BIOINFORMATICS APPLICATIONS NOTE Vol. 27 no. 19 2011, pages 2763–2764
doi:10.1093/bioinformatics/btr456

Systems biology

Advance Access publication August 8, 2011

METANNOGEN: annotation of biological reaction networks

Christoph Gille[1,*], Katrin Hübner[2], Andreas Hoppe[1] and Hermann-Georg Holzhütter[1]

[1]Institute of Biochemistry, Computational Systems Biochemistry Group, Universitätsmedizin Berlin (Charité), Reinickendorfer Strasse 61, D-13347 Berlin and [2]COS Heidelberg, BioQuant, University of Heidelberg, Im Neuenheimer Feld 267, D-69120 Heidelberg, Germany

Associate Editor: Jonathan Wren

ABSTRACT

Motivation: Semantic annotations of the biochemical entities constituting a biological reaction network are indispensable to create biologically meaningful networks. They further heighten efficient exchange, reuse and merging of existing models which concern present-day systems biology research more often. Two types of tools for the reconstruction of biological networks currently exist: (i) several sophisticated programs support graphical network editing and visualization. (ii) Data management systems permit reconstruction and curation of huge networks in a team of scientists including data integration, annotation and cross-referencing. We seeked ways to combine the advantages of both approaches.

Results: METANNOGEN, which was previously developed for network reconstruction, has been considerably improved. From now on, METANNOGEN provides SBML import and annotation of networks created elsewhere. This permits users of other network reconstruction platforms or modeling software to annotate their networks using METANNOGEN's advanced information management. We implemented word-autocompletion, multipattern highlighting, spell check, brace-expansion and publication management, and improved annotation, cross-referencing and team work requirements. Unspecific enzymes and transporters acting on a spectrum of different substrates are efficiently handled. The network can be exported in SBML format where the annotations are embedded in line with the MIRIAM standard. For more comfort, METANNOGEN may be tightly coupled with the network editor such that METANNOGEN becomes an additional view for the focused reaction in the network editor. Finally, METANNOGEN provides local single user, shared password protected multiuser or public access to the annotation data.

Availability: METANNOGEN is available free of charge at: http://www.bioinformatics.org/strap/metannogen/ or http://d-alignment.eu/metannogen/.

Contact: christoph.gille@charite.de

Supplementary information: Supplementary data are available at *Bioinformatics* online.

Received on May 25, 2011; revised on July 14, 2011; accepted on July 31, 2011

1 INTRODUCTION

The first version of METANNOGEN (Gille *et al.*, 2007) was used exclusively to reconstruct metabolic networks as in the case of whole-cell model HEPATONET (Gille *et al.*, 2010). The program features of this early version are comparable to other programs like the recently published METNETMAKER (Forth *et al.*, 2010) or MEMOSys (Pabinger *et al.*, 2011). However, limitations became apparent with respect to information management, exchange of networks and representation of unspecific transport processes in eukaryotic cells which forced us to substantially improve METANNOGEN.

2 RESULTS

From all currently available tools for network reconstruction such as BIOLOGICALNETWORKS CELLDESIGNER, CYTOSCAPE, METNETMAKER, MEMOSYS, META-ALL, ONDEX, SEMANTICSBML, SYCAMORE, VANTED, probably METANNOGEN represents unspecific enzymes and transmembrane transport carriers most efficiently. Therefore, it is suitable for compartmentalized metabolic models of eukaryotic cells where many carriers transport not just one but several different substrates. To our knowledge, it provides the most advanced management for information attached to individual reactions. To use it, modellers do not need to take METANNOGEN as the primary network reconstruction tool, but can annotate biological networks created in any other program that provides SBML export.

Importing networks: the METANNOGEN demo session which can be web-started on the download page automatically contains the following three large-scale metabolic networks: KEGG (Kanehisa, 2002), RECON1 (Duarte *et al.*, 2007) and EHMN. In addition, any other network given in SBML can be imported either from a local file or from online resources. METANNOGEN provides direct access to the models stored in the BIOMODELS database and JWS online by model ID. All networks and their annotations can be viewed using browsable graphical trees. KEGG provides pathway maps and images of metabolite structures.

Cross-references between different networks: to cross-reference equivalent reactions from different networks, METANNOGEN defines two reactions as identical if they have identical metabolites in the left and right side of the biochemical equation, irrespectively of the order. However, different identifier systems hinder automated workflows (Hucka and Le Novere, 2010; Wang *et al.*, 2010). Here, METANNOGEN works with dictionaries, to synchronize different metabolite identifier vocabularies and to interchange compounds assumed to be different due to pH-dependent protonation (e.g. carbon dioxide versus hydrogen carbonate) or due to different ontology levels (e.g. α-D-glucose versus D-glucose).

Annotating networks: the user can type annotations or may provide automatically generated annotations as a list file. Typing

*To whom correspondence should be addressed.

of complicated reaction and compound names is facilitated by word-auto-completion and spell check. Any single word or a group of selected words in an annotation text has a context menu that exhibits at least two fundamental functions: pattern-highlighting and forwarding the text to external applications. Words with special meaning like names or identifiers of network components have additional menu items. For example, the denoted network component can be spotted in the graphical pathway maps. The annotation text can also contain certain expressions that act directly as cross-links for online resources (POST or GET), annotations or network components. Using STRAP (http://www.bioinformatics.org/strap/ or http://3d-alignment.eu/), certain cross-references display results of predictions from amino acid sequences (Gille *et al.*, 2008). MIRIAM annotations are defined according to Le Novere *et al.* (2005).

Simultaneous network design and network annotation: when the SBML file changes on hard disk, METANNOGEN will automatically reload the model. This happens when the new network topology is exported from another network editor. The workflow can be further stream lined. While the interprogram communication with another modeling software is already prepared in METANNOGEN, it will require adaptations of the modeling software at program code level. The code changes would enable the network editor to open a METANNOGEN view for the reaction under consideration to allow editing of annotations. Vice versa, the context menu in METANNOGEN could provide actions of the other program for example highlighting the correspondingr graph node. This mutual interapplication communication has already been tested in the prerelease of the network modeling environment SYCAMORE (Weidemann *et al.*, 2008).

Exporting networks: any network created or annotated with METANNOGEN can be exported in SBML format and, thus, used by other tools for further topological analysis, visualization and model simulation as in FASIMU (Hoppe *et al.*, 2011) or OPTFLUX. Optionally, the library LIBSBML (Bornstein *et al.*, 2008) is used for SBML-export. The annotations are embedded in the SBML file in line with the MIRIAM standard. METANNOGEN is highly customizable and users can easily implement other standards for their workflow.

Team work: as in the network annotation platform SEMANTICSBML (Krause *et al.*, 2010), annotations are stored independently of the exported SBML-file in a METANNOGEN-specific file format either in a local or shared data repository (Fig. 1). Users can either create a central password protected data store at http://www.bioinformatics.org/strap/metannogen/ or they can set up a central repository on their own web server. If a password-protected central repository is used, all curators can simultaneously work on the same project. Loss of data can potentially occur through concurrent modification by two curators. This is prevented by mechanisms known from modern Wiki systems. Since there is no access control at dataset level, curators may change text from their colleagues. However, data deleted in error are not completely lost, since all data are stored on the client with date and time after each session.

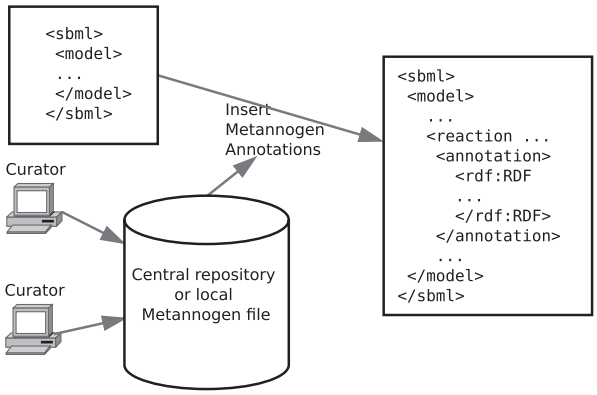

Fig. 1. Annotating reaction networks with METANNOGEN.

Publication management: METANNOGEN provides a central literature management. The publication full texts are shared in the team. Simultaneous highlighting of several specific text patterns in publication's abstracts and full texts as well as UNIPROT files helps with text mining. It is controlled by lists of text patterns.

Funding: Systems biology network HEPATOSYS of the German Ministry for Education and Research (grant 0313078A).

Conflict of Interest: none declared.

REFERENCES

Bornstein,B.J. *et al.* (2008) LibSBML: an API library for SBML. *Bioinformatics*, **24**, 880–881.

Duarte,N.C. *et al.* (2007) Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proc. Natl Acad. Sci. USA*, **104**, 1777–1782.

Forth,T. *et al.* (2010) MetNetMaker: a free and open-source tool for the creation of novel metabolic networks in SBML format. *Bioinformatics*, **26**, 2352–2353.

Gille,C. *et al.* (2010) HepatoNet1: a comprehensive metabolic reconstruction of the human hepatocyte for the analysis of liver physiology. *Mol. Syst. Biol.*, **6**, 411.

Gille,C. *et al.* (2007) METANNOGEN: compiling features of biochemical reactions needed for the reconstruction of metabolic networks. *BMC Syst. Biol.*, **1**, 5.

Gille,C. *et al.* (2008) Web-links as a means to document annotated sequence and 3D-structure alignments in systems biology. *Genome Inform.*, **20**, 277–284.

Hoppe,A. *et al.* (2011) FASIMU: flexible software for flux-balance computation series in large metabolic networks. *BMC Bioinformatics*, **12**, 28.

Hucka,M. and Le Novere,N. (2010) Software that goes with the flow in systems biology. *BMC Biol.*, **8**, 140.

Kanehisa,M. (2002) The KEGG database. *Novartis Found Symp.*, **247**, 91–52.

Krause,F. *et al.* (2010) Annotation and merging of SBML models with semanticSBML. *Bioinformatics*, **26**, 421–422.

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

Pabinger,S. *et al.* (2011) MEMOSys: Bioinformatics platform for genome-scale metabolic models. *BMC Syst. Biol.*, **5**, 20.

Wang,Y.T. *et al.* (2010) PINT: Pathways INtegration Tool. *Nucleic Acids Res.*, **38**, 124–131.

Weidemann,A. *et al.* (2008) SYCAMORE–a systems biology computational analysis and modeling research environment. *Bioinformatics*, **24**, 1463–1464.