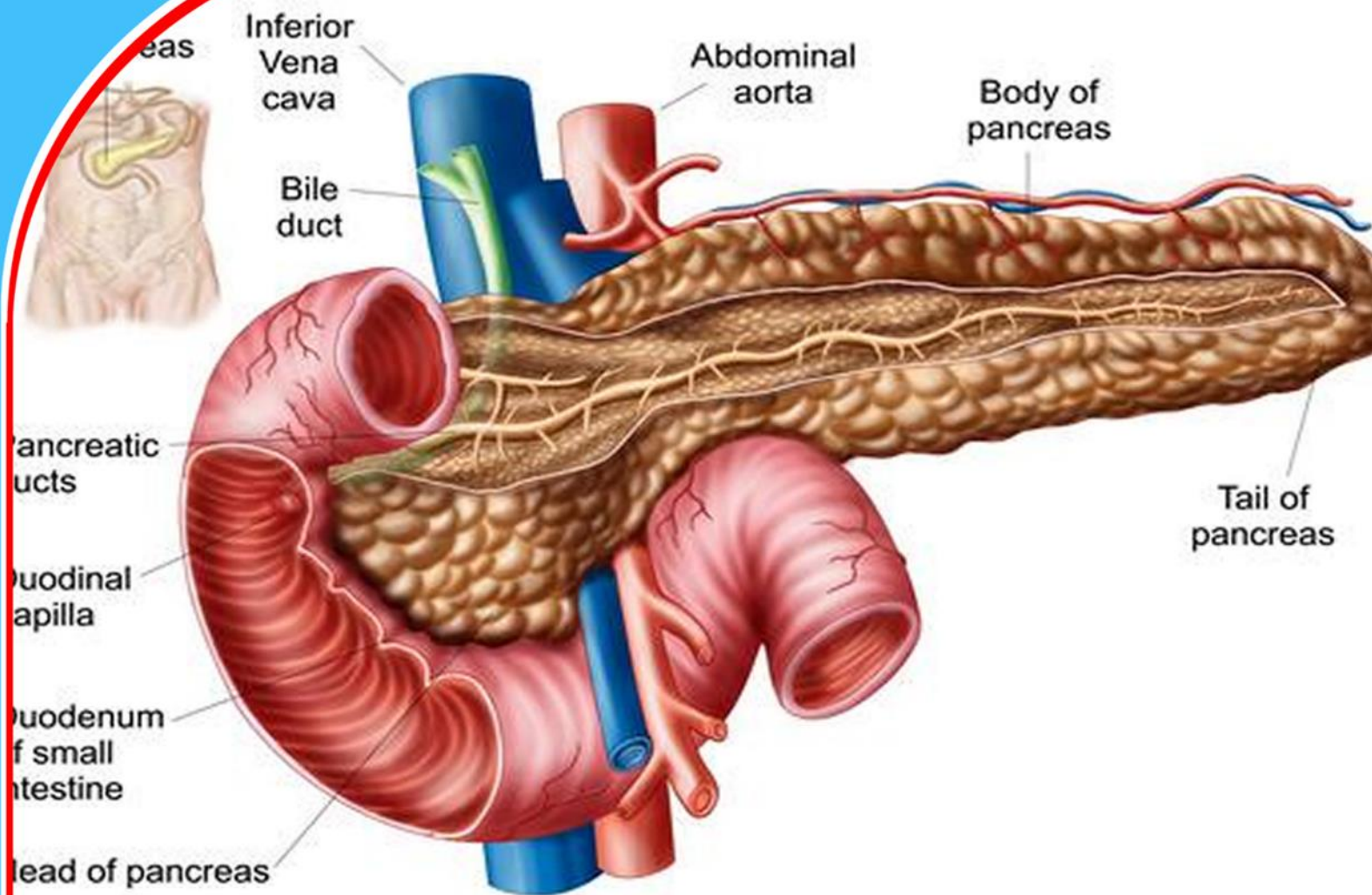


European Journal of Biology (EJB)



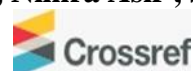
**The Identification of Common Druggable Targets For
Acute Dysentery in Four Pathogenic Bacteria, Shigella,
Salmonella, Campylobacter and E.Coli**

Iqra Shafique, Hira Shafique, Nimra Asif, Sadia Liaquat & Husnain Shahbaz



The Identification of Common Druggable Targets For Acute Dysentery in Four Pathogenic Bacteria, Shigella, Salmonella, Campylobacter and E.Coli

 Iqra Shafique^{1*}, Hira Shafique², Nimra Asif³, Sadia Liaquat⁴ & Husnain Shahbaz⁵



Article History

Submitted 28.08.2023 Revised Version Received 07.10.2023 Accepted 29.09.2023

Abstract

Purpose: Dysentery is a severe form of diarrhea and caused by four bacteria include *Shigella*, *Campylobacter*, *E.coli* and *Salmonella*. They are responsible for higher morbidity and mortality rates resulting from dysentery every year across the world. Antibiotic therapy of this disease plays a critical role in decreasing the prevalence as well as the fatality rate of this infection. However, the management of this disease remains challenging, owing to the overall increase in resistance against many antimicrobials. Hence, it has become important to identify as well as develop therapeutic methods presenting novel avenues for infections. In the current study, proteome based mapping was utilized to find the potential drug targets for dysentery. There is need to identify novel drug and vaccine targets to control this disease. This study is designed to identify new drug targets to develop drug and vaccines to battle bacillary dysentery. Subtractive genomics approach was used in this study to find novel drug targets.

Methodology: The proteomes of *Shigella*, *E. coli*, *Campylobacter* and *Salmonella* were retrieved from Uniprot. Paralogs in these proteomes were removed by CD-HIT. Gene essentiality was screened by Geptop server 2.0. Host-pathogen interaction was analyzed through

HPIDB database. To identify non-homologous proteins, the essential proteins were analyzed in Blastp against *Homo sapiens* (*H. sapiens*). The unique metabolic pathways were recognized by comparison of metabolic pathways of selected strains of bacteria and *H. sapiens* using the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway. Subcellular localization and functional classification of proteins involved in unique metabolic pathways were analyzed through PSORTb and SVMProt. Druggability potential of unique essential proteins was investigated using the DrugBank database.

Findings: Over all analysis showed 7 proteins that were common in all four bacteria. These proteins were essential to pathogens. Out of 7 proteins 6 proteins MURA, MURG, DAPA, MURB, DAPE, DnaA are reported as a drug target in different bacteria and one protein DAPD is novel. The study will enable the development of natural and cost-effective drugs against dysentery infections.

Recommendation: However, further validations are necessary to confirm their drug effectiveness and biocompatibility.

Keywords: KEGG, DEGG, PSORTb, Essential proteins, Metabolic pathways, SVMProt

1.0 INTRODUCTION

Dysentery also known as bloody diarrhea is usually initiated due to contamination by different bacteria and parasites. The most well-known parasitic reasons for dysentery are *Shigella*, *Campylobacter*, *E.coli* and *Salmonella*. These bacteria are well known to cause this disease in humans worldwide. Every year in the United States, it approximately causes around 450,000 cases, with around four cases for each 100,000 population (Kotloff et al., 2018). Every year around the world, they are evaluated to cause 165 million infections. In this study, *Shigella*, *Campylobacter*, *E.coli* and *Salmonella* bacteria were examined in order to find the drug targets which were common in all these four bacteria.

These are rod-shaped, non-motile, gram negative, non-spore-forming, and anaerobic bacteria. Infectious disease dysentery has become the most common cause of both developed and developing countries endemic disease. The main host and natural reservoir known for the bacteria is human gastrointestinal tract. They all mainly spread through the fecal route, sexual contact, and consumption of contaminated water or food. Nearly half or more strains are now resistant to multiple drugs throughout various parts of the world. There is currently no effective dysentery vaccine or drugs, but frequent hand washing can often prevent a person to person transmission. To combat this infection, there is a dire need to develop drugs that are common in all these bacteria and can be used against dysentery.

Shigella is commonly known as microorganism that is gram-negative, facultative, anaerobic, non-spore forming, non-motile, rod-shaped as well as extremely related to *E. coli*. It is named after Kiyoshi Shiga, who previously found it in 1897 (Yabuuchi, 2002). It is the major agent that causes human dysentery, *Shigella* produces disease in mammals. It is normally found in humans and gorillas (Erken,2008). During disease, it regularly produces dysentery. *Shigella* is also known as one of the foremost bacterial causes for bacterial dysentery worldwide, causing expected 80–165 million cases yearly. The number increases every year and it is estimated at the range of 74,000 and 600,000(Mani et al., 2016; McGuire et al., 2018).

It relies in the list of four microbes which is the root of moderate-to-extreme dysentery in south African and Asian children (Kotloff et al., 2018). The main species of *Shigella* are categorized by four serogroups. *S. dysenteriae* *S. flexneri* *S. boydii* and *S. sonnei* are four groups of this bacteria. *Shigella dysenteriae* and *boydii* are similar while *S. sonnei* could be distinguished on bases of biochemical assesses (Baron et al., 1996)

No vaccines or drugs are available for dysentery. The best way to prevent this disease is washing hands with soap and water thoroughly, repeatedly and cautiously, before handling the food, after and before using the washroom (Sati et al., 2019). Strict compliance with standard food and water safety precautions is also important. Water from lakes, untreated swimming pools, or ponds should also not be swallowed. Avoid having sex with diarrhea patients (Kotloff et al., 2018).

Various approaches like comparative along with the subtractive genomics have been utilized to find the targets in several human microbes. It is the newest methodology used to detect novel pathogen drug targets. These approaches are used to select potential drug targets that are used in clinical studies. The drug targets are identified based on the determination of the non-homologous and essential proteins in pathogenic organisms. A crucial step in the process of computer-aided drug design is to identify drug targets. The need for a rapid search for small molecules bound to objects of biological interest is crucial in the drug discovery process. The accessibility of

pathogenic genome sequences has provided a wide range of knowledge that could be helpful in identifying targets for vaccines and drugs. The difference between host and pathogen proteins can be effectively used to design pathogen-specific drugs (Jamal et al., 2017). The main objective of this study is to identify the targeted essential genes that show no homology with the *Homo sapiens* so that the drug targets can be used with minimum off-target effects in humans. Here, we employed an integrated *in-silico* subtractive genomics approach to the entire proteome of four bacteria to find the cure of bacillary dysentery.

2.0 METHODOLOGY

The identification of common druggable targets for acute dysentery in four pathogenic bacteria, *Shigella*, *Salmonella*, *Campylobacter* and *E.Coli* was the main purpose of this study. The subtractive genomic approach was used in order to meet the results. The complete framework of the methodology used in this study was illustrated in figure 1.

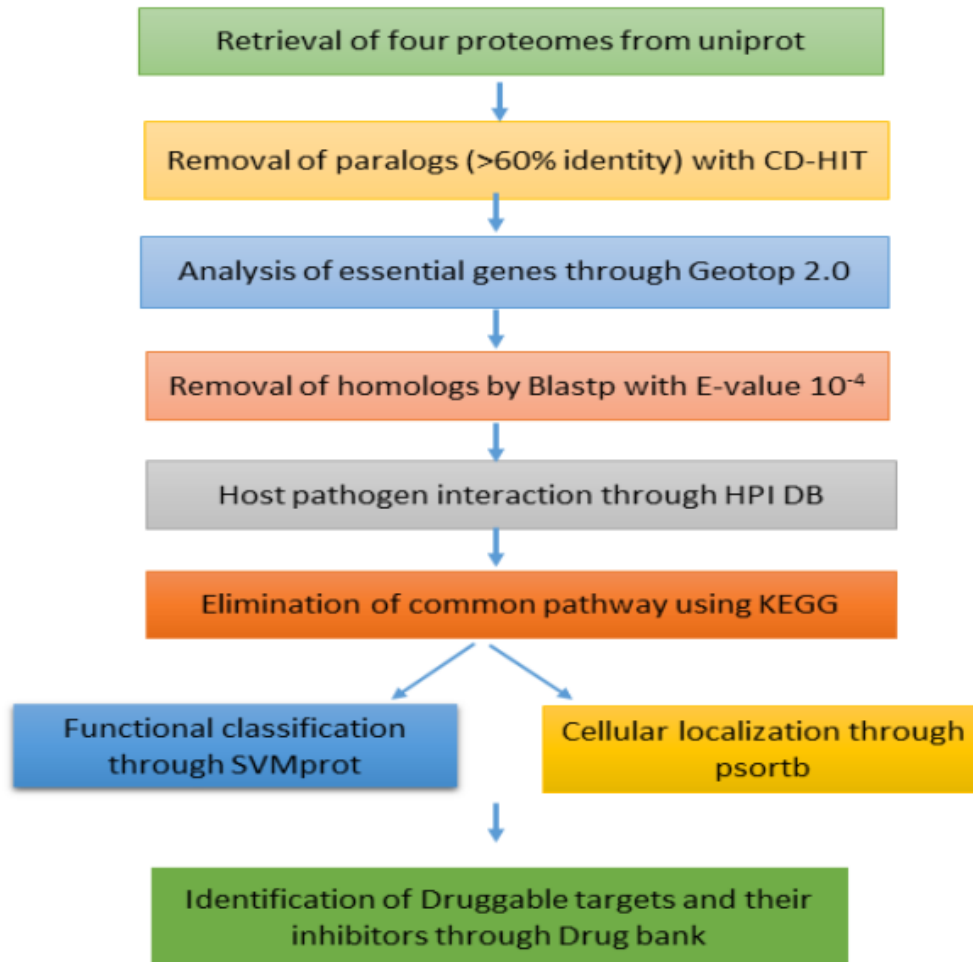


Figure 1: Schematic Representation of Workflow Used in this Research

Sequences Retrieval and Excluding Paralogs

Whole sequences of *Shigella*, *Campylobacter*, *E.coli* and *Salmonella* were retrieved from UniProtKB (<https://www.uniprot.org/help/uniprotkb>) in FASTA format (Bateman et al., 2017).

To exclude paralogous sequences from *Shigella*, *Campylobacter*, *E.coli* and *Salmonella* proteome, the sequences of set0 were subjected to CD-HIT analysis. The CD-HIT analysis was carried out with identity cutoff of 0.6 which was able to remove redundant sequences with more than 60% identity from set0 (Huang et al., 2010). The remaining proteins were listed as set1 as non-paralogous protein sequences.

Identification of Essential Proteins

Druggable targets are always made from essential genes. These essential genes are involved in the proper functioning of body. For this, druggable target cannot be made from non-essential genes. Hence non-essential genes were removed from proteomes by Geptop 2.0 server. Essential proteins are necessary to organisms for their survival and are considered the basis of life. Geptop 2.0 server (<http://cefg.uestc.cn/geptop/>) was used to retrieve essential proteins. The essential proteins were listed as set2 (Wen et al., 2019).

Identification of Proteins Sequences Non-Homologous to the Human Proteome

This step aimed to avoid functional similarity with human proteome to prevent bindings of therapeutic compounds to the active site of the host homologous proteins. Non-paralogous essential proteins of the pathogen were analyzed using BLASTp. The proteomes were then subjected to BLASTp against the *Homo sapiens* proteome using RefSeq database. The proteins having an E-value of 10^{-4} were removed as they are considered to have a certain level of homology with host genome. The resultant data consisting of sequences of pathogens having zero homology with host proteome were listed as set3 (Salgaonkar et al., 2011).

Analysis of Unique Metabolic Pathway

The analysis of metabolic pathways of the non-human homologous essential proteins were computed at KEGG. KEGG is the Kyoto Encyclopedia of Genes and Genomes that contains the complete metabolic pathways present in living organisms. Those metabolic pathways that were unique and are absent in humans were listed as set4 (Kanehisa et al., 2017).

Host-Pathogen Interaction

"HPIDB" a host-pathogen PPI database. It is used as integrated source to determine interaction between host and pathogen. Specially, HPIDB assimilates investigational PPIs from numerous open databases into single, non-redundant web available resource. This database could also be searched with diversity of options such as sequence identifiers, symbol, taxonomy, publication, author, or interaction type. HPIDB permits user to search protein sequences using BLASTP to retrieve homologous host/pathogen sequences. For high-throughput analysis, the user can search multiple protein sequences at a time using BLASTP and obtain results in tabular and sequence alignment formats. The resultant proteins were subjected to this database to analyze the level of interaction between host and pathogen (Kumar & Nanduri, 2010).

Subcellular Localization Analysis

The program PSORTb was used for subcellular localization prediction. It provides essential information concerning the function of proteins and is helpful in the annotation of genome and designing several proteome experiments as well as in the identification of certain diagnostic potential drug vaccine targets for the proteins of bacterial pathogens. The functional annotation of essential proteins was also retrieved. It was essential to determine subcellular localization as it

predicts the type of druggable targets. So, if the protein is present in outer membrane, it is possible that it can be highlighted as a vaccine. However, the proteins present in the cytoplasm (set6) were identified for further analysis (Yu et al., 2010).

Protein Functionality Analysis

The tool SVMprot <http://bidd.group/cgi-bin/svmprot/svmprot.cgi> was used to determine the molecular functions and biological processes of resultant proteins (Blum et al., 2020).

Identification of Druggable Targets

The proteins having cytoplasmic localization were subjected to DrugBank. DrugBank (<https://www.drugbank.com/>) is an online database used to find preferential druggable targets. E.value of 10^{-5} was set and the druggable targets for pathogens were obtained that are causing bacillary dysentery disease in humans (Wishart et al., 2018).

3.0 FINDINGS

Inquisition of novel therapeutic targets against *Shigella*, *Campylobacter*, *Salmonella* and *E.coli* was the preliminary purpose of this study, and here a subtractive genomic approach was used for the whole proteome by using several online databases and computational tools. A brief insight and findings of *Shigella* are shown in Table 1.

Table 1: Subtractive Genomic Analysis Scheme Towards the Identification of Novel Drug Target Against *Shigella*

| Sr No. | Subtractive Approach | Bioinformatics Tools | No. of Protein |
|--------|---|------------------------------------|----------------|
| 1. | Whole proteome of <i>shigella</i> | UniportKB | 5894 |
| 2. | Removal of paralogs | CD-HIT tool | 1892 |
| 3. | Essential genes | Geptop 2.0 server | 418 |
| 4. | Proteins non homologous to <i>H.sapiens</i> | BLASTP | 247 |
| 5. | Host pathogen interaction | Host pathogen interaction database | 56 |
| 6. | Unique metabolic pathway | KEGG | 14 |
| 7. | Subcellular localization (Cytoplasmic) | PSORTB | 14 |
| 8. | Druggability analysis | DrugBank | 12 |

Table 2: Subtractive Genomic Analysis Scheme Towards the Identification of Novel Drug Target Against *Campylobacter*

| Sr no. | Subtractive Approach | Bioinformatics Tools | No. of Protein |
|--------|---|------------------------------------|----------------|
| 1. | Whole proteome of <i>shigella</i> | UniportKB | 4048 |
| 2. | Removal of paralogs | CD-HIT tool | 1789 |
| 3. | Essential genes | Geptop 2.0 server | 474 |
| 4. | Proteins non homologous to <i>H.sapiens</i> | BLASTP | 118 |
| 5. | Host pathogen interaction | Host pathogen interaction database | 39 |
| 6. | Unique metabolic pathway | KEGG | 11 |
| 7. | Subcellular localization (Cytoplasmic) | PSORTB | 11 |
| 8. | Druggability analysis | DrugBank | 10 |

Table 3: Subtractive Genomic Analysis Scheme Towards The Identification of Novel Drug Target Against *E.Coli*.

| Sr no. | Subtractive Approach | Bioinformatics Tools | No. of Protein |
|--------|---|------------------------------------|----------------|
| 1. | Whole proteome of <i>shigella</i> | UniportKB | 26120 |
| 2. | Removal of paralogs | CD-HIT tool | 8965 |
| 3. | Essential genes | Geptop 2.0 server | 1489 |
| 4. | Proteins non homologous to <i>H.sapiens</i> | BLASTP | 549 |
| 5. | Host pathogen interaction | Host pathogen interaction database | 118 |
| 6. | Unique metabolic pathway | KEGG | 18 |
| 7. | Subcellular localization (Cytoplasmic) | PSORTB | 18 |
| 8. | Druggability analysis | DrugBank | 15 |

Table 4: Subtractive Genomic Analysis Scheme Towards the Identification of Novel Drug Target Against *Salmonella*

| Sr no. | Subtractive approach | Bioinformatics tools | No. of Protein |
|--------|---|------------------------------------|----------------|
| 1. | Whole proteome of <i>shigella</i> | UniportKB | 13314 |
| 2. | Removal of paralogs | CD-HIT tool | 5886 |
| 3. | Essential genes | Geptop 2.0 server | 1246 |
| 4. | Proteins non homologous to <i>H.sapiens</i> | BLASTP | 389 |
| 5. | Host pathogen interaction | Host pathogen interaction database | 78 |
| 6. | Unique metabolic pathway | KEGG | 8 |
| 7. | Subcellular localization (Cytoplasmic) | PSORTB | 8 |
| 8. | Druggability analysis | DrugBank | 7 |

Removal of Paralogous Sequence

The proteomes of *Shigella*, *Campylobacter*, *Salmonella* and *E.coli* were retrieved from uniprot consisted of 5894, 4048, 13314 and 26120, sequences respectively. The paralogs were removed from total proteome sequences retrieved from uniprot with the help of CD-HIT analysis one by one. The overall results showed 1892, 1789, 8965, 5886 non-redundant sequences in *Shigella*, *Campylobacter*, *E.coli* and *Salmonella* respectively.

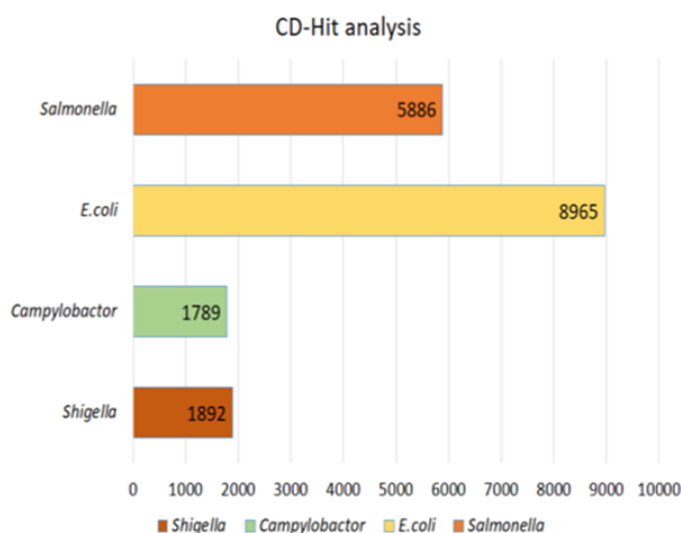


Figure 2: Paralogous Sequence Removed

Selection of Essential Genes

Antibacterial compounds are mainly prepared to dock and inhibit essential gene products. Essential proteins are therefore regarded as the most advantageous drug target. The Geptop 2.0 server was used to identify essential proteins. Four bacterial proteomes were subjected to the server one by one and results were obtained in a well precise manner. A total of four lists were acquired. The number of essential proteins in *Shigella*, *Campylobacter*, *Salmonella* and *E.coli* were 418, 474, 1246 and 1489 respectively.

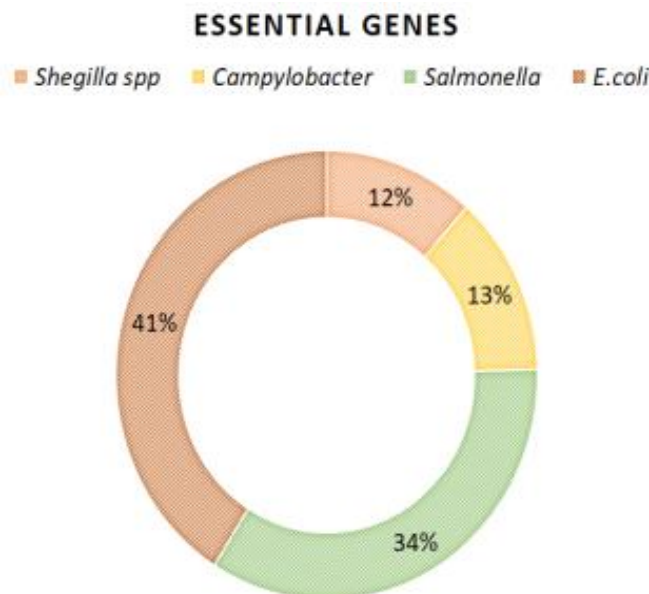


Figure 3: Essential Genes

Removal of Human Homologous Protein

Homology between host-pathogen protein results in undesirable cross-reaction as well as cytotoxicity. Therefore, drugs developed and helped to bind the target proteins of pathogens must evade cross-reactivity with host homologous proteins. The Blastp analysis was done with Blastp RefSeq database. The essential proteins were subjected to Blastp and analyzed with human genome. A total of 247, 118, 389 and 549 sequences were found among *Shigella*, *Campylobacter*, *Salmonella* and *E.coli* respectively.

Host-Pathogen Interaction

The non-homologous sequences were subjected to HPIDB database which is used to provide information about proteins which have interactions between host and pathogen proteomes. Thus, the proteins similar to humans were discarded. The total of 56, 39, 78 and 118 of *Shigella spp*, *Campylobacter*, *Salmonella* and *E.coli* proteins were left.

Analysis of Unique Metabolic Pathway

The KEGG database was helpful in identifying the proteins involved in common and uncommon pathways. The KEGG server was helpful to find the pathways which were unique to humans but present in pathogen. It provided the list of metabolic pathways that were common and unique in

humans. A total of 56, 39, 78 and 118 proteins were subjected to KEGG server and only 14, 11, 8 and 18 and were unique in *Shigella*, *campylobacter*, *Salmonella* and *E.coli*.

Analysis of Subcellular Localization

The localization of leftover unique proteins were analyzed and identified. PSORTb server was used to identify the localization. All of the 14, 11, 8 and 18 proteins were subjected to PSORTb one by one and a precise result was found. All proteins were cytoplasmic except MURG which was cytoplasmic membrane. Those proteins were preferred in this regard which were associated with localization in cytoplasm and cytoplasmic membrane.

Druggability Analysis

When the localization was identified, druggable targets were retrieved using Drugbank. Out of 14 proteins of *Shigella* 12 were found as targets. Out of 11 proteins of *Campylobacter* 10 were found as targets. In *Salmonella* 7 targets were found out of 8 proteins. Out of 18 proteins of *E.coli* 15 targets were found (Table 4.6). In all these proteins only 7 was found common in four bacteria. The proteins are MURA, MURG, DAPA, DAPE, DAPD, MURB, DnaA.

Discussion

Dysentery affects millions of people annually and causes several deaths. Mild cases are treated without antimicrobial agents and their recovery is quite rapid. Antibiotics control the spread of this disease inside the intestine as well as reduce the duration required for it to enter the entire body. The use of anti-diarrheal medications is not recommended as it may aggravate the disease (Sati et al., 2019). There is currently no effective vaccine for dysentery, but frequent and thorough hand washing can prevent person-to-person transmission.

To combat this life-threatening situation, there is a dire need to develop drugs against dysentery immediately. In this study, the subtractive genomics approach was employed to screen the drug targets against four bacteria that are causing dysentery. This approach was utilized to find the targets based on the determination of essential as well as non-homologous proteins within pathogenic organisms. Identifying drug targets is a crucial step in the computer-based drug designing procedures (Hosen et al., 2014). Recent advances in the disciplines of bioinformatics as well as computational biology have created a variety of approaches to drug design and *in silico* analysis, reducing the time and expense associated with trial and error of ions devoted to drug development (Barh et al., 2011).

In recent years, a large amount of preferential druggable targets are identified for those bacteria which either provide resistance against the drug or for which no vaccine is available and they are associated with a specific pathogen. The *in silico* subtractive genomics was applied to find drug targets in four bacteria that were *Shigella*, *Salmonella*, *E. coli* and *Campylobacter* as it was cost effective for target identification as well as easy approach completed in a short time, thus reducing the research time. As discussed above, subtractive genomics involves subtraction of sequences. So at the end those sequences were selected in this study which was unique, non-homologous and essential for bacterial pathogens *Shigella*, *Salmonella*, *E. coli* and *Campylobacter*. The result of this study was quite intriguing. Out of more than twenty thousand sequence proteomes downloaded, 7 potential drug able targets were found. Out of 7 targets, 6 were reported in many other disease-causing bacteria which explain 58 the significance of this approach. It limits the proteins in small numbers making the discovery of druggable targets easy.

REFERENCES

- Bateman, A., Martin, M. J., O'Donovan, C., Magrane, M., Alpi, E., Antunes, R., Bely, B., Bingley, M., Bonilla, C., Britto, R., Bursteinas, B., Bye-AJee, H., Cowley, A., Da Silva, A., De Giorgi, M., Dogan, T., Fazzini, F., Castro, L. G., Figueira, L., ... Zhang, J. (2017). UniProt: The universal protein knowledgebase. *Nucleic Acids Research*, *45*(D1), D158–D169. <https://doi.org/10.1093/nar/gkw1099>
- Blum, M., Chang, H.-Y., Chuguransky, S., Grego, T., Kandasaamy, S., Mitchell, A., Nuka, G., Paysan-Lafosse, T., Qureshi, M., Raj, S., Richardson, L., Salazar, G. A., Williams, L., Bork, P., Bridge, A., Gough, J., Haft, D. H., Letunic, I., Marchler-Bauer, A., ... Finn, R. D. (2020). The InterPro protein families and domains database: 20 years on. *Nucleic Acids Research*, 1–11. <https://doi.org/10.1093/nar/gkaa977>
- Hosen, M. I., Tanmoy, A. M., Mahbuba, D. Al, Salma, U., Nazim, M., Islam, M. T., & Akhteruzzaman, S. (2014). Application of a subtractive genomics approach for in silico identification and characterization of novel drug targets in *Mycobacterium tuberculosis* F11. *Interdisciplinary Sciences: Computational Life Sciences*, *6*(1), 48–56. <https://doi.org/10.1007/s12539-014-0188-y>
- Huang, Y., Niu, B., Gao, Y., Fu, L., & Li, W. (2010). CD-HIT Suite: A web server for clustering and comparing biological sequences. *Bioinformatics*, *26*(5), 680–682. <https://doi.org/10.1093/bioinformatics/btq003>
- Jamal, S. B., Hassan, S. S., Tiwari, S., Viana, M. V., De Jesus Benevides, L., Ullah, A., Turjanski, A. G., Barh, D., Ghosh, P., Costa, D. A., Silva, A., Röttger, R., Baumbach, J., & Azevedo, V. A. C. (2017). An integrative in-silico approach for therapeutic target identification in the human pathogen *Corynebacterium diphtheriae*. *PLoS ONE*, *12*(10), 1–25. <https://doi.org/10.1371/journal.pone.0186401>
- Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., & Morishima, K. (2017). KEGG: New perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Research*, *45*(D1), D353–D361. <https://doi.org/10.1093/nar/gkw1092>
- Kotloff, K. L., Riddle, M. S., Platts-Mills, J. A., Pavlinac, P., & Zaidi, A. K. M. (2018). Shigellosis. *The Lancet*, *391*(10122), 801–812. [https://doi.org/10.1016/S0140-6736\(17\)33296-8](https://doi.org/10.1016/S0140-6736(17)33296-8)
- Kumar, R., & Nanduri, B. (2010). HPIDB - a unified resource for host-pathogen interactions. *BMC Bioinformatics*, *11*(SUPPL. 6). <https://doi.org/10.1186/1471-2105-11-16>
- Salgaonkar, B. B., Kabilan, M., & Braganca, J. M. (2011). Basic local alignment search tool. In *World journal of microbiology & biotechnology* (Vol. 27, Issue 4, pp. 403–410). <https://doi.org/10.1038/nbt0388-282>
- Wen, Q. F., Liu, S., Dong, C., Guo, H. X., Gao, Y. Z., & Guo, F. B. (2019). Geptop 2.0: An updated, more precise, and faster Geptop server for identification of prokaryotic essential genes. *Frontiers in Microbiology*, *10*(JUN), 1–6. <https://doi.org/10.3389/fmicb.2019.01236>

- Wishart, D. S., Feunang, Y. D., Guo, A. C., Lo, E. J., Marcu, A., Grant, J. R., Sajed, T., Johnson, D., Li, C., Sayeeda, Z., Assempour, N., Iynkkaran, I., Liu, Y., MacIejewski, A., Gale, N., Wilson, A., Chin, L., Cummings, R., Le, Di., ... Wilson, M. (2018). DrugBank 5.0: A major update to the DrugBank database for 2018. *Nucleic Acids Research*, 46(D1), D1074–D1082. <https://doi.org/10.1093/nar/gkx1037>
- Yabuuchi, E. (2002). *Bacillus dysentericus* (sic) 1897 was the first rather than *Bacillus dysenteriae* 1898. *International Journal of Systematic and Evolutionary Microbiology*, 52(3), 1041. <https://doi.org/10.1099/ijs.0.02052-0>
- Yu, N. Y., Wagner, J. R., Laird, M. R., Melli, G., Rey, S., Lo, R., Dao, P., Cenk Sahinalp, S., Ester, M., Foster, L. J., & Brinkman, F. S. L. (2010). PSORTb 3.0: Improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics*, 26(13), 1608–1615. <https://doi.org/10.1093/bioinformatics/btq249>