LmTDRM Database: A Comprehensive Database on Thiol Metabolic Gene/Gene Products in *Listeria monocytogenes* EGDe

Vanishree Srinivas and Shubha Gopal*

Department of Studies in Microbiology, University of Mysore, Manasagangotri, Mysore 570006, Karnataka, India

Summary

There are a number of databases on the *Listeria* species and about their genome. However, these databases do not specifically address a set of network that is important in defence mechanism of the bacteria. *Listeria monocytogenes* EGDe is a well-established intracellular model organism to study host pathogenicity because of its versatility in the host environment. Here, we have focused on thiol disulphide redox metabolic network proteins, specifically in *L. monocytogenes* EGDe. The thiol redox metabolism is involved in oxidative stress mechanism and is found in all living cells. It functions to maintain the thiol disulphide balance required for protein folding by providing reducing power. Nevertheless, they are involved in the reversible oxidation of thiol groups in biomolecules by creating disulphide bonds; therefore, the term thiol disulphide redox metabolism (TDRM). TDRM network genes play an important role in oxidative stress mechanism and during host–pathogen interaction. Therefore, it is essential to have detailed information on these proteins with regard to other bacteria and its genome analysis to understand the presence of tRNA, transposons, and insertion elements for horizontal gene transfer. *LmTDRM* database is a new comprehensive web-based database on thiol proteins and their functions. It includes: Description, Search, TDRM analysis, and genome viewer. The quality of these data has been evaluated before they were aggregated to produce a final representation. The web interface allows for various queries to understand the protein function and their annotation with respect to their relationship with other bacteria. *LmTDRM* is a major step towards the development of databases on thiol disulphide redox proteins; it would definitely help researchers to understand the mechanism of these proteins and their interaction. Database URL: www.lmtdrm.com

1 Introduction

Bacteria during their interaction with the host or during other conditions generate reactive oxygen species (ROS) and reactive nitrogen species (RNS). These serve as secondary messengers and/or can also damage proteins, nucleic acids, and lipids [1, 2]. To avoid cellular damage by these processes, most bacteria have developed a series of antioxidants that can convert ROS and RNS to unreactive derivatives. They are (i) a number of enzymes such as catalase, glutathione peroxidase, glutathione-S-transferase, thiol-specific peroxidase, methionine sulphoxide reductase, thioredoxin peroxidase, glutathione reductase, and glutaredoxins; (ii) various metal binding proteins such as ceruloplasmin, ferritin, transferrin, various metabolites, and cofactors (NADs, lipoic acid, uric acid, bilirubin, etc); (iii) a number of dietary components (vitamin A, C, E, quercetin, etc.) and (iv) metal ions (Mg^{2+}, Mn^{2+}, Zn^{2+}) with enzymes or metalloenzymes like superoxide dismutase. Among them, thiol proteins are the major ones [3, 4].

* To whom correspondence should be addressed. Email: shubhagopal_mysore@yahoo.com

Thiols are organic compounds that contain sulphhydryl group. Protein thiols in the plasma include the protein sulphhydryl groups and protein mixed disulphides with homocysteine, cysteinylglycine, cysteine, and glutathione. Human plasma contains homocysteine (HcySH), cysteinylglycine (CysGlySH), cysteine (CysSH), and glutathione (GSH) as reduced thiols. These thiols are also found as low-molecular mass (symmetrical) disulphides i.e. homocysteine, cystinilglycine, cysteine, and glutathione disulphide [5,6].

Glutathione (GSH) acts as an important defence mechanism to protect cells from oxidative stress as it can forage free radicals, reduce peroxides, or bind electrophilic compounds [7,8]. Thioredoxins are a family of proteins with thioredoxin fold; they catalyze oxidation-reduction reactions by two cysteine residues in the active site (CXXC motif). Both the cysteine residues are separated by any amino acid (InterPro database, IPR015467, http://www.ebi.ac.uk/interpro/). Trx domain is present in N-terminal which helps in easy access to the redox reactions. Other proteins that include Trx folds are glutaredoxins and disulphide isomerases [1]. Thioredoxin and Glutaredoxins follow alpha–beta–alpha sandwich model in the 3D structure. They consist of four alpha helices and five beta sheets [9]. Both these enzymes are responsible for maintaining redox environment in the cell. Thioredoxin can also interact with other proteins to form functional protein complex [10, 11]. Apart from thioredoxins and glutaredoxins, several homologues are present wherein C/T terminal of the CXXC motif is replaced with serine or threonine; these proteins exhibit redox activity. One of such proteins having CXXS motif is methionine sulphoxide [10]. CXXS motifs are highly conserved and present in structurally distinct proteins [12]. The N-terminal cysteine can also be replaced with S/T [12]. Most of these homologues contain thioredoxin fold and perform redox reactions [10].

1.1 *Listeria monocytogenes* as model organism

L. monocytogenes is a gram positive facultative intracellular bacterium. It remarkably adapts to the host cell and multiplies in the macrophages and also in other cell types [13]. The bacterial surface proteins internalin A and internalin B induce the entry of *L. monocytogenes* into phagocytes; both the internalins are necessary for the bacterial entry into different cell types [14]. The entry of *L. monocytogenes* into mammalian cells requires actin polymerization and membrane remodelling, which is an active process and it is an example of how the bacterium utilizes the host cell signalling pathways to its advantage [15]. The study on *L. monocytogenes* infection emphasizes the primitive relationship between the bacterium and the host. For decades, *L. monocytogenes* has been a tool for clinicians and immunologists to study the host–pathogen interaction, bacterial pathogenesis, and cellular microbiology [16-18]. During *L. monocytogenes* infections, host IFN-gamma activated macrophages shows primary host defence mechanism inducing stress environment. Despite the stress, *L. monocytogenes* is able to survive. When whole genome base transcriptome was analyzed for IFN-γ-activated macrophages, host macrophage as well as intracellular replication bacteria showed elevated oxidative and nitrosative stress levels. 21 transcripts that were regulated included oxidative stress genes [19]. Dons et al. showed that 2-Cys peroxiredoxin (Prx) homologue (Lmo1604) showed protection against oxidative and nitrosative stress in vitro and in vivo. Lmo1604 also plays a role during infection to the host and protection against stress [20]. During host–pathogen interactions, oxidative stress genes were elevated indicating that these genes actively participate in this mechanism and are important for virulence [21]. *Staphylococcus aureus*, the related bacteria to *L. monocytogenes*, showed modifications of genes under stress conditions. When *S. aureus* genome was analyzed under oxidative and nitorsative stress, it was found that thioredoxin gene (NCTC 8325 Locus 02845) was upregulated with other iron and zinc regulators [22]. This information indicates that *L. monocytogenes* is a versatile organism and thiol genes are important in host pathogenicity.

The present databases on *L. monocytogenes* are *L. monocytogenes* database, *L. monocytogenes*—MLST Pasteur Web Site, VFDB Listeria and ListiList. All these databases are mainly on the *Listeria* species and their genome annotations; whereas, VFDB Listeria database is the only database on virulence factors of the organism and their role in pathogenicity. No particular database is available on TDRM gene/gene products of *L. monocytogenes* which are also important in host–pathogen mechanism. Hence, we have developed a comprehensive database of thiol genes/gene products in *L. monocytogenes* EGDe, which contains complete information on thiol genes/gene products called as *LmTDRM* Database. It contains data on 26 TDRM gene/gene products which are involved in thiol metabolism. It also includes comparison of TDRM gene/gene products across more than 500 bacterial genomes in presence and absence of Lmo2770 (bifunctional glutamate–cysteine ligase/glutathione synthetase). Genome viewer is a visualization platform for genome analysis for 26 TDRM gene transcription units across other bacteria. The database is available as an open electronic source through the user-friendly web interface which allows simple and complex queries such as: gene annotation, domain annotation, structure details, and information based on particular organism along with genome viewer. This database is the first attempt to systematically fetch information on *L. monocytogenes* EGDe metabolome with respect to thiol disulphide redox metabolism. As compared to other databases, *LmTDRM* database is very specific to the *L. monocytogenes* EGDe strain in connection to thiol disulphide redox metabolism and pathogenicity.

2 Methods

2.1 Database development

The development of *LmTDRM* database includes following major steps: Literature search, Data compilation, Data evaluation, Data aggregation, Development of cluster analysis, Genome viewer, and Final data export to Mysql database (Figure 1).

2.1.1 Literature search and Data compilation

Previously from our laboratory, we have identified 14 candidate proteins containing CXXC motif which is a hallmark for TDRM proteins [23]. Further, detailed analyses yielded five thiol-derived proteins like CXXT and TXXC [24]. All the 26 thiol proteins which were identified were searched for literature information in pubmed database (www.ncbi.nlm.nih.gov/pubmed). After confirmation, these set of genes/gene products were used for cluster analysis.

2.1.2 Data evaluation and data aggregation

26 gene/gene products identified and evaluated as thiol gene/gene products were based on their CXXC motif, Thioredoxin, and Glutaredoxin domains. These genes were grouped into Thioredoxin, Glutathione, Thioredoxin-Glutathione-Glutaredoxin dependent, transcriptional regulators, and others (Table 1).

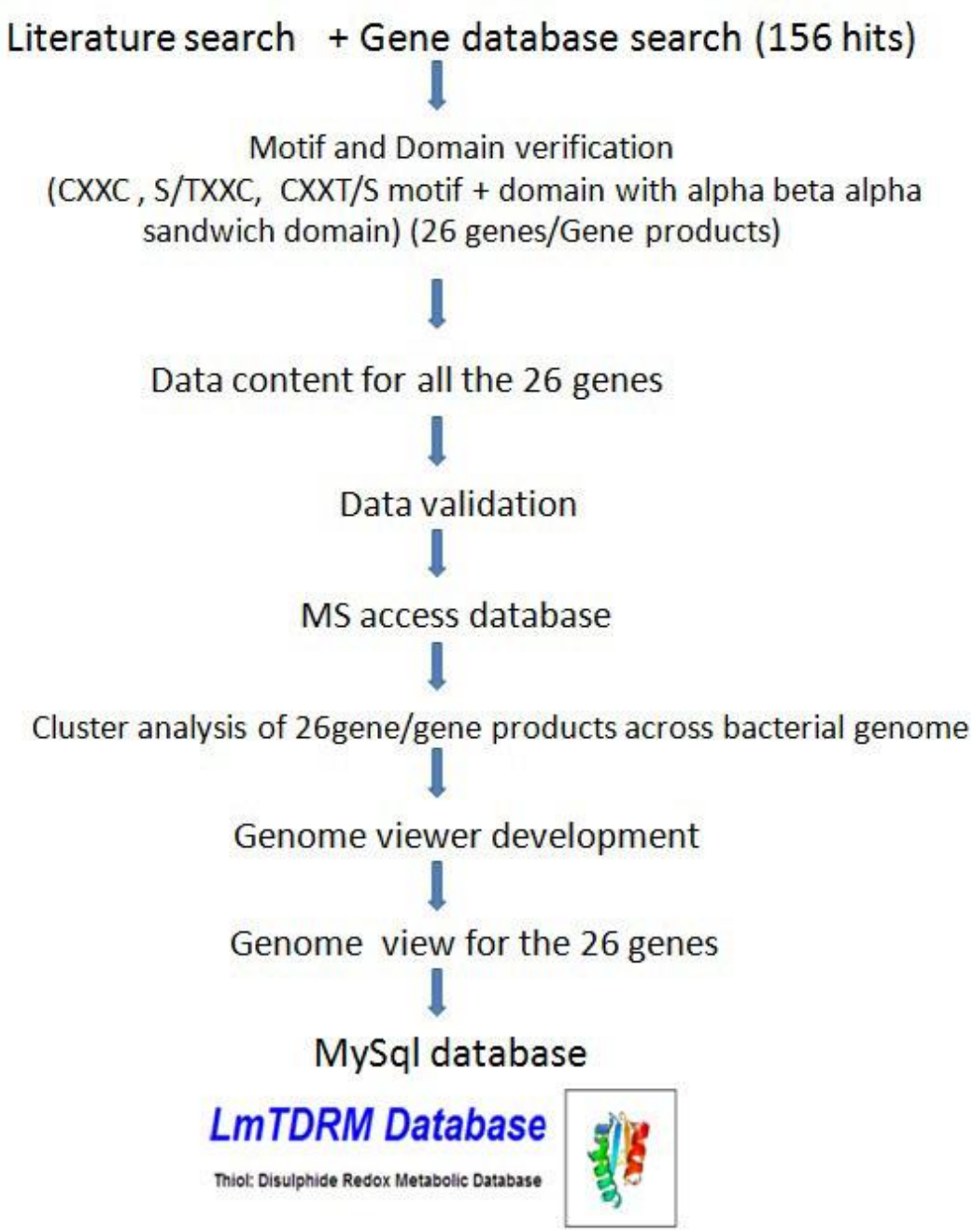

Figure 1: Workflow of the development of *LmTDRM* database

2.1.3 Development of cluster analysis

For understanding the presence or absence of 26 thiol genes across bacterial groups, cluster analysis was performed. All the thiol gene/gene products were aligned using BLAST programme. E-value with '0' other than *L.monocytogenes* EGDe was selected. All organisms were grouped based on bacterial groups. For each bacterial group, a bar graph was developed with different colours indicating presence and absence of the gene/gene products. The bar graph was plotted keeping strains of bacteria of particular group on X-axis and 26 gene/gene products on Y-axis. This type of cluster analysis was performed for 26 gene/gene products in presence and absence of GshF (Figure 2).

Table 1: Complete list of all 26 genes of TDRM network

Lmo	Protein Name	Motif	System
lmo1233	Thioredoxin	C-G-P-C	Thioredoxin
Lmo2478	Thioredoxin reductase	C-A-V-C	Thioredoxin
lmo1609	Thioredoxin family	C-G-D-C	Thioredoxin
lmo1903	Thioredoxin family	C-E-D-C	Thioredoxin
lmo2152	Thioredoxin family	C-P-N-C	Thioredoxin
lmo2424	Thioredoxin family	C-G-S-C	Thioredoxin
lmo2830	Thioredoxin family	C-A-P-C	Thioredoxin
lmo1583	Thiol peroxidase	T-S-V-C	Thioredoxin
lmo1604	2-cys-peroxiredoxin (= AhpC)	T-F-V-C	Thioredoxin
lmo1860	Methionine sulfoxide reductase	C-F-W-C	Thioredoxin
lmo0964	DsbA-like, unknown function	C-D-D-C	Thioredoxin
lmo2393	Thioredoxin family	C-A-S-C	Thioredoxin
lmo2426	ArsC family protein	C-S-T-C	Glutathione
lmo2770	Glutathione synthetase	**NA**	Glutathione
lmo0906	Glutathione reductase	**NA**	Glutathione
lmo1433	Glutathione reductase 2 (putative)	**NA**	Glutathione
lmo2344	Glutaredoxin	C-H-V-C	Glutathione
lmo0983	Glutathione peroxidase	C-G-L-T	Glutathione
lmo2155	Aerobic ribonucleotide reductase subunit A	**NA**	Thioredoxine-Glutathione-Glutaredoxin dependant
lmo2154	Aerobic ribonucleotide reductase subunit B	**NA**	Thioredoxine-Glutathione-Glutaredoxin dependant
lmo2153	Regulatory subunit (putative)	**NA**	Thioredoxine-Glutathione-Glutaredoxin dependant
lmo0279	Anaerobic ribonucleotide reductase subunit D	T-N-N-C	Thioredoxine-Glutathione-Glutaredoxin dependant
lmo0280	Anaerobic ribonucleotide reductase subunit G	T-A-T-C	Thioredoxine-Glutathione-Glutaredoxin dependant
lmo0222	Redox-switch chaperone (“holdase”)	**NA**	Thioredoxine-Glutathione-Glutaredoxin dependant
lmo2191	Redox/peroxide-sensitive regulator	C-T-S-C	Transcriptional regulators
lmo1059	Thiol:disulfide oxidoreductase (protein folding)	C-P-F-C	others

NA – Not applicable

Figure 2: ***LmTDRM database:*** **screenshot of the cluster analysis. Graph represents the presence of gene/gene products across the bacterial strains. X-axis represents bacterial strain and Y-axis represents thiol genes.**

2.1.4 Development of Genome viewer

Genome viewer was developed for all the 26 thiol gene/gene products. The genome view is for the transcriptional unit of the gene. Here, in this case, the gene of interest was taken as "0" position and scanned for the complete transcriptional unit. The upstream and downstream genes are marked using + and – positions. During this proces, tRNA were searched in the transcriptional unit and marked as pyramid. Details of the gene and tRNA were also captured. The details of the genes were curated from NCBI databases (www.ncbi.nlm.nih.gov) and uploaded to the database. Genes are represented as arrows and tRNA as pyramid. Genes from *L.monocytogenes* EGDe has been kept as reference organism and compared with other bacterial genomes. The colour codes of the genes are based on the gene ontology (Figure 3).

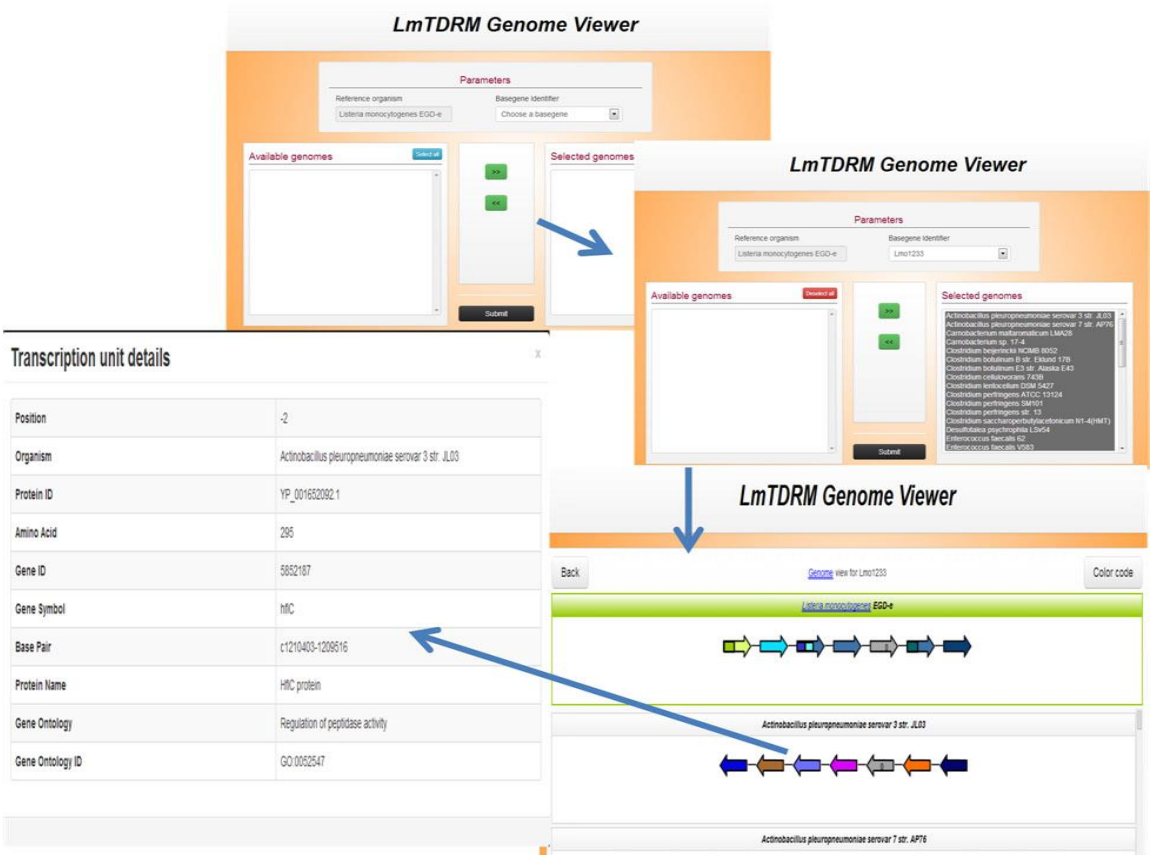

Figure 3: ***LmTDRM Database:*** **Screenshot of genome viewer with search page and graphical representation with organism details. Genes are represented as arrows and tRNA as pyramid. Color codes are based on Gene ontology. Details can be viewed in color code bar.**

LmTDRM database is first developed as a Microsoft Access® database containing 26 thiol genes involved in thiol redox metabolism pathway. After manual evaluation of the data sets, all the data were exported to Mysql database and a web-based graphical interface was developed to view the data.

The database is organized into 10 tables classified under three categories (Figure 4):

I. Gene Annotation/Search Interface

a. Structure_details: It contains structure details with structure id, gene id, active site type, fold type, PDB structure, and PDB structure details for the respective gene.

b. Gene_annotation: It contains gene annotation details with gene annotation id, gene id, protein id, GenBank id, sequence, function, motif, location number, and phylogenetic cluster for the 26 genes.

c. Prt_sequences: It contains protein sequence id, protein id, organism name, and protein sequence in FASTA format for the respective 26 gene products.

d. Complete_list: It contains a complete gene list with information on gene id, NCBI id, protein id, and the entire sequence of *Listeria monocytogenes* EGDe.

e. Domain_annotation: It contains domain annotation id, gene id, link to domain database, InterPro, and Pfam for the domain present in the protein. All the above tables mentioned are for searching gene, protein, and domain information for 26 TDRM genes/gene products.

II. **TDRM Analysis Interface**

a. **Cluster_with_bacteria**: It contains cluster id for all the 26 genes across bacterial genome and also with their properties like respiratory mechanism, gram staining, shape, pathogenicity, disease, mode of infection, intracellular/extracellular, host, URL, cluster, and hierarchy.

b. **Cluster_images**: Are pictorial representations of cluster_with_bacteria which contains cluster image id, cluster id, and cluster image for the respective gene. The tables above the graph explain the relationship for cluster analysis.

III. **Genome Viewer Interface**

a. **Structure**: It contains structure id, structure position, organism id, and gene id for the respective gene.

b. **Organism**: It contains organism id, organism name, and gene id for the respective gene.

c. **Genome**: It contains gene id and gene name for the respective gene. These tables explain the relationship of genome viewer for the complete transcriptional unit.

2.2 Implementation and architecture

The data from the Microsoft Access were imported to a Mysql database which is used by the web interface. The *LmTDRM* system is built in PHP (personal home page – server-sided scripting language) which employs smarty template engine as design pattern. Smarty (http://www.smarty.net/) is a template engine for PHP which facilitates the separation of presentation from application logic. Application logic is done using core PHP script. The publicly accessible *LmTDRM* web interface runs on an Apache 2.2.24 web server (http://httpd.apache.org/).

3 Results

3.1 Database querying

LmTDRM web page can be accessed using URL www.lmtdrm.com. The webpage consists of 6 tabs. They are:

I. **Home**

The first tab is about the home page, which explains the objective of the database.

II. **Search**

Search can be made from search page in the website. Here, information about gene, domain, and structure with respect to gene id or protein id can be retrieved. All the structures of thiol proteins were simulated in our laboratory and were validated by literature [23, 24]. To view the structural details, download the PDB file and open with any protein molecule visualization tool (Figure 5).

III. **TDRM analysis**

Here, we can view the clustered graphs and details of the graphs based on bacterial group. They are:

a. Thiol genes in presence of GshF (multidomain bifunctional glutathione synthase protein)

b. Thiol genes in absence of GshF

For example, if an organism is selected from the list, a new window pops up with the organism details, their properties, and the cluster analysis across bacterial genomes containing that particular gene (Figure 2).

IV. Genome viewer

It is a graphical interface that displays the information from a biological database for genomic data. Here, when a gene is selected from the dropdown, similar genes available in other bacterial genomes are listed. User can select all the bacterial genome or select few, based on the preference and use forward arrow >>. The genomes get populated on the right hand side. On click of submit button, the transcription unit can be viewed and analyzed. Details of the gene/tRNA can be obtained by clicking on the gene or tRNA (Figure 3).

V. Help

This tab explains the database for browsing purpose. Any comment or suggestion from the end user is tracked through a suggestion box which automatically directs to lmtdrm@gmail.com**.**

VI. References

This tab includes all the references used in understanding and analyzing the lmTDRM gene and gene products.

4 Discussion and Future enhancements

LmTDRM is the first comprehensive database on thiol disulphide redox metabolism proteins in *L. monocytogenes* EGDe. From data mining, it is evident that thiols play a major role in health care. Increased or decreased levels of thiol proteins in humans have been noted in various medical disorders like in chronic renal failure and other disorders related to kidney, cardiovascular disorders, stroke, and other neurological disorders; diabetes mellitus and alcoholic cirrhosis. Therapy using thiols has been under investigation for certain disorders. Hence, *LmTDRM* database facilitates the scientific community in understanding the thiol proteins in *L. monocytogenes* and their role in pathogenicity. It includes structure details, gene annotation, complete list, cluster analysis, cluster images, genome details, and genome viewer. All data are available using an open access user-friendly web interface and the data are helpful in understanding the genes in context to horizontal gene transfer of virulence genes and their mechanism of combating oxidative stress. Different queries can be made in the website with various display options. This new database is a first database on thiol proteins which allows fetching more information on gene annotation, structure details; to understand the oxidative stress mechanism and involvement of these gene products in host–pathogen interaction. This database is extended to include information on protein interaction network, host protein–bacterial protein interaction, network analysis, and docking studies to identify novel targets. Further enhancement would also include the changes in the host/bacterial genome due to infection, as the export module is kept simple at this point. Enhancements on data export modules using xml formats will be implemented.

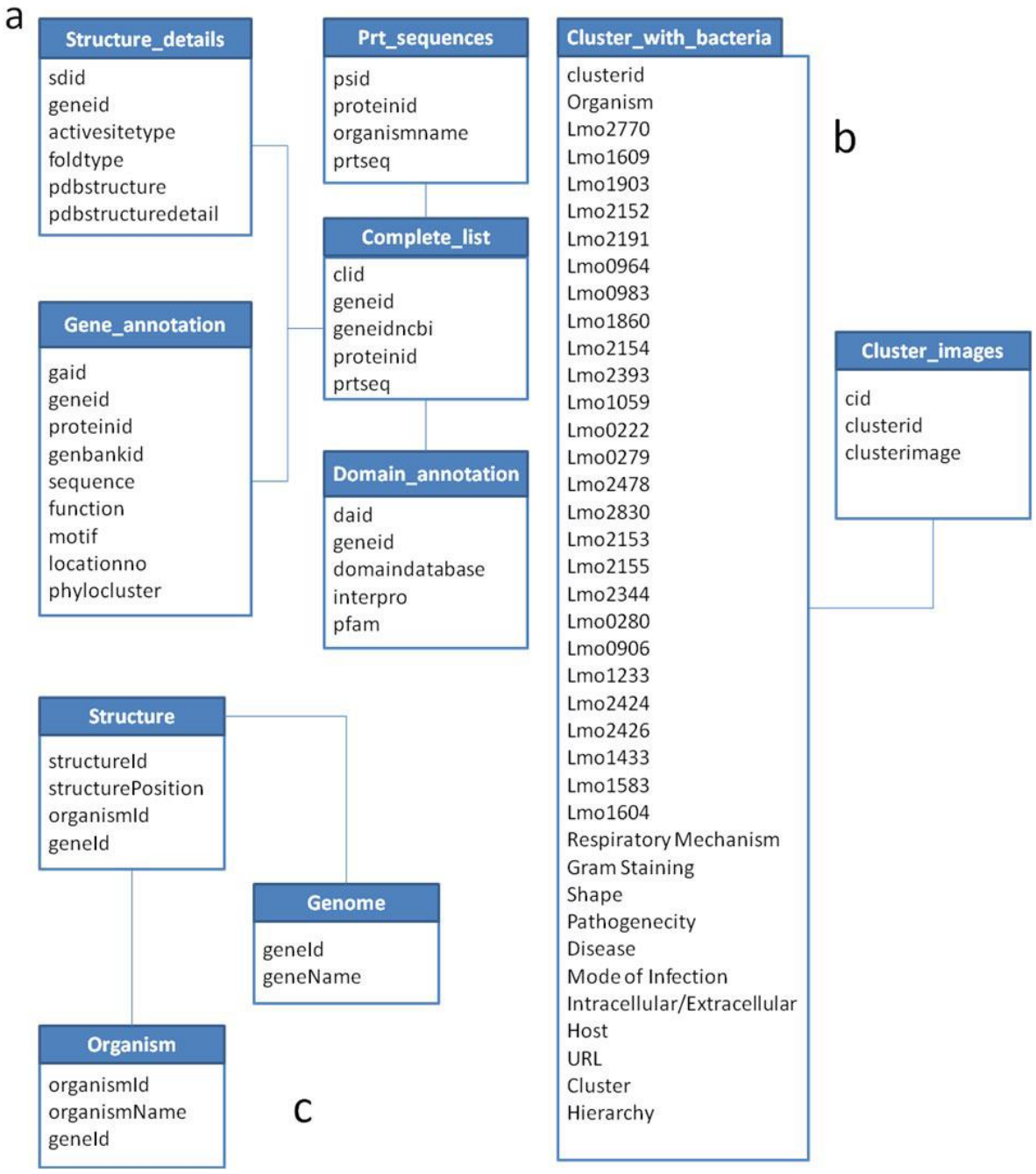

Figure 4: Schema of *LmTDRM* database. Schema has been developed in 3 parts. a. Relative tables for search. b. Relative tables for TDRM analysis. c. Relative tables for genome viewer.

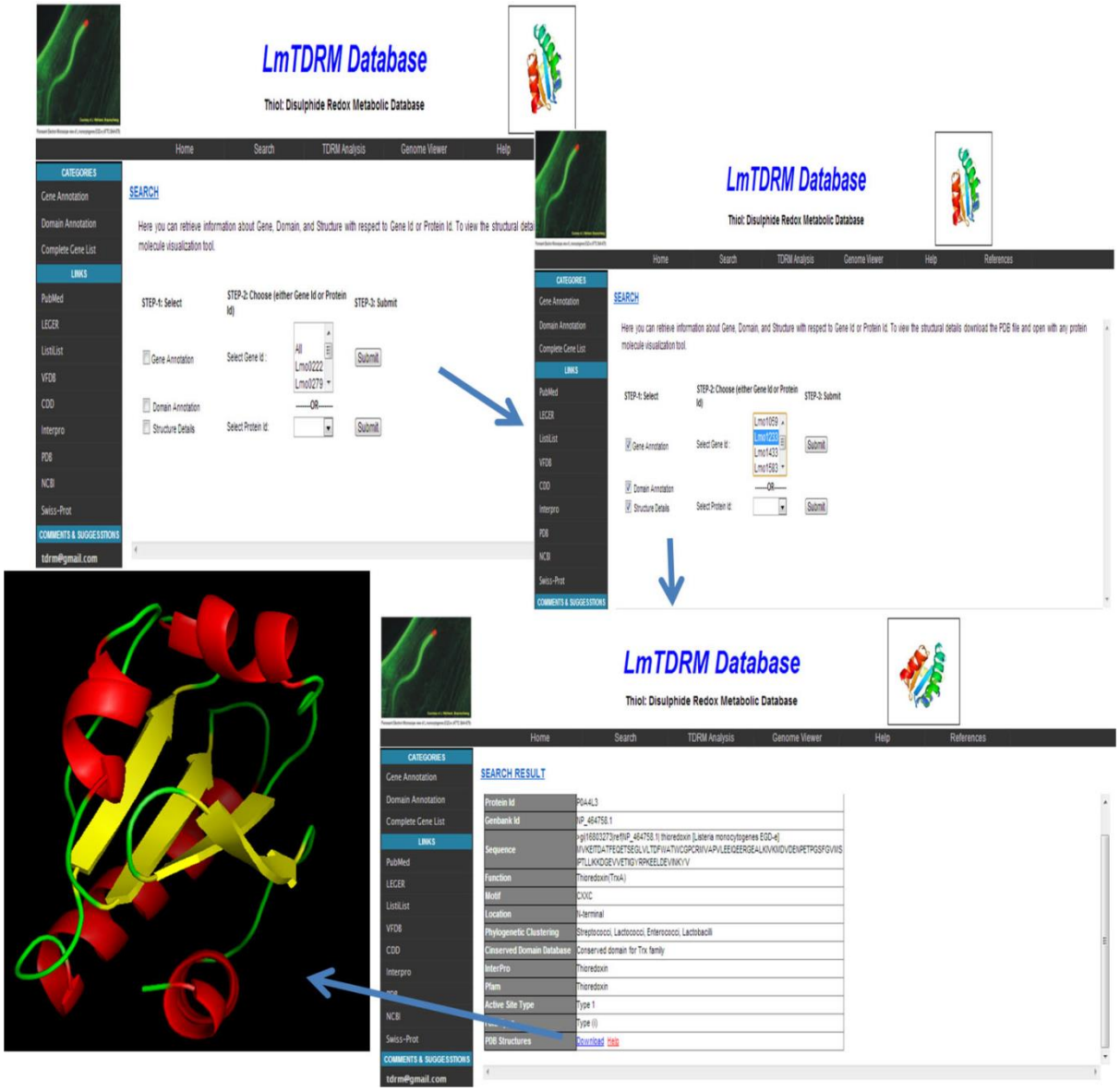

Figure 5: ***LmTDRM database:*** **Screenshot of search page with details and structure. Query results represent the complete information of the gene and structure details. PDB files can be downloaded and viewed using PYMOL.**

References

[1] E. Finkelstein, G.M. Rosen, and G.J. Rauckmann. Spin trapping of superoxide and hydroxyl radical: practical aspects. *Archives of Biochemistry and Biophysics*, 200:1–16, 1980.

[2] C.E. Presnell, G. Bhatti, L.S. Numan, M. Lerche, S.K. Alkhateeb, M. Ghalib, M. Shammaa and M. Kavida Computational insights into the role of glutathione in oxidative stress. *Current Neurovascular Research*, 10(2):185–194, 2013.

[3] F. Dénès, C.H. Schiesser, and P. Renaud. Thiols, thioethers, and related compounds as sources of C-centred radicals. *Chemical Society Reviews*, 42(19):7900-7942, 2013.

[4] M. Heindorf, M. Kadari, C. Heider, E. Skiebe, and G. Wilharm. Impact of *Acinetobacter baumannii* Superoxide Dismutase on Motility, Virulence, Oxidative Stress Resistance and Susceptibility to Antibiotics. *PLoS One*, 9(7):e101033, 2014.

[5] R. Rossi, D. Giustarini, A. Milzani et al. Cysteinylation and homocysteinylation of plasma protein thiols during ageing of healthy humans. *Journal of Cellular and Molecular Medicine*, 10:1582–4934, 2008.

[6] P. Ghezzi. Protein glutathionylation in health and disease. *Biochimica et Biophysica Acta*, 1830:3165–3172, 2013.

[7] P. Ghezzi. Role of glutathione in immunity and inflammation in the lung. *International Journal of General Medicine*, 25:105–113, 2011.

[8] Y. Meyer, B.B. Buchanan, F. Vignols and J.-P. Reichheld. Thioredoxins and glutaredoxins: unifying elements in redox biology. *Annual Review of Genetics*, 43:335–367, 2009.

[9] Y. Qi and N.V. Grishin. Structural classification of thioredoxin-like fold proteins. *Proteins*, 58:376–388, 2005.

[10] A. Holmgren. Thioredoxin and glutaredoxin systems. *Journal of Biological Chemistry*, 264:13963–13966, 1989.

[11] R.A. Kumar, A. Koc, R.L. Cerny et al. Reaction mechanism, evolutionary analysis, and role of zinc in *Drosophila* methionine-R-sulfoxide reductase. *Journal of Biological Chemistry*, 277:37527–37535, 2002.

[12] D.E. Fomenko and V.N. Gladyshev. CxxS: Fold independent redox motif revealed by genome-wide searches for thiol/disulfide oxidoreductase function. *Protein Science*, 11:2285–2296, 2002.

[13] J.S. Fetrow, N. Siew, J.A. Di Gennaro et al. Genomic-scale comparison of sequence- and structure-based methods of function prediction: does structure provide additional insight? *Protein Science*, 10:1005–1014, 2001.

[14] E.G. Pamer. Immune responses to *Listeria monocytogenes*. *Nature Reviews Immunology*, 4:812–823, 2004.

[15] R. Stachowiak, J. Wiśniewski, O. Osińska et al. Contribution of cysteine residue to the properties of *Listeria monocytogenes* listeriolysin O. *Canadian Journal of Microbiology*, 55:1153–1159, 2009.

[16] R.L. McCaffrey, P. Fawcett, M. O'Riordan et al. A specific gene expression program triggered by Gram-positive bacteria in the cytosol. *Proceedings of the National Academy of Sciences*, 101:11386–11391, 2004.

[17] P. Cossart and A. Toledo-Arana. *Listeria monocytogenes*, a unique model in infection biology: an overview. *Microbes and Infection*, 10(9):1041–1050, 2008.

[18] M. Hamon, H. Bierne, and P. Cossart. *Listeria monocytogenes*: a multifaceted model. *Nature Reviews Microbiology*, 4:423–434, 2006.

[19] M.A. Mraheil, A. Billion, W. Mohamed, D. Rawool, T. Hain, and T. Chakraborty. Adaptation of *Listeria monocytogenes* to oxidative and nitrosative stress in IFN-γ-activated macrophages. *International Journal of Medical Microbiology*, 301:547–555, 2011.

[20] L.E. Dons, A. Mosa, M.E. Rottenberg, J.T. Rosenkrantz, K. Kristensson, and J.E. Olsen. Role of the *Listeria monocytogenes* 2-Cys peroxiredoxin homologue in protection against oxidative and nitrosative stress and in virulence. *Pathogens and Disease*, 70:70–74, 2014.

[21] B. Xayarath and N.E. Freitag. Optimizing the balance between host and environmental survival skills: lessons learned from *Listeria monocytogenes*. *Future Microbiology*, 7:839–852, 2012.

[22] L.S. Nobre and L.M. Saraiva. Effect of combined oxidative and nitrosative stresses on *Staphylococcus aureus* transcriptome. *Applied Microbiology and Biotechnology*, 97(6):2563–2573. 2013.

[23] S. Gopal, V. Srinivas, F. Zameer, and J. Kreft. Prediction of proteins putatively involved in the thiol: disulfide redox metabolism of a bacterium (Listeria): the CXXC motif as query sequence. *Insilico Biology*, 9(5-6):407–414, 2009.

[24] V. Srinivas and S. Gopal. Computational analysis of thiol derived proteins in *Listeria monocytogenes* EGD-e. *International Conference on Chemical, Biological and Medical Sciences* (ICCBMS'2012) August 25–26, 2012 Kuala Lumpur (Malaysia), 2012.