**GENET, A DATABASE OF GENETIC NETWORKS.**

SAMSONOVA M.G.+, SAVOSTYANOVA E.G., SEROV V.N., SPIROV A.V., REINITZ J.#

Institute for High performance Computing and Databases, 118 Fontanka emb., St.Petersburg, 198005, Russia; samson@fn.csa.rul;

#Brookdale Center for Molecular Biology, Box 1126, Mt.Sinai School of Medicine, One Gustave L. Levi Place, New York, NY 10029-6574, USA; reinitz@ekruppel.molbio.mssm.edu

+Corresponding author

Keywords: database, genetic networks, sea urchin, Drosophila, vertebrates, java applets, expression data

We designed the genetic network database GeNet. GeNet is a hypertext database. The concept of genetic networks forms a basis for information structuring in this database - each of the thus far characterized genes is considered as a node of a genetic network, while the links between nodes represent the interactions of genes or their products. The are two parts in GeNet. The EmbryoNet part holds information on genetic networks in sea urchin, *Drosophila* and vertebrates and contains 6 types of data: genetic network maps, gene entries, gene sequence entries, regulatory region entries, bibliographies and regulatory interactions. The NetModel part of GeNet holds the models of genetic networks. Three entry points allow the user to browse the database, to search the database and to work with genetic network maps. GeNet is available on-line at <http://www.csa.ru/Inst/gorb_dep/inbios/genet/genet.htm> and its mirror site at http://www.mssm.edu /molbio/genet.htm.

**1. Introduction**

Recently there has been a growing awareness in the genomics community of the importance of studies that go beyond straightforward questions about nucleotide and amino acid sequence and address larger problems of biological function.

Genetic networks provide a useful framework for considering these questions. Progress in molecular biology has led to the realization that in order to reveal the mechanisms of cell function it is necessary to consider ensembles of interacting genes [1,2]. The central role in these ensembles, known as genetic networks [3,4,5], is played by genes encoding transcription factors, which activate or repress other genes in the network. The products of these genes in turn act on other target genes, which ultimately act on structural genes. Thus the genetic network can be considered as a complex web of genes turning each other on and off [6,7].

It is evident that understanding the organization and behavior of genetic networks will be a massive task, the accomplishment of which will require the design of appropriate databases, as well as development of new mathematical theories and algorithms. The structure of such databases should be oriented towards the representation of results of the analysis of the genetic network's organization, function, and evolution. Thus the genetic network databases should contain maps of the functional connections within the network, the experimental results from which these connections were inferred, as well as numerical and graphical data on expression patterns of network genes at different times during development or cell cycle. The presence of time series information in the database implies that it should also contain dynamical models of this data. Finally, all of this information should be presented through an easy to use interface that will allow the user to visualize the organization of the genetic network and its spatio-temporal behavior.

In this paper we describe the construction of a prototype of such a database, called GeNet. GeNet satisfies virtually all of the criteria listed above [see also 8] for a selected set of critical developmental networks. In this paper we discuss the implementation of GeNet, its organization, and its user interface. GeNet is located at [www.csa.ru/Inst/gorb\_dep/inbios/genet/](http://www.csa.ru/Inst/gorb_dep/inbios/genet/) genet.htm and its mirror site [www.mssm.edu/molbio/genet/genet.htm](http://www.mssm.edu/molbio/genet/genet.htm).

**2. Implementation**

**2.1 System and methods**

GeNet is a hypertext database designed for the World Wide Web. The database consists of text files in HTML, image files in the GIF or JPEG formats, a suite of programs for searching and generating reports, as well as user interface implemented in JAVA and as a server-side CGI-script. The database is currently running under CONVEX OS and Solaris.

**2.2 Java applets.**

There are two types of applets in the database. The graphical interface to GeNet is implemented as a set of applets, developed on the basis of Genes\_Graph applet [9]. The other applet NetWork enables the user to work with genetic network maps and to simulate the genetic nerwork dynamics in frame of Boolean network model. For presentation of genes and gene interactions, we extended the preexisting Nodes and Edges classes. The methods for simulation of genetic network dynamics are included in the java.applet.Applet class extension. The source code is available upon request to the authors.

**2.3 The Search engine**

The search engine, called Finder, treats GeNet as a array of key words allocated to database sections, definite types of entries and definite fields. The search engine can process simultaneously up to 30 key words with consideration for intersection and negation operations applied to matched documents. The search engine consists of three files describing the database structure and two programs IndexCreate and FindsLinks. IndexCreate performs the indexation of the words in the database with consideration to their occurence in the database structural units: sections, entries and fields. FindsLinks runs the search of matching documents, executes operations of joining and intersection for key word apiece and generates a report as html file. The programs are available by request from the authors.

**3. Overall organization**

The concept of genetic networks is the basis for information structuring in the GeNet database. In this database each of the thus far characterized genes is considered as a node of the genetic network, while the links between nodes represent interactions of genes or their products.

GeNet is subdivided in 2 parts, which hold information on genetic networks controlling development (EmbryoNet) and models of genetic networks (NetModel). The EmbryoNet part consists of 3 sections, which hold information on developmental networks in sea urchin, *Drosophila* and vertebrates. Each section contains 6 types of data: genetic networks maps, gene entries, gene sequence entries, regulatory regions, gene interactions and bibliography.

**3.1 Network maps**

The maps of genetic networks are shown as flow diagrams as well as Java applets. Both methods of presentation depict genes as rectangles and their interactions as arrows. However in the case of Java applets these arrows differ in shape and colour depending on interaction type. Red arrows connect a gene with its regulators, while blue arrows connect to regulatory targets. Filled and hollow arrows reflect the mode of gene action - activation and repression correspondingly.

**3.2 Gene entries**

Each gene entry contains information on the gene's function, subcellular localization, encoded protein, expression pattern, regulatory interactions (upstream and downstream genes), as well as links to other databases.

**3.3 Expression data**

In the EmbryoNet subdivision SegNet the expression pattern field incorporates the images of expression pattern of segmentation genes in the fruit fly *D.melanogaster* (eg., [http://www.csa.ru/ Inst/gorb\_dep/inbios/genet/krueppel.html](http://www.csa.ru/%20Inst/gorb_dep/inbios/genet/krueppel.html)). These images are obtained in experiments by the group of one of us (J.R.) [7,10].

**3.4 Regulatory and interaction entries**

The regulatory element entry in GeNet contains data on the organism source, a bibliography, the element's sequence, the coordinates of sites for transcription factors binding, as well as key words and definitions. We compile the gene interactions entries for each gene a piece. In this entry for each upstream and downstream gene of the given gene we present the information on mechanism of genes interactions, methods, by which these interaction are proven and references.

**3.4 NetModel**

The NetModel part of GeNet contains the description of different models used to study the genetic network structure and function. At present this part of GeNet, which is under development, contains description of models based on reaction-diffusion approach, Boolean network models and kinetic models. The option is provided which enables the user to model the dynamics of user-defined network in frame of Boolean network model.

**4. User interface**

Three entry points are provided allowing the user to browse the database, to search the database, and to work with genetic network maps. While browsing the EmbryoNet the user sequentially moves from the page containing the list of database sections to the maps of the genetic networks which control development in the organism under consideration. Each gene of a genetic network map is linked to its gene entry, which in turn holds hypertext links to data on gene sequence, regulatory regions, gene interactions, and references in the literature. Thus by clicking on a gene name in a genetic network map the user gets detailed information about the gene and the mechanisms of its regulation.

In addition to browsing the database, the interface is available for composing arbitrary searches. The user first selects the database part, section, entry type and entry field. After specifying the key words and submitting a query the user retrieves the html file hyperlinked to entries in GeNet matching the search criteria.

Another entry point into GeNet enables the user to work with genetic network maps. The user is presented with the option to work with genetic network of interest or with predefined genetic networks of the database. The following operations can be performed on the genetic network maps:

* the interactive construction of genetic network of interest;
* the reposition of a gene to a new place as well as the display of the selected gene interactions with regulators and targets in one window. This is sometimes required for better visualization of links between genes in cases where the network is large;
* the selection of gene by the mouse click. As the result the regulatory interactions with regulators and targets of the selected gene become highlighted.
* the selection of subnetwork of interest by sequential clicking on genes of interest while pressing the SHIFT key followed by pressing the PATH ONLY button. As the result only the selected genes remain on the screen. These procedures permits the visualization of the interactions of genes related by function (i.e. acting at definite time intervals or controlling the development of particular embryo parts) or structure.
* the interactive editing of the genetic network by deletion or addition of nodes or links. This procedure enables the user to model the effect of mutations in the network.
* the simulation of the genetic network dynamics in frame of different theoretical models.

The current GeNet version holds approximately 1200 files in \*.html and \*.gif format. It contains the information on 300 genes and 80 regulatory elements. GeNet holds 10 genetic network maps and 30 images of genes expression patterns.

**4. Discussion**

The large-scale projects currently in progress to sequence the genomes of humans and several model organisms have large and rapidly growing amounts of biological information [1,11,12]. A great many databases are designed for storage, processing and retrieval of molecular biology data.

At present as the emphasis of current biomedical research shifts from the identification of genes to the characterization of their function the design of databases containing functional information becomes crucial. GeNet contains functional information on mechanisms of gene action in embryogenesis. However the distinctive feature of GeNet is a model for information presentation, which is based on the concept of genetic networks. Such a database structure enables end users to retrieve information on the functional organization and evolutionary conservation of the whole ensemble of interacting genes.

At present GeNet contains about 30 digital images of expression patterns of genes controlling segmentation in Drosophila. These images are particularly useful, since they make possible the evaluation of gene expression at the level of discrete nuclei. We are working now on incorporating numerical data on segmentation genes expression [10] into GeNet. Our aim is to construct a quantitative atlas of D.melanogaster segmentation gene expression, containing image data, 3D reconstructions, and numerical data.

We believe that GeNet is the first database to contain the genetic network maps. Our presentation of genetic network maps as Java applets enables the user to visualize the genetic network, to simulate its dynamics and to model the effect of mutations in the network. The information in GeNet can be used to derive information about the function of genetic networks and digital images of Drosophila segmentation genes expression patterns.

In future we plan to broaden the content of GeNet by adding the information on genetic networks controlling later embryonic stages, as well as well as genetic networks controlling response of eukaryotic cells to stress. Besides that we will further incorporate in GeNet graphical and numerical information on expression patterns of genes controlling morphogenesis in Drosophila. The new version of GeNet will allow the Internet end users to access and process this information on-line.

**References**

1. Lander, «The new genomics: Global view of biology» Science **274**, 536 (1996)
2. Hunter, «Oncoprotein networks» Cell **88**, 333 (1997)
3. Kauffman, «Gene regulation networks: a theory for their global structure and behavior» Current Topics in Dev.Biol. **6**, 145 (1971)
4. Mjolsness, D.H. Sharp and J. Reinit, « A connectionist model of development» J. Theor. Biol **152** , 429 (1991)
5. Somogyi and A. Sniegorski, «Modeling the complexity of genetic networks: understanding multigenic and pleiotropic regulation» Complexity **1**, 45 (1996)
6. Goodwin and S.A. Kauffman, «Spatial harmonics and pattern specification in early Drosophila development. Part I. Bifurcation sequences and gene expression» J Theor Biol **144**, 303 (1990)
7. Reinitz and D.H. Sharp, «Mechanism of eve stripe formation» Mech Dev **49**, 133 (1995)
8. Spirov and M.G. Samsonova, «The GeNet database as a tool for the analysis of regulatory gene networks» Proc. of Int. Workshop on Information Processing in Cells and Tissues, Sheffield UK, 1-4 Sept., 300 (1997)
9. Serov, A.V. Spirov and M.G. Samsonova, «Development of tools for presentation of knowledge on genetic networks» Bioinformatics, in press
10. Kosman, J. Reinitz and D.H. Sharp, «Automated assay of gene expression at cellular resolution» Proc.of PSB'98, 6 (1998)
11. G.D.Schuler, M.S. Boguski, L.D. Stewart et al., «A gene map of the human genome» Science **274**, 540 (1996)
12. Nowak, «Entering in postgenome era» Science, **270**, 368 (1995)