ORIGINAL ARTICLE

Computational Prediction of Novel Broad-Spectrum Drug Targets Against Vibrio Cholerae by Integrated Genomics and Proteomics Approach

Suresh Kumar, Meera Ramanujam

Department of Diagnostic and Allied Health Science, Faculty of Health & Life Sciences, Management & Science University, 40100 Shah Alam, Selangor, Malaysia

ABSTRACT

Introduction: *Vibrio cholerae* is a motile, Gram-negative curved rod belonging to the Vibrionaceae family. It is the causative agent of cholera. The acute diarrheal disease cholera causes about 120 000 casualties annually and has a significant effect on the health of young kids between the ages of 1 and 5. The main cause of death is due to resistance to antibiotics. As a result, new drug targets need to be identified immediately. The study's goal is to identify *Vibrio Cholerae*'s putative drug target through an integrated approach to genomics and proteomics. **Methods:** Through this study, 2241 core protein sequence of *Vibrio Cholerae* were retrieved from the Panx tool. The sequence decreased to 173 druggable sequences by undergoing different phases of the process such as determining the non-homologous sequence against human proteome by using the BlastP tool, identifying the essential genes by using the DEG database, and determining the sequence of virulent proteins by using Virulent prediction tool. **Results:** 11 potential drug targets were identified through molecular weight, and sub-cellular localization analysis. **Conclusion:** Through pan-genome analysis, we can able to find potential drug targets. This study also helps to identify the potential drug targets against *Vibrio cholerae* and to increase the efforts of drug and vaccine developments.

Keywords: *Vibrio cholerae*, Drug Resistance, Microbial, Proteomics, Genomics

Corresponding Author:
Suresh Kumar, PhD
Email: sureshkumar@msu.edu.my
Tel: +60-14-2734893

INTRODUCTION

Vibrio cholerae is a gram-negative bacteria which causes serious and sometimes deadly diarrheal illness. Cholera is projected to remain a significant healthcare issue with an annual rate of 2.9 million instances (1). In developing nations where sanitation is poor, health care is restricted, and drinking water is unsafe, cholera has remained a severe threat over the previous 40 years (2). With adequate rehydration therapy, cholera can be handled. But in phases of disease severity, antibiotic therapy is needed for fast recovery. Studies, however, indicate that patients with cholera have documented bacterial clearance errors. In a couple of instances, *vibrio cholerae* have become resistant to most of the antibiotics available and are about to become untreatable. Without therapy, cholera can lead to a fatality rate of up to 70%. Currently the most alarming problem for human health and the health care system is antimicrobial resistance (AMR) (3). Globally, AMR already leads to 700,000 deaths / year, and the prediction is that 10 million deaths / year will occur in 2050, greater than today's 8.2 million cancer deaths (4). Over the years, owing to overuse of present antibiotics on the market, they are becoming less efficient or losing efficacy, making the quest for new molecules with distinct action mechanisms a priority.

In regards to bacterial pathogenicity, the experimental method for determining gene function has its own constraints. Identifying virulent genes from the entire genome sequence accessible in public databases and thus identifying novel drug targets has become a major thrust area, especially with the rise of swiftly acquiring pathogens of drug resistance. (5). The first step in the process of modern drug discovery is to identify potential drug targets subject to their validation and drug development.. A trend in the quest for drug targets in pathogens using computational methods has been noted recently, with a focus on genomic and proteomic information (6,7). Many researchers have used comparative / differential and subtractive genomics together with proteomics to identify drug targets in different pathogenic bacteria (8-12). More recently, sequence information from a single pathogen's various isolates given fresh insights into a species ' microevolution and helped scientists decipher its

virulence processes (13-14). Subtractive genomics has been used in the present research to define essential druggable proteins available in *Vibrio cholerae* pan-genome core proteins.

MATERIALS AND METHODS

Retrieval of the core protein of Vibirio cholerae

Complete proteome sequence of 44 strains of Vibrio cholera was downloaded from Panx:pang-genome analysis & Exploration database (15). Core proteins were retrieved from the pan-genome alignments. The core proteins were 101800 amino acid sequences. The core proteins were further pre-processed for paralogue removal at 60% identity by the cd-hit tool (16). Then less than 100 amino acids were removed. After pre-processing the sequences were reduced to 2241 amino acid sequences.

Prediction of non-homologus host proteins

BlastP was subjected to the collection of essential protein sequences of vibrio cholera against the human proteome database in order to discover the bacterial proteins that are not similar to a human host (Uniprot release 2016). A 10–5 cut-off expectation (E-value) was set as the standard for the identification of non-homologous proteins, an 11-gap penalty and a 1-gap extension penalty. The cut-off of e-value (10-5) was regarded on the basis of reported previous study procedure (17).

Evaluation of essential proteins

Major cellular functions of microbes are controlled by essential genes, which makes essentialness check a significant parameter for identifying potential durg targets. For Essentiality, the non-human homology sequences were verified using data acquired by sequence similarity from the Essential Genes Database (DEG) (18). Basic parameters used against DEG for BLAST were set as default, including 100 bit score and 1x 10-4 cut-off for E-value.

Druggability analysis

Drugability analysis will be useful to see the identified target have ability to bind a drug-like molecule. The drugability is analysed by similarity search with already existing drug molecules deposited in drug bank database by using BlastP with default parameters (19). Through blast search e-value more than 10-25 considered as druggability targets.

Sub-cellular localization Prediction

Drug-preferred proteins are expressed in cytoplasm based on subcellular localization, while vaccine applicants preferred at the surface of bacteria. The sub-cellular predictors like PSORTb 3.0 (20) and CELLO v.2.5 (21) was used. The prediction of sub-cellular will also helps in functional analysis of protein (22).

Molecular weight estimation

Protein with molecular weight less than 100 KDa are preferred for drug targets as it will be easy to purify and crystallize. Expasy tool (23) was used to deteremine molecular weight.

Evaluation of virulent factors

Virulence is the most important variable in the research of *Vibrio cholerae* pathogenesis Virulence factors of pathogenic bacteria were recognized using "Prediction of Virulence Factors, Information Molecules, and Cell Process and Metabolism Molecules in bacterial Proteins and VICMpred Tool" (24). The virulent variables allow a microorganism in a host of species to develop itself. Virulence factors include bacterial toxins, a protein on the cell surface that mediates bacterial attachment, carbohydrates on the cell surface, and proteins that safeguard bacterial attachment.

Prediction and analysis of 3D structures

Structural data is essential to protein objectives before immunogenic domains are predicted. Experimental structural database Protein Data Bank (PDB) verified the availability of crystalline structures for the chosen protein [25]. PDB obtained the accessible crystalline structures for the chosen proteins. Comparative homology structures were anticipated using Phyre2 server (26) for proteins lacking crystal structures.

RESULTS

Data retrieved from the Panx database revealed that vibirio cholera contains 2,452 core genes from 6,434 total genes. The core proteins from the core gene alignments contain 101800 amino acid sequences. This further processed with cd-hit resulted in 2423 sequences and sequences with less 100 amino acids were removed resulted in 2241 core proteins (Figure 1). These pre-processed core proteins were subjected to subtractive genomics approach through series of analysis as shown in Figure 2. 178 proteins were discovered to be non-homologous to proteins expressed by humans and remaining proteins were excluded through similarity search using BLAST against human proteome. Essential protein analysis was carried out on the sequences

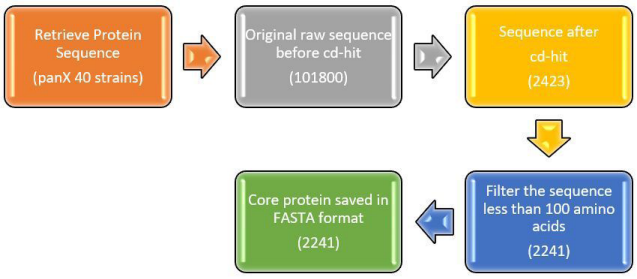

Figure 1: Retrieval and pre-processing of core proteins from the pan genome using Panx database and cd-hit tools

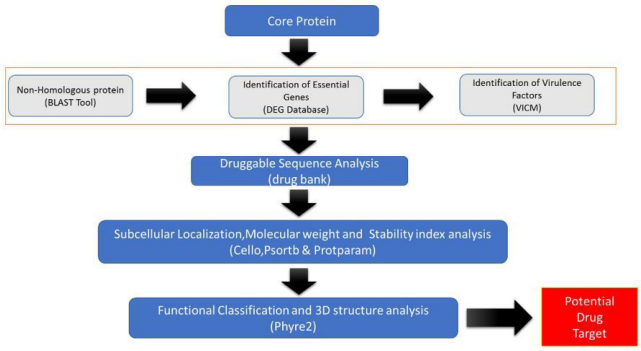

Figure 2: Workflow for Identification of drug target using subtractive genomics approach

extracted from non-homology against human proteome. Essential proteins are proteins that are necessary for any situation for a species ' survival. Using the Blast algorithm against Essential Genes Database (DEG), important proteins of vibrio cholera are recognized through similarity search. It led to the identification of 173 sequences of essential proteins and remaining as non-essential proteins. The druggability of non-homologous essential vibrio cholera proteins was evaluated by sequence similarity to small-molecule drug targets and similarity search were carried out using blastp against recognized non-homologous essential proteins with a database of drug bank. For subcellular localization, the druggable proteins were further analyzed using the prediction tool Cello and Psortb, which resulted in 108 proteins being kept located in the cytoplasm, excluding other membranes and unknown proteins. Then priority was calculated for the molecular weight for the expected drug targets. Proteins with less than 100 kDa molecular weight are regarded as more suitable targets. For virulence factor analysis, the anticipated drug targets were tested using the VICM pred tool, resulting in 24 sequences containing virulence and excluding other non-virulent proteins (Figure 3). Common proteins would be antibiotic targets of broad spectrum among several species. By blast searching against proteome data from 181 pathogenic bacteria (27) these 24 proteins were tested for broad-spectrum assessment, resulting in 11 proteins recognized as broad-spectrum proteins. These 11 proteins were structurally characterised using phyre2 server (Figure 4).

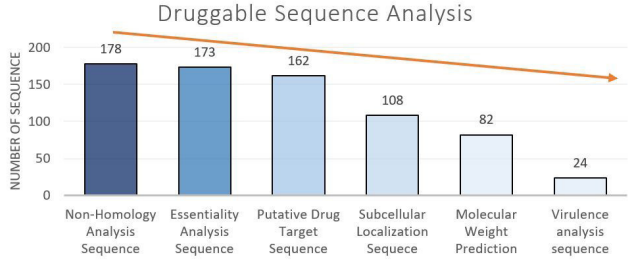

Figure 3: Number of proteins identified in each step of subtractive genomics

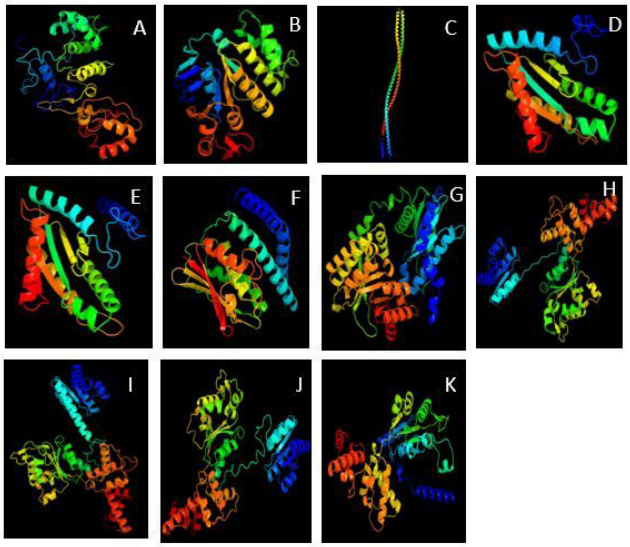

Figure 4: Structurally characterised 11 broad spectrum proteins using Phyre2 server. A. NZ_CP010812V-AB027-Transport Protein, B. NZ_CP026647C4E16 Transport Protein, C. NZ_CP020408A6J62 Signalling Protein, D. NZ_LT897797VCBK1071 Signalling Protein, E. NZ_CP013313ASZ79-Signalling Protein, F. NC_012668VCD-Transferase, G. NZ_LT897797VCBK1071 Transcription, H. NZ_CP009042IR04 Signalling Protein, I. NZ_CP009042IR04 Signalling Protein, J. NZ_CP013309ASZ87 Signalling Protein, K. NZ_CP025936C1H56

DISCUSSION

In this study, core proteins have been analysed from a complete proteome sequence of 44 Vibrio cholera strains. Core proteins have been acquired downloaded from Panx, extensive pan-genome analysis and exploration database (Figure 1). Using subtractive genomics, core proteins were used to identify drug targets as shown in figure 2. One of the comparative genomics approaches used to define new drug targets in the pathogenic organism is subtractive genomics or differential genome analysis (28,29). Identifying the non-host protein was the first step in subtractive genomics. Non-host proteins refer to bacterial proteins that have no human protein homology. If the homologous proteins are targeted, due to similarity with host proteins, they could severely affect the metabolism of the host. In vibirio cholera 178 proteins have no homology to human proteins have been recognized and these sequences are preferred as better objectives for drugs, as side effects and cross-reactivity induced by the use of antibiotics can be avoided to harm the host (30).

This computational identification of essential proteins is a good alternative technique over time-consuming and laborious experimental approach such as knockout of a single gene, mutagenesis of transposons and techniques of RNA interference. A query protein is regarded as vital in the method of fundamental proteome mining using

homology-based techniques if it is also present in another bacterium that is experimentally recognized as vital for survival. Analysis of essentiality shows some significant genes that the pathogen requires to perform significant tasks for survival, adhesion, host entry, infection, and host persistence. In *vibrio cholerae*, 173 vital proteins have been recognized in the pathogen life cycle that is connected to important metabolic pathways and are necessary for their survival. The essential proteins recognized play an important role in a host's pathogen survival. In identifying drug targets, these criteria are significant factors (31) .

DrugBank is a distinctive resource in bioinformatics and cheminformatics that combines detailed drug data with extensive drug target information. The druggability of non-homologous essential proteins of *V.cholerae* was identified by blastp similarity search to targets available in drug bank. A total of 162 proteins identified as druggable proteins which show high similarities with experimental compounds and approved drugs. The cell's sub-cellular location of proteins is a significant factor in identifying appropriate and efficient objectives for drugs. Identified 108 proteins from the CELLO and PsortB study as cytoplasmic proteins among the recognized druggable sequences. Due to the ease of purification and assay research, cytoplasmic proteins were preferred to membrane-localized proteins (32). Low molecular weight (< 100 kDa) was also chosen as the target protein's availability value (33). The drug target recognized was analyzed using expasy protparam and 82 proteins with low molecular weight and stable protein were predicted. Virulence is the pathogens ' main factor responsible for certain serious human illnesses. In our study, these characteristics were provided high priority to calculate future drug targets. In *V.cholerae*, 24 proteins were assessed as vital, and virulent proteins tend to be more effectively accountable for initiating the infection pathway than others (Figure 3). Also identified 11 broad-spectrum antibiotic candidates among 24 drug targets recognized by searching for similarity with common proteins among several species of bacteria (Table 1). These identified broad-spectrum proteins would be better targets as their inhibition of activity will hamper more than one system in the pathogen (34). The identified broad-spectrum targets were homology

Table I: Identified druggable drug target sequence by similarity search against drugbank

No.	ID	VICM	VIRULENTPRED	PREDICTION	PSORTB	CELLO	Drug Bank	
1	>NZ_CP010812VAB027_RS-164251ABC_transporter_ATP-binding_protein	0.4520658	0.8815	VIRULENT	Cytoplasmic Membrane / 7.88	Cytoplasmic/ 3.445	I. II. III. IV. V.	Cisplatin Rifampicin Nifedipine Sulfinpyrazone Verapamil
2	>NZ_CP026647C4E16_RS172652hemin_import_AT-Pbinding_protein_Hmu	0.087582334	1.0092	VIRULENT	CytoplasmicMem-brane /9.82	Cytoplasmic/3.788	I. II. III. IV. V.	Cisplatin Cholic Acid Glyburide Methotrexate Etoposide
3	>NZ_CP020408A6J62_RS-005852methylaccepting_che-motaxis_protein	1.4481392	0.7985	VIRULENT	Cytoplasmic Membrane / 7.88	Cytoplasmic / 2.375	I.	1,10-Phenanth-roline
4	>NZ_LT897797VCBK1071_RS035351diguanylate_cy-clase	0.12406278	0.4540	VIRULENT	Cytoplasmic / 9.97	Cytoplasmic / 1.329	I.	Guano-sine-5'-Mono-phosphate
5	>NZ_CP013313ASZ79_RS-048501GGDEF_domaincon-taining_protein	0.045789141	0.8274	VIRULENT	Cytoplasmic / 9.97	Cytoplasmic / 2.216	I.	Guano-sine-5'-Mono-phosphate
6	>NC_012668VCD_RS099101hybrid_sensor_his-tidine_kinase/response_reg-ulator	0.64389618	1.0030	VIRULENT	Cytoplasmic Membrane / 9.99	Cytoplasmic / 3.010	I.	Phosphoaspar-tate
7	>NZ_AP014524MS6_RS-028851biofilm_architecture_maintenance_protein_Mba	0.084949346	1.1270	VIRULENT	Cytoplasmic Membrane / 9.99	Cytoplasmic / 1.202	I.	Guano-sine-5'-Mono-phosphate
8	>NZ_CP009042IR04_RS-158651sigma54dependent_Fis_family_transcriptional_regulator	0.5402182	0.6897	VIRULENT	Cytoplasmic / 9.97	Cytoplasmic/ 4.040	I.	Glycerin
9	>NZ_CP009042IR04_RS-148701sigma54dependent_Fis_family_transcriptional_regulator	0.52212268	0.0843	VIRULENT	Cytoplasmic / 9.97	Cytoplasmic/ 3.640	I.	Glycerin
10	>NZ_CP013309ASZ87_RS-155952sigma54dependent_Fis_family_transcriptional_regulator	0.19514339	0.9708	VIRULENT	Cytoplasmic / 9.97	Cytoplasmic/ 3.162	I.	Glycerin
11	>NZ_CP025936C1H56_RS147602ATPase_	0.76224091	0.9505	VIRULENT	Cytoplasmic / 9.97	Cytoplasmic/ 3.117	I.	.Glycerin

modelled by using Phyre2 server (Figure 4).

CONCLUSION

In this study, we analysed core proteins of pan-genome of 41 strains of vibrio cholera and utilized the information in identifying the potential drug targets by integrating subtractive genomics approach and various proteomic bioinformatics tools as filters. Through this approach we identified 11 potential drug targets.

ACKNOWLEDGEMENTS

We would like to thank the Department of Diagnostic and Allied Health Sciences, Faculty of Health and Life Sciences, Management & Science University, Shah Alam, Selangor Darul Ehsan, Malaysia for providing us research facilities.

REFERENCES

1. Mandal S, Mandal MD, Pal NK. Cholera: a great global concern. Asian Pacific journal of tropical medicine. 2011 Jul 1;4(7):573-80.
2. Ali M, Nelson AR, Lopez AL, Sack DA. Updated global burden of cholera in endemic countries. PLoS neglected tropical diseases. 2015 Jun 4;9(6):e0003832.
3. Robinson TP, Bu DP, Carrique-Mas J, Fиvre EM, Gilbert M, Grace D, Hay SI, Jiwakanon J, Kakkar M, Kariuki S, Laxminarayan R. Antibiotic resistance is the quintessential One Health issue. Transactions of the Royal Society of Tropical Medicine and Hygiene. 2016 Jul 1;110(7):377-80.
4. Tagliabue A, Rappuoli R. Changing priorities in vaccinology: antibiotic resistance moving to the top. Frontiers in immunology. 2018;9.
5. Kuser CU, Ellington MJ, Cartwright EJ, Gillespie SH, Brown NM, Farrington M, Holden MT, Dougan G, Bentley SD, Parkhill J, Peacock SJ. Routine use of microbial whole genome sequencing in diagnostic and public health microbiology. PLoS pathogens. 2012 Aug 2;8(8):e1002824.
6. Katara P, Grover A, Kuntal H, Sharma V. In silico prediction of drug targets in *Vibrio cholerae*. Protoplasma. 2011 Oct 1;248(4):799-804.
7. Tang Y, Zhu W, Chen K, Jiang H. New technologies in computer-aided drug design: toward target identification and new chemical entity discovery. Drug discovery today: technologies. 2006 Sep 1;3(3):307-13.
8. Barh D, Tiwari S, Jain N, Ali A, Santos AR, Misra AN, Azevedo V, Kumar A. In silico subtractive genomics for target identification in human bacterial pathogens. Drug Development Research. 2011 Mar;72(2):162-77.
9. Jamal SB, Hassan SS, Tiwari S, Viana MV, de Jesus Benevides L, Ullah A, Turjanski AG, Barh D, Ghosh P, Costa DA, Silva A. An integrative in-silico approach for therapeutic target identification in the human pathogen Corynebacterium diphtheriae. PloS one. 2017 Oct 19;12(10):e0186401.
10. Muzzi A, Masignani V, Rappuoli R. The pan-genome: towards a knowledge-based discovery of novel targets for vaccines and antibacterials. Drug discovery today. 2007 Jun 1;12(11-12):429-39.
11. 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(5):255-8.
12. Koteswara RG, Nagamalleswara RK, Phani R, Krishna B, Aravind S. in silico identification of potential therapeutic targets inclostridium botulinumby the approach subtractive genomics. International Journal of Bioinformatics Research. 2010 Jul 1;2(2):12-6.
13. Tettelin H, Riley D, Cattuto C, Medini D. Comparative genomics: the bacterial pan-genome. Current opinion in microbiology. 2008 Oct 1;11(5):472-7.
14. Vernikos G, Medini D, Riley DR, Tettelin H. Ten years of pan-genome analyses. Current opinion in microbiology. 2015 Feb 1;23:148-54.
15. Ding W, Baumdicker F, Neher RA. panX: pan-genome analysis and exploration. Nucleic acids research. 2017 Oct 25;46(1):e5-.
16. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics. 2012 Oct 11;28(23):3150-2.
17. Perumal D, Lim CS, Sakharkar KR, Sakharkar MK. Differential genome analyses of metabolic enzymes in Pseudomonas aeruginosa for drug target identification. In silico biology. 2007 Jan 1;7(4, 5):453-65.
18. Zhang R, Lin Y. DEG 5.0, a database of essential genes in both prokaryotes and eukaryotes. Nucleic acids research. 2008 Oct 30;37(suppl_1):D455-8.
19. Wishart DS, Knox C, Guo AC, Shrivastava S, Hassanali M, Stothard P, Chang Z, Woolsey J. DrugBank: a comprehensive resource for in silico drug discovery and exploration. Nucleic acids research. 2006 Jan 1;34(suppl_1):D668-72.
20. Yu NY, Wagner JR, Laird MR, Melli G, Rey S, Lo R, Dao P, Sahinalp SC, Ester M, Foster LJ, Brinkman FS. PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. Bioinformatics. 2010 May 13;26(13):1608-15.
21. Yu CS, Chen YC, Lu CH, Hwang JK. Prediction of protein subcellular localization. Proteins: Structure, Function, and Bioinformatics. 2006 Aug 15;64(3):643-51.
22. Hossain MU, Omar TM, Alam I, Das KC, Mohiuddin AK, Keya CA, Salimullah M. Pathway based therapeutic targets identification and development of an interactive database CampyNIBase of

Campylobacter jejuni RM1221 through non-redundant protein dataset. PloS one. 2018 Jun 8;13(6):e0198170.
23. Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A. ExPASy: the proteomics server for in-depth protein knowledge and analysis. Nucleic acids research. 2003 Jul 1;31(13):3784-8.
24. Rashid MI, Naz A, Ali A, Andleeb S. Prediction of vaccine candidates against Pseudomonas aeruginosa: An integrated genomics and proteomics approach. Genomics. 2017 Jul 1;109(3-4):274-83.
25. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The protein data bank. Nucleic acids research. 2000 Jan 1;28(1):235-42.
26. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ. The Phyre2 web portal for protein modeling, prediction and analysis. Nature protocols. 2015 Jun;10(6):845.
27. Huynen M, Dandekar T, Bork P. Differential genome analysis applied to the species-specific features of Helicobacter pylori. FEBS letters. 1998 Apr 10;426(1):1-5.
28. Gleckman RA, Borrego F. Adverse reactions to antibiotics: clues for recognizing, understanding, and avoiding them. Postgraduate medicine. 1997 Apr 1;101(4):97-108.
29. Gustafson AM, Snitkin ES, Parker SC, DeLisi C, Kasif S. Towards the identification of essential genes using targeted genome sequencing and comparative analysis. Bmc Genomics. 2006 Dec;7(1):265.
30. Jamal SB, Hassan SS, Tiwari S, Viana MV, de Jesus Benevides L, Ullah A, Turjanski AG, Barh D, Ghosh P, Costa DA, Silva A. An integrative in-silico approach for therapeutic target identification in the human pathogen Corynebacterium diphtheriae. PloS one. 2017 Oct 19;12(10):e0186401.
31. Shanmugham B, Pan A. Identification and characterization of potential therapeutic candidates in emerging human pathogen Mycobacterium abscessus: a novel hierarchical in silico approach. PloS one. 2013 Mar 19;8(3):e59126.
32. Naz A, Awan FM, Obaid A, Muhammad SA, Paracha RZ, Ahmad J, Ali A. Identification of putative vaccine candidates against Helicobacter pylori exploiting exoproteome and secretome: a reverse vaccinology based approach. Infection, Genetics and Evolution. 2015 Jun 1;32:280-91.
33. Barh D, Kumar A. In silico identification of candidate drug and vaccine targets from various pathways in Neisseria gonorrhoeae. In silico biology. 2009 Jan 1;9(4):225-31.
34. Hadizadeh M, Tabatabaiepour SN, Tabatabaiepour SZ, Hosseini Nave H, Mohammadi M, Sohrabi SM. Genome-wide identification of potential drug target in enterobacteriaceae family: a homology-based method. Microbial Drug Resistance. 2018 Jan 1;24(1):8-17.