ORIGINAL RESEARCH

In silico identification and characterization of novel drug targets and outer membrane proteins in the fish pathogen *Edwardsiella tarda*

Mohammed Neema
Iddya Karunasagar
Indrani Karunasagar

UNESCO Microbial Resource Centre for Marine Biotechnology, Department of Fishery Microbiology, Karnataka Veterinary Animal and Fisheries Sciences University, College of Fisheries, Mangalore, Karnataka, India

Correspondence: Indrani Karunasagar
Professor and Head, Department of Microbiology, and Director, UNESCO Microbial Resource Centre for Marine Biotechnology, College of Fisheries, Mangalore 575 002, Karnataka, India
Tel +91 824 2246384
Fax +91 824 2246384
Email karuna8sagar@yahoo.com

Abstract: The complete genome of the fish pathogen *Edwardsiella tarda* and the host *Danio rerio* (zebrafish) was subjected to subtractive genome analysis with the objective of identifying potential drug targets and vaccine candidates. The homologous proteins were excluded to prevent any chance of cytotoxic cross-reaction with the host. The nonparalogous proteins were analyzed for sequence homology with the Database of Essential Genes to determine the essentiality of the proteins for the bacteria. These essential proteins were further analyzed to predict the metabolic pathways in which they were involved. Cellular localization analysis was carried out to determine the possibility of the protein being present in the outer membrane. Outer membrane proteins may play an important role in the interaction with hosts and in bacterial pathogenicity. The study identified 171 of 3986 proteins as potential drug targets. The study also identified seven novel outer membrane proteins.

Keywords: subtractive genomics, CD-HIT, DEG, KEGG, homology modelling, *Edwardsiella tarda*, OMPs

Introduction

Edwardsiella tarda is a Gram-negative Enterobacteriaceae member that causes edwardsiellosis in fish. This systemic disease affects a variety of fish taxa including channel catfish, striped bass, tilapia, and flounder[1] and has a worldwide distribution in fresh and marine waters.[2] Fish mortality caused by *E. tarda* infection has led to severe losses in the aquaculture industry.[3] *E. tarda* is a potential pathogen capable of causing disease and stimulating the inflammatory immune response in both zebrafish embryos and adults.[4] The affected fish show clinical signs such as abdominal distension, vent protrusion, enteritis, liver congestion, and abscess-like lesions in enlarged spleens and kidneys.[5] In a few cases, the fish reveal lesions on the skin, pale gills, tumefaction of the eye, excessive mucus secretion, scale erosion, and ulcers.[6] The pathogen enters the fish body through the gastrointestinal tract, gills, and body surface.[7] It was found that under unfavorable conditions the bacteria become nonculturable but still remain viable and gradually decrease in size and change in shape from short rods to coccoid forms. But when favorable conditions are provided they become infective,[8] showing the ability of the bacteria to survive in the environment.

A commercial vaccine for protecting fish against *E. tarda* is not available, though vaccine against *E. ictaluri*, which affects American catfish, is available. Numerous experimental studies on the identification of candidate molecules for vaccine development are reported in the literature, and these include lipopolysaccharides,[9] outer

membrane proteins (OMPs),[10] and glyceraldehyde 3-phosphate dehydrogenase.[11] Thus, chemotherapy is commonly used in cases of outbreaks of enteric septicemia caused by this pathogen in warm water aquaculture.

Laboratory experiments to identify candidate molecules as drug targets are time consuming and expensive. Bioinformatics-based methodologies provide alternative approaches and have the potential to cut costs in identifying drug targets. A subtractive genomics approach utilizes the whole proteome of host and pathogen to identify proteins exclusively present in the pathogen by deducing the homologous proteins. This procedure has been successfully used to identify novel drug targets in human pathogens like *Pseudomonas aeruginosa*,[12] *Helicobacter pylori*,[13] *Salmonella typhi*,[14] and *Neisseria meningitides*.[15] In this study we employ a technique that makes use of the systems of the Cluster Database at High Identity with Tolerance (CD-HIT), Basic Local Alignment Search Tool for Proteins (BLASTP), Database of Essential Genes (DEG), KEGG Automatic Annotation Server (KAAS), Kyoto Encyclopedia of Genes and Genomes (KEGG), PSORTb, and SWISS-MODEL to identify, characterize, and analyze the essential genes of the fish pathogen *E. tarda* that could be suitable vaccine or drug candidates.

Materials and methods

The sequential diagram to identify the essential genes of *E. tarda* is illustrated in Figure 1. The complete proteome of *E. tarda* was retrieved from Swiss-Prot (http://us.expasy.org/sprot/). The proteins were subjected to CD-HIT analysis (http://weizhong-lab.ucsd.edu/cdhit_suite/cgi-bin/index.cgi).[16] The program takes a fasta format sequence database as input and produces a set of nonredundant, representative sequences as output. The process was carried out with a sequence identity cutoff of 0.6, thus eliminating redundant sequences with more than 60% identity.[15] The resultant proteins were grouped as Set1 proteome. Because the complete genome sequence of zebrafish, *D. rerio*, is available, this was taken to represent the fish host for the subtractive genomics study. BLASTP (http://blast.ncbi.nlm.nih.gov/Blast.cgi) analysis was carried out for the Set1 proteome against the proteome of *D. rerio*. Proteins with an E-value (expectation value) $<10^{-4}$ were eliminated, assuming that they have a certain level of homology with the host genome.[15] The resultant data set (Set2 proteome) of *E. tarda* had no homology with the zebrafish proteome.

The next step was to identify the genes involved in the metabolic pathway in the pathogen. For this, the Set2 proteome was subjected to BLASTP analysis with DEG (http://tubic.tju.edu.cn/deg/), which contains all the essential genes currently available.[17] An E-value of 10^{-100} was set as the cutoff value.[14] Thus, a Set3 proteome of *E. tarda* was obtained by grouping the proteins that showed an E-value $<10^{-100}$. Set3 proteome contains the proteins of *E. tarda* that could be considered as novel drug targets because they are not present in the host and are not involved in essential metabolic
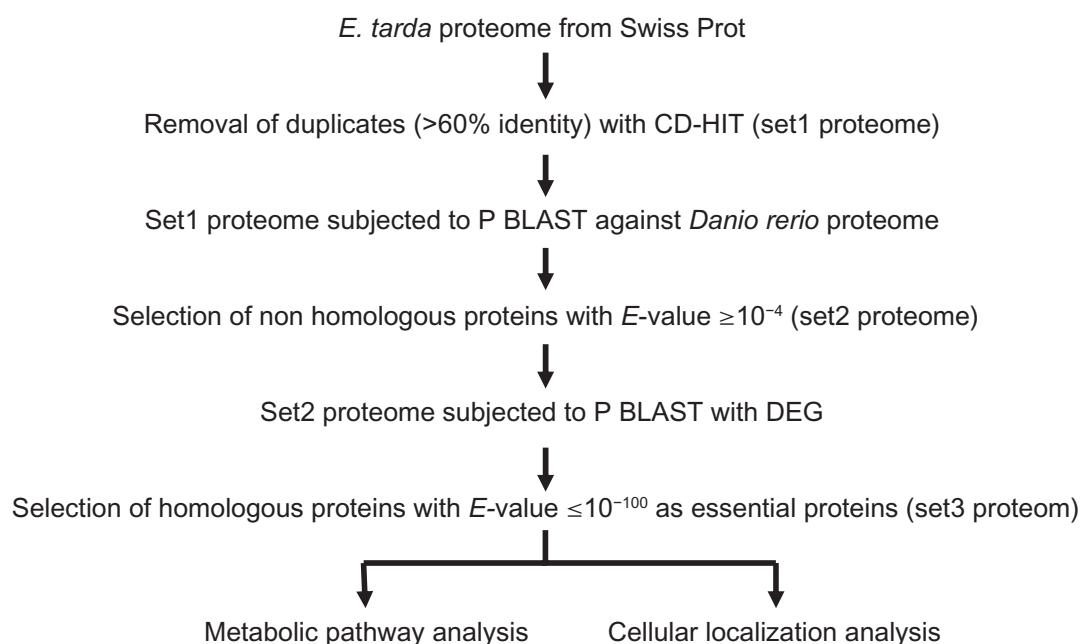
E. tarda proteome from Swiss Prot

↓

Removal of duplicates (>60% identity) with CD-HIT (set1 proteome)

↓

Set1 proteome subjected to P BLAST against *Danio rerio* proteome

↓

Selection of non homologous proteins with E-value $\geq 10^{-4}$ (set2 proteome)

↓

Set2 proteome subjected to P BLAST with DEG

↓

Selection of homologous proteins with E-value $\leq 10^{-100}$ as essential proteins (set3 proteom)

↓

Metabolic pathway analysis   Cellular localization analysis

Figure 1 Steps used to identify essential protein targets in *E. tarda*.

functions in the bacterium. Further analysis was carried out on these essential proteins to determine their subcellular localization and function. The KAAS (http://www.genome.jp/tools/kaas/) of the Kyoto Encyclopedia of Genes and Genomes was used to determine the functions. The server provides functional annotation of genes by BLAST comparisons against the manually curated KEGG GENES database. The result contains KO (KEGG Orthology) assignments and automatically generated KEGG pathways.[18] Computational tools are involved in the prediction of subcellular localization, which illustrates where the protein resides in the cell. So if the protein is present in the outer membrane, there is more possibility that it can be highlighted as a potential drug target. The program PSORTb V.3.0 (http://www.psort.org/psortb/index.html) was used for subcellular localization prediction.[19] Homology modelling of the outer membrane was done by the server SWISS-MODEL (swissmodel.expasy.org) to obtain a three-dimensional (3D) structure of the OMPs. The server models the protein based on the alignment between the target and template and builds the structure based on the homology.[20] A two-dimensional (2D) structure prediction of OMPs was done by PRED-TMBB (http://biophysics.biol.uoa.gr/PRED-TMBB/). The method predicts the transmembrane beta-strands of the Gram-negative bacteria OMPs and also discriminates such proteins from water-soluble ones when screening large datasets.[21] The biological significance of the homology-modeled OMPs was determined using the ProFunc server[26] and Prints server.[27]

Results and discussion

The proteome of *E. tarda* had a total of 3986 proteins. Of these, 436 proteins were excluded due to redundancy, as analyzed by the CD-HIT program at 60% identity. So there were 3550 proteins with nonparalogous sequence. BLASTP analysis of these proteins against the zebrafish genome revealed 2494 proteins that were nonhomologous with the host genome. A definite homology between the host and pathogen protein chosen as drug targets might lead to unwanted cross-reactions and cytotoxicity. Hence, nonhomologous proteins were selected for further identification of essential genes. The essential genes are those that are indispensable to support cellular life. Because most antibiotics target essential cellular processes in bacteria, essential gene products of microbial cells are promising new targets for antibacterial drugs.[17] Essential genes unique to an organism can be considered as species-specific drug targets.[22] The DEG hosts records of currently available essential genes among a wide range of organisms. The 2494 host nonhomologous

Table 1 Results of substractive proteomic and metabolic pathway analysis for *Edwardsiella tarda*

Outputs	Number
Total number of proteins	3986
Duplicates (>60% identical) in CD-HIT	436
Nonparalogs	3550
Non-zebrafish homologous proteins (E-value $\geq 10^{-4}$)	2494
Essential protein in DEG (E-value $\leq 10^{-100}$)	171
Essential proteins involved in metabolic pathways	62
Pathways unique to the organism (*E. tarda*)	21
Membrane associated non-fish homologs of essential genes	7

Abbreviations: CD-HIT, Cluster Database at High Identity with Tolerance; DEG, Database of Essential Genes.

proteins were subjected to BLASTP against DEG with an E-value cutoff score of 10^{-100} to obtain 171 proteins that were essential to the enterobacteria. This is in agreement with the experimentally determined essential genes of 150 numbers in the enterobacteria *E. coli*.[23] A detailed analysis of *E. tarda* proteome is provided in Table 1.

Metabolic pathway analyses of the 171 essential proteins of *E. tarda* revealed that 28 proteins are involved in genetic information processing, 23 in environmental information
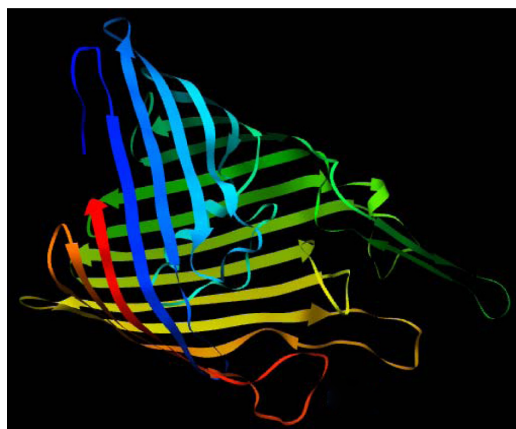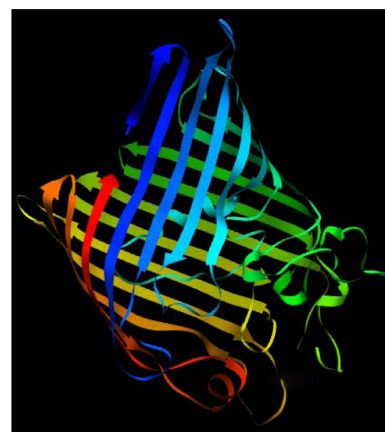
Table 2 Identification of the uncharacterized essential proteins in *Edwardsiella tarda*

Sl no.	Accession no.	Function
1	tr|D0ZBY3|D0ZBY3_EDWTE	Beta-barrel porins
2	tr|D0ZCR8|D0ZCR8_EDWTE	All lipid-binding proteins
3	tr|D0Z8D1|D0Z8D1_EDWTE	Cobalt-binding
4	tr|D0ZHD5|D0ZHD5_EDWTE	DNA-binding
6	tr|D0Z9A0|D0Z9A0_EDWTE	Electrochemical potential-driven transporters
7	tr|D0ZF96|D0ZF96_EDWTE	Electrochemical potential-driven transporters
8	tr|D0ZDN7|D0ZDN7_EDWTE	Hydrolases – acting on acid anhydrides
9	tr|D0ZDB8|D0ZDB8_EDWTE	Iron-binding
10	tr|D0Z8D0|D0Z8D0_EDWTE	Iron-binding
11	tr|D0ZBG3|D0ZBG3_EDWTE	Iron-binding
12	tr|D0Z9Y7|D0Z9Y7_EDWTE	Iron-binding
13	tr|D0Z9Z4|D0Z9Z4_EDWTE	Iron-binding
14	tr|D0ZDU5|D0ZDU5_EDWTE	Lyases – carbon–carbon lyases
15	tr|D0ZG57|D0ZG57_EDWTE	Repressor
16	tr|Q4ACC8|Q4ACC8_EDWTA	Repressor
17	tr|Q8GF10|Q8GF10_EDWTA	Transferases – glycosyltransferases
18	tr|D0ZG02|D0ZG02_EDWTE	Transmembrane
19	tr|D0Z8J1|D0Z8J1_EDWTE	Transmembrane
20	tr|D0Z933|D0Z933_EDWTE	Transmembrane
21	tr|D0ZE91|D0ZE91_EDWTE	Transmembrane
22	tr|D0ZGG7|D0ZGG7_EDWTE	Transmembrane
23	tr|D0ZAM0|D0ZAM0_EDWTE	Zinc-binding
24	tr|D0Z8E4|D0Z8E4_EDWTE	Zinc-binding
25	tr|D0ZAW6|D0ZAW6_EDWTE	Zinc-binding

Table 3 List of the outer membrane proteins of *Edwardsiella tarda* identified by PSORTb

Accession no.	Name of protein	Subcellular localization	Biological significance
Q4ACC8	Outer membrane protein S1 precursor	Outer membrane	Porin
Q8GFI0	Outer membrane protein	Outer membrane	Porin
D0ZC49	Organic solvent tolerance protein	Outer membrane	Porin
D0ZAR5	Putative outer membrane protein	Outer membrane	Surface antigen
D0ZDN7	Lipoprotein	Outer membrane	ABC transporter
D0ZFR4	Outer membrane usher protein	Outer membrane	Fimbrial usher
D0ZB52	Long-chain fatty acid transport protein	Outer membrane	Beta-lactamase A

processing, 20 in amino acid metabolism, 14 in carbohydrate metabolism, 13 in metabolism of cofactors and vitamins, 12 in cellular processes and signalling, 11 in energy metabolism, 8 in glycan biosynthesis and metabolism, 7 in nucleotide metabolism, 5 in different enzyme families, 3 in biosynthesis of polyketides and terpenoids, 1 in lipid metabolism, and 1 in xenobiotics biodegradation and metabolism. The features of the essential genes of *E. tarda* are listed in Supplementary Table 1. Genetic information processing involves the process of transcription, translation, folding, sorting, and degrading proteins, as well as DNA replication and repair. Environmental information processing involves membrane transport, signal transduction, and interaction of signalling molecules. Out of these 171 proteins, 25 were uncharacterized. The functional classification of these 25 uncharacterized essential proteins was performed using the SVM-Prot Web server.[24] The Web server classifies these 25 proteins as transmembrane proteins (5), iron-binding proteins (5), zinc-binding proteins (3), DNA-binding proteins (2), repressor proteins (2), cobalt-binding proteins (1), lipid-binding proteins (1), transferases (1), carbon–carbon lyases (1), hydrolases (1), electrochemical potential driven transporter proteins, (2) and beta-barrel porins (1). Details are shown in Table 2.

Figure 3 Homology based model of OMPs of *E. tarda* (Q8GFI0).
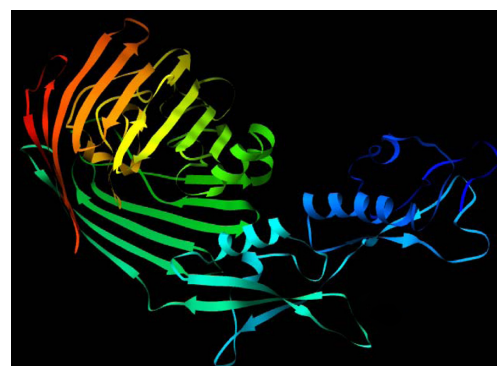
Figure 4 Homology based model of OMPs of *E. tarda* (DOZAR5).

Figure 2 Homology based model of OMPs of *E. tarda* (Q4ACC8).
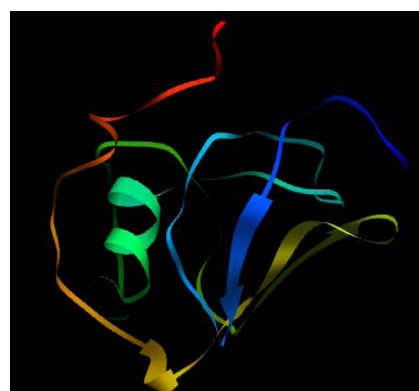
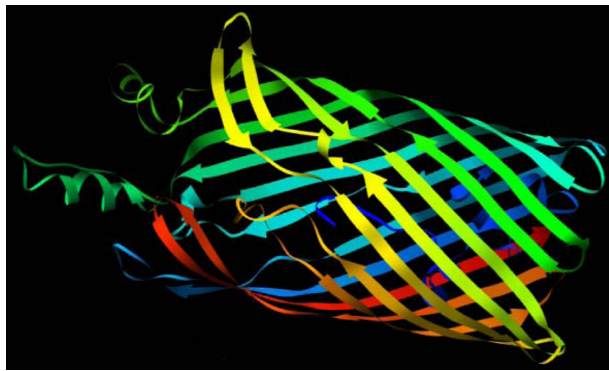**Figure 5** Homology based model of OMPs of *E. tarda* (DOZFR4).

Figure 6 Homology based model of OMPs of *E. tarda* (DOZB52).

Comparative analysis of the metabolic pathways of the pathogen *E. tarda* and the host *D. rerio* revealed 20 metabolic pathways that are unique to the pathogen. These pathways were further screened to obtain enzymes and proteins that were specific to *E. tarda* (Supplementary Table 2). Further investigations on the predicted genes are required to verify the reliability of the data. The outer membrane of Gram-negative pathogenic bacteria has an important role in the interaction with hosts in bacterial pathogenicity, playing a role in adherence, uptake of nutrients from the host, and countering host defense mechanisms.[25] They could be protective antigens because the components of the outer membrane are easily recognized as foreign substances by immunological defense systems of hosts. The essential proteins were therefore analyzed for their localization in the outer membrane. The programme PSORTb identified seven proteins that reside in the outer membrane (Table 3). Figures 2–6 show the homology modeled structure of the OMPs of *E. tarda.* Because Protein Data Bank structure homologous to the proteins D0ZC49 and D0ZDN7 was not available, the 2D structure prediction of these proteins was carried out using PRED-TMBB (Figures 7 and 8), which predicts the number of transmembrane beta-barrels in the
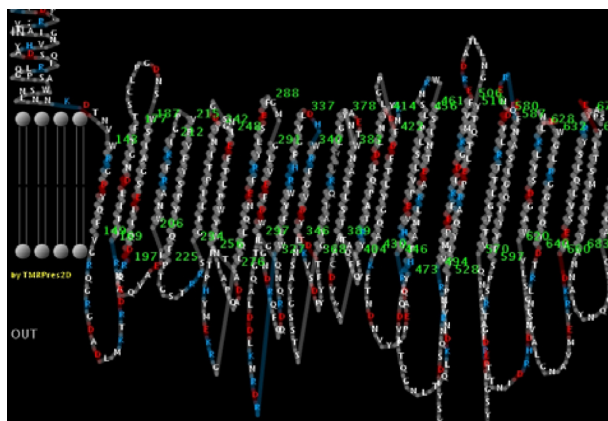

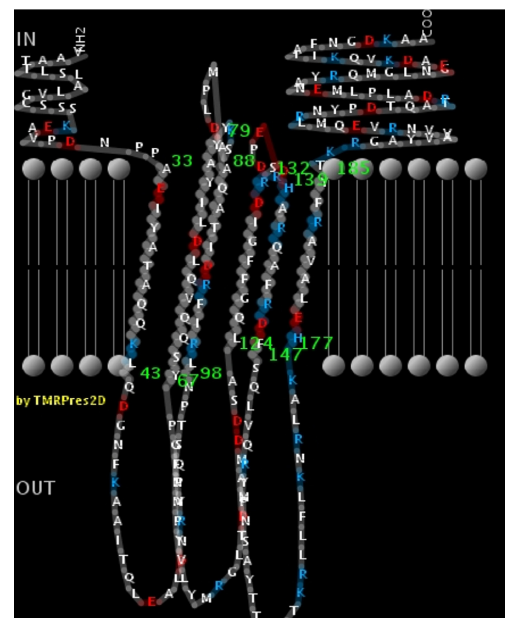**Figure 7** 2D structure prediction by PRED TMBB (D0ZC49).

Figure 8 2D structure prediction by PRED TMBB (D0ZDN7).

outer membrane. The results show that the proteins identified as OMPs by PSORTb had beta-barrels characteristic of such proteins, thus supporting the prediction.

Conclusion

This study is the first report of the application of subtractive genome analysis of any fish pathogen and its potential host. Using this approach, a total of 3986 proteins of *E. tarda* were reduced to 171 proteins, which could further be investigated experimentally as potential drug targets. The study also identified OMPs that could be drug targets or vaccine candidates. Virtual screening against these target proteins might be useful in the discovery of novel therapeutic compounds against the fish pathogen *E. tarda.*

Acknowledgments

Disclosure

The authors report no conflicts of interest in this work.

References

1. Plumb JA. *Edwardsiella* septicaemia. In: Inglis V, Roberts RJ, Bromage NR, editors. *Bacterial Diseases of Fish*. Oxford: Blackwell Scientific Publications; 1993;61–79.
2. Austin B, Austin DA. Characteristics of the diseases. In: Austin B, Austin DA, editors. *Bacterial Fish Pathogens: Diseases of Farmed and Wild Fish*. 3rd ed. Heidelberg: Springer-Verlag; 1999;13–15.

3. Thune RL, Stanley LA, Cooper RK. Pathogenesis of gram negative bacterial infections in warm water fish. *Annu Rev Fish Dis*. 1993;3: 37–68.

4. Pressley ME, Phelan PE, Witten PE, et al. Pathogenesis and inflammatory response to *Edwardsiella tarda* infection in the zebrafish. *Dev Comp Immunol*. 2005;29(6):501–513.

5. Yu JH, Han JJ, Park KS, et al. *Edwardsiella tarda* infection in Korean catfish, *Silurus asotus*, in a Korean fish farm. *Aquacult Res*. 2009;41(1): 19–26.

6. Mohanty BR, Sahoo PK. Edwardsiellosis in fish: a brief review. *J Biosci*. 2007;32(7):1331–1344.

7. Ling SHM, Wang XH, Lim TM, Leung KY. Green fluorescent protein-tagged *Edwardsiella tarda* reveals portal of entry in fish. *FEMS Microbiol Lett*. 2001;194(2):239–243.

8. Du M, Chen J, Zhang X, et al. Retention of virulence in a viable but nonculturable *Edwardsiella tarda* isolate. *Appl Environ Microbiol*. 2007;73(4):1349–1354.

9. Salati F, Kawai K, Kusuda R. Immune response of eel to *Edwardsiella tarda* lipopolysaccharide. *Fish Pathol*. 1984;19(3):187–192.

10. Kawai K, Liu Y, Ohnishi K, Oshima S. A conserved 37 kDa outer membrane protein of *Edwardsiella tarda* is an effective vaccine candidate. *Vaccine*. 2004;22(25–26):3411–3418.

11. Liu Y, Oshima S, Kurohara K, et al. Vaccine efficacy of recombinant GAPDH of *Edwardsiella tarda* against Edwardsiellosis. *Microbiol Immunol*. 2005;49(7):605–612.

12. Sakharkar KR, Sakharkar MK, Chow VTK. A novel genomics approach for the identification of drug targets in pathogens, with special reference to *Pseudomonas aeruginosa*. *In Silico Biol*. 2004;4:0028.

13. Dutta A, Singh SK, Ghosh P, et al. In silico identification of potential therapeutic targets in the human pathogen *Helicobacter pylori*. *In Silico Biol*. 2006;6:0005.

14. Rathi B, Sarangi AN, Trivedi N. Genome subtraction for novel target definition in *Salmonella typhi*. *Bioinformation*. 2009;4(4):143–150.

15. Sarangi AN, Aggarwal R, Rahman Q, Trivedi N. Subtractive genomics approach for *in silico* identification and characterization of novel drug targets in *Neisseria Meningitides* Serogroup B. *J Comput Sci Syst Biol*. 2009;2:255–258.

16. Huang Y, Niu B, Gao Y, et al. CD-HIT suite: a web server for clustering and comparing biological sequences. *Bioinformatics*. 2010; 26:680–682.

17. Zhang R, Ou HU, Zhang CT. DEG, a Database of Essential Genes. *Nucleic Acids Res*. 2004;32 Suppl 1:D271–D272.

18. Moriya Y, Itoh M, Okuda S, et al. KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res*. 2007;35 Suppl 2:W182–W185.

19. Gardy JL, Laird MR, Chen F, et al. PSORTb v.2.0: expanded prediction of bacterial protein subcellular localization and insights gained from comparative proteome analysis. *Bioinformatics*. 2005;21(5):617–623.

20. Arnold K, Bordoli L, Kopp J, Schwede T. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics*. 2006;22(2):195–201.

21. Bagos PG, Liakopoulos TD, Spyropoulos IC, Hamodrakas SJ. PRED-TMBB: a web server for predicting the topology of beta-barrel outer membrane proteins. *Nucleic Acids Res*. 2004;32 Suppl 2: W400–W404.

22. Judson N, Mekalanos JJ. TnAraOut, a transposon-based approach to identify and characterize essential bacterial genes. *Nat Biotechnol*. 2000;18(7):740–745.

23. Jordan K, Rogozin IB, Wolf YI, Koonin EV. Essential genes are more evolutionarily conserved than are nonessential genes in bacteria. *Genome Res*. 2002;12:962–968.

24. Cai CZ, Han LY, Ji ZL, et al. SVM-Prot: web-based support vector machine software for functional classification of a protein from its primary sequence. *Nucleic Acids Res*. 2003;31:3692–3697.

25. Seltmann G, Holst O. *The Bacterial Cell Wall*. Heidelberg: Springer Berlin; 2002.

26. Laskowski RA, Watson JD, Thornton JM. ProFunc: a server for predicting protein function from 3D structure. *Nucleic Acids Res*. 2005; 33:W89–W93.

27. Attwood TK, Kell DB, McDermott P, et al. Utopia documents: linking scholarly literature with research data. *Bioinformatics*. 2010;26: i568–i574.