

## Network System for Animal Genome Informatics

Yasuhiko WADA and Hiroshi YASUE

Department of Animal Breeding and Genetics, National Institute of Animal Industry  
(Tsukuba, Ibaraki, 305 Japan)

### Abstract

An animal genome database for comparative mapping was developed on a UNIX workstation. Linkage map, cytogenetic map, reference, PCR primer information and mapping homology between species were included in the database. These data can be retrieved by electronic mail system and World Wide Web (WWW) client through the Internet. In addition, an online linkage analysis system was developed using CRI-MAP 2.4; this system includes the management of linkage data, computing of lod score for each pair of genetic markers and construction of linkage map. We have also developed a gene mapping workbench designated as gRanch. The system displays an integrated gene map and 2 kinds of homology maps between species. The functions for linkage data input and the linkage map analysis are included in the gRanch. Furthermore, Mouse Genome Informatics mirror site in Japan has been opened since August 23, 1996. The users of Mouse Genome Database and Gene Expression Database in Asia are invited to connect with the Japan mirror site.

**Discipline:** Animal industry/Biotechnology

**Additional key words:** gene mapping, linkage analysis, WWW

### Introduction

Genomic data, such as linkage map, cytogenetic map and PCR primer data, have been accumulated progressively with a growing complexity. For effective genome analysis, attempts to save all the genomic data into a genome database have been made for several species: The GDB<sup>2)</sup> is for human; GBASE<sup>4)</sup> and MGD (<http://www.informatics.jax.org/mgd.html>) for mouse; PiGBASE (<http://www.ri.bbsrc.ac.uk/pigmap/pigmap.html>) for pig; BovMaP (<http://locus.jouy.inra.fr/cgi-bin/bovmap/intro.pl>) for cattle; and US Meat Animal Research Center<sup>5)</sup> for pig and cattle. These databases have been maintained by each mapping consortium.

We are currently isolating the porcine genetic markers for linkage analysis to identify the gene responsible for phenotype in collaboration with the National Institute of Animal Industry, National Institute of Animal Health and STAFF Institute. In connection with this project, we have developed an animal genome database focusing on comparative mapping among pig, cattle, mouse and human. In addition, this database has provided information about the

linkage map, cytogenetic map, PCR primers, and references on the above species through the World Wide Web (WWW)<sup>1)</sup> and electronic mail system.

In addition, the online linkage analysis system has been developed. The system manages the linkage database and constructs the linkage map using the parallel computing technique through WWW clients. We have also developed a gene mapping workbench designated as gRanch. The system displays the integrated gene map and 2 kinds of homology maps among species. The pedigree or allele data registration and the linkage analysis are included in the gRanch. The system overview is shown in Fig. 1.

### Animal Genome Database<sup>9,11)</sup>

The animal genome database has been constructed in a UNIX workstation using the relational database system. The genome data of human and mouse in our database are constructed by automatically converting GDB and MGD to our format every month. We have collected the genome data on pig and cattle from published reports, and we are updating these data at intervals of 3 months. Linkage map, cytogenetic map, references, PCR primer information and cytogenetic homology among species are

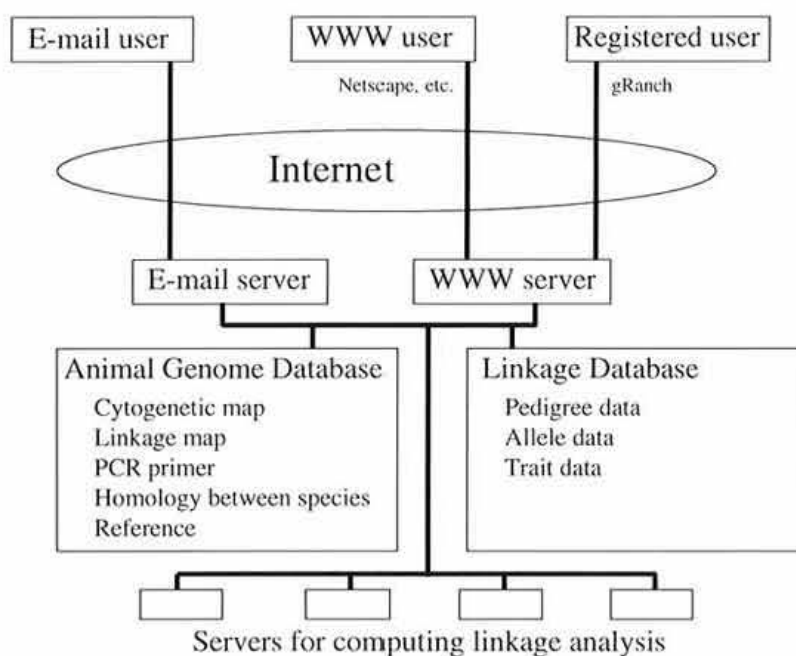


Fig. 1. Overview of animal genome database and online linkage analysis system

included in the database. The MGD linkage map for mouse, USDA linkage map for pig and cattle, and PIGM linkage map for pig are available in our system. The NIAI linkage map for pig will be available in 1997. The latest dates of data update have been displayed in the menu page (<http://ws4.niai.affrc.go.jp/jgbase2.html>). The present number of records of each table is shown in Table 1.

### Search system using World Wide Web

Clickable cytogenetic map, clickable linkage map, keyword search for locus symbol, locus name and chromosome, and reference search for authors, year of publication and title are available through the WWW.

Users first run a WWW client, such as NCSA Mosaic or Netscape, to connect with the Internet

and then open the home page of our genome database (<http://ws4.niai.affrc.go.jp/>). The news, a brief explanation of our database, and mode of search request are described in the home page. Users can display their favorite page simply by clicking the anchors for those pages. The keyword search, locus search and reference search are available for all the animal species registered in our genome database. Clickable cytogenetic maps are available for pig, cattle and mouse. Clickable linkage maps are available for pig and cattle. Java version of clickable linkage maps is available for pig, cattle and mouse. For non-graphical clients, although access to the above described pages is impossible, a linkage map viewer for pig, cattle, and mouse is available.

The clickable cytogenetic map of swine chromosome 1 showing an R-band pattern is depicted in Fig. 2 as an example. This page is displayed when the users click the chromosome 1 on the page of

Table 1. Number of records in various tables of the animal genome database

(as of February 17, 1997)

Table	Pig	Cattle	Mouse	Human	Total
Locus	1,404	1,454	24,298	11,603	38,759
Reference	284	227	4,485	874	5,870
PCR primer	631	443	2,006	67,756	70,836
Homology					2,172

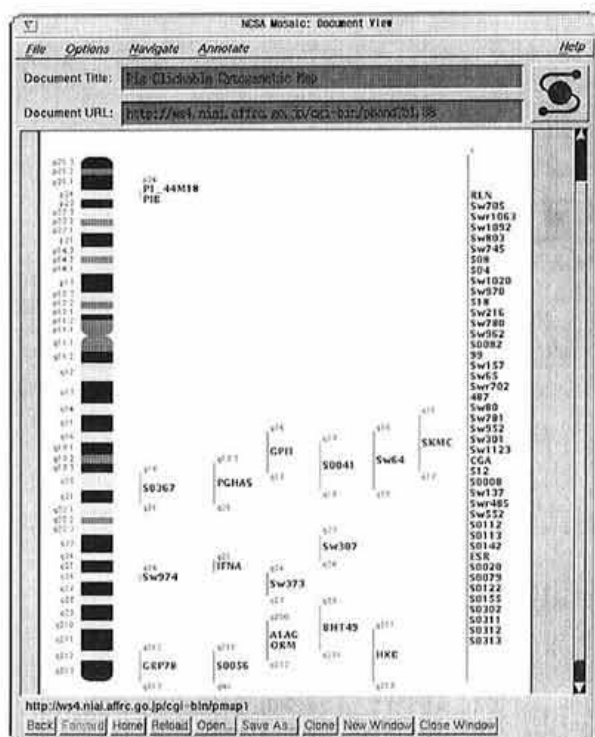


Fig. 2. Image of clickable cytogenetic map of swine chromosome 1

Locus with red letters was reported by our group; that with black letters, by other groups.

the swine chromosome R-band image panel. Detailed information on a given locus, including comparative mapping data among species, is displayed by clicking the locus symbol in the image. For example, the detailed locus information of IFNA gene is shown in Fig. 3. This system provides the primer sequences for IFNA gene analysis, homology information (localization of IFNA on cattle, human and mouse chromosomes) and IFNA reference.

Keyword search for locus symbol, locus name and chromosome is possible through WWW. Reference search for author name, year of publication and title, and locus search for locus symbol are also available. The binary tree index files for each search item are included in the database. The intersection set of data which contain the keyword in each item is displayed for WWW anchor.

For example, in the keyword search page, the input forms for locus symbol, locus name and chromosome are displayed. When users input the keywords into the forms, and then click the submit button, the anchor lines for selected data are displayed. The users can get more detailed informa-

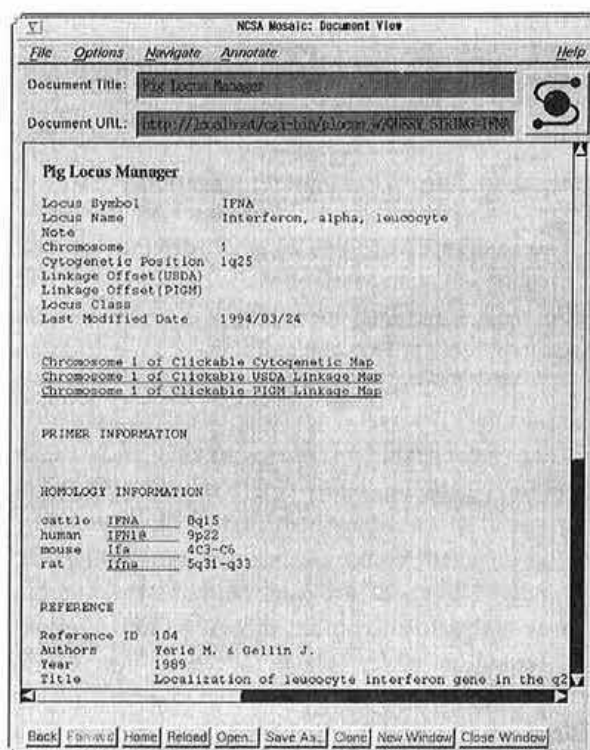


Fig. 3. Detailed information on the locus "IFNA" When the locus "IFNA" shown in Fig. 2 is clicked, detailed information on the locus shown in this figure is displayed.

tion about the entity by clicking the anchor line.

### Online linkage analysis system

The linkage database contains pedigree data, allele data and traits data for the linkage analysis. These data have been separated into each mapping project and made accessible to persons involved in the project. The system recognizes the persons with the user ID and the password. Three user classes are set for the linkage database. The class one users can input the data, but not update nor delete them. The class two or class three users can input, update and delete the data. Furthermore, the class three users can register and delete users in the project.

The users can input, update and delete these data through WWW clients. The input data are stored in the part of user's project. The users can not access the other part of the database.

The linkage analysis requires a great deal of computational power. However, since a super-computer is expensive, we constructed an analysis system comprising several servers to perform parallel

computing. We use CRIMAP-PVM<sup>6)</sup> in our online linkage analysis system for constructing genetic linkage map. Lod score analysis is executed using CRIMAP<sup>3)</sup> without parallel computing.

### Gene mapping workbench: gRanch<sup>10,12)</sup>

The interfaces of WWW are restricted for complicated search such as cytogenetic homology between species. The management of the genetic linkage data is not connected with the animal genome database in our WWW search system. We therefore developed a new client-server system designed as gRanch in order to determine the genetic linkage and to query the animal genome database effectively. The client program of gRanch is written in Tcl/Tk language<sup>7,8)</sup>. The server-client communication is based on HTTP protocol. The client program contains the mapping data available locally and receives the new mapping data from the server, when the client accesses the server. The program receives the old pedigree or allele data and puts the new pedigree or allele data, if necessary.

As for the animal genome database, the gRanch displays the gene map integrating the cytogenetic and the linkage maps, 2 kinds of cytogenetic homology maps among species, references and the information of PCR primer. The cytogenetic mapping data of pig, cattle, mouse and human are included in the gRanch. The USDA linkage map for pig and cattle, and the MGD linkage map for mouse are also available.

In the linkage analysis, each time the client accesses the server, the client program obtains the pedigree and/or allele data already deposited in the server. When new data are added to the data for the client, the program feeds these data, introducing even more new data, back to the server at the end of the program. Lod scores and recombination frequencies are calculated by the CRIMAP<sup>3)</sup> installed in this system.

### Utilization of Animal Genome Database using gRanch

The window of the integrated gene map is divided into 4 parts (Fig. 4). In the first part, the chromosome banding image is located in the far left. The second panel from the left-hand side displays the number of linkage markers mapped along with the chromosome banding position. In the third part, the partial linkage map is located on the lower right-

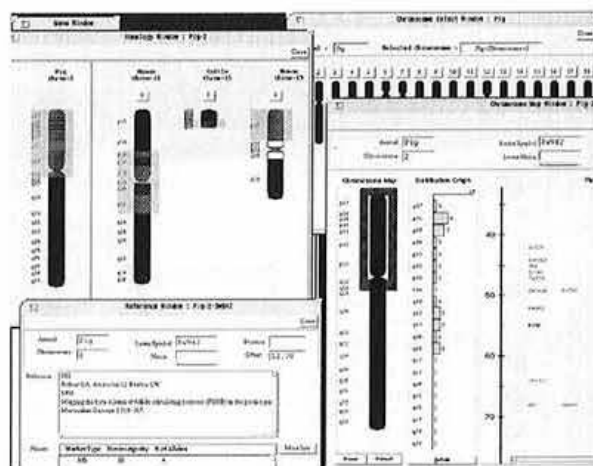


Fig. 4. Sample image of the use of gRanch

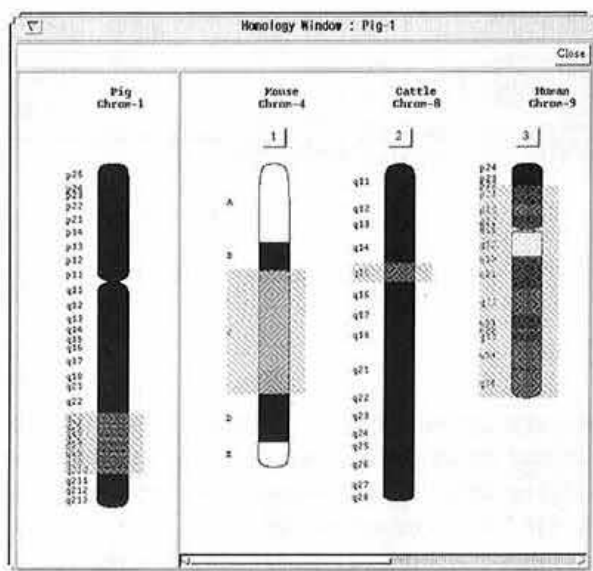


Fig. 5. Sample image of a cytogenetic homology map of gRanch

hand side. The map displays the linkage region corresponding to the chromosomal region covered with the frame shown in the chromosome map on the far left. The fourth part located in the upper area displays the locus symbol, locus name, etc. when a given locus symbol is clicked.

Two kinds of cytogenetic homology maps among species can be displayed. The chromosomal regions of other species showing a homology with the chromosomal regions in the frame of the integrated map are displayed (Fig. 5) and the chromosome regions showing a homology with those of other species are indicated by bars with the corresponding chromosome number.

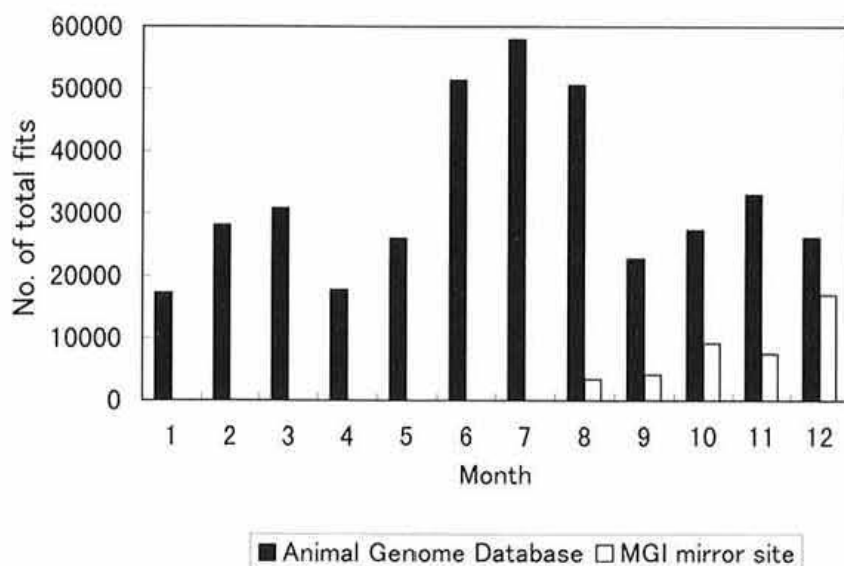


Fig. 6. Number of total fits in 1996

### Online linkage analysis using gRanch

Each time the client accesses the server, the client obtains pedigree and/or allele data already deposited in the server. When new data are added for the client, the program feeds these data, including even more new data, back to the server at the end of the program. Lod scores and recombination frequencies are calculated by the CRI-MAP 2.4<sup>3)</sup>.

### Mouse Genome Informatics mirror site in Japan

Mouse Genome Informatics (MGI) (<http://www.informatics.jax.org>) is a World Wide Web service for mouse gene mapping. It has been developed by the Jackson Laboratory and includes Mouse Genome Database (MGD) and Gene Expression Database (GXD). Users can easily search the genomic data in MGD and GXD.

Since the Pacific communication line to USA is very busy during working hours, we developed a public mirror site of Mouse Genome Informatics (<http://mgd.niai.affrc.go.jp/>). It has been opened since August 23, 1996 and all the data for mirror site have been updated daily. MGD and GXD users in Asia are invited to connect with the Japan mirror site to access the databases.

### Present situation and future prospects

The prototype of the Animal Genome Database

was first constructed in June 1994 and the public system has been experimentally open to Internet users since October 1994. About 500 clients access the animal genome database each month, and over 10,000 WWW events have been sent to the users on a monthly basis. The numbers of total fits of the Animal Genome Database and MGI mirror site in Japan are shown in Fig. 6.

The aim of our system is to assist in the construction of gene maps of domestic animals. A more useful system for comparative mapping will be available using Java Applet, and it will be possible to display the DNA sequence of a gene based on the detailed locus information page of the Animal Genome Database.

### References

- 1) Berners-Lee, T. J. et al. (1992): The world-wide web. *Comput. Networks ISDN Syst.*, **25**, 454-459.
- 2) Fasman, K. H. et al. (1994): The GDB<sup>TM</sup> human genome data base anno 1994. *Nucleic Acids Res.*, **22**, 3462-3469.
- 3) Green, P. et al. (1990): Documentation for CRI-MAP, version 2.4 (3/26/90).
- 4) The Jackson Laboratory (1992): GBASE; The genomic database of the mouse. User manual, version 2.0.
- 5) Keele, J. W. et al. (1994): A conceptual database model for genomic research. *J. Comp. Biol.*, **1**, 65-76.
- 6) Matise, T. C. et al. (1995): Parallele computation of genetic likelihoods using CRI-MAP, PVM, and a network of distributed workstations. *Hum. Hered.*, **45**, 103-116.

- 7) Ousterhout, J. K. (1990): TCL: An embeddable command language. *In Proc. 1990 winter USENIX conference*, 133-146.
- 8) Ousterhout, J. K. (1991): An X11 toolkit based on the TCL language. *In Proc. 1991 winter USENIX conference*, 105-115.
- 9) Wada, Y. et al. (1994): Animal genome database for network users. *In Proc. genome informatics workshop 1994*, 214-215.
- 10) Wada, Y. et al. (1995): gRanch: Gene mapping workbench. *In Proc. genome informatics workshop 1995*, 148-149.
- 11) Wada, Y. & Yasue, H. (1996): Development of animal genome database and its search system. *Comput. Applic. Biosci.*, **12**(3), 231-235.
- 12) Wada, Y. et al. (1997): Software tool for gene mapping: gRanch. *Comput. Applic. Biosci.* **13**(3), 323-324.

(Received for publication, March 4, 1997)