ABSTRACT

Helicobacter pylori (H.pylori), a gram-negative bacterium, takes up residence in the human stomach and can potentially persist for an individual's entire life if not treated. It is responsible for various gastrointestinal disorders such as gastritis, peptic ulcers, and gastric cancer. The relationship between H. pylori and its host is intricate, resulting in varying clinical consequences; some people show no symptoms while others face serious complications. The rise of antibiotic resistance in H.pylorire requires immediate attention to find an effective drug target. Fortunately, the use of in silico genomics approach can help prevent the challenges faced in developing new drugs and encountering host toxicity. In this study, we utilized an advanced genome subtraction technique to identify potential drug targets specifically targeting H.pylori pathogens. In this study, a set of subtraction techniques were employed to narrow down the whole genome of the pathogen. The methods involved focusing on specific attributes such as paralogous proteins that exhibit significant similarity with human proteins, essential functions, drug-like characteristics, non-virulent properties, and anti-target proteins. Through these steps, only 14 highly promising drug targets were identified from the entire genome analysis. The protein that has been identified provides a foundation for finding a potential drug candidate which could potentially inhibit it, leading to the eradication of otitis media caused by drug-resistant H.pylori. Despite this, this study establishes a framework for identifying future targets in order to aid wet-lab research going forward.

KEYWORDS: *drug targets, H.pylori, subtractive genomics, resistance*

1.0 INTRODUCTION

The spiral microorganism known as Helicobacter pylori (H. pylori) was initially discovered by Giulio Bizzozero in 1892 when he identified it in the stomachs of dogs [1]. In 1983, Barry Marshall and Robin Warren classified these organisms as Campylobacter-like spirals and named them Campylobacter pyloridis [2]. However, in 1989 Goodwin et al. renamed them "Helicobacter pylori" due to their helical structure and predominant presence in the stomach's pyloric region [3]. H. Pylori is a Gram-negative microorganism with a short-helical or S-shaped form that measures approximately 0.5-1 µm wide and 2-4 µm long; It infects over half of the global population [4]. The distinctive morphology of H. pylori makes it highly capable of thriving in the harsh conditions of the stomach. With its spiral shape, polar flagella, and curved rod structure, this bacterium can efficiently maneuver, attach to surfaces, and engage with host cells. These morphological features are crucial for H. pylori's ability to cause different gastrointestinal disorders and exemplify its intricate interactions within the human body.

In 1991 and 1994, researchers investigated the relationship between H. pylori and gastric cancer. The International Agency for Research on Cancer, a branch of the World Health Organization, concluded that H. pylori is carcinogenic in humans based on epidemiological data [5,6]. In 2009, this finding was further confirmed. In the United States in 1994, the National Institute of Health reported that treating H.pylori may be essential as it could potentially be the main cause of peptic ulcer disease. Moreover, H.pylori also plays a significant part. The bacterium can lead to various diseases including gastritis, and mucosa-associated lymphoid tissue (MALT) lymphoma. [7].

The exact mode of transmission for H. pylori is not known, but it is believed that the common routes are through fecal-oral or oral-oral contact via consumption of water or food [8]. The prevalence of H. pylori infection increases with age and is higher in societies with low socioeconomic status [9]. Its ability to survive in the stomach and cause chronic inflammation indicates resistance against both immune response and acid levels [10]. Various antibiotic treatments are used for treating H. pylori infections; however, studies indicate a rapid increase in strains resistant to these antibiotics [11,12]. Consequently, researchers have been exploring alternative agents that can provide safer and more effective results when combined with antibiotic treatments [13].

The existence of antibiotic resistance in H. pylori is influenced by a complex interplay of multiple factors. One key factor is the bacterium's ability to undergo genetic variation, fueled by its high mutation rate and diverse strains, which provides a breeding ground for mutations that result in antibiotic resistance. At the same time, the widespread use of antibiotics across various domains creates selective pressure on H. pylori populations, favoring the survival and spread of strains with traits resistant to these drugs. Additionally, H.pylori's capacity for horizontal gene transfer - exchanging genetic material with other bacteria - further expedites the acquisition of genes conferring antibiotic resistance and contributes to an increased emergence of this kind of resistance.

The emergence of antibiotic resistance in H. pylori necessitates the urgent prioritization of an effective drug target. Fortunately, by employing a subtractive genomics approach, issues such as new drug failures and host toxicity that are commonly associated with traditional methods of drug development can be avoided. In this study, an advanced in silico genome subtraction technique was utilized to identify potential targets specific to H. pylori for the development of suitable drugs against it.

2.0 METHODOLOGY

2.1 Sequence Retrieval

The National Center for Biotechnology Information (NCBI) database provided the dataset consisting of H. pylori sequences in FASTA format. To ensure relevance to this research focus, specific search criteria related to H.pylori were used during the retrieval process. The widely recognized standard for representing biological sequences known as the FASTA format was utilized for obtaining these sequences. Each sequence's header contained a unique identifier and the corresponding genetic sequence data followed this information.

2.2 Identification of Paralogous Sequence

In order to guarantee the integrity and accuracy of the data, a thorough process for cleaning was carried out. As a component of this procedure, batch processing using the NCBI Entrez system was employed to refine and eliminate any possible duplicates or irrelevant entries from the dataset. After undergoing the data refinement process, the dataset underwent streamlining. These enhancements are expected to strengthen subsequent research stages by providing more reliable and precise analyses.

2.3 Removal of paralogous sequence with CD-Hit

CD-Hit was used to apply a sequence similarity threshold of 0.75, meaning that sequences with 75% or higher similarity were identified as redundant and removed from the dataset. This stringent threshold aimed to preserve only the most distinct and informative sequences, while effectively reducing the likelihood of closely related sequences being overrepresented. (https://galaxy.pasteur.fr/)

2.4 Removal of sequence with less than 100 amino acids with Europe Galaxy

The primary aim of this step was to remove sequences that had less than 100 amino acids. This decision was based on the understanding that sequences below this threshold may not offer significant information or be dependable targets for the research goals. The European Galaxy platform made it easier to filter out sequences that did not meet the minimum length requirement, resulting in a more efficient and systematic removal process. By applying this filter, only those sequences with a significant length were included in the dataset, ensuring a stronger basis for further investigations. (https://usegalaxy.eu/)

2.5 Subtractive Genomics

Subtractive genomics is a computational strategy used in bioinformatics to identify distinct features of a particular target organism by systematically comparing its genetic information with that present in a reference database. The study utilized the Protein-Based Inference Tool (PBIT) to comprehensively assess various features of the target organism's genetic composition. PBIT is an advanced bioinformatics tool developed for analyzing genetic sequences and corresponding protein results. With its wide range of capabilities, it enables thorough examination of multiple aspects of genetic data. (Pipeline builder for identification of target (bicnirrh.res.in). The process begins by creating a step-by-step pipeline (online mode). The results obtained from the European Galaxy platform after removing extended protein sequences are utilized at this point. In continuation, specific pathways within this workflow are systematically selected to advance subtractive genomic analysis. These include conducting analyses that compare against non-human proteomes instead of human proteomes, assessing essentiality and virulence factors, evaluating drugability potential, differentiating from microbiota proteomes accordingly analyzing various spectrums as well investigating host-pathogen interactions in order to differentiate it further from non-human anti-targets.

2.6 KASS (KEGG Automatic Annotation Server) Tool

The KAAS tool was used to compare H. pylori and HSA (Homo sapiens). This analysis produced KO assignments and KEGG pathways automatically, offering insights into the functional characteristics and genetic information pathways of both organisms. After using the KAAS tool to compare H. pylori and HSA, a filtering process was applied to enhance the accuracy of the results obtained. This involved examining gene sequences that were assigned the same KO (KEGG Orthology) in both organisms. Genes with identical KO assignments were identified as having common functional traits. In order to refine the data, gene sequences that had the same KO assignments were excluded from further analysis.

2.7 Prediction of subcellular localization Psortb

After excluding gene sequences with identical KO assignments, the remaining ones were analyzed further using the Psortb tool. To emphasize unique functional characteristics, these refined sequences were converted to FASTA format and then entered into the Psortb platform for comprehensive localization prediction. Psortb specifically targeted gram-negative bacteria, selecting them for classification. The criteria matched the characteristics of H. pylori - a gram-negative bacterium - ensuring that the analysis focused on attributes unique to this particular organism. During the following stage of the process, further examination was conducted on the outcomes produced by Psortb. In particular, focus was given to sequences that were predicted to have only cytoplasmic localization. Identifying these sequences signified the concluding stage of the analysis.

3.0 RESULTS

3.1 Preprocessing Steps

Initial Number of Sequence obtained from NCBI	1427
Number of sequences after removal from Batch Entrez	1424
Number of sequences after the removal of gene duplication with CD-Hit Removal	1416
Number of sequences after the removal of the sequence below 100 amino acid	1294

3.2 Subtractive Genomics

Non-homology against human proteome	1228
Essentiality	907
Virulence factor	254
Druggability Analysis	151
Non-homology against gut microbiota proteomes	45
Broad Spectrum Analysis	45
Host-Pathogen Interaction	37
Non-homology against human anti-targets	37

3.3 Automatic annotation genes and subcellular localization

Removal of sequence with KEGG orthology	HSA	6	HPY	25
Removal after the subcellular localization of PsortB	14			

3.4 The genes selected are based on their subcellular localization, specifically those with a final cytoplasmic prediction.

No	ID	Name
1	WP_000133864.1	cag pathogenicity island type IV secretion system ATPase VirB11 [Helicobacter pylori]
2	WP_000688273.1	aminodeoxychorismate/anthranilate synthase component II [Helicobacter pylori]
3	WP_001169746.1	copper response regulator transcription factor CrdR [Helicobacter pylori]
4	WP_001959998.1	polysaccharide deacetylase [Helicobacter pylori]
5	WP_156534302.1	tryptophan synthase subunit alpha [Helicobacter pylori]
6	WP_209611414.1	UDP-4-amino-4,6-dideoxy-N-acetyl-beta-L-altrosamine transaminase [Helicobacter pylori]
7	WP_209611556.1	NADH-quinone oxidoreductase subunit G [Helicobacter pylori]
8	WP_209611663.1	type II/IV secretion system ATPase subunit [Helicobacter pylori]
9	WP_209611808.1	pyridoxine 5'-phosphate synthase [Helicobacter pylori]
10	WP_209612060.1	bifunctional anthranilate synthase component I family protein/aminotransferase class IV [Helicobacter pylori]
11	WP_209612279.1	pyridoxal phosphate-dependent aminotransferase family protein [Helicobacter pylori]
12	WP_209612466.1	flagellar biosynthesis protein FlhF [Helicobacter pylori]
13	WP_209612506.1	anthranilate synthase component I [Helicobacter pylori]
14	WP_209612511.1	HAMP domain-containing sensor histidine kinase [Helicobacter pylori]

4.0 DISCUSSIONS

This study utilizes a subtractive genomic method to identify and predict potential drug candidates. This approach has received recognition for its ability to accurately determine new targets for drugs among different disease-causing organisms. However, there is still a lack of high-throughput sequencing data available for many bacterial pathogens, which creates an area that requires further research. Considering this limitation in data availability, the subtractive genomic approach becomes even more valuable as it allows researchers to anticipate possible drug targets. These identified targets could then enable the pharmaceutical industry to develop antibiotics that effectively reduce risks and manage bacterial infections. By adopting this approach for H. pylori, the study seeks to lower infection rates and pave the way for reducing cases associated with this pathogen. This advancement holds potential as a solution for individuals suffering from an H.pylori infection.

A subtractive genomic approach was used to identify and shortlist the protein sequences. The non-homologous analysis conducted against the human proteome involves comparing the pathogen's proteins with those found in humans. The aim is to identify any proteins that do not bear resemblance to human counterparts, as their interaction could potentially cause undesired side effects. Upon completing this step, a total of 1228 protein sequences were obtained. The essentiality analysis yielded 907 protein sequences, which are crucial for the pathogen's survival and growth. By excluding non-essential genes from the pathogen’s genome, the search for suitable drug targets was narrowed down. Another analysis focused on virulence resulted in 254 proteins that may play a role in causing specific diseases like peptic ulser. Identifying these factors could potentially lead to new drug targets. Additionally, a druggability assessment identified 151 potential targets with higher chances of responding positively to drugs. This helps prioritize target selection for effective treatment options.

In addition to that, the proteomes of both non-homologous microbiota and broad-spectrum analysis were condensed into 45 protein sequences. By examining the hist pathogen interaction network, specific targets for pathogens can be identified. Furthermore, investigating host-pathogen interactions and non-human anti-targets yielded a total of 37 protein sequences. The purpose is to avoid proteins known as anti-targets in order to prevent unintentional effects from non-pathogenic organisms later on. Subsequently, gene annotation and determination of subcellular localization are performed automatically. Removing sequences with KEGG orthology left us with 6 human sequences and 25 bacterial protein sequences.As a final step,the obtained H.pylori sequences undergo PSORTb subcellular localization resulting in 14 sequencing results.

In summary, the process of subtractive genomic analysis is pivotal in determining appropriate medications.The application of this technology in drug design holds the potential to facilitate the development of a more tailored medication. The desired outcome is a target that possesses strong resistance against microbes.

5.0 CONCLUSION

The analysis of genomes and proteomes of pathogens has greatly improved the discovery process for therapeutic targets against them. In this study, a subtractive genomic approach was used to identify non-homologous essential druggable proteins in H. pylori that could be potential drug targets. These findings can assist in creating new antibiotics specifically designed to target H. pylori without affecting human genes, avoiding any allergic reactions or harm to the host (Homo sapiens). By targeting these proteins with novel drugs, researchers may find ways to eliminate infections caused by H.pylori from their hosts effectively. This research provides comprehensive information on crucial and powerful drug targets within H.pylori which can guide future studies aiming at developing efficient medications and vaccines tailored towards strain-specific strains of this pathogen.

References

1. Bizzozero G. Ueber die schlauchförmigen drüsen des magendarmkanals und die beziehungen ihres epithels zu dem oberflächenepithel der schleimhaut dritte mittheilung. Arch. Für Mikrosk. Anat. 1893;42:82–152. doi: 10.1007/BF02975307.
2. Marshall B.J., Warren J.R. Unidentified curved bacilli in the stomach of patients with gastritis and peptic ulceration. Lancet. 1984;1:1311–1315. doi: 10.1016/S0140-6736(84)91816-6.
3. Goodwin C.S., Worsley B.W. Microbiology of Helicobacter pylori. Gastroenterol. Clin. North Am. 1993;22:5–19. doi: 10.1016/S0889-8553(21)00260-0.
4. Kusters J.G., van Vliet A.H., Kuipers E.J. Pathogenesis of Helicobacter pylori infection. Clin. Microbiol. Rev. 2006;19:449–490. doi: 10.1128/CMR.00054-05.
5. International Agency for Research on Cancer (IARC) Schistosomes, Liver Flukes and Helicobacter Pylori, Monograph on the Evaluation of Carcinogenic Risks to Humans. Vol. 61. IARC; Lyon, France: 1994. pp. 1–241.
6. Ansari S., Yamaoka Y. Survival of Helicobacter pylori in gastric acidic territory. Helicobacter. 2017;22:e12386. doi: 10.1111/hel.12386.
7. Laszewicz W., Iwańczak F., Iwańczak B. Seroprevalence of Helicobacter pylori infection in Polish children and adults depending on socioeconomic status and living conditions. Adv. Med. Sci. 2014;59:147–150. doi: 10.1016/j.advms.2014.01.003.
8. Brown L.M. Helicobacter pylori: Epidemiology and routes of transmission. Epidemiol. Rev. 2000;22:283–297. doi: 10.1093/oxfordjournals.epirev.a018040.
9. Mégraud F. Epıdemıology of helıcobacter pylorı ınfectıon. Gastroenterol. Clin. North Am. 1993;22:73–88. doi: 10.1016/S0889-8553(21)00264-8.
10. Besiski F.S. Helicobacter pylori ınfection: Epidemiology and pathogenesis. Flora. 1996;3:160–166.
11. Raymond J., Lamarque D., Kalach N., Chaussade S., Burucoa C. High level of antimicrobial resistance in French Helicobacter pylori isolates. Helicobacter. 2010;15:21–27. doi: 10.1111/j.1523-5378.2009.00737.x.
12. Opekun A.R., El-Zaimaity H.M., Osato M.S., Gilger M.A., Malaty H.M., Terry M., Headon D.R., Graham D.Y. Novel therapies for Helicobacter pylori infection. Aliment. Pharm. Ther. 1999;13:35–42. doi: 10.1046/j.1365-2036.1999.00435.x.
13. Guttner Y., Windsor H.M., Viiala C.H., Marshall B.J. Human recombinant lactoferrin is ineffective in the treatment of human Helicobacter pylori infection. Aliment. Pharmacol. Ther. 2003;17:125–129. doi: 10.1046/j.1365-2036.2003.01395.x.