BIOINFORMATICS APPLICATIONS NOTE Vol. 29 no. 17 2013, pages 2216–2217
doi:10.1093/bioinformatics/btt370

Systems biology

Advance Access publication June 26, 2013

GRN2SBML: automated encoding and annotation of inferred gene regulatory networks complying with SBML

Sebastian Vlaic[1,*], Bianca Hoffmann[1], Peter Kupfer[1], Michael Weber[1] and Andreas Dräger[2,3,*]

[1]Leibnitz Institute for Natural Product Research and Infection Biology - Hans-Knöll-Institute, Beutenbergstr. 11a, D-07745 Jena, Germany, [2]Center for Bioinformatics Tuebingen (ZBIT), University of Tuebingen, Sand 1, 72076 Tübingen, Germany and [3]University of California, San Diego, 417 Powell-Focht Bioengineering Hall 9500, Gilman Drive, La Jolla, CA 92093-0412, USA

Associate Editor: Martin Bishop

ABSTRACT

Summary: GRN2SBML automatically encodes gene regulatory networks derived from several inference tools in *systems biology markup language*. Providing a graphical user interface, the networks can be annotated via the simple object access protocol (SOAP)-based application programming interface of BioMart Central Portal and *minimum information required in the annotation of models* registry. Additionally, we provide an R-package, which processes the output of supported inference algorithms and automatically passes all required parameters to GRN2SBML. Therefore, GRN2SBML closes a gap in the processing pipeline between the inference of gene regulatory networks and their subsequent analysis, visualization and storage.

Availability: GRN2SBML is freely available under the GNU Public License version 3 and can be downloaded from http://www.hki-jena.de/index.php/0/2/490.

Supplementary information: General information on GRN2SBML, examples and tutorials are available at the tool's web page.

Contact: Sebastian.Vlaic@hki-jena.de or andraeger@eng.ucsd.edu

Received on April 15, 2013; revised on June 3, 2013; accepted on June 21, 2013

1 INTRODUCTION

Understanding the complex interaction processes within biological networks is one of the driving forces in the research of systems biology. One of the important processes is the regulation of gene expression, which provides information about the function and behavior of the biological system under investigation. To this end, mathematical models are frequently used to elucidate the complex nature of these processes. To make these models accessible to other researchers, to ensure interoperability and to permit their further development and reuse, mathematical models should be represented in a standardized format. By now, the *de facto* standard for the storage, exchange and use of quantitative models is the XML-based *systems biology markup language* (SBML, Hucka *et al.*, 2003).

Typically, SBML is used to represent complex models at the biochemical reaction level, which describe the underlying biological processes in a mechanistic in-depth manner. However, in the field of dynamic *gene regulatory network* (GRN) inference the applied approaches often use phenomenological modeling concepts instead, disregarding the detailed mechanisms. This accounts for missing knowledge about the underlying molecular processes and the limited amount of data available (see Hecker *et al.*, 2009a for a review on different approaches).

Therefore, the adequate representation of such models in SBML, even if they are based on quantitative frameworks such as differential equations, becomes a difficult task because the SBML core package contains several parts that are not directly applicable to the qualitative character of GRNs and lacks necessary attributes for their interpretation (Chaouiya *et al.*, 2013).

A subsequent problem to the representation of GRNs in SBML is the appropriate annotation of such models. To make them completely understandable and reusable for others, all models should be annotated following the *minimum information required in the annotation of models* (MIRIAM, Le Novère *et al.*, 2005) standard. This includes annotation of the model itself as well as annotation of its components using ontologies such as the *gene ontology* (GO; Ashburner *et al.*, 2000), the *systems biology ontology* (SBO; Courtot *et al.*, 2013) and standardized uniform resource identifiers as provided by the MIRIAM Registry (Juty *et al.*, 2012).

In the field of automatic GRN inference, model annotation and representation in SBML is still a time-consuming manual task that requires multiple tools. In fact, there is a gap between the automated inference of networks and their subsequent analysis with tools such as Cytoscape (Shannon *et al.*, 2003) or the SBMLsimulator (Keller *et al.*, 2013) (Fig. 1).

2 IMPLEMENTATION

We present GRN2SBML, a Java™ program that encodes GRNs in SBML using the Java library JSBML (Dräger *et al.*, 2011). At the moment, GRN2SBML includes parsers for three successfully applied network inference algorithms, NetGenerator (Weber *et al.*, 2013), TILAR (Hecker *et al.*, 2009b) and ExTILAR (Vlaic *et al.*, 2012). GRN2SBML can be used in three different ways to encode GRNs in SBML. The tool can either be employed from the terminal, using the graphical user interface, or directly out of R passing the algorithm's specific output of the networks. Basic annotation with SBO- and GO-terms is automatically performed by GRN2SBML according to the algorithm-specific interpretation of the network model. Additional annotation of the genes can be added either by the

*To whom correspondence should be addressed.

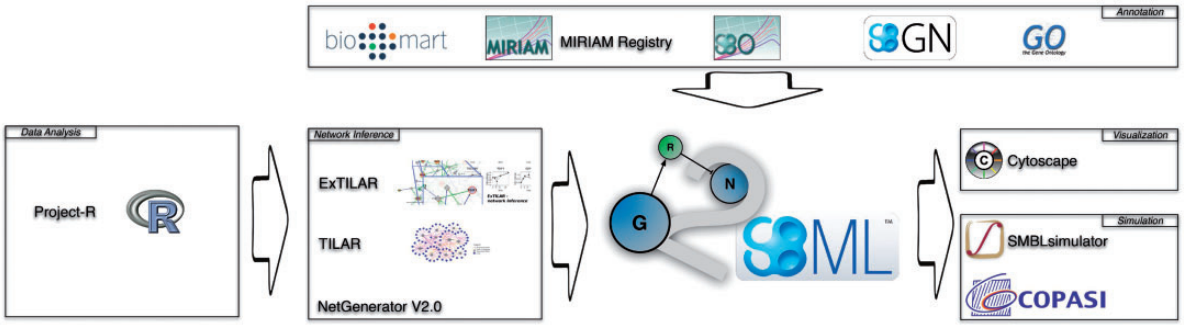

Fig. 1. In a typical network inference approach, gene expression that was pre-processed and analyzed (i.e. by using R) is used for automated network inference to derive new testable biological hypothesis. GRN2SBML offers the possibility to automatically encode the inferred models in SBML and, therefore, makes them accessible for diverse processing, visualization and simulation tools. The automated annotation with SBO-terms and GO-terms allows easy interpretation of the models. Additional annotation of the enclosed species can be added by accessing the BioMart Central Portal

import of text files or by using the simple object access protocol (SOAP)-based application programming interface of BioMart Central Portal (Guberman *et al*., 2011) and MIRIAM Registry. Relations between the genes in a network can also be annotated using PubMed identifiers. This is particularly important for networks that were inferred using prior knowledge because the annotation refers to the knowledge's original source.

Although the format to process networks is often similar across different algorithms, we found that the interpretation of the inferred networks often differs. Regarding the currently supported algorithms, species are used to denote for genes, while rules or events are used to represent the mathematical framework. In this way, we can use reactions for the visualization allowing to display structural features used, i.e. by TILAR-type of algorithms (see Supplementary Material for details). However, proper annotation that allows a correct interpretation of the network model still depends on algorithm-specific parsers. Therefore, GRN2SBML has a modular architecture that supports the implementation of parsers for both new and currently unsupported algorithms without the need to change core functions of the program.

3 CONCLUSION

A key achievement of this project is the development of GRN2SBML, a tool that provides a framework for the automated encoding of GRN-models in SBML in compliance with the specifications of the standard. This is a mandatory step toward interoperability, exchangeability and proper storing of the inferred networks. GRN2SBML automatically annotates the networks corresponding to the interpretation of the model and allows easy and up-to-date annotation of its components via diverse web resources. We used the SBML layout extension (Gauges *et al*., 2006) to provide visualization of different levels of abstraction of the models. This is especially helpful when working with large and complex networks. Therefore, GRN2SBML closes the gap in the processing pipeline between network inference and a wide range of subsequent network processing and visualization tools.

ACKNOWLEDGEMENT

The authors are grateful to Roland Keller (ZBIT, Tuebingen) and Sebastian Henkel (BioControl Jena GmbH) for valuable discussions and suggestions.

Funding: Federal Ministry of Education and Research (BMBF, Germany) in the projects Virtual Liver Network (FKZ 0315736, 0315756) and ERASysBio+ (FKZ 0315719).

Conflict of Interest: none declared.

REFERENCES

Ashburner,M. *et al*. (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet*, **25**, 25–29.

Chaouiya,OB. *et al*. (2013) Qualitative models, Version 1 Release 1. Technical report, Available from COMBINE http://identifiers.org/combine.specifications/sbml.level-3.version-1.qual.version-1.release-1 (23 July 2011, date last accessed).

Courtot,M. *et al*. (2013) Controlled vocabularies and semantics in systems biology. *Mol. Syst. Biol.*, **7**, 543.

Dräger,A. *et al*. (2011) JSBML: a flexible Java library for working with SBML. *Bioinformatics*, **27**, 2167–2168.

Gauges,R. *et al*. (2006) A model diagram layout extension for SBML. *Bioinformatics*, **22**, 1879–1885.

Guberman,J.M. *et al*. (2011) BioMart Central Portal: an open database network for the biological community. *Database*, **2011**, bar041.

Hecker,M. *et al*. (2009a) Gene regulatory network inference: data integration in dynamic models-a review. *Biosystems*, **96**, 86–103.

Hecker,M. *et al*. (2009b) Integrative modeling of transcriptional regulation in response to antirheumatic therapy. *BMC Bioinformatics*, **10**, 262.

Hucka,M. *et al*. (2003) The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics*, **19**, 524–531.

Juty,N. *et al*. (2012) Identifiers.org and MIRIAM Registry: community resources to provide persistent identification. *Nucleic Acids Res.*, **40**, D580–D586.

Keller,R. *et al*. (2013) The systems biology simulation core algorithm. *BMC Systems Biology*, **7**, 55.

Le Novère,N. *et al*. (2005) Minimum information requested in the annotation of biochemical models (MIRIAM). *Nat. Biotechnol.*, **23**, 1509–1515.

Shannon,P. *et al*. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.*, **13**, 2498–2504.

Vlaic,S. *et al*. (2012) The extended TILAR approach: a novel tool for dynamic modeling of the transcription factor network regulating the adaption to in vitro cultivation of murine hepatocytes. *BMC Syst. Biol.*, **6**, 147.

Weber,M. *et al*. (2013) Inference of dynamical gene-regulatory networks based on time-resolved multi-stimuli multi-experiment data applying NetGenerator V2.0. *BMC Syst. Biol.*, **7**, 1.