



LmSmdB: an integrated database for metabolic and gene regulatory network in *Leishmania major* and *Schistosoma mansoni*



Priyanka Patel, Vineetha Mandlik, Shailza Singh *

National Centre for Cell Science, SP Pune University Campus, Ganeshkhind Road, Pune 411007, India

ARTICLE INFO

Article history:

Received 10 November 2015

Accepted 17 December 2015

Available online 19 December 2015

Keywords:

L.major

S.mansoni

Regulatory networks

Transcription factors

Database

ABSTRACT

A database that integrates all the information required for biological processing is essential to be stored in one platform. We have attempted to create one such integrated database that can be a one stop shop for the essential features required to fetch valuable result. LmSmdB (*L. major* and *S. mansoni* database) is an integrated database that accounts for the biological networks and regulatory pathways computationally determined by integrating the knowledge of the genome sequences of the mentioned organisms. It is the first database of its kind that has together with the network designing showed the simulation pattern of the product. This database intends to create a comprehensive canopy for the regulation of lipid metabolism reaction in the parasite by integrating the transcription factors, regulatory genes and the protein products controlled by the transcription factors and hence operating the metabolism at genetic level.

© 2015 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

An integrated relational database of curated parasites namely *L. major* and *S.mansoni*, causative organism for leishmaniasis and schistosomiasis is hosted on the NCCS webportal (<http://www.nccs.res.in/LmSmdb/>). This database provides information from various 'omics' studies which integrates analyze and manage environments for systems biology research. This is possible with the advent of genome analysis tool that allows identification of genes related to pathogenesis and aid to determine genes as a diagnostic marker essential in metabolic pathways which may lead to drug target identification. Towards this end, we have compared the lipid metabolic pathways specifically for *Leishmania major* and *Schistosoma mansoni* by integrating the knowledge of genome sequences of the mentioned organisms. Molecular pathways that appear to be targeted during infection, and these results highlight pathways exhibiting responses specific for a given pathogen infection. The present database is an example of the usefulness of data integration techniques which enable the creation of a hypothesis generation platform for human diseases.

Identifying a drug target and modeling a disease network is straightforward if there is conservation between two species. This conservation between the species helps to search the homologs and orthologues in a model organism. On the basis of sequence conservation, function between two genes can be deciphered which estimates probability of conservation, but evolution does not always evolve parallel with respect to evolution of sequences. We therefore made an effort to integrate two

orthologues which are evolutionarily related for extracting information useful for cataloging details of interaction network particularly a gene regulatory network (GRN) and a protein interaction network (PIN) giving an insight into the relationship of protein/genes within the network. A gene regulatory network (GRN) was built in order to find the genes that are essential and highly connected with their corresponding partners. The highly connected nodes in the GRN or the hubs indicate the importance of the node in the interaction and a probable target.

In the present database, we have laid our focus on the lipid metabolism of parasites that cause infectious diseases. It is noted from [3,4] that lipid metabolism is important for maintenance of infectivity and viability of the parasite. Any change in the biochemical network results in a change in the intercellular lipid trafficking and change in the membrane composition. As lipids play an important role in the virulence and multiplication of the parasite, any approach to disrupt the lipid metabolism could result in a therapeutic strategy to stop parasite proliferation and growth [5]. In this regard, a database LmSmdB was created which studies the network of interactions, maps pathways across taxonomic branches and also incorporates data obtained from simulation studies. During the creation of LmSmdB we have simplified the process of integrating metabolic pathway interactions and protein sequence information for pathogen related studies. The main aim behind the construction of the metabolic network was to identify the important hubs which could act as drug targets in the network. We have also simplified the process of integrating metabolic pathway interactions and protein sequence information for pathogen related studies in order to address the problems related to drug target identification.

The parasite, *Leishmania major*, the causative organism for cutaneous leishmaniasis was fully sequenced in 2005 [1]. Similarly *S. mansoni*

* Corresponding author.

E-mail address: shailza_iitd@yahoo.com (S. Singh).

which causes schistosomiasis was sequenced in the year 2009. [2]. There are several databases like KEGG [8], BIOCYC [9], UNIPROT (The Uniprot Consortium) [10], NCBI [11] etc. that hold information about the proteins, genes and metabolites related to various biochemical pathways present in *L.major* and *S.mansoni*. In additions, several organism specific databases provide information on both metabolic and regulatory network whereas none of the said database provides simulation of the constructed network. In order to understand network architecture and simulation pattern of the biochemical network, we have simplified and streamlined the process of integration of molecules, genes, and protein structure in order to perform simulation of the built network. The contents on the server are exactly the same as the source code of the web page on the client side. It is the first known attempt for integrating the metabolic network and simulation of the network in a database. The different metabolic and regulatory networks were constructed using different tools [6]; the detailed information is discussed as below.

1.1. Metabolic network reconstruction

Metabolic networks of the lipid metabolism of *L.major* and *S.mansoni* genome were constructed using Cell designer 4.3 software [7], manually to curate network of the metabolism of an organism with all the genes, proteins and reactions assembled from a functionally annotated genome, biochemical data, and literature that are compiled into a stoichiometric matrix which serves as an opportunity for systematic identification of metabolic targets in the pathogen.

The regulatory network analysis of *L.major* and *S.mansoni* was dealt using the gene regulatory network interfaces with GeneNetWeaver. As the state of metabolic networks are sensed by association with the gene regulatory network, the initial data was drawn from KEGG database, Biocyc database and literature survey.

Different types of simulation like stochastic, deterministic and hybrid simulation was applied on the generated regulatory network. The advantage of using *in-silico* network is the ease to carry the perturbation experiments that can be easily simulated to produce expression

data unlike *in vivo* experiments, which are usually expensive and time consuming. Moreover, both quantity and quality of the expression data generated can be controlled (e.g. by varying the amount of molecular and/or measurement noise).

2. Materials and methods

2.1. Database creation

WAMP (WAMPserver) a Windows web development environment allows the creation of web applications with Apache2, HTML and a MySQL database. WAMP 5 was used to create the LmSmdB, where HTML and PHP were used as an administrator to handle the administration of MySQL over the Web.

2.2. Querying database

The LmSmdB web page can be searched using keywords or queried using gene name or the NCBI/KEGG cluster identifier. A collection of gene aliases was assembled based on their availability on different databases. One of the key features of the database is its ability to extract data for several genes together in a batch, eliminating the need for cross referencing of the data from external databases. This batch search is useful for genomic studies where frequent updations associated with genes or proteins have to be examined frequently. The list of genes that serves as an input is in the form of a text file. This can be uploaded on the server or the list can be directly pasted into the search box. Grouping of the documents or web pages relevant to the query is done by the data sources and by the similarity searches among these documents. The entire architecture of LmSmdB database has been presented in Fig. 1. The similarity searches between two documents are based on the sequence of 'words' like '*Leishmania*', or an abbreviation, like '*Lmj*', a sequence word 'ATGCTGGG' or 'MEPQSDVL', a number, or a combination of letters and numbers like 'Smp_124730'. A binary vector is then built on the descriptive set of words which actually define the biological

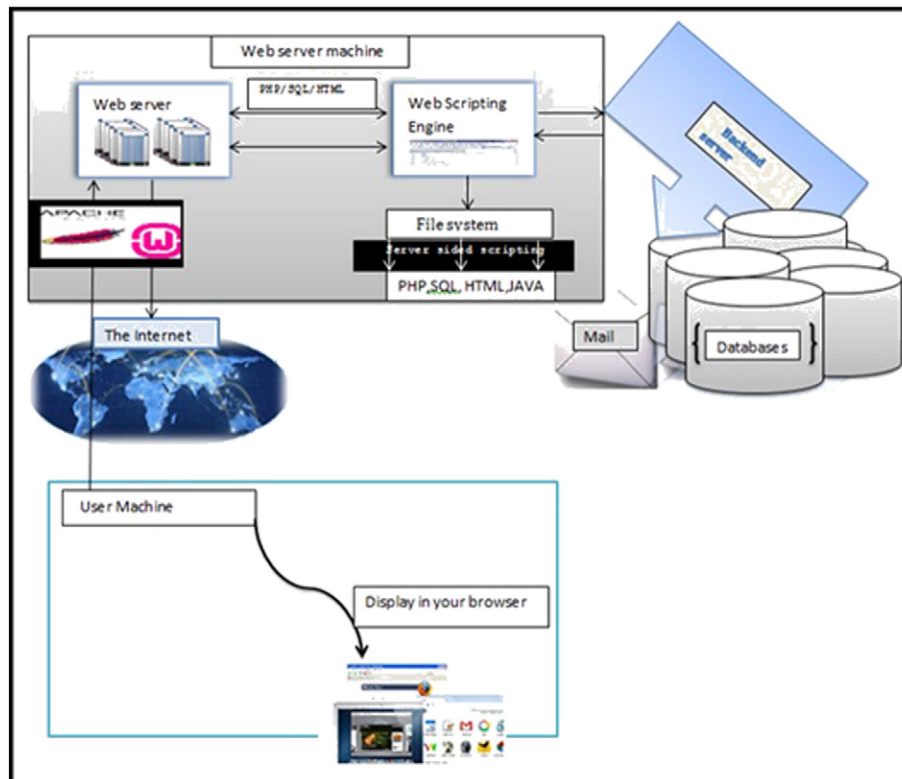


Fig. 1. The architecture of LmSmdB database consists of a backend server, web-server machine and a user machine.

Download English Version:

<https://daneshyari.com/en/article/2822151>

Download Persian Version:

<https://daneshyari.com/article/2822151>

[Daneshyari.com](https://daneshyari.com)