB (A *Caenorhabditis elegans* Data Base) developed by the researchers of the Laboratory of Molecular Biology, MRC, which contains information on genetic maps, mapped clones, sequences and bibliography, is distributed free of charge via Internet [Contact : Richard Durbin, e-mail rd@cele.mrc-lmb.cam.ac.uk].

AAAtDB (An *Arabidopsis thaliana* Data Base) is developed by the scientists at the Massachusetts General Hospital (MGH) in Boston on the basis of ACEDB. The database is one of the objectives of the Multinational Coordinated *Arabidopsis thaliana* Genome Research Project. It contains a physical map, genetic map, list of seed stored in Nottingham, literature and sequences from GenBank [Contact : Howard Goodman, Fax +1-617-726-6893 ; e-mail curator@frodo.mgh.harvard.edu].

Sequence Database

Sequence database was initiated based on the publication of *Atlas of Protein Sequences and Structure* by M. O. Dayhoff at Georgetown University in 1965. In 1978, this Chemical Biology group of the NBRF (National Biomedical Research Foundation) at Georgetown University started to develop computer retrieval systems for the two databases, the Nucleic Acid Sequence Database System and the Protein Sequence Database System. In 1984, the NBRF group became the Protein Identification Resource (PIR), offering a variety of services, including custom searches and on-line access. The PIR project is supported by the National Library of Medicine, NIH. The protein sequence database is currently an international project implemented in collaboration with the NBRF, the Martinsried Institute for Protein Sequences (MIPS) and the International Protein Information Database in Japan (JIPID) [Contact : W. C. Baker, Fax +1-202-687-1662 ; e-mail barker@gunbrf.bitnet]. In addition to the PIR-international, in Europe, A. Bairoch at the University of Geneva developed SWISS-PROT, the database of protein and peptide sequences, and PROSITE which stores sequence motifs and patterns characteristic of enzyme active sites, etc. SWISS-PROT and PROSITE are linked together.

Since 1979, Los Alamos National Laboratory of DOE (Department of Energy) has collected all the

published nucleotide sequences in a computer retrieval database. In 1982, NIH started the GenBank project with LANL through contract with BBN (Balt Bernek and Newman, Inc), later with IntelliGenetics, Inc. (GenBank is a registered trademark of the U. S. Department of Health and Human Services). Since 1992, main part of the activity of GenBank was transferred to NCBI (National Center of Biotechnology Information) at NLM (National Library of Medicine)/NIH.

In Europe, EMBL (European Molecular Biology Laboratory) Data Library was established in 1980 to collect the nucleotide sequences in a database, to arrange and to distribute them. Since then, GenBank and EMBL have collected sequence data and related data and arranged them into a computer database. Later in 1986, DNA Data Bank of Japan (DDBJ) which was established at the National Institute of Genetics (NIG) at Mishima, joined these collaboration. Since 1991, all the nucleotide sequence data have been deposited in single public-distributing DDBJ/EMBL/GenBank databases. [Contact: DDBJ, Fax +81-559-75-6040, e-mail ddbj@ddbj.nig.ac.jp; EMBL, Fax +49-6221-387-306, e-mail datalib@EMBL-Heidelberg.DE; NCBI, Fax +1-301-480-9241, e-mail info@ncbi.nlm.nih.gov]

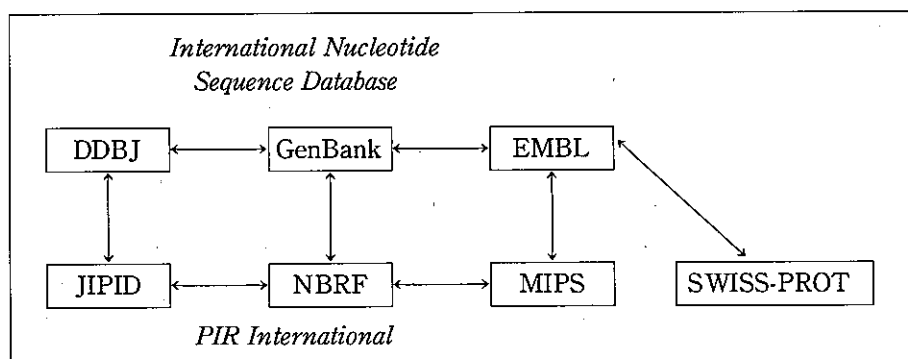


Fig. 2 Relationship of the International Sequence Databases

Plant Genome Data and Information Center (PGDIC) in USDA

USDA (U. S. Department of Agriculture) started Plant Genome Research Program in October, 1990 to promote the improvement of plants, in which agronomic, horticultural, and forest species were included, by locating important genes and markers on chromosomes, and determining the sequences of those genes.

At the same time, NAL (National Agricultural Library) opened the PGDIC (Plant Genome Data and Information Center), which offers the information on plant genome research in readily available format to users. PGDIC is developing a public database which includes information on the genetic, molecular, physical and cytogenetic maps and their mappable objects or site for four agricultural species, i. e. maize, soybean, wheat, and loblolly pine. Recently, *Arabidopsis* has been added. PGDIC aims at improving the AG-RICOLA, which is the bibliographic database produced by the NAL, in relation to biotechnology and plant genetics [Contact: S. MacCarthy, Fax +1-301-504-7098, e-mail smccarthy@nalusda.gov].

DNA Information and Stock Center in Japan

Recently in Japan, DNA-related research has been carried out actively in agricultural fields. Thus, the Agriculture, Forestry and Fisheries Research Council (AFFRC)/MAFF decided to establish a DNA Information and Stock Center (so-called DNA Bank, tentatively named "DISC") at NIAR (National Institute of Agrobiological Resources) as the sixth center of the Gene Bank Project to assist the DNA-related research community in these fields. In the Gene Bank Project, there are five Genetic Resources Centers, i. e. Plants, Microorganisms, Animals, Forest Trees and Aquatic Organisms. These centers have each sub-banks and deal with genetic materials mainly. On the other hand, DISC is scheduled to play a role as a stock center of DNA materials and an information center, to prepare the sequence databases and provided

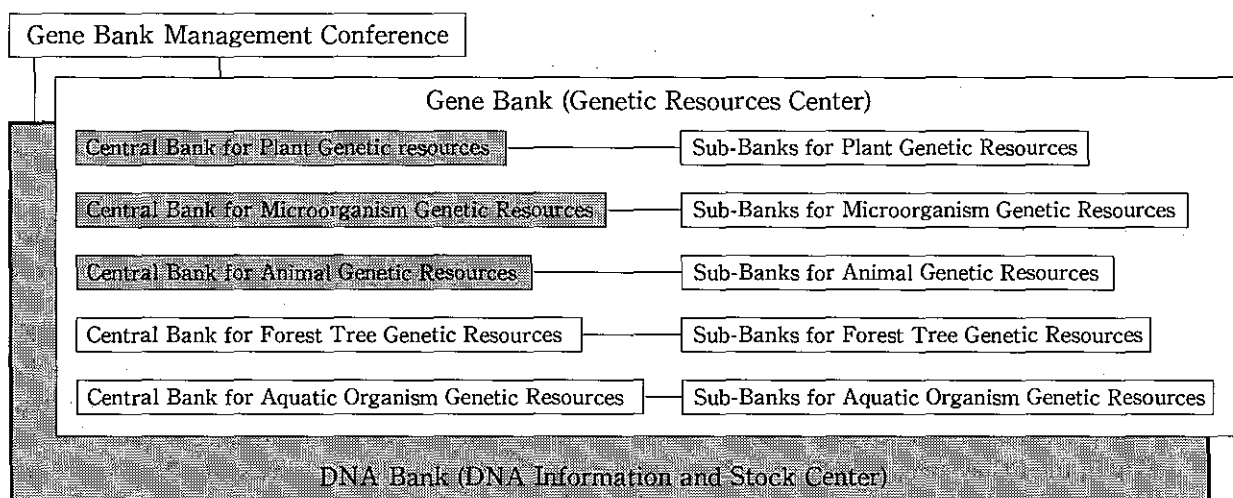


Fig. 3 Gene Bank Project of MAFF

tools to enable researchers to search up-to-date data and to analyze them.

DISC will start on 1st of October, 1993 and will be completed year by year. In the first step, a faculty of information is being set up this year in collaboration with DNA Data Bank in Japan at Mishima. A massively parallel super-computer will be introduced to run one of the full dynamic programming algorithms for sensitive search of homologous sequences. A high performance server will be used for fast search of homologous sequences. Two database servers will serve international public databases as the satellite of DDBJ. A graphic workstation will execute a 3D graphic software to display PDB (Protein Data Base) data. In addition, several UNIX workstations will be equipped to manage DISC. These computers will be connected to FDDI (Fiber Distributed Data Interface) network of Computer Center/AFFRC/MAFF in Tsukuba and via CC connected to Internet. Then researchers whose computer has internet environments can access DISC and use them.

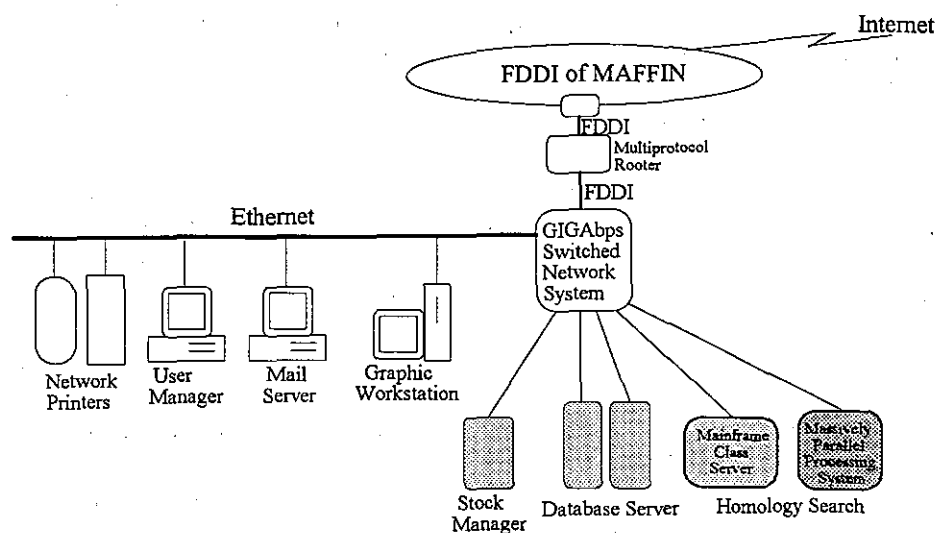


Fig. 4 Computer Facilities of DISC

In the near future, DISC will prepare mapping databases of agriculturally important organisms. DISC still plans to set up a faculty of materials where collection, evaluation, multiplication and distribution of DNA clones such as cDNA or RFLP markers will be conducted.