



HAL
open science

Visualization of time-series data in the context of metabolic networks with Systrip Software

Jonathan Dubois, Ludovic Cottret, Amine Ghozlane, David Auber, Frédéric Bringaud, Patricia Thébault, Fabien Jourdan, Romain Bourqui

► **To cite this version:**

Jonathan Dubois, Ludovic Cottret, Amine Ghozlane, David Auber, Frédéric Bringaud, et al.. Visualization of time-series data in the context of metabolic networks with Systrip Software. Journées Ouvertes Biologie Informatique Mathématiques (JOBIM 2013), Jul 2013, France. hal-00838725

HAL Id: hal-00838725

<https://hal.science/hal-00838725v1>

Submitted on 26 Jun 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Visualization of time-series data in the context of metabolic networks with Systrip Software

Jonathan DUBOIS¹, Ludovic COTTRET², Amine GHOZLANE³, David AUBER¹, Frédéric BRINGAUD⁴,
Patricia THÉBAULT^{1,5}, Fabien JOURDAN⁶ and Romain BOURQUI¹

¹ LaBRI, CNRS UMR5800, Université de Bordeaux, 351, cours de la Libération F-33405, Talence cedex, France
{jonathan.dubois, david.auber, patricia.thebault, romain.bourqui}@labri.fr

² LIPM, UMR INRA-CNRS, BP 52627, 31326 Castanet Tolosan Cedex, France
ludovic.cottret@toulouse.inra.fr

³ DSIMB, INSERM UMR-S 665, Université Paris 7, INTS, 6, rue Alexandre Cabanel, 75739 Paris Cedex 15, France
amine.ghozlane@inserm.fr

⁴ RMSB, UMR 5536 CNRS, Université Bordeaux Segalen, 146, rue Léo Saignat, 33076 Bordeaux cedex, France
bringaud@rmsb.u-bordeaux2.fr

⁵ CGFB - CBiB, Université Victor Segalen Bordeaux 2, 146, rue Léo Saignat, 33076 Bordeaux, France

⁶ INRA UMR1331 Toxalim, 180, chemin de Tournefeuille, BP 3 31931 Toulouse CEDEX, France
fabien.jourdan@toulouse.inra.fr

Abstract *Technological advances in biology lead to a profusion of quantitative data, raising analytical challenges. Information visualization is particularly well suited to address these difficulties as it helps to interactively move through the different levels of analysis and to simultaneously investigate data with different point of views. In this article we present, Systrip, a visual environment for the analysis of time-series data in the context of metabolic networks. Systrip gathers bioinformatics and graph theoretical algorithms that can be assembled in different ways to help bioinformaticians/biologists in their visual mining process.*

Keywords metabolic network, time-series data, contextual visualization.

1 Introduction

These last years, a technological gap has been filled in molecular biology with the development of high sequencing technologies. While it took years to sequence the genome of a single organism (e.g. fourteen for the human genome), it can now be done within a few days (and soon a few hours). It opens a new area in terms of biological investigation since the so called *omics* technologies can generate a large amount of high throughput data for the newly sequenced organisms. Global biological studies are oftenly designed to investigate the adaptation of the organism metabolism to an environmental stress (e.g. drug treatment). Samples are collected all along the adaptation process. The output data will be, for each biological entity (e.g. metabolite), a vector of numerical values corresponding to measures made at different time points. These kinds of data are called time series data. However, most general biological network visualization software only superimposes biological measurements on nodes for one time point (e.g. *Cytoscape* [11], *Pathway Tools* [10], or *BiologicalNetworks* [2]). Other tools, such as *VANTED* [8] and *VisANT* [7], support the visualization of the whole experimental data by embedding an expression profile plot, a bar chart or line chart inside the nodes.

In this article we briefly present Systrip, a visual environment for the analysis of time-series data in the context of metabolic networks (for more details please refer to [4]). The originality of Systrip is to combine bioinformatics methods (e.g. scope detection [6], chokepoint detection [12]) with well-known theoretical graph ones (e.g. strahler, strength). Systrip also supports exploration awareness as it enables the biologists to manually create views on the data at each important filtration step of their analysis process. We demonstrated some measurable benefits when using Systrip to monitor the metabolism of *Trypanosoma brucei*, which is the causative agent of human African trypanosomiasis (sleeping sickness).

2 Software System

Systrip is developed in C++ (see <http://tulip.labri.fr/TulipDrupal/?q=systrip> for executable and source code) and is based on Tulip [1]. Tulip is an information visualization framework dedicated to the analysis and visualization of relational data. It aims at providing to the developer a complete library supporting the design of interactive information visualization applications.

2.1 Main Features

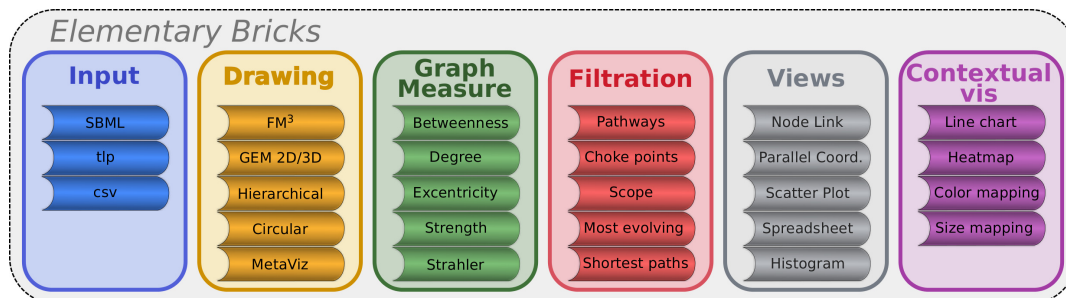


Figure 1. Systrip provides an environment gathering input facilities, graph drawing, graph measure and filtration algorithms, and supports several visualization techniques.

Primary aim of Systrip is to support visual analysis of time-series data in the context of the complete metabolic network of an organism. To achieve that goal, Systrip provides 5 classes of functionalities (that we call *bricks*): input, drawing, graph measure, filtration and views (see Fig. 1). These elementary bricks can also be classified into network related (e.g. betweenness centrality or excentricity), bioinformatics related (e.g. scope selection [6] or chokepoint selection [12]) and multi-dimensional data related (e.g. parallel coordinates or scatterplot). By combining several simple but also complex bricks in a meaningful pipeline, the user can define his own specifications and therefore build an image that answers the original biological question. Unlike tools like Cytoscape, Systrip disconnects the original data from the focussed subnetworks and related views. As the subnetworks are fully disconnected from the original network, the user can easily but also safely modify them (e.g. adding/removing elements, computing graph measures or applying metabolic-dedicated algorithms) without altering the original data. That feature is particularly suited to the top-down analysis process used in biology and bioinformatics. It also enables to perform simultaneously several analytics pipelines on the same data by applying different combinations of elementary bricks. Fig. 2 shows a screenshot of Systrip. In this figure, one can see on the left side a panel containing (top-left) the elements and pathways of the metabolic network and (bottom-left) all subnetworks created by the user (also containing a list of elements and pathways). The middle part of the tool contains the different views created by the user on the metabolic network, the time-series data but also database queries results. Since all the data are related to elements of the metabolic network, these views (except the information views) can be synchronized. We support a linking and brushing technique that allows to highlight selected elements in all views.

Visualizing metabolic network: In Fig. 2, one can see two node-link diagram views on the *Trypanosoma brucei* metabolic network. In this representation, a node is either a metabolite or a reaction, and, a metabolite and a reaction are linked if and only if the metabolite is consumed or produced by the reaction. Systrip supports various drawing algorithms (hierarchical [1], 2D/3D force-directed [5] and MetaViz [3] layout algorithms), bioinformatics algorithms (scope selection [6], chokepoints selection [12]) but also graph theoretical measures (betweenness centrality, eccentricity, strahler, strength).

Visualizing time-series data: Systrip supports different visualization methods to mine time-series -or any stamped- data. When focussing on individual time points, the user may use a spreadsheet or an histogram to visualize the values associated to the monitored elements (e.g. concentration or expression levels). If the user is interested in comparing values over several time points, Sytrip offers a scatterplot matrix view or parallel coordinates one. While scatterplot supports pairwise comparison and therefore evaluation of correlations

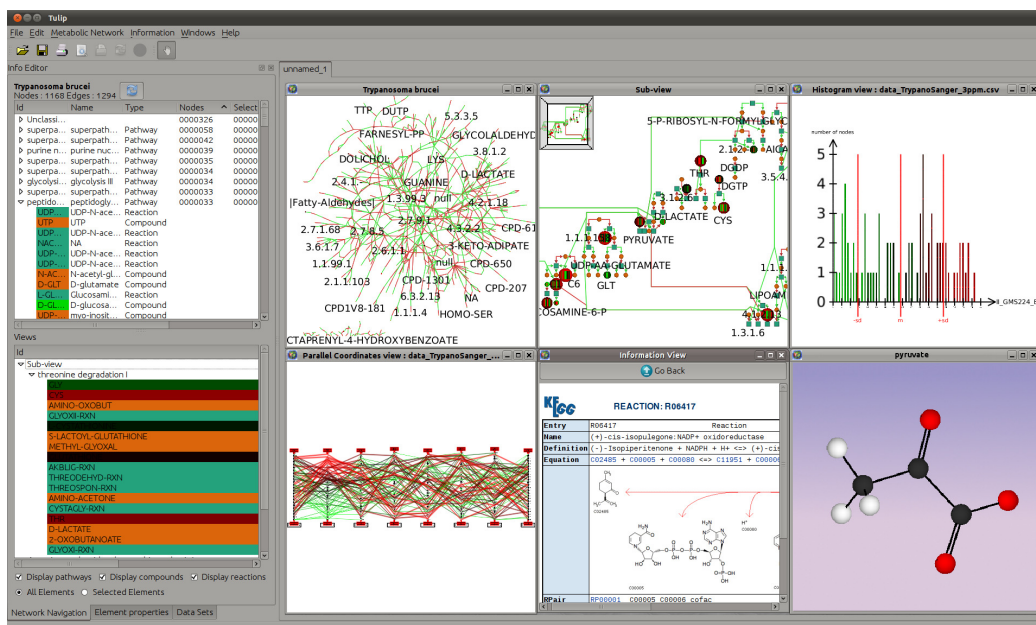


Figure 2. Screenshot of Systrip software. (Left) A panel containing (top-left) the elements and pathways of the metabolic network and (bottom-left) all subnetworks created by the user (also containing a list of elements and pathways); (Right) Representations of the metabolic network and subnetworks, time-series data and also database queries results.

between two dimensions (or time points), parallel coordinates allows to compare simultaneously elements over an arbitrary number of dimensions.

Contextual visualization: The user can also visualize time-series data in the context of the metabolic network. In that case, one can again distinguish two tasks in the analysis (i) the comparison of values of the elements at a single time point (ii) their evolutions over the entire experiment. For the comparison of values of the elements at a single time point, Systrip supports both node color and node size mapping to encode the node values for that particular time point (in Fig. 2, size and color mappings are used). To show the evolutions over the entire experiments, Systrip supports the visualization of the whole time-series of each element by rendering a heatmap or curvemap [9] glyph inside it. Another option is to navigate through the time points and to follow the evolution of the node sizes and/or colors. To preserve the user mental map, a smooth animation is performed during each transition between a time point to the next one by interpolating the node sizes and/or colors.

Visualizing database queries: The elements of the metabolic network are related to a large range of databases. In Systrip, the user can query both KEGG and PublicHouse databases. To visualize database results, Systrip offers an information view embedding a web browser (e.g. Fig. 2 shows *NADP⁺ oxidoreductase* as described in KEGG) as well as a 3D molecular geometry visualization obtained from the OpenBabel library (Fig. 2 shows the 3D geometry of *pyruvate*).

2.2 Simple vs Advanced user interface

Systrip integrates many features that can be complex to correctly set for a non-expert in graph visualization and analysis. According to the bioinformatician feedbacks, we decided to set a simple user interface as default, and, on demand, to display advanced options. The simple interface contains the most useful algorithms, i.e. mostly algorithms dedicated to biological networks. To make the interface even simpler, the parameters are automatically set to default values (these values were either set according to their biological meaning or according to empirical tests). The advanced user interface is designed for expert users and offers more functionalities than the simple one. First of all, an “Algorithms” menu, containing Tulip plug-ins installed on the computer, is added in the menu bar. Using that advanced interface, the user can set manually the needed parameters when applying an algorithm. The left panel (see Fig. 2) is also enriched: a new tab to manage the views and interaction tools settings is inserted and a configuration widget is added in the “Data Sets” tab.

3 Conclusion and future work

We presented Systrip, a visualization environment dedicated to metabolic networks. Selection and extraction of subnetworks based on the classification of the nodes in reference to metabolic pathways or to biological features such as biosynthetic capability have been facilitated. One of the main contributions of Systrip for the metabolic network visualization is the powerful interface to support contextual visualization of time-series data. The entire time series can be visualized as heatmap and curvemaps while single time points can be mapped to node size and/or color. Additional ways to explore time-series data are proposed: spreadsheets, histograms, scatterplots and parallel coordinates. Each way to explore data can be used at the same time and mapped onto the metabolic network and eventually the extracted subnetworks. Selections on any of the representations are reflected in the other ones.

An interesting direction is to offer the ability of loading metabolic networks directly from public databases. Moreover, an actual challenge in system biology is to integrate various sources of omics data with the objective to improve the realistic significance of bioinformatics analyses. Regarding Systrip, the integration of new methods dedicated to the visualization of specific and various omics data would be on purpose and may be carried by extending the actual well defined framework.

Acknowledgements

This work was partially done under the Systry project, supported by the ANR (France) and BBSRC (UK); and under the EVIDEN project (ANR 2010 JCJC 0201 01), supported by the ANR (France).

References

- [1] D. Auber. Tulip : A huge graph visualisation framework. In P. Mutzel and M. Jünger, editors, *Graph Drawing Softwares*, Mathematics and Visualization, pages 105–126. Springer-Verlag, 2003.
- [2] M. Baitaluk, M. Sedova, A. Ray, and A. Gupta. BiologicalNetworks: Visualization and analysis tool for systems biology. *Nucleic Acids Res.*, 34:W466–W471, 2006.
- [3] R. Bourqui, V. Lacroix, L. Cottret, D. Auber, P. Mary, M.-F. Sagot, and F. Jourdan. Metabolic network visualization eliminating node redundancy and preserving metabolic pathways. *BMC Systems Biology*, 1(29), 2007.
- [4] J. Dubois, L. Cottret, A. Ghodzlane, D. Auber, F. Bringaud, P. Thebault, F. Jourdan, and R. Bourqui. Systrip: a visual environment for the investigation of time-series data in the context of metabolic networks. In *Proc. of the 16th International Conference on Information Visualization (IV'12)*, 2012.
- [5] S. Hachul and M. Jünger. Drawing large graphs with a potential-field-based multilevel algorithm. In *Proc. Graph Drawing 2004*, pages 285–295, 2004.
- [6] T. Handorf, O. Ebenhöf, and R. Heinrich. Expanding metabolic networks: scopes of compounds, robustness, and evolution. *J Mol Evol*, 61(4):498–512, Oct 2005.
- [7] Z. Hu, D. M. Ng, T. Yamada, C. Chen, S. Kawashima, J. Mellor, B. Linghu, M. Kanehisa, J. M. Stuart, and C. DeLisi. VisANT 3.0: new modules for pathway visualization, editing, prediction and construction. *Nucleic Acids Res.*, 35:W625–W632, 2007.
- [8] B. H. Junker, C. Klukas, and F. Schreiber. VANTED: A system for advanced data analysis and visualization in the context of biological networks. *BMC Bioinformatics*, 7:109, 2006.
- [9] M. Meyer, B. Wong, M. Styczynski, T. Munzner, and H. Pfister. Pathline: A Tool for Comparative Functional Genomics. *Computer Graphics Forum*, 29, 2010.
- [10] S. Pailey and P. Karp. The Pathway Tools cellular overview diagram and Omics Viewer. *Nucleic Acids Research*, 34(13):3771–3778, 2006.
- [11] P. Shannon, A. Markiel, O. Ozierand, N. Baliga, J. Wang, D. Ramage, N. Amin, B. Schwikowski, and T. Ideker. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Research*, 13:2498–2504, 2003.
- [12] I. Yeh, T. Hanekamp, S. Tsoka, P. Karp, and R. Altman. Computational analysis of plasmodium falciparum metabolism: organizing genomic information to facilitate drug discovery. *Genome Res*, 14(5):917–924, May 2004.