GeneNet: a database on structure and functional organisation of gene networks

E. A. Ananko*, N. L. Podkolodny, I. L. Stepanenko, E. V. Ignatieva, O. A. Podkolodnaya and N. A. Kolchanov

Institute of Cytology and Genetics (Siberian Branch of the Russian Academy of Sciences), 10 Lavrentiev Avenue, Novosibirsk, 630090, Russia

ABSTRACT

The GeneNet database is designed for accumulation of information on gene networks. Original technology applied in GeneNet enables description of not only a gene network structure and functional relationships between components, but also metabolic and signal transduction pathways. Specialised software, GeneNet Viewer, automatically displays the graphical diagram of gene networks described in the database. Current release 3.0 of GeneNet database contains descriptions of 25 gene networks, 945 proteins, 567 genes, 151 other substances and 1364 relationships between components of gene networks. Information distributed between 14 interlinked tables was obtained by annotating 968 scientific publications. The SRS-version of GeneNet database is freely available (http://wwwmgs.bionet.nsc.ru/mgs/systems/genenet/).

INTRODUCTION

The GeneNet database is designed for accumulation of information about the structure of gene networks and functional relationships between their components (1). As a gene network, we understand an ensemble of genes functioning in a coordinated manner to control vital functions, fine regulation of physiological processes or responses to external stimuli (2). Currently, the following classes of gene network elementary structures are arranged in the GeneNet database: *Genes*, *RNAs*, *Proteins* and other *Substances* (for example, steroid hormones, metabolites, lipids, small regulatory molecules, etc.). If necessary, there is a possibility of adding novel classes of objects. Elementary relationships in gene networks include *Reactions* (interactions between the entities that lead to appearance of new entities) and *Regulatory Events*. Regulatory events of four types are distinguished depending on the effect they produce on reaction: (i) switch on, (ii) switch off a process, (iii) positive effect, or (iv) negative effect of a regulator in case the process proceeds without this regulator (1). GeneNet format enables the user to take into account distribution of gene network components between different organs, tissues, cells and cell compartments. Applying GeneNet, one may represent the gene networks at different levels, from a particular cell or cell

compartment to a whole organism. Moreover, within the frames of technology suggested, it is possible to deal with symbiont gene networks, that is, co-ordinately functioning networks referring to different organisms (2). For a formalised description of gene networks, a specialised language was developed (1). This language is suitable not only for description of a gene network, but also for signal transduction pathways and metabolic pathways.

The GeneNet database is being developed at the Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences, and has been since 1998 (1). GeneNet release 3.0 is implemented in SRS and is available at http://wwwmgs.bionet.nsc.ru/mgs/systems/genenet/.

DATABASE STRUCTURE AND FORMAT

The information stored in the GeneNet database is distributed between 14 interlinked tables. Descriptions of the format and number of entries of each GeneNet table are represented in Table 1.

GRAPHIC USER INTERFACE

The Java applet GeneNet Viewer provides a visualisation of the data in a graphical form (Fig. 1). GeneNet Viewer is activated by clicking specialised links from the entries in GN_SCHEME. GeneNet Viewer works properly only with operation systems Microsoft Windows 95, 98, 2000 and Windows NT by using Internet Explorer 4.2 and higher, or Netscape Communicator 4.7 and higher. The current version of GeneNet Viewer is unusable under UNIX operation system.

The gene network diagram consists of interactive components. Each gene network component is visualised by its own image that reflects some features of a component (1). For example, the shape of the image depicting a protein gives information about the multimerisation state of a protein, whereas functional state of a protein (active or inactive) is marked by colour (pink or green, respectively), etc. By clicking an image, you may retrieve the description of an entry from the GeneNet database in a special text box (Fig. 1).

The data obtained in different species are summarised in the diagram. As a result, the diagram may contain several equivalent objects displayed by a single node (for example, homologous genes of different species). The system of filters enables the user to select visualisation of only those entities and relations

---

*To whom correspondence should be addressed. Tel: +7 3832 333119; Fax: +7 3832 331278; Email: eananko@bionet.nsc.ru

Table 1. Tables of the GeneNet database

Table name (and description)	Information fields	Number of entries
GN_GENE (genes)	ID, identifier; IC, species abbreviation: gene abbreviation; DT, data of the entry creating/editing; annotator; created/updated; OS, species; SN, abbreviated gene name; NM, complete gene name; SY, gene synonymous name; SO, cells, tissues, organs; CH, chromosome; RE, inducers, repressors; PN, ID of the encoded protein in the GeneNet database; DR, links to other databases; RF, references to publications; CC, comments.	567
GN_RNA (RNAs)	ID, identifier; IC, species abbreviation: RNA abbreviation; DT, data of the entry creating/editing; annotator; created/updated; OS, species; SN, abbreviated RNA name; NM, RNA full name; SY, gene synonymous name; DR, links to other databases; RE, inducers, repressors; SO, cells, tissues, organs; TP, RNA type; GN, link to the GN_GENE SRS table; RF, reference to publications; CC, comments.	82
GN_PROTEIN (proteins)	ID, identifier; IC, species abbreviation: protein abbreviation; DT, data of the entry creating/editing; annotator; created/updated; OS, species; SN, abbreviated protein name; NM, complete protein name; SY, protein synonymous name; DR, links to other databases; RE, inducers, repressors; SO, cells, tissues, organs; FN, functional state; MM, multimerisation level; MD, phosphorylated/dephosphorylated state; GN, ID of the gene encoding this protein in the GeneNet database; RF, references to publications; CC, comments.	945
GN_SUBSTANCE (other substances)	ID, identifier; IC, substance abbreviation; DT, data of the entry creating/editing; annotator; created/updated; SN, abbreviated protein name; NM, complete protein name; RF, references to publications; CC, comments.	151
GN_RELATION (relationships between entities)	ID, identifier; DT, data of the entry creating/editing; annotator; created/updated; RE, relation code; IN, input entity; TY, relation class; OU, output entity; CO, controlled relation; EF, direct/indirect; AT, type of regulatory event; RF, references to publications; CC, comments.	1364
GN_SCHEME (descriptions of gene networks)	ID, identifier; NM, gene network name (GeneNet Viewer link); DT, data of the entry creating/editing; annotator; created/updated; TP, representation level (cell/organism); MD, GeneNet Dynamic Model link; DE, gene network description; OS, species; EN, link to the list of entities included in the gene network; RE, link to the list of regulatory events in the gene network; RA, link to the list of reactions in the gene network; DR, links to other databases; RR, references to publications; CC, comments.	25
GN_SCHEME_ENTITY [entities (elementary structures)]	ID, identifier; SC, GN_SCHEME identifier; NM, gene network name; ET, entity type; OS, species; EN, entity name; SU, compartment localisation.	1532
GN_SCHEME_RELATION (relationships in gene networks)	ID, identifier; SC, GN_SCHEME identifier; NM, gene network name; RT, relation type; IN, input; OU, output entity; CO, output relation.	1523
GN_COMPARTMENT (compartments)	ID, identifier; IC, compartment code; DT, data of the entry creating/editing; annotator; created/updated; CL, colour, dimension (X), dimension (Y), NM, compartment name; CC, comments.	83
GN_ORGANISM (species)	ID, identifier; IC, species Latin name; AC, species English name; DT, data of the entry creating/editing; SN, species abbreviation; OC, classification.	74
GN_PROCESS (input and output processes)	ID, identifier; IC, process code; DT, data of the entry creating/editing; annotator; created/updated; SN, process name; NM, complete process name; CC, comments.	82
GN_CELL (cells, tissues, organs)	ID, identifier; IC, species abbreviation: item abbreviation; DT, data of the entry creating/editing; annotator; created/updated; OS, species; SN, abbreviated name; NM, complete name; RF, references to publications; CC, comments.	300
GN_BIBLIOGRAPHY (references to the papers annotated)	ID, identifier; IC, paper code; DT, data of the entry creating/editing; annotator; created/updated; AU, authors; TI, title of the paper; SO, journal; VL, volume; IS, issue; YR; year; PG, pages; ML, MEDLINE UI.	968
GN_EXPERT (GeneNet annotators)	ID, identifier; IC, annotator code; NM, complete annotator's name; LB, laboratory; OR, organisation; CT, city; CN, country; EM, email.	21

that were experimentally identified for the organism specified by a user.

GeneNet Viewer is supplemented by the tools for diagram zooming, data navigation, online help, interactive cross-references within the GeneNet database and references to other databases [TRRD (3), EMBL (4), SWISS-PROT (5), TRANSFAC (6), MEDLINE], which are displayed in the browser window (Fig. 1)

DATABASE CONTENT

Informational content of the GeneNet database is shown in Table 2. By August 1, 2001, the GeneNet database stored descriptions of 25 gene networks classified between six thematic sections (Table 2), 567 genes, 945 proteins, 1364 relationships between entities. This information is obtained by annotating 968 scientific publications.

GeneNet USAGE

The GeneNet database contains useful information for studying molecular processes, pharmaceuticals designing, prediction of drug by-effect, etc. GeneNet Viewer enables the user to produce a generalised view of a gene network, whereas a special tool, GeneNet Modeling (http://wwwmgs.bionet.nsc.ru/mgs/gnw/gn_model/), is designed for studying the impact of a hypothetical mutation on a gene network functioning.
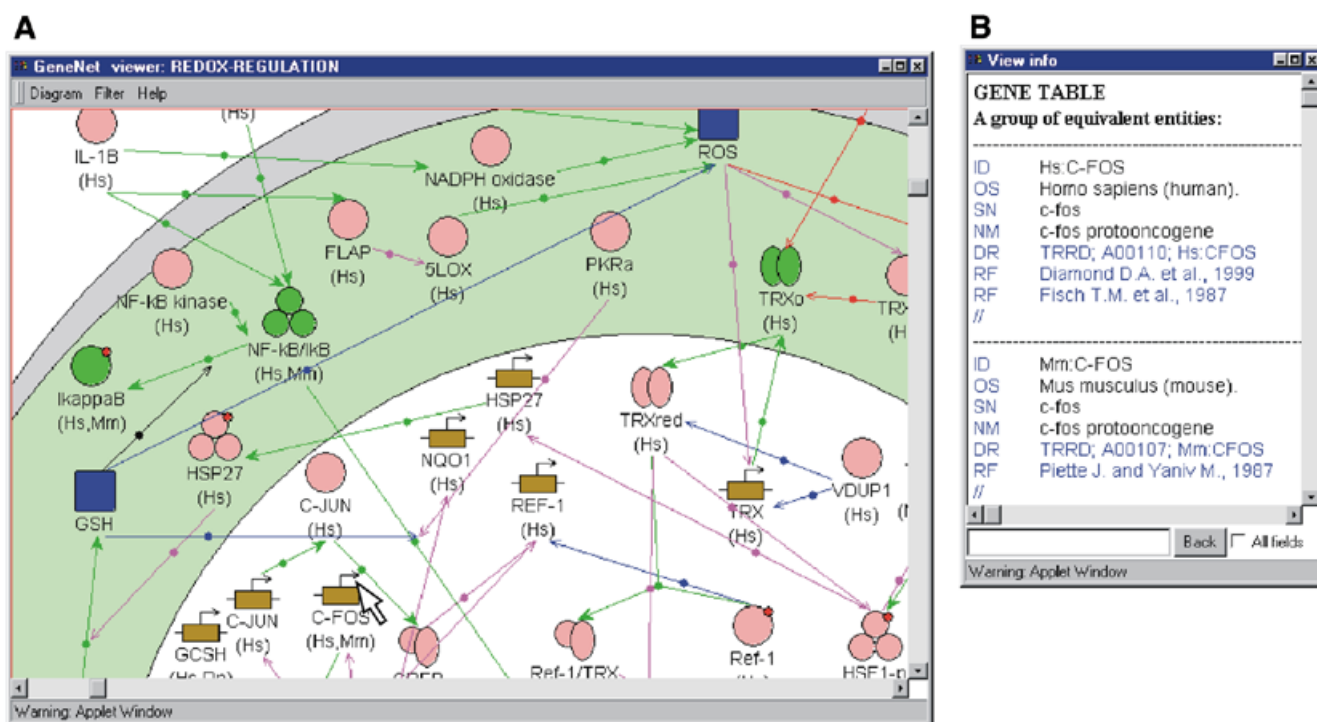
Figure 1. GeneNet Viewer. (**A**) A fragment of the graphical diagram 'REDOX-REGULATION' obtained by means of GeneNet Viewer. Designations: brown rectangles, genes; pink and green circles, protein molecules; blue squares, other substances; arrows, relationships between entities. (**B**) Description of the group of homological c-*fos* genes (human and murine), illustrated at the diagram as a single object. This description is displayed in a special text window by clicking the relevant gene image.

For description of processes in the table GN_RELATION, the following information is formalised: input (IN line code) and output components (for entities, OU line code, and for reactions, CO), relation class (discriminating reactions and regulatory events, line code TY), type of the process (direct or indirect, EF); type of regulatory event (switch on, switch off, increase or decrease, line code AT), references to publications (RF). The formalised data about location of a component in a definite compartment are contained in the field RE (GN_RELATION), as well as in the field SU of the table GN_SCHEME_ENTITY (Table 1).

The GeneNet database is helpful for performing the following tasks:

1. To extract the list of entities that are involved in functioning of a particular gene network and select the items by species, compartment, type of an entity.
2. To extract the list of all reactions and regulatory relations for a particular gene network.
3. To browse information about all relationships that involve the protein of interest.
4. To extract the list of genes, transcription of which is induced by a particular transcription factor.
5. To view reactions that involve a protein, as well as the role of this protein in these reactions, etc.

FUTURE PROSPECTS

In future, we plan to develop special approaches for operating with the complex hierarchical gene networks aimed at data mining, knowledge discovery and developing of the methods of gene network computer modeling on the basis of information stored in the GeneNet database.

In addition, we plan to develop the relational version of the GeneNet database. The GeneNet database format for representation of complex hierarchical gene networks with descriptions worked out in various details will be modified. The format of description of relationships between network components will be also modified in order to describe the processes in more detail, in particular, to accumulate quantitative and qualitative data about gene network dynamics, etc. The format of the database will be adopted for more complete integration of the GeneNet database with the TRRD database (3). This will enable the user to use structural and functional characteristics of transcription regulation described in TRRD for visualisation of gene network diagrams and modeling of their dynamics.

Extension of the GeneNet database will be mainly produced by describing gene networks controlling vitally important processes in a normal state and under genetic disorders.

AVAILABILITY

SRS-version of GeneNet 3.0 is freely available at http:// wwwmgs.bionet.nsc.ru/mgs/systems/genenet/. To order the licensed GeneNet version including GeneNet viewer, SRS and XML versions of databases, please email the supervisor of GeneNet, Prof. Nikolay A. Kolchanov (kol@bionet.nsc.ru). All rights reserved. We kindly ask that this and a previously

Table 2. Informational content of the GeneNet database (August 1, 2001)

GeneNet section	Entry name in GN_SCHEME	Number of components		
		Genes	Proteins	Relationships
Lipid metabolism	Cholesterol	6	11	36
	Cholesterol_MODEL	5	18	43
	Leptin (organism level)	43	19	89
Endocrine regulation	Principal cell of CCD	3	15	34
	Steroidogenesis (adrenal cortex)	15	39	80
	Steroidogenesis (sex steroids)	12	41	78
	Thyroid system	23	66	110
Erythrocyte maturation	Erythroid differentiation	41	51	98
Immune system	Antiviral response	12	51	53
	Macrophage activation (model)	37	70	124
Plant gene networks	Germination (endosperm)	5	21	25
	LEA program	13	32	27
	Plant-pathogen	31	34	65
	Seed reserve mobilization (i) carbohydrates	7	7	34
	Seed reserve mobilization (ii) lipids and phosphates	5	8	31
	Seed reserve mobilization (iii) proteins	5	11	42
	Seed reserve mobilization (iv) regulatory relationships	12	22	62
	Seed reserve mobilization (v) general diagram	11	27	59
	Seed reserve mobilization (organism level)	7	23	46
	Storage protein biosynthesis (dicot)	8	21	31
	Storage protein biosynthesis (monocot)	14	35	33
Heat shock response	HSP70 autoregulation	6	19	37
	Heat shock response	36	41	112
	Thermotolerance	4	40	64
Redox-regulation	REDOX-regulation	48	43	111

published article (1) be cited when reporting results based on GeneNet usage.

REFERENCES

1. Kolpakov,F.A., Ananko,E.A., Kolesov,G.B. and Kolchanov,N.A. (1998) GeneNet: a database for gene networks and its automated visualization. *Bioinformatics*, **14**, 529–537.
2. Kolchanov,N.A., Anan'ko,E.A., Kolpakov,F.A., Podkolodnaya,O.A., Ignat'eva,E.V., Goryachkovskaya,T.N. and Stepanenko,I.L. (2000) Gene Networks. *Mol. Biol. (Mosk.)*, **34**, 449–460.
3. Kolchanov,N.A., Podkolodnaya,O.A., Ananko,E.A., Ignatieva,E.V., Stepanenko,I.L., Kel-Margoulis,O.V., Kel,A.E., Merkulova,T.I., Goryachkovskaya,T.N. and Busygina,T.V. *et al.* (2000) Transcription Regulatory Regions Database (TRRD): its status in 2000. *Nucleic Acids Res.*, **28**, 298–301. Updated article in this issue: *Nucleic Acids Res.* (2002), **30**, 312–317.
4. Stoesser,G., Baker,W., van den Broek,A., Camon,E., Garcia-Pastor,M., Kanz,C., Kulikova,T., Lombard,V., Lopez,R., Parkinson,H. *et al.* (2001) The EMBL nucleotide sequence database. *Nucleic Acids Res.*, **29**, 17–21. Updated article in this issue: *Nucleic Acids Res.* (2002), **30**, 21–26.
5. Bairoch,A. and Apweiler,R. (2000) The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res.*, **28**, 45–48.
6. Wingender,E., Chen,X., Fricke,E., Geffers,R., Hehl,R., Liebich,I., Krull,M., Matys,V., Michael,H., Ohnhauser,R. *et al.* (2001) The TRANSFAC system on gene expression regulation. *Nucleic Acids Res.*, **29**, 281–283.