BaSyLiCA[★] : A web interface for automatic process of Live Cell Array data using R.

Leslie Aïchaoui-Denève[1*], Vincent Fromion [1], Stéphane Aymerich[2], Matthieu Jules[2], Ludovic Le-Chat[2], Anne Goelzer[1].

1. INRA- Mathématique, Informatique et Génome – Jouy-en-Josas, FRANCE
2. INRA- Microbiologie et Génétique Moléculaire – Grignon, FRANCE
* Contact author: laichaoui@inra.jouy.fr

Keywords: Web Interface, Polynomial Regression, Live Cell Array.

As new technologies in biology produces large amounts of data, the necessity of developing dedicated tools for management and processing increases.

In this study, we present our recent efforts in developing a user-friendly tool allowing biologists to handle time series data associated to Live Cell Array (LCA; Zaslaver A. and al., 2006). It consists of a triple tool combining a database for storage, a WEB interface to manage database and specific *R* functions for processing data.

The LCA is a library of fluorescent transcriptional reporters, which is cultivated in microtiter plates (96 wells, 12 x 8, named A1 to H12). It allows to monitor live cell (for example Bacillus subtilis) transcriptional changes over time (e.g. every 1 to 10 minutes for hours). Two quantities are measured over time by the robot (a multi-detection microplate reader), the optical density (which is proportional to the number of individuals) and the fluorescence (which informs on the expression of a gene thanks to the fluorescent reporter).

The first problem was to store the data automatically acquired by the robot, as well as all information related to experimental conditions (temperature, injection ...) and the description of strains used. To do so, we chose a *MySQL* database. Automatic insertion and formatting of data were performed with the *R* package **RMySQL** (David A. and al., 2009). Manual insertions of experimental conditions were made through the interface.

The goal of this experiment is to evaluate the activity of the promoter (Zaslaver A. and al., 2004). This activity is defined as the derivative of fluorescence with respect to time, divided by the optical density. We have to take into account that data are noisy, bacteria have a natural auto-fluorescence, and shifted with respect to time because bacteria do not start growth at the same time in all wells (so-called "lag-phase"). To obtain this derivative we proposed several methods of treatment of curves. One of them is to smooth the curves then adjust the offset curves and remove automatically problematic ones. Another method uses the Kalman filter to estimate the fluorescence derivative.

A *PHP* interface is available to enable biologist to upload files and to enter information in the database in a few clicks. Furthermore, different results and graphs can be produced through *R*, according to different criteria, and downloaded as CSV file (Comma-Separated Values) and PDF format. Users can also simply click and view the contents of a plate (strains, culture medium, injected products ...). The interface has an admin part that can manage access rights to the data according to their property (public or private) and the level of the user (administrator or simple user).

The tool has been applied on a data set of 80 plates in the European project BaSysBio (Bacillus Systems Biology). The aim of this project is to study the global regulation of gene transcription in a model bacterium: Bacillus subtilis.

[★]BaSyLiCA : BaSysBio Live Cell Array.

References

David A. James and Saikat DebRoy (2009). *RMySQL: R interface to the MySQL database*, http://biostat.mc.vanderbilt.edu/wiki/Main/RMySQL.

Zaslaver A, Bren A, Ronen M, Itzkovitz S, Kikoin I, Shavit S, Liebermeister W, Surette MG, Alon U. (2006), A comprehensive library of fluorescent transcriptional reporters for *Escherichia coli. Nat Methods*. 3, 623-8.

Zaslaver A, Mayo AE, Rosenberg R, Bashkin P, Sberro H, Tsalyuk M, Surette MG, Alon U. (2004). Just-in-time transcription program in metabolic pathways. *Nat Genet*. 36, 486-91.