

Systems biology

GEMMER: GEnome-wide tool for Multi-scale Modeling data Extraction and Representation for *Saccharomyces cerevisiae*

Thierry D. G. A. Mondeel, Frédéric Crémazy and Matteo Barberis*

Synthetic Systems Biology and Nuclear Organization, Swammerdam Institute for Life Sciences, University of Amsterdam, 1098 XH Amsterdam, The Netherlands

*To whom correspondence should be addressed.

Associate Editor: Jonathan Wren

Received on October 6, 2017; revised on January 4, 2018; editorial decision on January 28, 2018; accepted on January 30, 2018

Abstract

Motivation: Multi-scale modeling of biological systems requires integration of various information about genes and proteins that are connected together in networks. Spatial, temporal and functional information is available; however, it is still a challenge to retrieve and explore this knowledge in an integrated, quick and user-friendly manner.

Results: We present GEMMER (GEnome-wide tool for Multi-scale Modeling data Extraction and Representation), a web-based data-integration tool that facilitates high quality visualization of physical, regulatory and genetic interactions between proteins/genes in *Saccharomyces cerevisiae*. GEMMER creates network visualizations that integrate information on function, temporal expression, localization and abundance from various existing databases. GEMMER supports modeling efforts by effortlessly gathering this information and providing convenient export options for images and their underlying data.

Availability and implementation: GEMMER is freely available at <http://gemmer.barberislab.com>. Source code, written in Python, JavaScript library D3js, PHP and JSON, is freely available at <https://github.com/barberislab/GEMMER>.

Contact: M.Barberis@uva.nl

Supplementary information: [Supplementary data](#) are available at *Bioinformatics* online.

1 Introduction

Biological systems are complex systems: they exist in space and time, and their behavior results from the coherent integration of functionally diverse elements that interact selectively and nonlinearly (Kitano, 2002). The understanding has emerged that a cross-talk between molecular pathways is crucial to achieve the system's functions. In this context, generation of multi-scale models of biological systems, spanning multiple spatial, temporal and functional scales, is currently a major challenge in Systems Biology (Castiglione *et al.*, 2014).

Crucial steps in multi-scale modeling are the identification and visualization of the biological function, and spatial localization of interactions that occur among a set of molecules. Tools that retrieve and visualize such interaction networks for several organisms exist.

However, these are not specific for the budding yeast *Saccharomyces cerevisiae*, and do not combine the features of: (i) being web-based instead of a desktop application, (ii) allowing visual exploration through simultaneous clustering, colouring and filtering of molecules and their interactions that are (iii) based on function, localization, abundance and timing at which they occur.

Here, we present GEMMER, a novel web-based data-integration and visualization tool for budding yeast that satisfies these three requirements. The tool provides unique features as compared to existing web-based visualization tools and databases (see [Supplementary Table S1](#) for a detailed comparison). Furthermore, through its export options, GEMMER conveniently integrates with external tools that may be used to build and simulate multi-scale models.

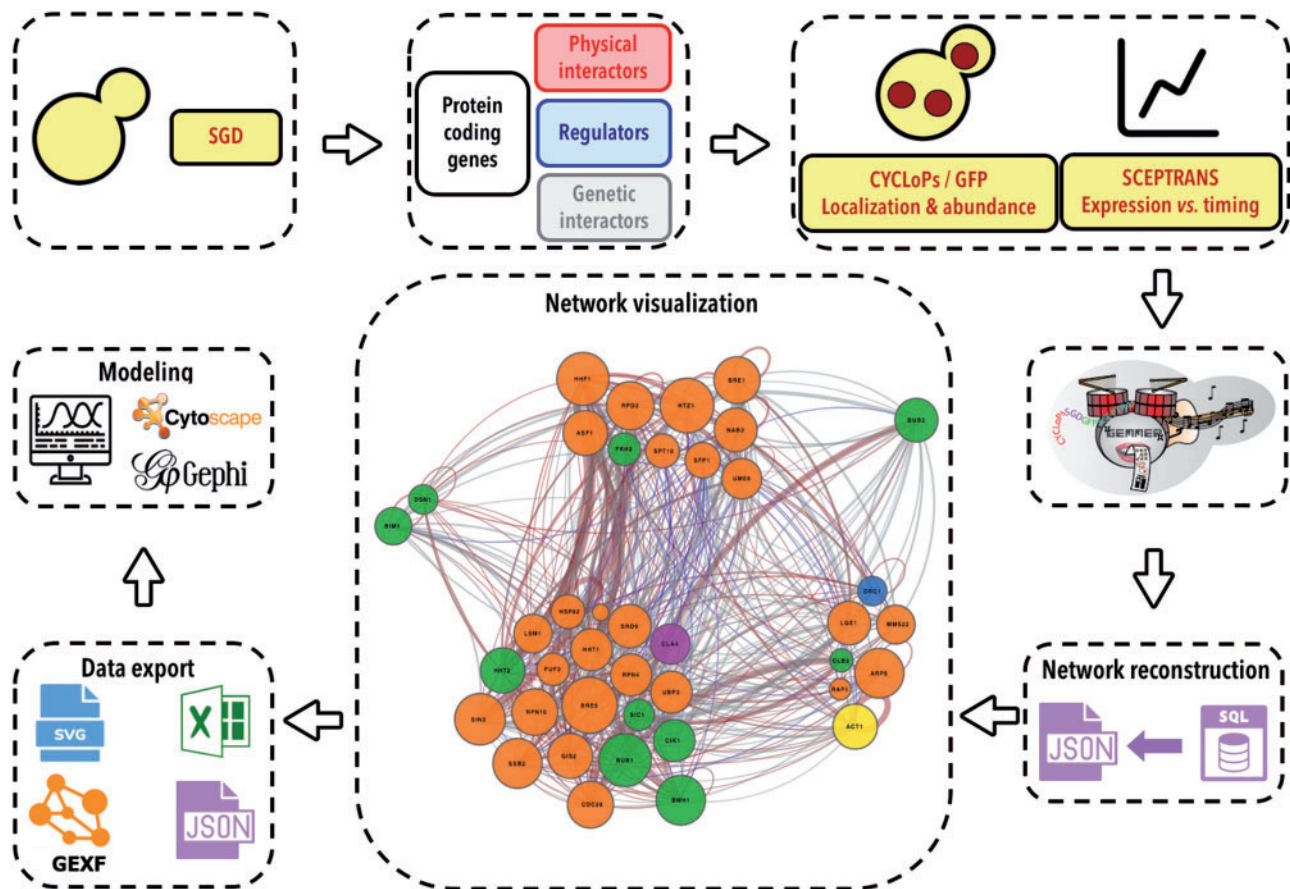


Fig. 1. GEMMER workflow. GEMMER integrates data from the SGD, CYCLOPs, GFP and SCEPTRANS databases and provides a user-friendly web interface to interact with this data. Physical, genetic and regulatory interactions for protein-coding genes and their annotations from SGD are integrated with the localization and abundance data from CYCLOPs and YeastGFP and with the timing and assignment of specific cell cycle phases of peak transcription retrieved from SCEPTRANS. The user queries GEMMER for specific interaction networks, and GEMMER interfaces with the SQL database to produce a JSON file representing an interaction network corresponding to the user's requirements. By using D3.js, Cola.js and Cytoscape.js libraries, GEMMER provides interactive, publication-quality visualizations of the interaction networks that may be exported to SVG, Excel, JSON and GEXF formats. GEMMER has been designed to aid in generation of multi-scale visualizations and models for the budding yeast *Saccharomyces cerevisiae*. To this end, the exported data may be imported into desktop applications such as Cytoscape and Gephi

2 Features

GEMMER integrates together (i) protein-coding genes, interactions and general and functional annotation from the *Saccharomyces* Genome Database (SGD) (Cherry *et al.*, 2012), (ii) localization and abundance data from both the CYCLOPs (Koh *et al.*, 2015) and Yeast GFP Fusion Localization, YeastGFP (Ghaemmaghami *et al.*, 2003; Huh *et al.*, 2003) databases and (iii) the timing and cell cycle phase of peak occurrence of RNA transcript levels (Kudlicki *et al.*, 2007; Rowicka *et al.*, 2007). GEMMER provides distinct webpages for each protein-coding gene, where this information may be viewed.

Features and information flow of GEMMER are summarized in Figure 1. After the user selects one or more genes, for which the aforementioned information is retrieved, GEMMER generates an interaction network, which varies across functional, spatial and temporal scales. Nodes in this interaction network may be clustered and coloured based on their localization in a number of cellular compartments and their functional classification. In addition, interactions may be filtered out based on: (i) type of interaction (physical, genetic or regulation), (ii) total number of experiments, (iii) unique experimental method, (iv) type of experimental evidence and (v) number of publications showing an interaction. Similarly, nodes may be filtered out based on function (process or GO term), cellular compartment, and cell cycle phase where the peak of

transcription occurs. As a result, the user receives as output an interaction network which is generated by using up-to-date literature data and filtered for their specific needs.

GEMMER provides a set of unique features as compared to existing web-based tools that allow visualization of budding yeast-specific data: STRING (Szklarczyk *et al.*, 2017), BIOGRID (Chatr-Aryamontri *et al.*, 2017), APID (Alonso-López *et al.*, 2016) and IntACT (Orchard *et al.*, 2014) (see Supplementary Table S1 for a detailed comparison). These are: (i) Generating interaction networks seeded by > 1 protein; (ii) Filtering interactions on a number of unique experimental methods that have been employed to prove an interaction; (iii) Clustering and colouring interactions based on cellular compartments or GO terms; (iv) Displaying protein expression levels; (v) Filtering nodes based on network characteristics: degree, eigenvector and Katz centrality. Conversely, GEMMER currently lacks different visual layouts and certain export formats, such as PNG, JPEG and XML, features that are instead available in some of the aforementioned tools.

3 Implementation

GEMMER stores the integrated data from the external databases in an SQLite database. This is updated by using a Python script

that downloads data from the latest available releases of SGD, CYCLOPs, YeastGFP and SCEPTRANS databases. Periodic running of the update script provides GEMMER with up-to-date literature data.

The GEMMER front-end provides a user-friendly interface with a set of menus that facilitate user input. This includes, but is not limited to, the gene(s) of interest to build an interaction network, and the filtering, clustering, colouring and scaling of nodes in the visualized network. Upon querying, the input is processed by a PHP script that executes the core application. This is written in Python and interfaces with the SQLite database, ultimately generating a JSON file of the network to be visualized. GEMMER then visualizes the network as a force-directed graph by using the JavaScript library D3js, which reads the JSON file. In addition, alternative visualizations such as hierarchical edge bundling and a circular layout are provided, together with a constraint-based layout that implements compartment separation with coloured boxes. The latter two make use of Cytoscape.js (Franz *et al.*, 2016) and Cola.js (<http://marvl.infotech.monash.edu.au/webcola/>), respectively. Accompanying the visualization(s), tables are provided with information about each protein and interaction within the network as well as links via the PubMed search engine to publications with experimental evidence.

Export options provided are: SVG for the network visualization, JSON and GEXF for the interaction network and Excel workbook for the raw data. The Excel workbook and the GEXF network may be imported into Cytoscape (Shannon *et al.*, 2003) and Gephi (Bastian *et al.*, 2009), respectively, for further analysis and model building. The webpage design utilizes the Bootstrap library, which, together with the universality of the D3js JavaScript library, allows the user to run GEMMER on any of the modern browsers such as Firefox, Google Chrome and Safari.

4 Conclusions

GEMMER is being developed to integrate existing data on proteins in budding yeast, by providing publication-quality visualizations of their interactions. The tool serves as a data-integration hub, and its visualizations aid exploration and understanding of complex networks encountered in multi-scale models. The currently available data and the implemented features, expandable in the future, achieve this goal. We aim for GEMMER to become a go-to tool to support the yeast community.

Author's Contributions

M.B. conceived the idea, the tool name and designed the strategy of the study. M.B., T.D.G.A.M. and F.C. designed the tool and its features. T.D.G.A.M. and F.C. programmed the source code. T.D.G.A.M. and M.B. implemented the tool features. T.D.G.A.M.

and M.B. wrote the paper, with contribution from F.C.. M.B. provided scientific leadership and supervised the study.

Acknowledgements

We thank Brenda J. Andrews and her lab members for providing the corrected Excel files linked to the CYCLOPs database; Dominique Groenveld for assistance with the server setup; Paul Verbruggen for help with the final layout of the GEMMER logo; and Lucas van der Zee for valuable discussions.

Funding

This work was supported by the Swammerdam Institute for Life Science Starting Grant of the University of Amsterdam and by the Systems Biology Research Priority Area Grant of the University of Amsterdam to M.B.

Conflict of Interest: none declared.

References

- Alonso-López, D. *et al.* (2016) APID interactomes: providing proteome-based interactomes with controlled quality for multiple species and derived networks. *Nucleic Acids Res.*, **44**, W529–W535.
- Bastian, M. *et al.* (2009) Gephi: an open source software for exploring and manipulating networks. In: *Int. AAAI Conf. Weblogs Soc. Media*, 361–362.
- Castiglione, F. *et al.* (2014) Modeling biology spanning different scales: an open challenge. *Biomed Res. Int.*, **2014**, 902545.
- Chatr-Aryamontri, A. *et al.* (2017) The BioGRID interaction database: 2017 update. *Nucleic Acids Res.*, **45**, D369–D379.
- Cherry, J.M. *et al.* (2012) Saccharomyces Genome Database: the genomics resource of budding yeast. *Nucleic Acids Res.*, **40**, D700–D705.
- Franz, M. *et al.* (2016) Cytoscape.js: a graph theory library for visualisation and analysis. *Bioinformatics*, **32**, 309–311.
- Ghaemmaghami, S. *et al.* (2003) Global analysis of protein expression in yeast. *Nature*, **425**, 737–741.
- Huh, W. *et al.* (2003) Global analysis of protein localization in budding yeast. *Nature*, **425**, 686–691.
- Kitano, H. (2002) Computational systems biology. *Nature*, **420**, 206–210.
- Koh, J.L.Y. *et al.* (2015) CYCLOPs: a comprehensive database constructed from automated analysis of protein abundance and subcellular localization patterns in *Saccharomyces cerevisiae*. *G3 (Bethesda)*, **5**, 1223–1232.
- Kudlicki, A. *et al.* (2007) SCEPTRANS: an online tool for analyzing periodic transcription in yeast. *Bioinformatics*, **23**, 1559–1561.
- Orchard, S. *et al.* (2014) The MIntAct project–IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Res.*, **42**, D358–D363.
- Rowicka, M. *et al.* (2007) High-resolution timing of cell cycle-regulated gene expression. *Proc. Natl. Acad. Sci. USA*, **104**, 16892–16897.
- Shannon, P. *et al.* (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.*, **13**, 2498–2504.
- Szklarczyk, D. *et al.* (2017) The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible. *Nucleic Acids Res.*, **45**, D362–D368.