

GeNetDB: a bioinformatic platform for the study of Genetic Regulatory Networks

> www.genetdb.net



A. Mazurie, J. Lambert, V. Bernard, J. Mallet
 Laboratoire LGN, CNRS UMR 7091, Hôpital Pitié-Salpêtrière, Paris
aurelien.mazurie@free.fr

CONTEXT

> **Genetic regulatory networks (GRN)** are composed by all of the regulatory links among genes. The study of these networks must permit the identification of **regulation anomalies** in pathologies having a genetic component, as well as the search for **structures** having a biological function. More generally this study aims to discover a **gene level organization** which could explain the main cells' behavior: growth, differentiation, environment adaptation.

For a long time such studies have remained in the field of theoretical biology, because of the **impossibility to observe on a large scale the in vivo behavior** of the network components, namely genes and their transcripts. Thanks to technologies such as DNA microarrays (for expression data) or CHIP-chips (for genetic networks), such information is now at our disposal.

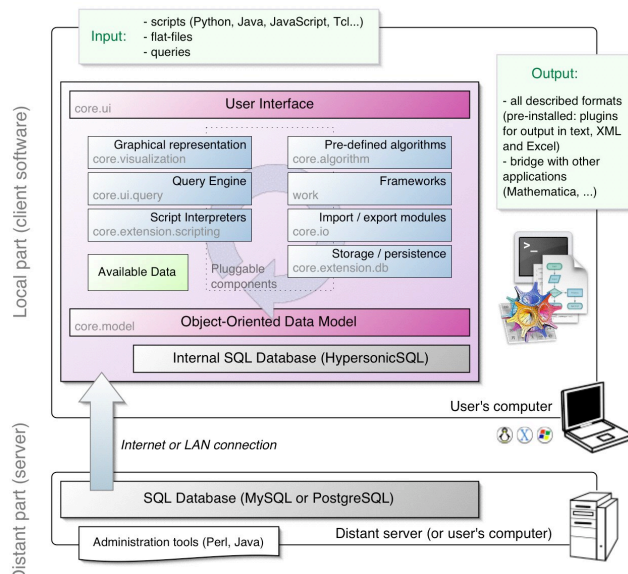
But the other main difficulty is to **piece together genetic networks** without extensive reading of the literature: they remain **scattered in specialized databases**, are not very accessible to the non specialist, which moreover are often little or non-formatted, and only hardly available for **consultation by automated querying tools**, making them only accessible "by hand".

> The primary goal of GeNetDB is to **simplify** these studies by proposing a **unique and homogeneous source** for these data, in an annotated and standardized form, and to permit **easy application** of any **processing, research and visualization algorithm** that can be developed by its users. A secondary goal is to provide **complementary means** for molecular biologists to study genes of interest in a more **global** fashion, by responding to this question: **what do we know about my gene neighbors, and how can they influence its behavior?**

1. The tool

GeNetDB is a software platform organized into a **client / server** architecture: the **client** is a lightweight software, written in **Java**, installed on the user's computer. It has access to the **server** which contains the data themselves (see *Features*). This can be either the one hosted in our laboratory, or a **stand-alone version** installed by the user (for speed and confidentiality purpose).

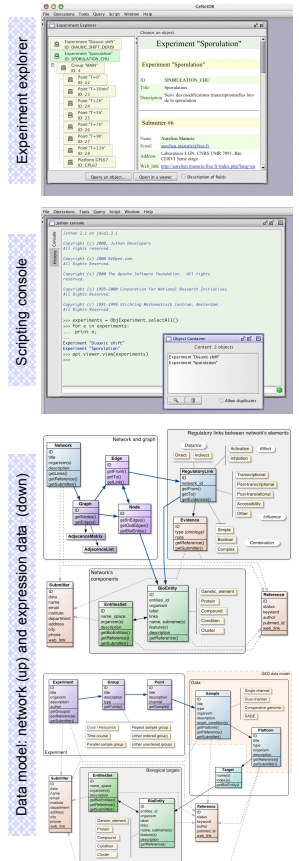
The client software provides all **representation, querying and processing** algorithms needed for works on genetic networks, as well as a programming environment (by means of API and scripting frameworks) for **automation and refinement** of these works. Finally, most of the client components are **pluggable**, and the client itself, available under an open-source license, can be **modified** at will.



> **Schematic view of the GeNetDB platform architecture**
 The client / server distinction is shown here. The **server** contains all data needed for works (genetic networks, expression data, annotations, etc). The **client** is a lightweight Java software (running on any platform) providing representation, querying and processing algorithms.

2. Features

- > **First available database of genetic regulatory networks**
- > **Availability of main data classes enabling their study**
 - networks' component **behavior** (gene expression data)
 - **factual data** about these components (annotations, etc)
- > **Interface adapted to biologists as well as computer scientists**
 Data can be manipulated by means of
 - a **graphical interface** (visualization and querying tools, transfer of data by drag-and-drop from one task to another)
 - **console and script files** allowing the refinement and automation of works (several languages available)
- > **Object-oriented data model (organism-independent)**
 All data are represented by **objects**. These include among others
 - **networks of genetic interactions**
 - four descriptive aspects for genetic links (distance, effect, influence, combination)
 - concept of **interaction complexes** (boolean combinations)
 - linked to an **ontology of evidence** of these interactions
 - **expression data** (DNA microarrays, SAGE, CGH...)
 - **gene annotations**
 - **ontologies** (*Gene Ontology, MultiFun...*)
- > **Advanced querying functions**
 - classical research (by fields) on all objects
 - research of **genetic sub-networks**:
 - by **topological parameters**
 - by a search for a **motif** (built on screen, or pre-defined)
 - search for **expression data**:
 - by global parameters (SD, min/max values, resolution)
 - by approximation, in drawing a **desired profile** on screen
- > **Bioinformatic dedicated workspace**
 Following resources are made available:
 - **internal database**, accessible without limit for user own **processing needs** (storage, SQL queries...)
 - in addition to the data model, **study frameworks** which can serve as a basis for scripts (works on information correlation, on expression data processing, etc.)
- > **Data and results output**
 - by means of **file writing** (text, XML, Excel... All other formats can be supported by **plugin additions**)
 - by **software-to-software bridges** (*Mathematica*, soon *R* and *Matlab*)



3. Data

	S. cerevisiae	E. coli
Genetic regulatory networks	- Lee <i>et al.</i> - Guelzim <i>et al.</i> - Futcher <i>et al.</i> ... = +4000 interactions* between 3500 genes	Content of <i>RegulonDB</i> + literature browsing (Alon <i>et al.</i>) = 577 interactions* between 424 genes
Expression data	All datasets from GEO (NCBI) Soon: <i>SMD</i> (Stanford), <i>ArrayExpress</i> (EBI)	
Annotations	All datasets from SGD (Stanford)	All datasets from <i>GenProtEC</i> , <i>PEC</i> (Japan)

(Planned: *Drosophila melanogaster*, *Mus musculus*, *Rattus norvegicus*)

*: with all corresponding evidence, organized into a semantic tree

CONCLUSION

> GeNetDB is conceived to facilitate **interdisciplinary studies** of **gene networks** (modeling, inference...) and **biological studies** using them (biological process characterization, target identification...) thanks to a simple and exhaustive access to **relevant data** and **needed tools**.

For advanced users (having mastered rudiments of programming), GeNetDB offers the possibility to **create their own tools** from pre-existing skeletons and provide an **easy access** to common tools used in bioinformatics (SQL databases, script interpreters, etc).

> GeNetDB aims to become a **data source**, as well as a **working environment**, for pluridisciplinary teams working on biological networks.

> One address: www.genetdb.net