Original Research Paper

Volume-8 | Issue-8 | August-2018 | PRINT ISSN No 2249-555X

Biological Science

A STUDY ON PROTEIN TARGET VALIDATION AND FUNCTIONAL ANNOTATION OF HYPOTHETICAL PROTEIN FOR *SHIGELLA DYSENTERIAE*

K. Shrivastava JNRM college portblair AN , MIG 30 Triveni complex, New Market Bhopal

A. Shrivastava* Govt holkar science college Indore, Plot No 301 A flat no 206 vimalnath apartment siliconcity Indore *Corresponding Author

ABSTRACT Significant developments have been made in the ground of biological and computer science in past few decades which can be utilized to develop biological data. The study was performed on Shigella dysenteriae (S.dysenteriae). Shigella has four types of species. S.dysenteriae is one of them which cause dysentery. It destroys immunity of Homo sapiens because the survival capacity affects all age groups and it most commonly spreads in new born child and travelers.

There are many drugs available in the market but they are not much effective, so we intended to identify the innovative drug target which is efficient and effective in curing the disease. The whole analysis is based on comparative study in S.dysenteriae (Query sequence) and homosapiens by online software to achieve the target.

Entire evaluation is based on few steps:-

A) Comparative study between S. dysenteriae and Homo sapiens protein sequence.

B) Functional annotation of the hypothetical protein, for which KEGG software was used, which can satisfy the process of validation of drug target and in retrieving protein sequences.

Major work is to establish the functional annotation of the hypothetical sequences from some conserved domain database, and for novel medicine, the high confidence limit standards are always considered.

KEYWORDS :

Introduction:

Shigella is a gram negative bacteria and it has four types of species viz. boydii, flexneri, dysenteriae and sonnei. This study is based on *Shigella dysenteriae* which is known to cause dysentery in humans. This bacterium is found in water and soil and causes shigellosis, which is an intestinal infection characterized by pain in abdomen and bloody diarrhea. In India, every year a significant number of neonates die from diarrhea, and large number of adults also suffer from diarrhea. Currently available drugs don't involve targeted approach and result in drug resistance, so the requirement to generate and indentify the new drug target arises, with which the novel drug can be designed. This novel drug shows no or less failure in clinical test.

This research is based on comparative study in S. dysenteriae with Homo sapiens genome through the Basic Local Alignment Search Tool (BLAST). After complete analysis, bacterial protein (query sequence) which is dissimilar to human protein is considered as potent drug target.

After this analysis, the functional annotation of the hypothetical protein is done. It means complete study of functionality of hypothetical proteins of S. dysenteriae by using some web tools. This study is done by comparative proteomics using the available database for conserved domain, motif and supper families, determining the enzymatic classification of the hypothetical protein and the functionality.

In this survey the most advance genomics and proteomics databases help to know about the molecular interaction, mechanism of metabolism and pathogencity of bacterial dysentery. All the data through Kyoto Encyclopedia of Gene and Genomes (KEGG) online software are retrieved and sorted out power of local alignment involved in BLASTp is used for searching the similarity of sequences between S. dysenteriae and Homo sapiens, and finding drug target to make a novel drug design.

Material and method

Collection of Protein Sequence:-

In this study, complete genome of S.dysenteriae and total 4616 protein sequences were obtained with the help of Kyoto Encyclopedia of gene and genome database (www.genome.jp/kegg) . The pseudo gene and gene codes from rRNA and tRNA were not considered in the analysis; hence we got 3553 protein sequences. These protein sequences were saved in fasta format in various files and all protein sequences worked as query sequence for this study.

Validation of protein target:-

Now a comparative analysis between 3553 amino acid sequences (query) of S.dysenteriae and Homo sapiens (taxid: 9606) was performed using BLASTp (the BLAST P algorithm applicable for protein-protein BLAST, using all non redundant protein databases). With the similarity search tool, we categorized the sequences into two

1) Dissimilar ones (which do not show similarity with human sequences)- these are Drug target, and are total 647 (KEGG ID) in number (detailed in Table no 1).
2) Similar sequences between two species- Remaining 2906 sequence. These are Non Drug Target.

Table No 1: Protein Drug Target (KEGG ID)

1	330	671	1029	1374	1743	2086	2475	2870	3276	3639	4079	4443
3	331	679	1032	1382	1744	2088	2478	2919	3279	3643	4096	4453
6	336	682	1035	1389	1750	2092	2497	2925	3286	3646	4103	4454
7	341	694	1037	1393	1751	2094	2516	2934	3287	3655	4105	4484
10	346	698	1043	1397	1752	2097	2532	2937	3289	3680	4107	4492
15	347	708	1046	1409	1754	2101	2533	2940	3290	3681	4112	4494
18	352	709	1050	1411	1755	2102	2539	2941	3296	3684	4114	4502
21	353	717	1052	1413	1757	2152	2546	2953	3298	3724	4118	4506
23	367	719	1053	1417	1772	2160	2548	2957	3299	3730	4129	4527
29	369	735	1065	1470	1773	2174	2550	2959	3301	3733	4140	4529
32	371	739	1067	1477	1821	2183	2551	2962	3302	3747	4146	4535
42	379	752	1082	1504	1822	2187	2558	2970	3304	3765	4170	4553
44	402	753	1133	1505	1826	2191	2576	2981	3308	3766	4171	4557
45	422	757	1137	1511	1833	2200	2578	2982	3314	3770	4174	4560
46	423	802	1144	1513	1839	2209	2607	2984	3318	3778	4187	4570
47	425	803	1149	1515	1844	2225	2612	2991	3323	3782	4192	4581
53	426	804	1154	1522	1848	2241	2616	2992	3329	3789	4197	4582
62	431	826	1159	1538	1852	2243	2619	2993	3333	3796	4204	4583
72	436	828	1171	1541	1853	2246	2633	2995	3338	3798	4208	4589
73	437	833	1182	1545	1854	2252	2642	2999	3339	3801	4212	4591
79	462	837	1183	1546	1872	2253	2644	3009	3345	3803	4216	4596
86	468	842	1200	1560	1875	2260	2653	3013	3348	3804	4218	4607
97	476	860	1205	1563	1878	2262	2665	3026	3357	3808	4231	4616
103	490	861	1209	1573	1888	2265	2671	3041	3359	3812	4233	
106	497	867	1215	1584	1896	2274	2672	3042	3361	3815	4236	
107	498	879	1217	1592	1904	2277	2674	3045	3366	3819	4242	
110	500	880	1222	1600	1907	2281	2675	3050	3397	3825	4244	
115	503	887	1231	1641	1923	2282	2685	3092	3417	3827	4246	
118	509	889	1235	1650	1924	2328	2696	3093	3445	3840	4249	
121	512	901	1237	1652	1927	2334	2697	3098	3456	3844	4253	
126	514	905	1241	1653	1929	2345	2699	3099	3464	3853	4254	

130	516	907	1255	1659	1931	2352	2702	3105	3466	3877	4265	
132	530	919	1298	1665	1933	2353	2705	3113	3468	3888	4268	
149	534	927	1300	1666	1937	2355	2712	3131	3469	3895	4273	
155	544	962	1303	1669	1979	2358	2720	3144	3476	3905	4283	
156	549	963	1306	1670	1980	2365	2751	3147	3478	3926	4339	
168	553	968	1313	1676	1993	2384	2764	3152	3482	3927	4340	
175	556	969	1314	1681	2000	2386	2765	3158	3498	3934	4346	
187	557	973	1316	1685	2004	2389	2768	3159	3504	3935	4352	
203	562	976	1319	1686	2010	2413	2769	3161	3526	3937	4362	
207	569	980	1320	1693	2014	2416	2777	3181	3571	3963	4363	
208	578	983	1322	1696	2022	2430	2781	3187	3575	3964	4375	
217	579	991	1330	1697	2027	2439	2782	3190	3583	3965	4380	
236	583	992	1337	1703	2037	2442	2791	3197	3587	3967	4391	
238	588	997	1342	1715	2038	2444	2794	3203	3589	3979	4392	
258	641	998	1343	1716	2043	2447	2795	3238	3595	3981	4395	
263	647	1007	1348	1722	2048	2457	2796	3242	3602	4024	4401	
266	650	1008	1349	1726	2056	2467	2803	3249	3606	4030	4405	
267	651	1011	1359	1728	2064	2470	2812	3250	3620	4033	4408	
272	655	1012	1362	1729	2067	2472	2849	3251	3624	4047	4413	
279	656	1014	1365	1734	2075	2473	2862	3273	3631	4052	4436	
323	660	1016	1370	1740	2080	2474	2865	3275	3638	4054	4437	

Retrieval of Hypothetical Proteins

KEGG database was used again for retrieval of hypothetical protein of S.dysenteriae and a search limit of "sdy" type sequence was used, so eventually, 1004 hypothetical protein sequence were obtained, out of which randomly selected 424 hypothetical proteins were considered for functional annotation.

Functional Annotation

On the basis of presence of enzymatic conserved domain, we performed functional annotation on selected 500 hypothetical protein sequences. With all default parameter settings, four different web tools were used for this purpose-

- CDD BLAST (http://www.ncbi.nlm.nih.gov/BLAST/) [6,11,17]
- Interproscan (http//www.abi.ac.uk/interpro)[12]
- Pfam (https://pfam.xfam.org/)
- COGs (https://www.ncbi.nlm.nih.gov/COG/) [8]

The result of functional annotation analysis is expressed in confidence limits (in percentage), which depends on the extent of similarity between each query sequence and human sequences, as determined by these tools. Specific Criteria is represented in Table No 1 to evaluate the confidence value.

Table No.2:- Confidence value in percentage where 1 = similarity, 0=Dissimilarity

CDD Blast	Interproscan	PFAM	COGs	Confidence value in percentage
1	1	1	1	100%
1	1	1	0	75%
0	1	1	1	
1	0	1	1	
1	1	0	1	
1	1	0	0	50%
1	0	0	1	
1	0	1	0	
0	1	0	1	
0	0	1	1	
0	1	1	0	
1	0	0	0	25%
0	1	0	0	
0	0	1	0	
0	0	0	1	
0	0	0	0	0%

Result and discussion:

From the comparative study of between S. dysenteriae and homo sapience, we found total 647 proteins as a Drug targets represented in table no 1(KEGG ID) these are not homologous with each other. These drug targets were act as a novel drug for the Shigellosis.

On the other hand we performed functional annotation on randomly selected 424 hypothetical proteins with the help of conserved domain databases (CDD-BLAST, INTERPROSCAN, PFAM and COGs) and the ratio of the functional annotation to the different confidence limits represented on table No 3

Table no 3: functional percentage of classification

Functional classification	
Similarity percentage	Count of protein
100%	31
75%	61
50%	43
25%	101
0%	188

Table No. 4 Enzymatic Classification

Sno.	KEGG No.	NCBI Blast	Interproscan	PFAM	COG	%
1	sdy:SDY_0040	Polysaccharide deacetylase	Polysaccharide deacetylase	Polysaccharide deacetylase	Polysaccharide deacetylase	100%
2	sdy:SDY_0042	Transposase_31, YhgA	Transposase_31,YhgA	Transposase_31,YhgA	NO	75%
3	sdy:SDY_0163	2',5' RNA ligase biogenesis	2',5' RNA ligase biogenesis	2',5' RNA ligase biogenesis	2',5' RNA ligase biogenesis	100%
4	sdy:SDY_0205	Glyoxalase	Glyoxalase	Glyoxalase	Lactoylglutathione lyase	75%
5	sdy:SDY_0228	Exo_endo_phos, family	Exo_endo_phos, family	Exo_endo_phos, family	Uncharacterized BCR	75%
6	sdy:SDY_0232	NADB_Rossmann superfamily	Methyltransferase	Methyltransferase	SAM-dependent methyltransferases	75%
7	sdy:SDY_0243	Transposase, Mutator	Transposase, Mutator	Transposase, Mutator	,Predicted transposase	100%
8	sdy:SDY_0245	Carbon-nitrogen hydrolase	Carbon-nitrogen hydrolase	Carbon-nitrogen hydrolase	Predicted amidohydrolase	75%
9	sdy:SDY_0250	Ykud domain	Ykud domain	Ykud domain	Uncharacterized BCR	75%
10	sdy:SDY_0287	4HBT	4HBT	4HBT	Predicted thioesterase	100%
11	sdy:SDY_0306	GATase1_DJ-1	DJ-1/PfpI family	DJ-1/PfpI family	Putative intracellular protease/amidase	100%
12	sdy:SDY_0380	ftrA	Helix-turn-helix, AraC type	Helix-turn-helix, AraC type	AraC-type	75%
13	sdy:SDY_0391	ATase	MethylDNA_cys_mtrans_DNA_bd	MethylDNA_cys_mtrans_DNA_bd	MethylDNA_cys_mtrans_DNA_bd	75%
14	sdy:SDY_0393	EAL	EAL	EAL	EAL	100%
15	sdy:SDY_0437	YbaK	, YbaK	YbaK	Uncharacterized ACR	75%
16	sdy:SDY_0494	GCS2,	GCS2,	GCS2,	Uncharacterized BCR	75%
17	sdy:SDY_0497	Pectin lyase fold	Pectin lyase fold	Pectin lyase fold	NO	75%

18	sdy:SDY_0528	PaaI_thioesterase	4HBT	4HBT	aromatic compounds catabolism	75%
19	sdy:SDY_0533	ParB-like nuclease domain	ParB-like nuclease domain	ParB-like nuclease domain	Predicted transcriptional regulators	75%
20	sdy:SDY_0537	PAPS_reductase	PAPS_reductase	PP-loop family	PAPS_reductase	75%
21	sdy:SDY_0558	SPOUT_MTase	SPOUT_MTase	SPOUT_MTase	Uncharacterized ACR	75%
22	sdy:SDY_0618	OprD family	OprD family	OprD family	NO	75%
23	sdy:SDY_0626	Esterases and lipases	Abhydrolase_1	Abhydrolase_1	Abhydrolase_1	75%
24	sdy:SDY_0639	Transposase_11	Transposase_11	Transposase_11	NO	75%
25	sdy:SDY_0748	Asp-Al_Ex	Asp-Al_Ex	Asp-Al_Ex	Asp-Al_Ex	100%
26	sdy:SDY_0752	Hydrolases-3	Hydrolases-3	Hydrolases-3	Hydrolases-3	100%
27	sdy:SDY_0765	Hydrolases-3	Hydrolases-3	Hydrolases-3	Hydrolases-3	100%
28	sdy:SDY_0768	Mu_DNA_bind	Putativ_DNA_bind	Mu_DNA_bind	NO	75%
29	sdy:SDY_0777	YkuD	YkuD	YkuD	Uncharacterized BCR	75%
30	sdy:SDY_0799	Glycos_trans_3N	Glycos_trans_3N	Glycos_trans_3N	Glycos_trans_3N	100%
31	sdy:SDY_0813	Exo_endo_phos	Exo_endo_phos	Exo_endo_phos	NO	75%
32	sdy:SDY_0837	GSDH,	GSDH,	GSDH,	GSDH,	100%
33	sdy:SDY_0903	EAL domain	EAL domain	EAL domain	EAL domain	100%
34	sdy:SDY_0905	PAP2_like	PAP2_like	PAP2_like	phospholipid phosphatase	100%
35	sdy:SDY_0906	P-loop NTPase	CobW	CobW	G3E family	100%
36	sdy:SDY_0940	NADB_Rossmann,	CoA binding	CoA binding	CoA binding	75%
37	sdy:SDY_0942	SAM methyltransferases	PUA	SAM methyltransferases	SAM methyltransferases	75%
38	sdy:SDY_0943	Acylphosphatase	Acylphosphatase	Acylphosphatase	Acylphosphatase	100%
39	sdy:SDY_0982	Flavin reductase	Flavin reductase	Flavin reductase	flavoprotein oxygenases	100%
40	sdy:SDY_0983	Nitroreductase family	Nitroreductase family	Nitroreductase family	Nitroreductase family	100%
41	sdy:SDY_1038	HIUase/Transthyretin family	HIUase/Transthyretin family	HIUase/Transthyretin family	HIUase/Transthyretin family	100%
42	sdy:SDY_1043	HD domain	HD domain	HD domain	HD domain	100%
43	sdy:SDY_1044	HD domain	HD domain	HD domain	HD domain	100%
44	sdy:SDY_1074	LysM	unintegrated	LysM	LysM	75%
45	sdy:SDY_1129	CN_hydrolase	CN_hydrolase	CN_hydrolase	CN_hydrolase	100%
46	sdy:SDY_1140	SIS	SIS	SIS	NO	75%
47	sdy:SDY_1143	Peptidase_M23	Peptidase_M23	Peptidase_M23	metalloendopeptidases	100%
48	sdy:SDY_1154	Cysteine hydrolases	Isochorismatase(HYDROLASE)	Isochorismatas (HYDROLASE)	nicotinamidase	75%
49	sdy:SDY_1169	SAM-methyltransferases	Methyltransf-12	Methyltransf-12	SAM-methyltransferases	100%
50	sdy:SDY_1210	PPC	PPC	PPC	NO	75%
51	sdy:SDY_1410	MS_channel	MS_channel	MS_channel	MS_channel	100%
52	sdy:SDY_1427	DNA_BRE_C	Phage integrase	Phage integrase	Phage integrase	75%
53	sdy:SDY_1477	NO	MFS_gen_substrate_transporter	MFS_gen_substrate_transporter	Permeases of the major facilitator	75%
54	sdy:SDY_1485	MipA	MipA	MipA	NO	75%
55	sdy:SDY_1487	Aldose_epim	Aldose_epim	Aldose_epim	Aldose_epim	100%
56	sdy:SDY_1496	Aldo/keto reductase	Aldo/keto reductase	Aldo/keto reductase	oxidoreductases	100%
57	sdy:SDY_1497	Aldo/keto reductase	Aldo/keto reductase	Aldo/keto reductase	oxidoreductases	100%
58	sdy:SDY_1508	Nitroreductase	Nitroreductase	Nitroreductase	Nitroreductase	100%
59	sdy:SDY_1523	CMD	CMD	CMD	Uncharacterized ACR	75%
60	sdy:SDY_1636	Amino acid permease	Amino acid permease	Amino acid permease	Amino acid permease	100%
61	sdy:SDY_1661	Sulfatase	Sulfatase	Sulfatase	Arylsulfatase A	100%
62	sdy:SDY_1674	PaaI_thioesterase	4HBT	4HBT	aromatic compounds catabolism	75%
63	sdy:SDY_1685	YbaK	YbaK	YbaK	Uncharacterized ACR	75%
64	sdy:SDY_1711	ROK	Glycoprotease	Glycoprotease	metal-dependent proteases	75%
65	sdy:SDY_1719	Nudix hydrolase	Nudix hydrolase	Nudix hydrolase	NTP pyrophosphohydrolases	75%
66	sdy:SDY_1802	EAL	EAL	EAL	EAL	100%
67	sdy:SDY_1818	Protein Kinases	Fructosamine-kinase	Fructosamine-kinase	Fructosamine-3-kinase	75%
68	sdy:SDY_1821	metal-dependent hydrolase	metal-dependent hydrolase	metal-dependent hydrolase	metal-dependent hydrolase	100%
69	sdy:SDY_1871	Aldo/keto reductase	Aldo/keto reductase	Aldo/keto reductase	Aldo/keto reductase	100%
70	sdy:SDY_1898	Ni_hydr_CYTB	Ni_hydr_CYTB	Ni_hydr_CYTB	NO	75%

71	sdy:SDY_1909	YkuD	YkuD	YkuD	Uncharacterized BCR	75%
72	sdy:SDY_1911	Cysteine desulfurase	Cysteine desulfurase	Aminotransferase	Selenocysteine lyase	75%
73	sdy:SDY_1922	DEAD-like helicases	DEAD-like helicases	Type III restriction enzyme	DNA and RNA helicases	75%
74	sdy:SDY_1924	Cro	Cro	Cro	NO	75%
75	sdy:SDY_1961	EAL	EAL	EAL	EAL	100%

Table no 4 represented enzymatics classification of total 75 hypothetical proteins which are not less than 75%. Confidence value, these protein sequences shows involvement in lifecycle and survival of Shigella dysenteriae in future and can be the future drug targets for controlling Shigellosis.

Conclusion

This study helped to search functionality in the hypothetical proteins of Shigella dysenteriae using the power of pair-wise alignment tool BLAST, which suggests that many uncharacterized proteins are available in the Shigella dysenteriae and their exact role in bacterial lifecycle still remains unclear. And these show some other unknown function of the hypothetical protein carries valuable information related to the biological activity of the S.dysenteriae which is very helpful to control bacterial Disease shigellosis.

References

1. Ewing, W. H., R. W. Reavis, and B. R. Davis. Provisional Shigella serotypes. Can. J. Microbiol. 4:89-107. 1958.
2. Clerc, P. L., A. Ryter, J. Mounier, and P. J. Sansonetti. Plasmid mediated early killing of eukaryotic cells by Shigella flexneri as studied by infection of J774 macrophages. Infect. Immun. 55:521–527. 1987.
3. Ronsmans, C., M. L. Bennish, and T. Wierzba. Diagnosis and management of dysentery by community health workers. Lancet ii: 552–555. 1988.
4. Altschul, S.F., Gish,W., Miller, W., Myers, E.W. and Lipman, D.J. Basic local alignment search tool. J. Mol. Biol., 215, 403–410. (1990).
5. Salyers, A. A., and D. D. Whitt (ed.)., Dysentery caused by Shigella species,. In Bacterial pathogenesis: a molecular approach. American Society for Microbiology, Washington, D.C, p. 169–181,1994.
6. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25 (17): 3389-402.
7. Edward E, et al. Biological function made crystal clear- annotation of hypothetical proteins via structural genomics. Current Opinion in Biotechnology 11: 25-30. (2000).
8. Roman L, et al, the COG database: a tool for genome –scale analysis of protein functions and evolution. Nucleic Acid Research. 28:33-36.(2000).
9. Wendy B, et al., The EMBL Nucleotide Sequence Database. Nucleic Acid Research. 28: 19-23. (2000).
10. Fernandez-Prada, C. M., et al Shigella flexneri IpaH(7.8) facilitates escape of virulent bacteria from the endocytic vacuoles of mouse and human macrophages. Infect. Immun. 68:3608–3619. 2000.
11. Schaffer AA, Aravind L, Madden TL, Shavirin S, Spouge JL, Wolf YI, Koonin EV, Altschul SF (2001). Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements. Nucleic Acids Res. 29(14):2994-3005.
12. Zdobnov EM, Rolf A Interproscan- an integration platform for the signatures recognition methods in InterPro. Bioinformatics 17: 847-848. (2001).
13. Shah M. Faruque, et al, Comparative Molecular Analysis of Environmental Shigella Isolates versus Clinical Strains, APPLIED AND ENVIRONMENTAL MICROBIOLOGY, p. 3908–3913, Aug. 2002.
14. James F. Koterski, Virulent Shigella flexneri Causes Damage to Mitochondria and Triggers Necrosis in Infected Human Monocyte-Derived Macrophages, INFECTION AND IMMUNITY, 73(1) p. 504–513 Jan. 2005.
15. Yang F et al. "Genome dynamics and diversity of Shigella species, the etiologic agents of bacillary dysentery". Nucleic Acids Res 33 (19): 6445–58. (2005).
16. Silveira NJ, et al"Molecular Models of protein targets from Mycobacterium tuberculosis", J.Mol Model, 11(2):160-6, Mar 2005.
17. Marchler-Bauer A et al. ,"CDD: a conserved domain database for interactive domain family analysis." Nucleic Acids Res. 35(D): 237-240. (2007),
18. Yakugaku Zasshi. Okuno Y. In silico drug discovery based on the integration of bioinformatics and chemoinformatics; 128(11):1645-51. 2008 Nov.
19. S. Anandkumar, et al., Computational annotation for hypothetical proteins of Mycobacterium tuberculosis, JCSB, vol.1:050-062 (2008).
20. Gross RJ, Rowe B, Cheasty T, and Thomas LV. Increase in drug resistance among Shigella dysenteriae, Sh flexneri, and Sh boydii. Br Med J (Clin Res Ed). 1981 Aug 29;283(6291):575–576.
21. R J Gross, E J Threlfall, L R Ward, and B Rowe. Drug resistance in Shigella dysenteriae, S flexneri and S boydii in England and Wales: increasing incidence of resistance to trimethoprim. Br Med J (Clin Res Ed). 1984 March 10; 288(6419): 784–786.
22. Yang F et al .Genome dynamics and diversity of Shigella species, the etiologic agents of bacillary dysentery. Nucleic Acids Res. 2005 Nov 7; 33(19):6445-58.
23. Seema Namboori, Natasha Mhatre, Sentivel Sujatha , Narayanaswamy Srinivasan and Shashi Bhushan Pandit. Enhanced functional and structural domain assignments using remote similarity detection procedures for proteins encoded in the genome of Mycobacterium tuberculosis H37Rv. Volume 29, Number 3 / September, 2004 245-259.
24. Jin Q et al Genome sequence of Shigella flexneri 2a: insights into pathogenicity through comparison with genomes of Escherichia coli K12 and O157. Nucleic Acids Res. 2002 Oct 15; 30(20):4432-41.