

SALMONELLABASE - An online database of druggable targets of Salmonella species

Ophilia Benedict Pushpa & Muthaiyan Xavier Suresh*

Department of Bioinformatics, Sathyabama University, Jeppiaar Nagar, Rajiv Gandhi Salai, Chennai - 600 119; M Xavier Suresh - Email: xaviersuresh@gmail.com; *Corresponding author

Received July 18, 2012; Accepted July 30, 2012; Published August 03, 2012

Abstract:

Salmonellosis is one of the most common and widely distributed food borne diseases caused by Salmonella serovars. The emergence of multi drug resistant strains has become a threatening public health problem and targeting unique effectors of this pathogen can be considered as a powerful strategy for drug design. SalmonellaBase is an online web portal serving as an integrated source of information about Salmonella serovars with the data required for the structural and functional studies and the analysis of druggable targets in Salmonella. We have identified several target proteins, which helps in the pathogenicity of the organism and predicted their structures. The database will have the information on completely sequenced genomes of Salmonella species with the complete set of protein sequences of the respective strains, determined structures, predicted protein structures and biochemical pathways of the respective strains. In addition, we have provided information about name and source of the protein, Uniprot and Protein Data Bank codes and literature information. Furthermore, SalmonellaBase is linked to related databases and other resources. We have set up a web interface with different search and display options so that users have the ability to get the data in several ways. SalmonellaBase is a freely available database.

Availability: <http://www.salmonellabase.com/>

Keywords: SalmonellaBase, Database, Drug Design, Salmonella, Druggable Targets

Background:

Salmonellosis is one of the most common and widely distributed food borne diseases caused by Salmonella serovars. *Salmonella enterica serovar Typhi* is a human-specific pathogen causing enteric typhoid fever, a severe infection of the reticuloendothelial system. The early administration of antibiotic treatment has proven to be highly effective in eliminating infections, but indiscriminate use of antibiotics has led to the emergence of multidrug-resistant strains of *S. enterica* serovar Typhi [1-4]. The emergence of multi drug resistant strains has become a threatening public health problem. Salmonella is responsible for an estimated 3 billion human infections worldwide and killing 217,000 people every year [5]. Since typhoid is becoming difficult to treat with conventional drugs, information about the whole genome sequence and genes of *S. enterica* serovar Typhi will help to reveal more specific targets for drugs aimed at disease treatment and

vaccines for prevention. Targeting unique effectors of this pathogen can be considered as a powerful strategy for drug design against bacterial variations to drug resistance [6]. Over the past few decades the nucleotide and protein sequence data have been accumulated and available in number of public databases [7, 8]. However, such databases are broad in scope and there is a gap between the public databases and the small curated databases focusing on a particular organism or a type of data. Therefore we have developed SalmonellaBase, an online database with the data required for the structural and functional studies of *Salmonella serovars'* druggable targets.

Methodology:

The database consists of the records of nucleotide sequences, protein sequences, biological pathways, determined 3D structures, and the predicted druggable protein targets of 16 different strains of Salmonella species. They are organized to

simplify the task of finding relevant data for proteins in the related strains (Figure 1). Complete genome and proteome sequences data was collected from NCBI. The protein records include all the functionally assigned protein sequences as well as the hypothetical proteins; each record when accessed returns the primary sequence in fasta format. Databases such as Database of Essential Genes (DEG) [9] (<http://tubic.tju.edu.cn/deg>) and Kyoto Encyclopedia of Genes and Genomes (KEGG) [10] pathway database were used to identify putative drug targets for the individual strains respectively. SalmonellaBase was constructed with standard HTML and JavaScript. The database used PHP as the frontend and MySQL as the backend. The database is available online with user-friendly search methods and graphical browsers (Figure 2).

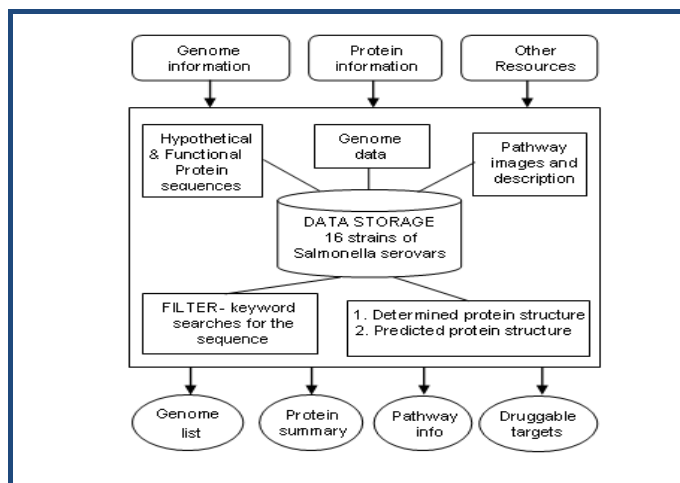


Figure 1: Organization of data in SalmonellaBase

3D Structures of druggable targets:

One of the main focuses of the database is to provide the structural information of the druggable targets found in the completely sequenced genomes of Salmonella strains. The database has a record of predicted structures of proteins which could serve as targets for novel drug design. The target proteins were identified by differential genomics approach and the structures are modeled using comparative and homology modeling approaches and are refined by energy minimization and validated by Ramachandran plot. Along with the predicted structures of protein targets, structure files of the experimentally determined proteins obtained from PDB are also curated. These structures were made available for download from respective pages in the database.

Features of the database:

The database consists of the collection of the primary sequences of all hypothetical proteins, functionally assigned proteins, complete genome sequences, biological pathways and the tertiary structure files of protein targets. The important feature of SalmonellaBase is the integrated availability of data. The completely sequenced genomes of over 16 strains of Salmonella serovars, are incorporated in this database Table 1 (see supplementary material). Moreover, each of the protein record shows, (i) The start position and end position of the protein in the genome; (ii) The length of the nucleotide sequence and the length of the amino acid sequence; (iii) Displays the primary protein sequence in fasta format; (iv) The protein summary page also has the hypothetical protein sequences of all the strains with their position in the genome; (v) The database also allows the users to search the protein of interest from the data by giving a keyword in the filter option; (vi) The biological pathway record includes all the pathway names with its image.

Accession	Start	End	Length	AA Length	Product
NC_011149	367	2799	2433	810	Aspartokinase / Homoserine dehydrogenase
NC_011148	1160	1363	204	67	haemolysin expression modulating protein
NC_011149	2801	3730	930	309	Homoserine kinase
NC_011148	3474	3668	195	64	micrococcal nuclease-like protein
NC_011149	3734	5020	1287	428	Threonine synthase
NC_011149	5113	5886	774	257	UPF0246 protein YaaA
NC_011148	5581	6366	786	261	zinc metalloproteinase Mtr protein
NC_011149	5965	7422	1458	485	Putative alanine/glycine transport protein
NC_011148	7270	7695	426	141	DNA-binding protein H-NS
NC_011149	7664	8617	954	317	Transaldolase

Figure 2: Homepage of SalmonellaBase and filter option in protein summary page.

Conclusion:

The three dimensional structural information of several undetermined proteins will give an in-depth understanding which in turn paves ways to novel drug design strategies. Screening of functional inhibitors against these targets may also result in discovery of novel therapeutic compounds that can be effective against multidrug resistant strains of Salmonella. The data on hypothetical proteins will be helpful to annotate the genes of interest and to predict functional regions.

Future development:

We plan to further refine and update this database with links to known drug related data in the near future.

Acknowledgement:

The bioinformatics computational facilities available at Department of Bioinformatics, Sathyabama University are greatly acknowledged. The authors also thank the database providers whose links have been used and the anonymous reviewers for their valuable comments and suggestions.

References:

- [1] Tajbakhsh M *et al.* *Folia Microbiol (Praha)*. 2012 **57**: 91 [PMID: 22307834]
- [2] Adhikary R & Joshi S, *Indian J Pathol Microbiol*. 2011 **54**: 849 [PMID: 22234135]
- [3] Sánchez-Vargas FM *et al.* *Travel Med Infect Dis*. 2011 **9**: 263 [PMID: 22118951]
- [4] Rowe B *et al.* *Clin Infect Dis*. 1997 **24**: S106 [PMID: 8994789]
- [5] Harish BN & Menezes GA, *Indian J Med Microbiol*. 2011 **29**: 223 [PMID: 21860101]
- [6] Chaudhuri RR *et al.* *PLoS Pathog*. 2009 **5**: e1000529 [PMID: 19649318]
- [7] <http://www.ncbi.nlm.nih.gov/>
- [8] Chaudhuri RR *et al.* *Nucleic Acids Res*. 2008 **36**: D543 [PMID: 17984072]
- [9] Zhang R *et al.* *Nucleic Acids Res*. 2004 **32**: D271 [PMID: 14681410]
- [10] Ogata H *et al.* *Nucleic Acids Res*. 1999 **27**: 29 [PMID: 9847135]

Edited by P Kanguane

Citation: Pushpa & Suresh, *Bioinformation* 8(15): 734-737 (2012)

License statement: This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited

Supplementary material:

Table 1: The details of data available in SalmonellaBase

S. No	Name of strains	No of functionally assigned proteins	No of hypothetical proteins	Biological pathways	Protein structures	
					Determined structures	Predicted structures
1	<i>Salmonella enterica</i> subspecies <i>enterica</i> serovar <i>Agona</i> str. <i>SL483</i>	3261	831	71	17	24
2	<i>Salmonella enterica</i> subspecies <i>enterica</i> serovar <i>arizonae</i> str. <i>62:Z4,Z23</i>	3034	796	72	12	24
3	<i>Salmonella enterica</i> subspecies <i>enterica</i> serovar <i>Choleraesuis</i> str. <i>SC-B67</i>	3530	826	76	21	24
4	<i>Salmonella enterica</i> subspecies <i>enterica</i> serovar <i>Dublin</i> str. <i>CT_02021853</i>	3231	853	103	2	24
5	<i>Salmonella enterica</i> subspecies <i>enterica</i> serovar <i>Enteritidis</i> str. <i>P125109</i>	3182	781	81	8	24
6	<i>Salmonella enterica</i> subspecies <i>enterica</i> serovar <i>Gallinarum</i> str. <i>287/91</i>	3771	812	93	8	24
7	<i>Salmonella enterica</i> subspecies <i>enterica</i> serovar <i>Heidelberg</i> str. <i>SL476</i>	3771	846	70	7	24
8	<i>Salmonella enterica</i> subspecies <i>enterica</i> serovar <i>Newport</i> str. <i>SL254</i>	3392	873	57	2	24
9	<i>Salmonella enterica</i> subspecies <i>enterica</i> serovar <i>paratyphi A</i> str. <i>AKU_12601</i>	3155	776	53	2	24
10	<i>Salmonella enterica</i> subspecies <i>enterica</i> serovar <i>paratyphi A</i> str. <i>ATCC9150</i>	3158	822	64	2	24
11	<i>Salmonella enterica</i> subspecies <i>enterica</i> serovar <i>paratyphi B</i> str. <i>SPB7</i>	3116	881	78	2	24
12	<i>Salmonella enterica</i> subspecies <i>enterica</i> serovar <i>paratyphi C</i> str. <i>RKS4594</i>	2598	807	64	2	24
13	<i>Salmonella enterica</i> subspecies <i>enterica</i> serovar <i>schwarzengrund</i> str. <i>CVM19633</i>	3104	833	90	4	24
14	<i>Salmonella enterica</i> subspecies <i>enterica</i> serovar <i>typhi</i> <i>CT18</i>	3547	863	57	5	24
15	<i>Salmonella enterica</i> subspecies <i>enterica</i> serovar <i>typhi</i> str. <i>TY2</i>	3421	819	60	30	24
16	<i>Salmonella enterica</i> subspecies <i>enterica</i> serovar <i>typhimurium</i> str. <i>LT2</i>	3563	785	22	20	24