Bioinformatics, 35(24), 2019, 5385–5388
doi: 10.1093/bioinformatics/btz509
Advance Access Publication Date: 24 June 2019
Applications Note

Databases and ontologies

Interactive exploration of heterogeneous biological networks with Biomine Explorer

Vid Podpečan[1,]*, Živa Ramšak[2], Kristina Gruden[2], Hannu Toivonen[3] and Nada Lavrač[1,4]

[1]Department of Knowledge Technologies, Jožef Stefan Institute, Ljubljana, Slovenia, [2]Department of Biotechnology and Systems Biology, National Institute of Biology, Ljubljana, Slovenia, [3]Department of Computer Science, University of Helsinki, Helsinki, Finland and [4]Centre for Information Technologies and Applied Mathematics, University of Nova Gorica, Nova Gorica, Slovenia

*To whom correspondence should be addressed.

Associate Editor: Jonathan Wren

Received on December 20, 2018; revised on April 18, 2019; editorial decision on June 10, 2019; accepted on June 19, 2019

Abstract

Summary: Biomine Explorer is a web application that enables interactive exploration of large heterogeneous biological networks constructed from selected publicly available biological knowledge sources. It is built on top of Biomine, a system which integrates cross-references from several biological databases into a large heterogeneous probabilistic network. Biomine Explorer offers user-friendly interfaces for search, visualization, exploration and manipulation as well as public and private storage of discovered subnetworks with permanent links suitable for inclusion into scientific publications. A JSON-based web API for network search queries is also available for advanced users.

Availability and implementation: Biomine Explorer is implemented as a web application, which is publicly available at https://biomine.ijs.si. Registration is not required but registered users can benefit from additional features such as private network repositories.

Contact: vid.podpecan@ijs.si

1 Introduction

The integration of data sources into heterogeneous networks and the application of appropriate search algorithms enable the discovery of hidden connections, new links and yet unknown patterns and laws (Shi *et al.*, 2017). Cytoscape (Shannon *et al.*, 2003), Ondex (Köhler *et al.*, 2006), GeneMANIA (Warde-Farley *et al.*, 2010), KnetMaps (Singh *et al.*, 2018) and OmicsNet (Zhou and Xia, 2018) are examples of tools for the visualization and analysis of heterogeneous biological networks.

This paper presents Biomine Explorer, a web application which enables interactive search, visualization and exploration in large biological networks integrated from several public data sources. Biomine Explorer uses Biomine (Eronen and Toivonen, 2012) as the backend data integration and graph search engine and implements a user-friendly search interface and an interactive network visualization interface, which also provides unlimited network exploration by in-place search and network management options such as editing, saving, downloading and sharing.

The following biological databases are used as the source of data for constructing heterogeneous networks: Entrez Gene, UniProt, Gene Ontology, InterPro, STRING, OMIM, PubMed, GoMapMan and CKN, which is a manually curated network of plant data, compiled from different sources (Ramšak *et al.*, 2018).

Biomine Explorer currently enables the exploration of three large heterogeneous networks. The first and the largest one contains data about human, mouse, rat, fruit fly and nematode. It has 1 520 673 nodes and 32 761 889 links. The second one is a subset of the first one and contains only human data. The third one consists of 1 740 270 nodes and 4 125 274 edges and contains data about the following plant organisms: arabidopsis, potato, rice, tomato, tobacco, beet, cacao, pearl millet, bread wheat, green algae, black cottonwood and wine grape.

2 Implementation

Biomine Explorer is implemented as a modern web application and works in all recent web browsers. The implementation consists of three main components: the front end, the back end and the middleware layer.

The front end implements the web user interface for search and network exploration and communicates with the back end. The back end handles a variety of basic and advanced tasks such as URL resolution, HTML template rendering, serving static content, user management, security, management of graph repositories, API calls, graph algorithms and data transformation and preparation routines. The middleware layer provides a set of functions which invoke and interact with different parts of Biomine, such as the search engine, the cache server, the database and name resolution utilities. The middleware layer is loosely coupled with Biomine and integration of other network mining engines is possible by implementing the appropriate functions. The front end of Biomine Explorer which runs in the browser uses the *vis.js* library for high-performance graph rendering and layout with HTML5 canvas. The *Bootstrap* Javascript library is used for responsive web page layout and user interface components. Several other Javascript libraries such as *Select2*, *Vex* and *jQuery* are also used. The back end (application server and web server) of Biomine Explorer is based on the *Django* web development framework, *Nginx* web server and various libraries such as *NetworkX* for graph analysis and *MySQL* for relational data management.

The Biomine system is an important part of the Biomine Explorer software stack. It is implemented as a standalone system providing a collection of programs and scripts. There are sets of tools for data download, parsing, extraction and integration, database import and management, implementations of several graph algorithms and several other data management utilities. Biomine components are implemented in a variety of programming languages and require a 64-bit Linux system.

3 Features

Biomine Explorer implements several features which are focused on enabling exploration of relevant biological knowledge hidden in large heterogeneous networks. The search interface supports three query modes and provides autosuggest on the query input field, name resolution for imported query terms, source network selector, desired output network size, grouping of equivalent nodes and an interactive preview of the discovered network with the option to open it in the advanced network exploration interface or download it in one of the several supported formats.

Visualization of heterogeneous networks is performed using the physics module of the *vis.js* library. It is based on a force-directed layout and uses the Barnes Hut algorithm for n-body simulation. A damping factor is also used to ensure stabilization of node positions. However, when a node is moved or deleted the simulation is restarted so the network responds to user manipulation. There is also an option to freeze node positions which is useful when preparing a networks for publication where a more compact layout is required. When the network is saved, node positions are also saved which ensures that the visitors following the link to the network will see exactly the same picture.

The advanced network exploration interface supports interactive manipulation of nodes, edges and the whole network. Interactive exploration is made possible by the *expand* function which enables in-place querying with the selected node(s) and merging the resulting subnetwork with the currently visualized subnetwork. This way, the most interesting nodes or groups of nodes, their neighbors and context can be explored further without limitations. In addition, direct links to data sources are provided for all nodes. The exploration interface also offers editing functions which include removal of nodes and edges by type, removal of isolated nodes, saving modifications, loading the original (unmodified) subnetwork, downloading the network and creating a private copy of the current network.

4 Examples

4.1 Cholesterol biosynthesis

Cholesterol is an important biological molecule with several key roles. Its synthesis is a complex multi-step process which is tightly regulated. Here we demonstrate how Biomine Explorer can help us explore interrelations between genes important in human cholesterol biosynthesis and release.

We used LDLR, PCSK9 and HMGCR genes as the source query nodes and the default (largest) network to obtain a subnetwork which is partially shown in Figure 1. Gene coding for the low density lipoprotein receptor (LDLR) is shown to interact with three proteins (a multi-membrane spanning protein RNF139, a transmembrane receptor CD209 and GABRA6, an inhibitory neurotransmitter), which in turn directly interact with 3-hydroxy-3-

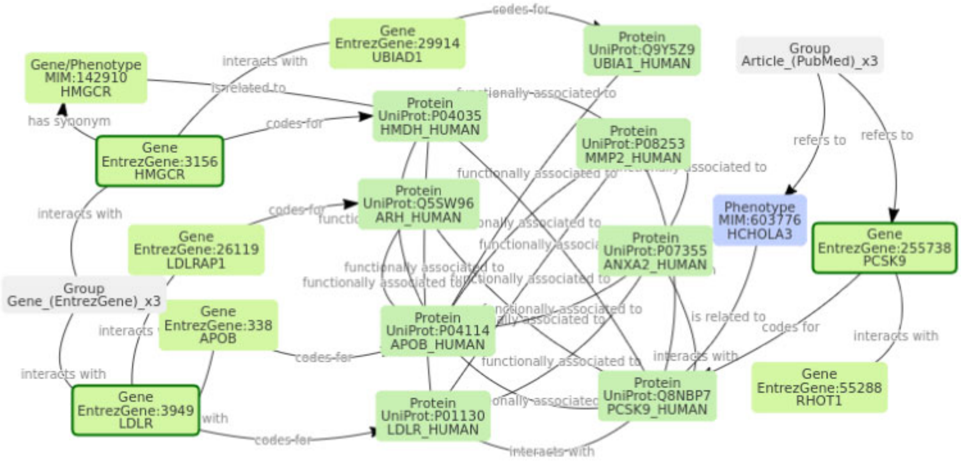

Fig. 1. Biomine Explorer's visualization of a part of the cholesterol biosynthesis network. The network was edited in Biomine Explorer to expose the most relevant parts. The edited and the original network are available at https://biomine.ijs.si/visualize/biomine_3_oct_2018/Wed-17-Apr-2019-15-11-55–69490/

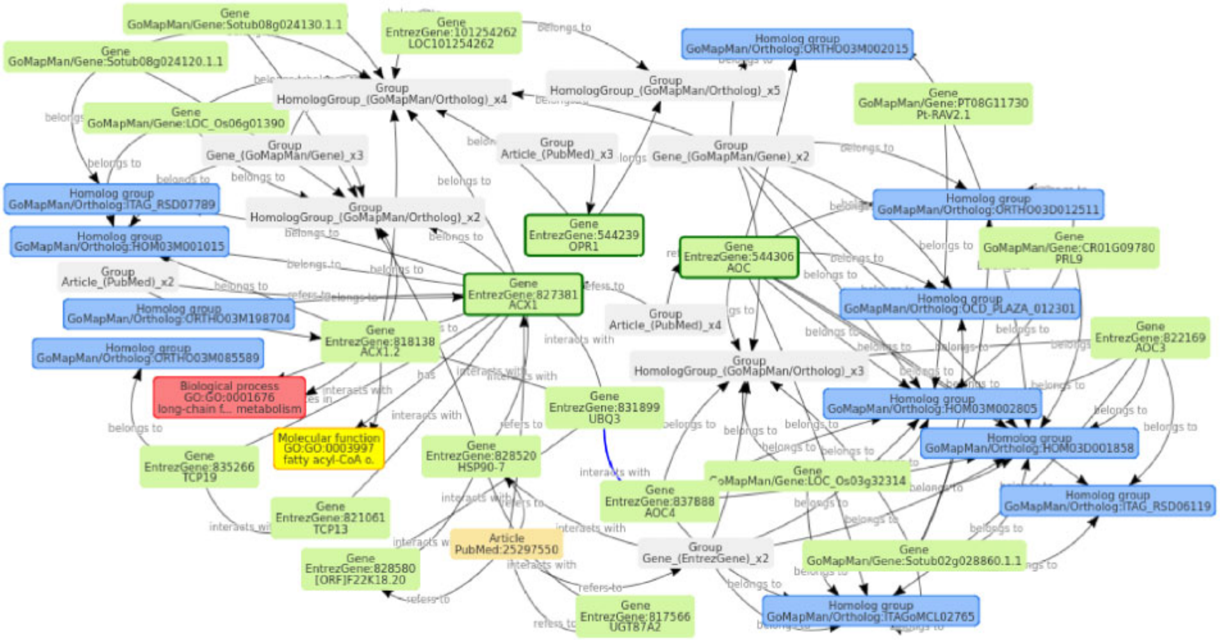

Fig. 2. Biomine Explorer's visualization of the synthesis of jasmonic acid in plants. The query nodes are shown with a thick border. The network is available at https://biomine.ijs.si/visualize/plants_2_aug_2018/Wed-17-Apr-2019-15-46-44–63457/

methylglutaryl-CoA reductase (HMGCR), the rate-limiting enzyme for cholesterol synthesis, whose action in mammalian cells is suppressed by cholesterol derived internalized degraded remnants of low density lipoproteins (LDLs). In addition, HMGCR gene is connected with the OMIM database (OMIM: 142910), informing us of HMGCR protein function in sterol-dependent suppression and statin-dependent induction and the connection to familial hypercholesterolemia. HMGCR gene also interacts with a gene that codes for a protein thought to be involved in cholesterol and phospholipid metabolism (UBIA1). On the other hand, the gene interact with two genes (LDLRAP1, APOB) that code for a protein found to interact with the cytoplasmic tail of the LDL receptor (ARH) and the main apolipoprotein of chylomicrons and low density proteins (APOB). Mutations in both genes are known to cause several cholesterol related disease pathologies (e.g. hypercholesterolaemia; OMIM: 603776). Proprotein convertase subtilisin/kexin type 9 (PCSK9), a crucial player in the regulation of plasma cholesterol homeostasis, acts on hepatic LDLR (clathrin LDLRAP1/ARH-mediated pathway) and inhibits intracellular degradation of APOB (autophagosome/lysosome pathway). Protein products of genes HMGCR (HMDH), LDLRAP1 (ARH), APOB (APOB) and LDLR (LDLR) represent links to the PCSK9 gene via two proteins: a metalloproteinase (MMP2) involved in diverse functions (vasculature remodeling, angiogenesis, tissue repair, tumor invasion, inflammation) and a calcium-regulated membrane-binding protein (ANXA2) that inhibits PCSK9-enhanced LDLR degradation. Lastly, PSCK9 gene interacts with ras homolog family member T1 (RHOT1), an atypical Rho GTPase with a role in apoptosis and mitochondrial homeostasis.

4.2 Jasmonic acid synthesis

The synthesis of hormone jasmonic acid in plants is well understood. It requires the release of linolenic acid from phosphatidylcholine by lipoxygenase. Furthermore, the action of enzymes allene oxide synthase (AOS), allene oxide cyclase (AOC), OPR1 and ACX1 is required before the metabolite enters the beta-oxidation cycles.

We used AOS, AOC and ACX1 as source query nodes against the network containing only plant-related data. The resulting subnetwork is shown in Figure 2. It gives an overview of biological processes related to jasmonic acid synthesis using GO as well as MapMan terms, information on publications on query genes and information on orthologues in other species. It also shows information on interaction of target proteins with other proteins thus creating a basis for generating novel biological hypotheses.

5 Availability and documentation

The Biomine Explorer web application is publicly available at https://biomine.ijs.si. The project is currently developed as closed source. A user manual available at https://biomine.ijs.si/howto provides information about how to use the search interface to perform queries and how to use the network visualization and exploration interface. The API which is documented at https://biomine.ijs.si/api_documentation/ lists the available functions and parameters, and provides an example in the Python programming language.

Acknowledgements

The authors wish to thank members of the former Biomine development team: Atte Hinkka, Kimmo Kulovesi, Lauri Eronen, Petteri Hintsanen and Laura Langohr. The authors would also like to acknowledge the help of biology experts Helena Motaln and Blaž Škrlj.

Funding

This work was supported by the Slovenian Research Agency grants N2-0078, J7-7303, P2-0103, J6-9372 and P4-0165.

Conflict of Interest: none declared.

References

Eronen,L. and Toivonen,H. (2012) Biomine: predicting links between biological entities using network models of heterogeneous databases. *BMC Bioinformatics*, **13**, 1–21.

Köhler,J. *et al.* (2006) Graph-based analysis and visualization of experimental results with ONDEX. *Bioinformatics*, **22**, 1383–1390.

Ramšak,Ž. *et al.* (2018) Network modeling unravels mechanisms of crosstalk between ethylene and salicylate signaling in potato. *Plant Physiol.*, **178**, 488–499.

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

Shi,C. *et al.* (2017) A survey of heterogeneous information network analysis. *IEEE Trans. Knowl. Data Eng.*, **29**, 17–37.

Singh,A. *et al.* (2018) KnetMaps: a BioJS component to visualize biological knowledge networks. *F1000Research*, **7**, 1651.

Warde-Farley,D. *et al.* (2010) The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res.*, **38**, W214–W220.

Zhou,G. and Xia,J. (2018) OmicsNet: a web-based tool for creation and visual analysis of biological networks in 3D space. *Nucleic Acids Res.*, **46**, W514–W522.