Bioinformation by Biomedical Informatics Publishing Group *open access*

www.bioinformation.net **Hypothesis**

Comparative sequence analysis of acid sensitive/resistance proteins in *Escherichia coli* and *Shigella flexneri*

Selvaraj Manikandan[1, $], Seetharaaman Balaji [2, $], Anil Kumar[1] and Rita Kumar[1, *]

[1] Institute of Genomics and Integrative Biology, Mall Road, Delhi - 110007, India; [2] Department of Biotechnology, Manipal Institute of Technology, Manipal University, Manipal - 576104, India; [$] Both the authors contributed equally
Rita Kumar* - E-mail: rita@igib.res.in; Phone: 91 11 27662133; Fax: 91 11 27667471; * Corresponding author

received September 17, 2007; revised October 25, 2007; accepted November 22, 2007; published online December 11, 2007

Abstract:
The molecular basis for the survival of bacteria under extreme conditions in which growth is inhibited is a question of great current interest. A preliminary study was carried out to determine residue pattern conservation among the antiporters of enteric bacteria, responsible for extreme acid sensitivity especially in *Escherichia coli* and *Shigella flexneri.* Here we found the molecular evidence that proved the relationship between *E.coli* and *S.flexneri.* Multiple sequence alignment of the *gadC* coded acid sensitive antiporter showed many conserved residue patterns at regular intervals at the N-terminal region. It was observed that as the alignment approaches towards the C-terminal, the number of conserved residues decreases, indicating that the N-terminal region of this protein has much active role when compared to the carboxyl terminal. The motif, FHLVFFLLLGG, is well conserved within the entire *gadC* coded protein at the amino terminal. The motif is also partially conserved among other antiporters (which are not coded by *gadC*) but involved in acid sensitive/resistance mechanism. Phylogenetic cluster analysis proves the relationship of *Escherichia coli* and *Shigella flexneri*. The *gadC* coded proteins are converged as a clade and diverged from other antiporters belongs to the amino acid-polyamine-organocation (APC) superfamily.

Keywords: amino acid -polyamine-organocation (APC); Glutamate decarboxylase (GadC); bacteria; sequence; proteins

Background:
Microbes are not always boned to have the favorable condition for their survival. So as to tackle the unfavorable conditions, they adopt certain mechanisms to overcome it. All enteric pathogens are required to bypass the acidic environment of stomach before infecting the intestinal mucosa, where luminal pH approaches neutrality. **[1]** Enteric micro-organisms have developed several inducible mechanisms for surviving transient periods of extreme acid stress. **[2]** Though such acid resistance mechanism is found in *Enterobacteriaceae* family but it is not the characteristic feature of all microbes of the family. *Escherichia coli* and *Shigella flexneri* have been reported to possess the acid resistance mechanism **[3]** where gadA and gadB genes code for the isoforms of glutamate decarboxylase (GAD). The gad system is based on the coordinated action of these two homologues of glutamate decarboxylase and of a specific glutamate/gamma-aminobutyrate antiporter (GadC) **[4]**, in which glutamate is internalized and converted to γ-aminobutyrate (consuming an intracellular proton) that is subsequently exchanged for another extracellular glutamate via a membrane-located antiporter. **[5]** Gale and Epps **[6]** as well as others **[7, 8]** demonstrated that there are a variety of decarboxylases that respond to low pH. The putative glutamate/GABA antiporter which is encoded by the gadC gene is responsible for importing the glutamate inside the cell and simultaneously exporting the GABA to the acidic environment. This helps for neutralization and survival in the acidic environment. The acid sensitivity inner membrane antiporter protein plays a pivotal role in the acid resistance indirectly, it is also found that mutation in the inner membrane antiporter protein makes the organism acid sensitive as neither intake of glutamate nor export of GABA takes place, which pave an acidic environment where the microbes will undergo death phase. This specific GABA antiporter belongs to the amino acid -polyamine-organocation (APC) super family. Gad A, B, C, hde AB, all are essential for the expression of acid resistance strains and mutations in any of these regions may block glu-dependent systems. **[9]** These genes encode a glutamate-dependent acid resistance mechanism that is optimally active under conditions in which it is needed to maintain viability. **[10]**

Present study deals with gadC encoded inner membrane antiporter due to its importance in transporting glutamate across inner membrane through *gadC* and making favourable environment for surviving in extreme condition. **[9]** Here we tried to decipher, is there any evidence hidden in the antiporter protein, of *Escherichia coli* K-12, O157:H7 and *Shigella flexneri*? Because this mechanism was not found in any other *Enterobacteriaceae.* We suspect that there must be some sequence conservation which was not detected in other enterobacteriaceae. GadC of *Listeria monocytogenes* has a motif FHLVFFLLLGG that corresponds to the *Shigella flexneri* GadC FSLVFFLLLGG and is considered to play an important role in the recognition of the glutamate. **[5]** Here we also address this pattern in the rest of the gadC coded acid sensitive/resistance proteins and other antiporters of APC super family as well.

Methodology:

The key word 'gamma aminobutyrate antiporter' yielded 290 hits of protein sequences from GenBank [URL http://www.ncbi.nlm.nih.gov/Genbank/]; Synonyms to gadC-XasA coded antiporter proteins were also retrieved. **[11]** From the 290 hits, gadC coding proteins were selectively chosen; besides, a few amino acid antiporters and arginine/ornithine antiporters were also included for analysis (Table 1 under supplementary material). The other antiporters such as, putrescine-ornithine antiporters, lysine:cadaverine antiporters, histidine/histamine antiporters were omitted from our analysis data.

A multiple sequence alignment was done by using Clustal X Ver.1.83 **[12]**, the gap opening was set at 10.00, the gap extension at 0.20 with 30% delay divergent sequences and Gonnet series weight matrix was used. From the multiple sequence alignment, the guide tree was derived. To justify the confidence of the clades, re-sampling method (bootstrap) was used with 10000 trails. Web logo (ver 2.8.2) was used to identify the conserved pattern in the gadC coded antiporters of *Enterobacteriaceae*. Alignments were analysed and phylogenetic relationships among the sequences were established using different procedures: Neighbour-Joining (NJ) **[13]**, Fast Minimum Evolution (FastME) **[14]** Unweighted Pair Group Method with Arithmetic Mean (UPGMA). **[15]** The final tree was displayed by using MEGA 3.1 **[16]**, the nodes and clades of gadC antiporters were traced out by visual examination.

Results and discussion:

Tracing the gadC cluster among the antiporters

A preliminary multiple sequence alignment was carried out among all antiporters of enteric bacteria belonging to the APC superfamily. Based on the multiple sequence alignment and tree construction with 10000 bootstrap trials Figure 1 shows that the gadC coded proteins form a separate cluster from other antiporters which belong to APC super family.

The similar trend was also observed in phylogenies obtained by using different methods (NJ, UPGMA and FastME). The antiporter sequence of *Rhodopirellula baltica* (NP_864077) which belongs to proteobacter was used as an out-group. The convergence of gadC coded antiporters stands separately from other antiporters which comprises of *Listeria monocytogenes*, Clostridium perfringens, *Lactococcus lactis E. coli*, *S. flexneri* and *S. dysenteriae.*

Evolutionary distance between antiporters

From Figure 2, it is clear that ten major proteins coded by gadC forms the root of the tree (0.0) which corresponds to the gadC cluster (shown in Figure 1). The out group used showed maximum deviation (0.90) and 100% confident divergence from other antiporters from other operational taxonomical units (OTUs). The root comprises antiporters from *S. flexeneri* M25-8A, *S. flexeneri*, *E.coli* 06, *E. coli* UT189, *E. coli* CFT073, *E. coli* K-12, *E. coli* K-12: W3110 and *E. coli* 0157:H7. This proves the very close relationship of *E. coli* and *S. flexneri.* Whereas the *S. dysenteriae* Serovar 1 was little diverged (0.01) form the root and shows the close relationship with the root.

A slightly deviated cluster from the root (0.43-0.47) which corresponds to the gadC cluster shown in Figure 1 comprises of *Listeria monocytogenes* EGD5, *L. monocytogenes* LO28 (0.44), *Clostridium perfringens* str.13 (0.43), *Lactococcus lactis* subsp. *Cremoris*, *L. lactis* subsp. cremoris MG1363, *L. lactis* subsp. *Lactis* str. IL1403 (0.47) shows a close relationship among each other. This is congruent with Sanders *et al* (1998), showed that *Lactococcus lactis* gadC is homologous to putative glutamate-gamma-aminobutyrate antiporters of *E. coli* and *S .flexneri* **[10]** and also with Cotter *et al* (2001), showed that *L. monocytogenes* GadC shares high homology, 65% and 51% identity (77% and 68% similarity) with the equivalent transporters in the *L. lactis* and *E. coli.* **[5]**

The root (0.0) and the closely related cluster (0.43-0.47) have the conserved LVFFLLLGCC motif. The conservation goes on decreases with respect to other clusters or distantly related antiporters and reveals that electrochemical-potential-driven transporters essential for the expression of acid resistance, could not be detected in other family members of the *Enterobacteriaceae.* **[17]**

In contrast to neutralophilic bacteria such as *Salmonella typhimurium*, *E. coli* and *Shigella* have acid resistance systems which are unique. **[2]** Moreover the acid resistance in *E.coli* and *Shigella* species is similarly regulated. **[18]** Hence our main focus lies on gadC; the gadC clade was analyzed separately so as to determine the relationship between *E. coli* and *S. flexneri* because studies conducted by Lin *et al.* **[3]** showed that the toxic strain of *E.coli*, H10407 did adapt well at pH 4.3, although not quite as well as *S. typhimurium* UK1. It was also considered possible that two other strains of *E. coli* and *S. flexneri* might respond better to an acid shock at a less acidic pH. Therefore a separate sequence analysis was carried out between *E. coli* strains and *S. flexneri*. Careful analysis of multiple sequence alignment of *E. coli* [P63235], *E. coli* 0157:H7 [P58229], *E. coli*O6 [Q8FHG6], showed that these organisms have 98-99% homology with their closely related clades of phylogeny (Figure 3). This high similarity may be due to the two glutamate decarboxylases, encoded by gadA and gadB, with gadB forming part of an operon with the antiporter determinant gadC. These homologues obviously resulted from a gene duplication event, given that they share 98% and 99% similarity at the DNA and protein levels respectively **[19]** as shown in Figure 3.

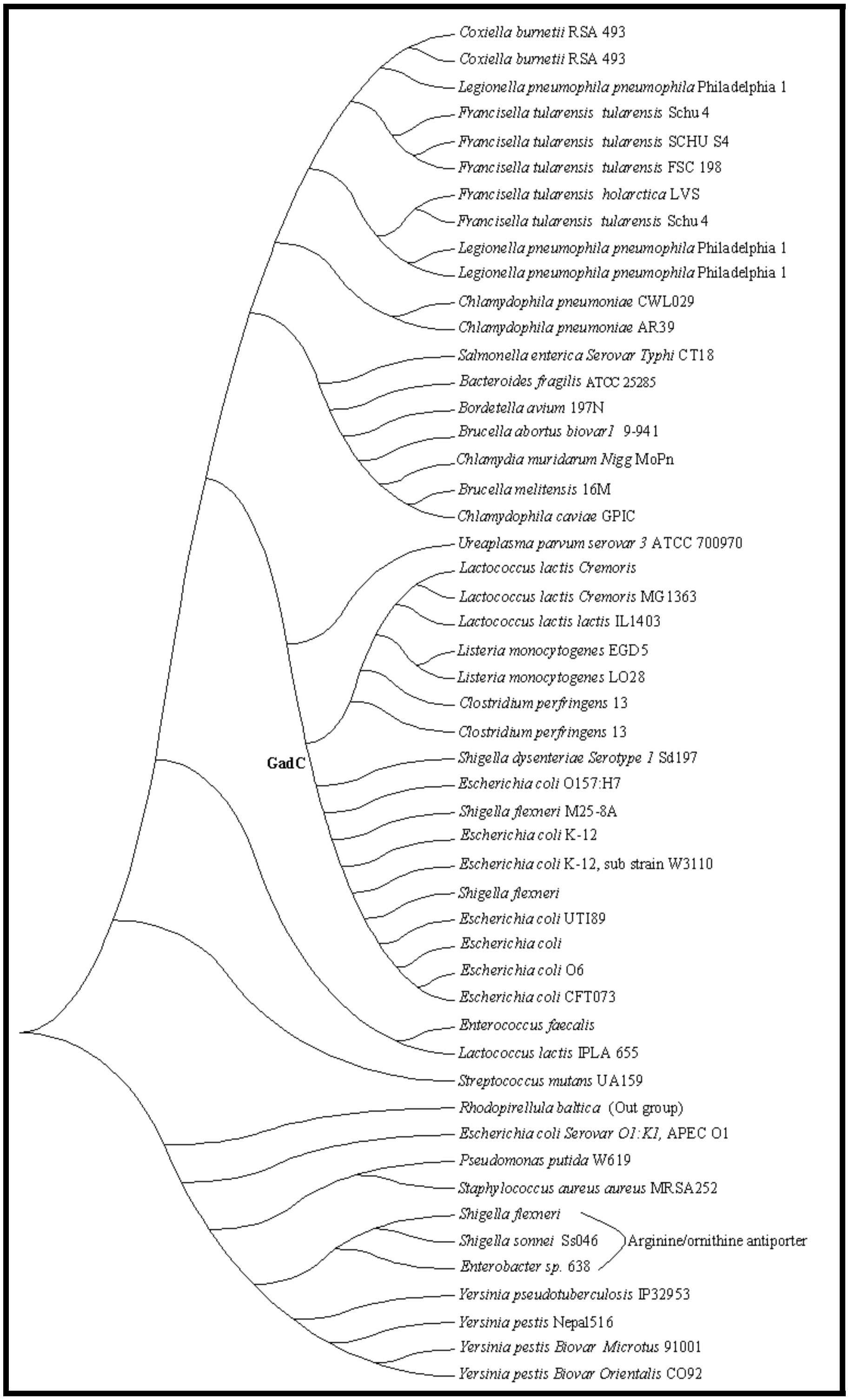

Figure 1: The phylogenetic tree with 10000 bootstrap trials shows a separate cluster of *gadC* coded proteins among other antiporters belong to amino-acid-polyamine-organocation (APC) superfamily. The similar trend was also observed by using different methods (NJ, UPGMA, and FastME). Each branch shows the organism name followed by sub-species and strain.

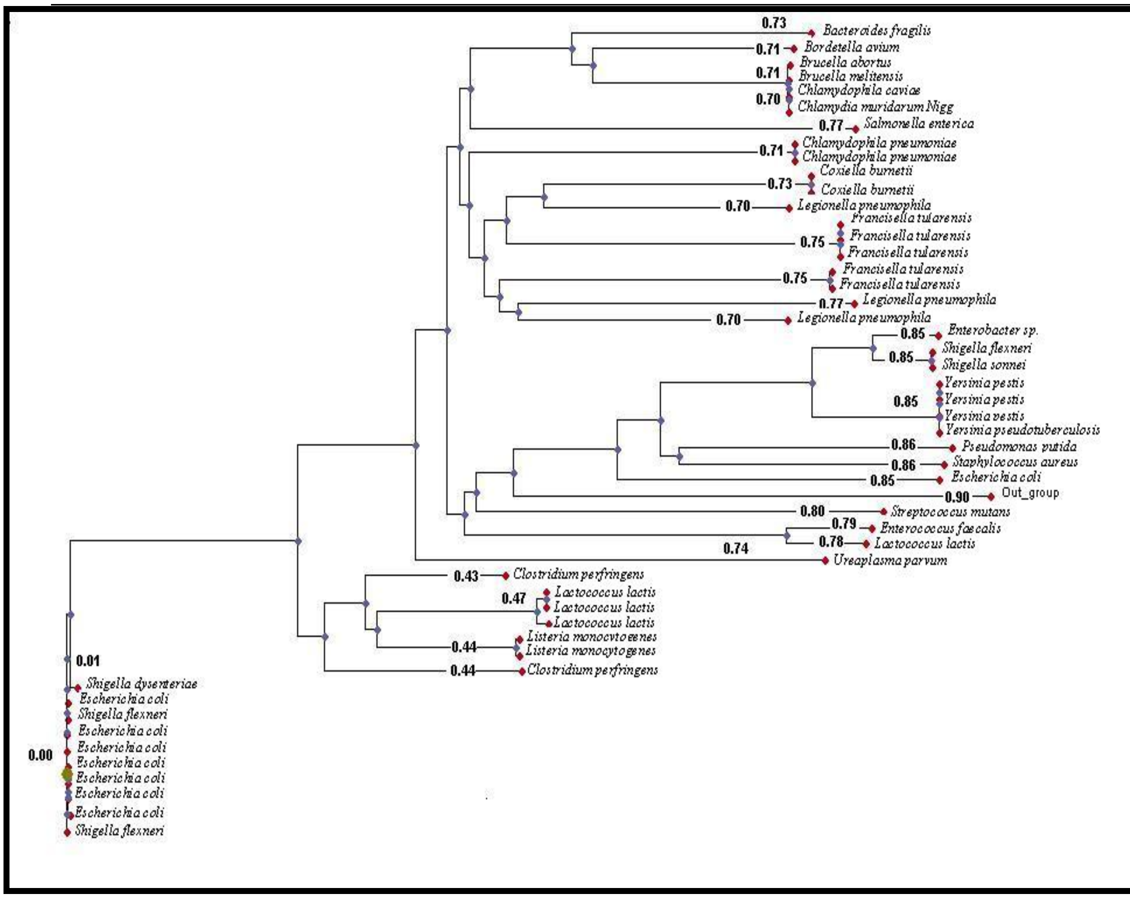

Figure 2: Phylogram shows the branch lengths / evolutionary distances among antiporters.

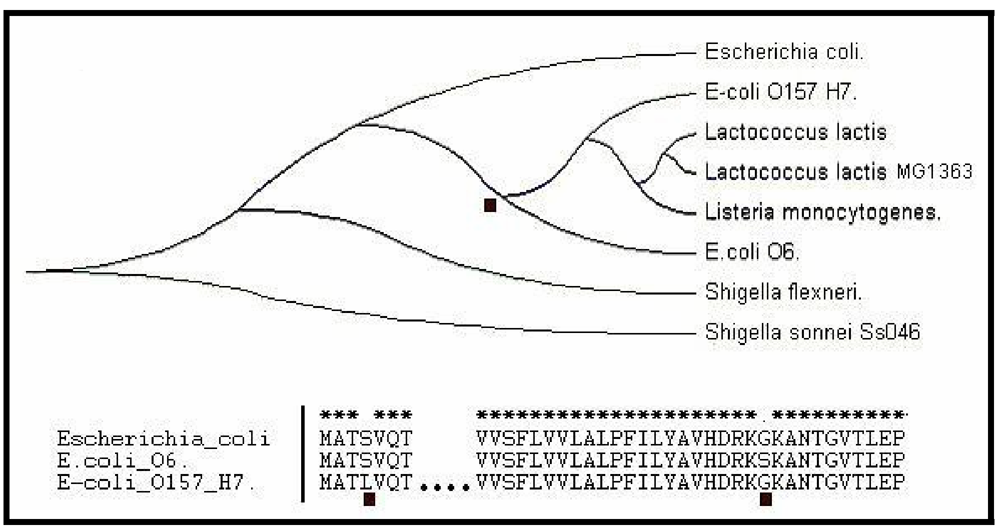

Figure 3: Relationship between the *E.coli* and its strains with *S.flexneri.* The dual mutations one at 4 th residue and the other at 470th amino acid are represented as blocks. The descendants of *E.coli* strains such as *E.colib0157:H7* and *E.coliO6* might have deviated at a particular evolutionary time period because of the dual mutation occurred in the sequence (indicated as blocks).

Bioinformation by Biomedical Informatics Publishing Group *open access*

www.bioinformation.net

Hypothesis

Figure 4: Web logo (ver 2.8.2) was used to identify the consensus region of the N-terminal among the *gadC* coded proteins which is shown in box.

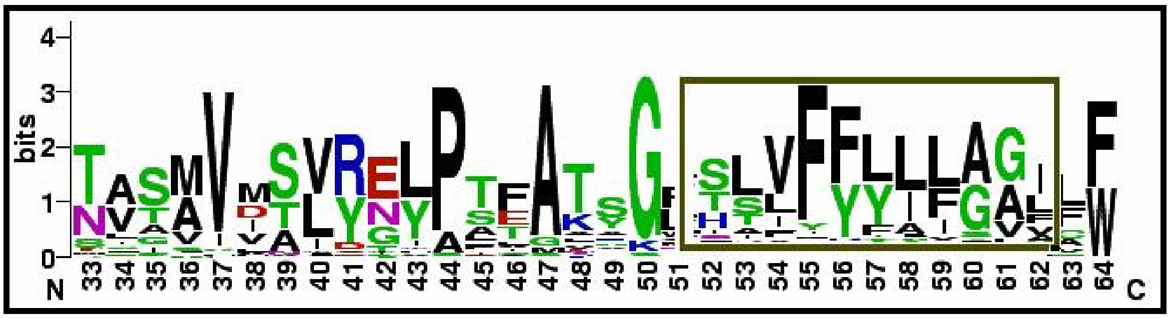

Figure 5: Web logo (ver 2.8.2) was used to identify the consensus region of the N-terminal among the antiporter proteins not coded by *gadC* which is shown in box.

Distinct pattern conservation of glutamate binding region

Multiple alignment of the gadC coded acid sensitive antiporter showed many conserved residue patterns in a regular interval at the N-terminal region and as the alignment approached the C-terminal end, the number of conserved residues decreased, indicating that the N-terminal region of this protein has a much active role when compared to the carboxyl terminal end. The motif FSLVFFLLLGG is considered to play an important role in the recognition of the glutamate and our alignment analysis (Figure 4) confirmed that the motif FHLVFFLLLGG was well conserved within the entire gadC coded proteins (at the amino terminal). It proved that the FHLVFFLLLGG motif is not only unique for *Shigella flexineri* but also for the other gadC coded bacteria such as *Escherichia coli*, *E. coli* O157:H7, *E. coli* O6, *Shigella sonnei* Ss046, *Lactococcus lactis* subsp. *lactis*, *Lactococcus lactis* subsp. *cremoris* and *Listeria monocytogenes*.

We also extended our analysis to address the pattern conservation among the other antiporters (which are not coded by gadC) involved in acid sensitive/resistance mechanism. Amazingly we found that the pattern is still partially conserved for the acid sensitive/ resistance mechanism (Figure 5). This pattern conservation also depicts that the function is highly dependant on the pattern used for the acid resistance. The motif 'FHLVFFLLLGG' was well conserved with the entire gadC coded proteins at the amino terminal where the binding residue could be found with in the first and second transmembrane helices. **[5]** The partial conservation of this motif among the other antiporters (not coded by gadC) is due to the poor acid resistance. The strong motif conservation could be the reason for the extreme acid resistance of *E. coli* and *S. flexneri.* Our pattern analysis shows the relationship of *Escherichia coli* and *Shigella flexneri*. This can be correlated with the claims of Waterman and Small (2003) **[19]**, for a strong-link between the possession of the gadC genes and the expression of stationary-phase acid resistance. This also correlates with the epidemiological data that associated these species with having a lower infective dose compared to other enteric pathogens and confirms the close evolutionary relationship between *Escherichia coli* and *Shigella flexneri* amongst the *Enterobacteriaceae.*

The overall analyses presented herein clearly confirm and adds support to the claim that *Shigella* species possess acid resistance because they are essentially *E. coli* **[20]** in agreement with the taxonomic criteria indicate that *Shigella* and *Escherichia* are actually the same genus **[21]** and have identical virulence determinants that cause clinically indistinguishable disease. **[22, 23]** The phylogenetic analysis of gadC cluster is in congruent with the high degree of identity between the coding regions of *rpoS* in *S. flexneri* and *E. coli* confirms the close taxonomic relationship between the species. **[24]** This close connection (observed from the acid resistance) may lead to the construction of acid resistant vaccine strains which would be effective at low dosages and would not require encapsulation or administration of bicarbonate to ensure passage through the stomach.

Acknowledgment:
This work was supported by Council of Scientific and Industrial Research (CSIR). SB thanks Mr. R.Aravind and Mr. V.Bhadramurthy, Senior Research Fellows, Pathology Division, Indian Institute of Spices Research for their useful interactions.

References:

[01] K. S. Gajiwala and S. K. Burley, *J Mol Biol.*, 295: 605 (2000) [PMID: 10623550]
[02] J. W. Foster, *et al.*, *Novartis Symp.*, 221: 55 (1999) [PMID: 10207913]
[03] J. Lin, *et al.*, *J Bacteriol.*, 177: 4097 (1995) [PMID: 7608084]
[04] M. P. Castanie-Cornet, *et al.*, *J Bacteriol.*, 181: 3525 (1999) [PMID: 10348866]
[05] P. D. Cotter, *Mol Microbiol.*, 40: 465 (2001) [PMID: 11309128]
[06] E. F. Gale, *et al.*, *Biochem. J.*, 36: 600 (1942) [PMID: 16747564]
[07] E. A. Auger, *et al.*, *Mol. Microbiol.*, 3: 609 (1989) [PMID: 2527331]
[08] J. L. Slonczewski, *et al.*, *J. Bacteriol.*, 169: 3001 (1987) [PMID: 2954947]
[09] S. R. Waterman and P. L. Small, *FEMS Microbiol Lett.*, 15: 224 (2003) [PMID: 12855178]
[10] J. W. Sanders, *et al.*, *Mol Microbiol.*, 27: 299 (1998) [PMID: 9484886]
[11] Y. Jiang, *et al.*, 54: 149 (2005) [PMID: 16122562]
[12] J. D. Thompson, *et al.*, *Nucleic Acids Res.*, 24: 4876 (1997) [PMID: 9396791]
[13] N. Saitou and M. Nei, *Molecular Biology and Evolution*, 4: 406 (1987) [PMID: 3447015]
[14] R. D. J. Desper, *et al.*, *J Comput Biol.*, 9: 687 (2002) [PMID: 12487758]
[15] R. R. Sokal and C. D. Michener, *Scientific Bulletin*, 38: 1409 (1958)
[16] S. Kumar, *et al.*, *Brief Bioinform.*, 5: 150 (2004) [PMID: 15260895]
[17] M. Giangrossi, *et al.*, *J Biol Chem.*, 280: 21498 (2005) [PMID: 15795232]
[18] P. L. C. Small and S. Falkow, *Abstr. Gen. Meet. Am. Soc. Microbiol.*, 74: 38 (1992)
[19] D. K. Smith, *et al.*, *J Bacteriol.*, 174: 5820 (1992) [PMID: 1522060]
[20] J. Gorden, *et al.*, *Infection and Immunity*, 61: 364 (1993) [PMID: 8418063]
[21] D. J. Brenner, *et al.*, *Infect Immun.*, 6: 308 (1972) [PMID: 4564889]
[22] T. L. Hale, *et al.*, *Microbiol. Rev.*, 55: 206 (1991) [PMID: 1886518]
[23] D. N. Taylor, *et al.*, *J. Clin. Microbiol.*, 26: 1362 (1988) [PMID: 3045152]
[24] P. Small, *et al.*, *J Bacteriol.*, 176: 1729 (1994) [PMID: 8132468]

Edited by P. Kangueane
Citation: Manikandan, *et al.*, Bioinformation 2(4): 145-152 (2007)

Bioinformation by Biomedical Informatics Publishing Group *open access*

www.bioinformation.net **Hypothesis**

Supplementary material

Sl No	Primary accession number	Organism	Sub-species	Strain	Protein length (amino acids)	Molecular weight (Da)
1	YP_210202	*Bacteroides fragilis*		ATCC 25285; NCTC 9343	532 aa	57130
2	CAJ50406	*Bordetella avium*		197N	491 aa	53194
3	YP_223596	*Brucella abortus*	*biovar 1*	9-941	510 aa	55080
4	NP_541887	*Brucella melitensis*		16M	510 aa	55150
5	NP_829364	*Chlamydophila caviae*		isolate="GPIC"	466 aa	51138
6	NP_296865	*Chlamydia muridarum Nigg*		MoPn	466 aa	50981
7	NP_224487	*Chlamydophila pneumoniae*		CWL029	468 aa	51508
8	NP_445023	*Chlamydophila pneumoniae*		AR39	468 aa	51508
9	NP_562976	*Clostridium perfringens*		13	472 aa	50926
10	AAO91508	*Coxiella burnetii*		RSA 493	476 aa	52577
11	NP_820994	*Coxiella burnetii*		RSA 493	476 aa	52446
12	ZP_01587708	*Enterobacter sp.*		638	460 aa	47394
13	AAM46084	*Enterococcus faecalis*			454 aa	49565
14	NP_753817	*Escherichia coli*		CFT073	511 aa	54976
15	NP_416009	*Escherichia coli*		K-12	511 aa	54946
16	AP_002115	*Escherichia coli*		K-12, sub strain W3110	511 aa	54946
17	YP_859796	*Escherichia coli*	*Serovar O1:K1*	APEC O1	489 aa	53280
18	P58229	*Escherichia coli*		O157:H7	511 aa	55103
19	Q8FHG6	*Escherichia coli*		O6	511 aa	55107
20	ABE07184	*Escherichia coli*		UTI89	511 aa	55091
21	NP_753817	*Escherichia coli*		CFT073	511 aa	54976
22	CAG45113	*Francisella tularensis*	*tularensis*	SCHU S4	469 aa	51642
23	YP_513914	*Francisella tularensis*	*holarctica*	LVS	473 aa	52696
24	YP_666650	*Francisella tularensis*	*tularensis*	FSC 198	469 aa	51511
25	YP_169518	*Francisella tularensis*	*tularensis*	Schu 4	469 aa	51511
26	YP_169957	*Francisella tularensis*	*subsp. tularensis*	Schu 4	471 aa	52479
27	CAF33981	*Lactococcus lactis*		IPLA 655	464 aa	50641
28	O30417	*Lactococcus lactis*	*Cremoris*		503 aa	55369
29	AAC46187	*Lactococcus lactis*	*cremoris*	MG1363	503 aa	55369
30	NP_562216	*Clostridium perfringens*		13	485 aa	52630
31	NP_267447	*Lactococcus lactis*	*lactis*	IL1403	503 aa	55434
32	YP_095718	*Legionella pneumophila*	*pneumophila*	Philadelphia 1	464 aa	50647
33	YP_095685	*Legionella pneumophila*	*pneumophila*	Philadelphia 1	445 aa	49056
34	YP_094448	*Legionella pneumophila*	*pneumophila*	Philadelphia 1	467 aa	50332

35	AAK17186	*Listeria monocytogenes*		EGD5	507 aa	55169
36	AAG22561	*Listeria monocytogenes*		LO28	507 aa	55154
37	ZP_01642072	*Pseudomonas putida*		W619	475 aa	47649
38	CAD07591	*Salmonella enterica*	*Sub sp. enterica Serovar Typhi*	CT18	473 aa	51854
39	YP_403230	*Shigella dysenteriae*	*Serotype 1*	Sd197	511 aa	54984
40	AAD14843	*Shigella flexneri*		M25-8A	511 aa	55077
41	P63236	*Shigella flexneri*			511 aa	55077
42	P0AAE7	*Shigella flexneri*			460 aa	49501
43	YP_310489	*Shigella sonnei*		Ss046	460 aa	49370
44	CAG41690	*Staphylococcus aureus*	*aureus*	MRSA252	478 aa	51918
45	NP_720726	*Streptococcus mutans*		UA159	452 aa	49472
46	NP_078056	*Ureaplasma parvum*	*serovar 3*	ATCC 700970	759 aa	84583
47	NP_993425	*Yersinia pestis*	*Biovar Microtus*	91001	463 aa	49740
48	NP_405843	*Yersinia pestis*	*Biovar Orientalis*	CO92	463 aa	49740
49	YP_647695	*Yersinia pestis*		Nepal516	463 aa	47684
50	YP_070743	*Yersinia pseudotuberculosis*		IP32953	463 aa	46696

Table 1: Acid sensitivity/resistance antiporter protein sequences retrieved from GenBank.