In Silico Comparative Analysis of Different *vacA* Genes of *Helicobacter pylori*

HALIMAT CHISOM ATANDA, Faculty of Medicine, The University of Queensland, Queensland, Australia; ELIJAH KOLAWOLE OLADIPO, Genomics Unit, Helix Biogen Institute, Ogbomoso, Oyo State, Nigeria, and Department of Microbiology, Laboratory of Molecular Biology, Bioinformatics and Immunology, Adeleke University, P. M. B. 250. Ede, Osun State, Nigeria; SEUN ELIJAH OLUFEMI[1], Genomics Unit, Helix Biogen Institute, Ogbomoso, Oyo State, Nigeria, and Department of Biochemistry, Ladoke Akintola University of Technology, Ogbomoso, Oyo State, Nigeria.

ABSTRACT. *Helicobacter pylori* is a class I carcinogen responsible for 90% of gastrointestinal and gastroduodenal disorders, including gastric cancer and peptic ulcer disease. The virulence and pathogenicity peculiar to *H. pylori* have been associated with several genes, including cytotoxin associated gene (*cagA*), vacuolating cytotoxin A (*vacA*), outer inflammatory protein A (*oipA*), and duodenal ulcer promoting (*dupA*). This study explored the relationship between African-generated *vacA* genes with genes from other regions with high gastrointestinal disorder prevalence. Nucleotide sequences of 228 *vacA* genes of *H. pylori* were retrieved from the National Centre for Biotechnology Information (NCBI). Pairwise and multiple sequence alignment was carried out on 228 *vacA* nucleotide sequences using MEGA 10.2.4 software to identify regions of similarities. Phylogenetic analysis, also using MEGA software, was carried out to establish the evolutionary relationships between all extracted sequences. Analysis for conserved domain was also performed on the NCBI Conserved Domain Database to better understand each geographical data's properties. After the evolutionary analysis, it was observed that South African *vacA* genes were more closely related to genes from Mexico, Italy, Spain, and Germany—with Italy having the highest occurring relationship. Conserved domain analysis showed 2 highly conserved superfamilies, cl20029 and cl22877, and 2 protein family models, pfam02691 and pfam03797. The results demonstrate relatedness of *vacA* genes from the African region to the European region; Italy, Mexico, and Spain. The study shows the biogeographical diversity among *vacA* genes and emphasizes the degree of domain conservation across each gene. It also shows the need for a holistic assessment of the virulent genes in *H. pylori*.

Publication Date: June 2024 **https://doi.org/10.18061/ojs.v123i2.9135** **OHIO J SCI 123(2):38-43**

INTRODUCTION

The *Helicobacter* genus is a group of bacterial species that affect the mucosa lining of the gastrointestinal system in humans and other mammals. *Helicobacter pylori* is a gram-negative bacterium and the causative agent of several gastrointestinal disorders, including gastritis, peptic ulcer disease, and gastric cancer (van Duynhoven and de Jonge 2001). In 1994, the World Health Organization categorized *H. pylori* as a class I carcinogen (Dong et al. 2009) as it affects about 50% of adults in developed countries and almost 90% in the developing world (van Duynhoven and de Jonge 2001). The variations in the occurrence rates could be attributed to the difference in public health measures practiced in both worlds. *H. pylori* pathogenesis involves the ability of the organism to survive in the acidic stomach environment, release its toxins, and cause tissue damage to the host. The process is facilitated by biochemical activities such as regulating urease activity (Kao et al. 2016), elevating gastrin level to regulate gastric acid secretion (Makola et al. 2007), and inducing inflammation by producing reactive oxygen species which overwhelms the protective mucosal layer and dysregulates glutathione availability (Bhattacharyya et al. 2014). *H. pylori* has several genes associated with its virulence and pathogenicity, including the cytotoxin associated gene (*cagA*), which is

[1]Address correspondence to Seun Elijah Olufemi, Genomics Unit, Helix Biogen Institute, Ogbomoso, Oyo State, Nigeria. Email: oluwaseunjr1@gmail.com

an oncogenic gene responsible for gastric cancer (Jiménez-Soto and Haas 2016); urease, which plays an active role in the development of gastric carcinoma (Olivera-Severo et al. 2017); and vacuolating cytotoxin A, *vacA*, which constitutes an increased risk to peptic ulcer and gastric cancer (Kauser et al. 2005; Gangwer et al. 2007). Other virulent genes of *H. pylori* are the blood group antigen binding adhesion (babA2) gene, duodenal ulcer promoting (*dupA*) gene, outer inflammatory protein A (*oipA*) gene, and sialic acid-binding adhesion (*sabA*) gene (Šterbenc et al. 2019).

VacA is a pore-forming toxin secreted by an autotransporter pathway that causes multiple alterations in human cells. It derives its name from its ability to produce vacuoles within eukaryotic cells, which is dependent on the presence of weak bases like ammonia and the toxin internalization. The expression of the *vacA* gene in *H. pylori* causes the release of toxins that affects human cells by causing alterations and increasing the occurrence of gastric ulcers and cancers. Some other cellular activities of the *vacA* gene include modulation of autophagy and cellular destruction, gastric epithelial cell apoptosis, disruption of endocytosis, and activation of mitogen-activated protein kinases (Cover et al. 2003; Šterbenc et al. 2019). The *vacA* gene has varying vacuolating capacities, which are dependent on the variations that occur in one or more of the following genetic regions: s-region, i-region, m-region, d-region, and c-region. The prevalence of each region is often dependent on geographical locations, with the i, m, and s regions having higher occurrences than the d and c regions. Its mode of action includes decreasing the intracellular levels of the enzyme glutathione and promoting apoptosis of host cells through severe levels of oxidative stress (Šterbenc et al. 2019).

Some studies have highlighted the differences in *vacA* gene sequences obtained from different geographical locations, including Europe, China, and Japan (Kauser et al. 2005; Šterbenc et al. 2019). The current study aims to explore and identify the evolutionary relationship between the *vacA* genes obtained and sequenced in African regions and those obtained from other regions with a predominance of peptic ulcer disease and other gastrointestinal disorders. It also compared the conserved regions of genes and superfamilies across the analyzed *vacA* genes.

MATERIALS AND METHODS

Data Assembly and Sequence Alignment

A total of 228 randomly selected nucleotide sequences of the *H. pylori vacA* gene were downloaded from the National Centre of Biotechnology Information (NCBI) database (February 2022). Source countries of the extracted genes were South Africa (25), India (20), United States (3), Spain (20), Mexico (18), Germany (20), Thailand (20), South Korea (37), Japan (20), Italy (17), Canada (20), and Pakistan (8). Multiple sequence alignment was performed on the retrieved sequences using MEGA 10.2.4.

Phylogenetic Analysis and Detection of Conserved Domains in *vacA* Gene of *H. pylori*

The phylogenetic relationship of the *vacA* gene sequences from NCBI was determined using a maximum likelihood tree model. Bootstrap analysis of 500 replications was used to determine the reliability of the inferred phylogenetic tree.

EMBOSS transeq (Madeira et al. 2024) was used to translate the sequences to their corresponding protein codes to identify the conserved domain regions of the nucleotide sequences. The protein sequences from each region were uploaded to the Conserved Domain Database in NCBI for analysis (Marchler-Bauer and Bryant 2004; Marchler-Bauer et al. 2011, 2015, 2017; Lu et al. 2020). Displayed results were recorded in tabular and pictorial formats.

RESULTS

Phylogenetic Analysis of *vacA* Genes

Pairwise and multiple sequence alignment was carried out for the sequences, and the aligned were directly used to construct a phylogenetic tree. The evolutionary history was inferred using the maximum likelihood method and Tamura-Nei model (Tamura and Nei 1993). The tree with the highest log likelihood (-50344.71) is shown in Fig. 1. Initial trees for the heuristic search were obtained automatically by applying neighbor-join and BIONJ algorithms to a matrix of pairwise distances estimated using the Tamura-Nei model and then selecting the topology with a superior log-likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. There were a total of 5,455 positions in the final dataset. Evolutionary analyses were conducted in MEGA 10 (Kumar et al. 2018).

The phylogenetic tree demonstrates the ancestral relationships between the *vacA* genes of 12 countries. The plot consists of the 11 major clades and is rooted at *vacA* HM047722.1, which originated from South Korea. Fig. 1 shows African-derived sequences in the color red. Of the 25 sequences from Africa, 6 (HQ650802, HQ650801, HQ650806, HQ650805, HQ650804, HQ650803) were closely related to sequences which originated from Italy. Sequences HQ709111, HQ709114, HQ709109, HQ709110, HQ709113, and HQ709112 have close ancestry with sequences DQ61144, which originated from Mexico. Eleven sequences show a close relationship with Spain-derived *vacA*. The African *vacA* gene that appears to be isolated from other sequences shows close relations with Germany-derived sequences.

These results show a wide diversity in the ancestral relationship of *vacA* genes of African sources to that of other regions.

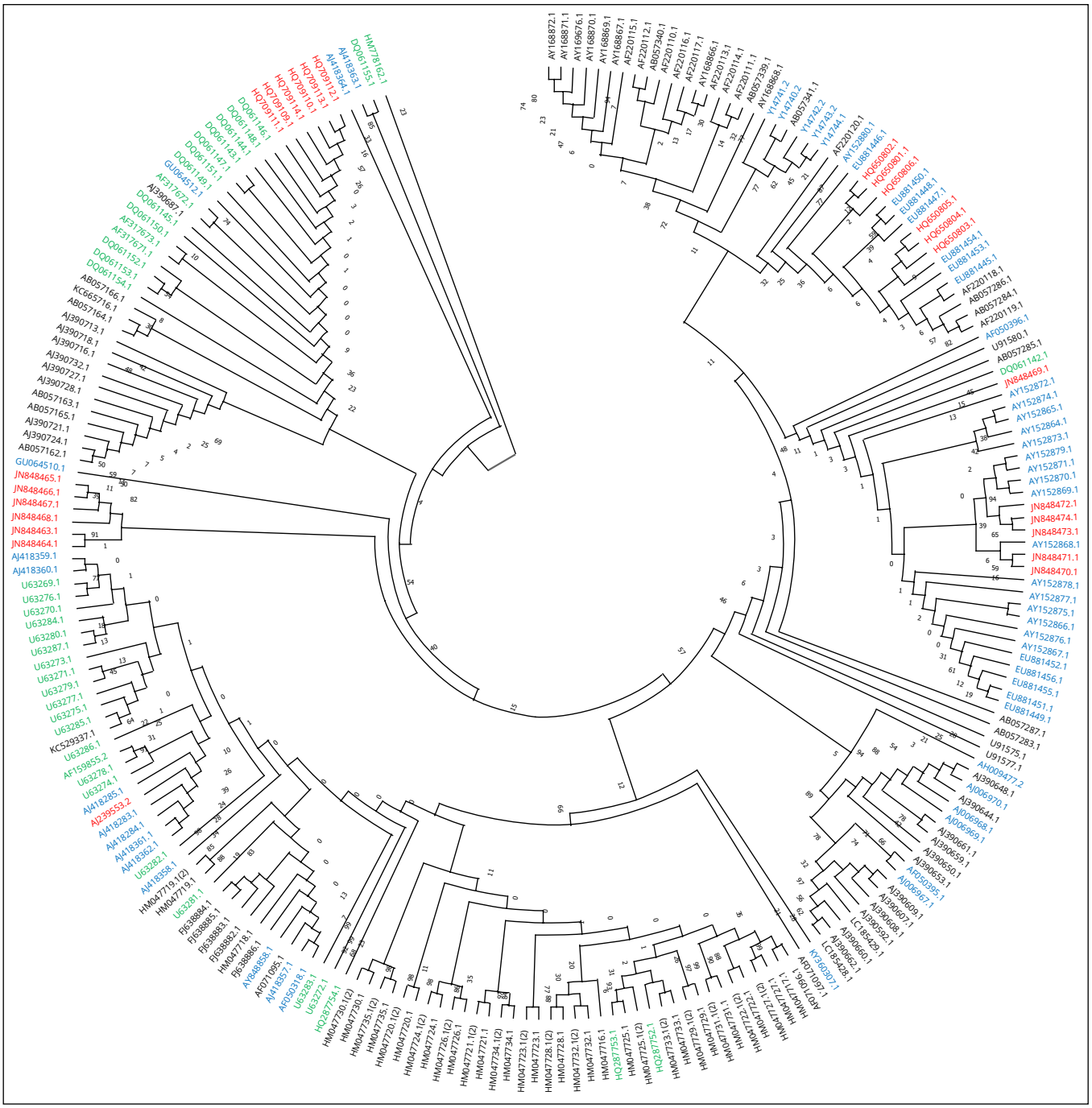

FIGURE 1. Circular view of phylogenetic tree showing the evolutionary relationship of *vacA* gene of *H. pylori* using maximum likelihood method. Samples are colored by region: red represents samples from Africa (South Africa), blue represents samples from Europe (Spain, Germany, Italy), green represents samples from North America (United States, Mexico, Canada), and black represents samples from Asia (Thailand, India, South Korea, Japan, Pakistan).

Conserved Domain Analysis

NCBI's Conserved Domain Database (CDD) (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) provides protein sequence annotation based on the functional sites inferred from the conserved domain footprints. CDD analysis (Table 1) showed that the majority of the sequences belonged to the superfamily cl20029 with no specific domain hits. One of the Japan sequences had a specific domain pfam02691, the only member of the superfamily mentioned earlier, which belongs to clan Pec_lyase-like characterized by a right-handed beta-helix. The pfam02691 is a group of autotransporter proteins that induces cytoplasmic vacuolation in mammalian cells and is specific to the *vacA* protein in *H.pylori* (Atherton et al. 1995; Marchler-Bauer et al. 2013). The other observed specific protein family model is the pfam03797, located in 3 of the Japan-retrieved sequences and 2 sequences sourced from the United States. It belongs to the superfamily cl22877 and is a member of the protein clan MBB, which stands for membrane beta-barrel that mediates the transfer and secretion of proteins across the outer membranes of cells. The pfam03797 is present across different categories of organisms, including archaea, bacteria, and fungi.

DISCUSSION

H. pylori is a widely-spread pathogenic microorganism that co-exists with humans and accounts for the death of millions, especially in developing and underdeveloped nations. It has a vast collection of genes that influence and contribute to its virulence and pathogenicity, including *vacA* (Kauser et al. 2005). *H. pylori* has a symbiotic relationship with humans, such that it has evolved to a level as robust and complicated as humans (Linz et al. 2007). This evolution has resulted in distinct genetic characteristics based on geographical regions. Hence, understanding the evolutionary relationship across regions is essential. The current study focused on the evolution of African-derived *vacA* sequences by comparing them against 11 other countries across the Asian, American, and European regions.

Phylogenetic analysis outlines the ancestral relationship between genes. It allows for a unique understanding of the behavior of these genes and, to an extent, predicts behavioral patterns for further analysis. Of the 11 countries compared to the African-sequenced *vacA* genes, 3 (Mexico, Italy, and Spain) had the closest relationship, which may suggest a high evolutionary connection between the

Table 1
Summary of conserved domain analysis across *vacA* sequences by country

Sample source	*VacA* superfamily accession	No. of specific domain hits
South Africa	cl20029	-
Canada	cl20029	-
Germany	cl20029	-
India	cl20029	-
Italy	cl20029	-
Japan	cl20029	4 (1 pfam02691, 3 pfam03797)
South Korea	cl20029	-
Mexico	cl20029	-
Pakistan	cl20029	-
Spain	cl20029	-
Thailand	cl20029	-
United States	cl20029	2 (pfam03797)

European and African infectious agent. However, all sequences point to an origin in Asia, specifically South Korea. Perhaps the high prevalence of *H. pylori* infection in the Asian region (Kauser et al. 2005) or the enormous amount of *H. Pylori* related data from Asia and Europe (Hooi et al. 2017) would account for the observed pattern of ancestry.

Domains are recurring units in a group of protein sequences that are related by descent. CDD houses over a hundred domain superfamilies curated from the NCBI domain model (Marchler-Bauer et al. 2013). The current study employed the standardized alignment and clustering methodologies of CDD to analyze query sequences by country and provide relevant annotations subject to interpretation. Since only 2 protein domains were identified among over 200 protein sequences, it may indicate that—regardless of geographical differences—the *vacA* genes of *H. pylori* have utilized similar building blocks or sequence patterns to maintain their virulence activities. The *vacA* gene has about 7 allelic variations across the signal, intermediate, and middle regions, all of which contribute to gastrointestinal infections, either in terms of occurrence or severity (Kauser et al. 2005; Qumar et al. 2021). The identification of only 2 domains could also indicate that the known allelic diversity is, to a great extent, uniform despite the continuous evolution.

Conclusions

The results of this study show the biogeographical diversity among *vacA* genes and emphasize the degree of domain conservation across each gene. Continuous and holistic analysis of *H. pylori*'s virulent genes is necessary to evaluate the burden of gastrointestinal infections and generate an effective strategy for managing each infection. Further analysis could involve comparing the *vacA* genes with similar genes from other *Helicobacter* species to further explore its evolution and investigating the correlation between gene expression level, geographical differences, and clinical outcomes. These analyses could contribute to a deeper understanding of the role of *vacA* genes in *H. pylori* infection and its clinical implication in different geographic regions.

LITERATURE CITED

Atherton JC, Cao P, Peek RM Jr, Tummuru MKR, Blaser MJ, Cover TL. 1995. Mosaicism in vacuolating cytotoxin alleles of *Helicobacter pylori*: association of specific *vacA* types with cytotoxin production and peptic ulceration. J Biol Chem. 270(30):17771-17777. https://doi.org/10.1074/jbc.270.30.17771

Bhattacharyya A, Chattopadhyay R, Mitra S, Crowe SE. 2014. Oxidative stress: an essential factor in the pathogenesis of gastrointestinal mucosal diseases. Physiol Rev. 94(2):329-354. https://doi.org/10.1152/physrev.00040.2012

Cover TL, Krishna US, Israel DA, Peek RM Jr. 2003. Induction of gastric epithelial cell apoptosis by *Helicobacter pylori* vacuolating cytotoxin. Cancer Res. 63(5):951-957. PMID: 12615708.

Dong QJ, Wang Q, Xin YN, Li N, Xuan SY. 2009. Comparative genomics of *Helicobacter pylori*. World J Gastroentero. 15(32):3984-3991. https://doi.org/10.3748/wjg.15.3984

Gangwer KA, Mushrush DJ, Stauff DL, Spiller B, McClain MS, Cover TL, Lacy DB. 2007. Crystal structure of the *Helicobacter pylori* vacuolating toxin p55 domain. P Natl Acad Sci USA. 104(41):16293-16298. https://doi.org/10.1073/pnas.0707447104

Hooi JKY, Lai WY, Ng WK, Suen MMY, Underwood FE, Tanyingoh D, Malfertheiner P, Graham DY, Wong VWS, Wu JCY, Chan FKL, Sung JJY, Kaplan GG, Ng SC. 2017. Global prevalence of *Helicobacter pylori* infection: systematic review and meta-analysis. Gastroenterology. 153(2):420-429. https://doi.org/10.1053/j.gastro.2017.04.022

Jiménez-Soto LF, Haas R. 2016. The *CagA* toxin of *Helicobacter pylori*: abundant production but relatively low amount translocated. Sci Rep-UK. 6:23227. https://doi.org/10.1038/srep23227

Kao CY, Sheu BS, Wu JJ. 2016. *Helicobacter pylori* infection: an overview of bacterial virulence factors and pathogenesis. Biomed J. 39(1):14-23. https://doi.org/10.1016/j.bj.2015.06.002

Kauser F, Hussain MA, Ahmed I, Srinivas S, Devi SM, Majeed AA, Rao KR, Khan AA, Sechi LA, Ahmed N. 2005. Comparative genomics of *Helicobacter pylori* isolates recovered from ulcer disease patients in England. BMC Microbiol. 5:32. https://doi.org/10.1186/1471-2180-5-32

Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. Mol Biol Evol. 35(6):1547-1549. https://doi.org/10.1093/molbev/msy096

Linz B, Balloux F, Moodley Y, Manica A, Liu H, Roumagnac P, Falush D, Stamer C, Prugnolle F, van der Merwe SW, Yamaoka Y, Graham DY, Perez-Trallero E, Wadstrom T, Suerbaum S, Achtman M. 2007. An African origin for the intimate association between humans and *Helicobacter pylori*. Nature. 445:915-918. https://doi.org/10.1038/nature05562

Lu S, Wang J, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Marchler GH, Song JS, et al. 2020. CDD/SPARCLE: the Conserved Domain Database in 2020. Nucleic Acids Res. 48(D1):D265-D268. https://doi.org/10.1093/nar/gkz991

Madeira F, Madhusoodanan N, Lee J, Eusebi A, Niewielska A, Tivey ARN, Lopez R, Butcher S. 2024. The EMBL-EBI Job Dispatcher sequence analysis tools framework in 2024. Nucleic Acids Res. gkae241:1-5. https://doi.org/10.1093/nar/gkae241

Makola D, Peura DA, Crowe SE. 2007. *Helicobacter pylori* infection and related gastrointestinal diseases. J Clin Gastroenterol. 41(6):548-558. https://doi.org/10.1097/MCG.0b013e318030e3c3

Marchler-Bauer A, Bryant SH. 2004. CD-Search: protein domain annotations on the fly. Nucleic Acids Res. 32(2):W327-W331. https://doi.org/10.1093/nar/gkh454

Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, Fong JH, Geer LY, Geer RC, Gonzales NR, et al. 2011. CDD: a Conserved Domain Database for the functional annotation of proteins. Nucleic Acids Res. 39(1):D225-D229. https://doi.org/10.1093/nar/gkq1189

Marchler-Bauer A, Zheng C, Chitsaz F, Derbyshire MK, Geer LY, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Lanczycki CJ, et al. 2013. CDD: conserved domains and protein three-dimensional structure. Nucleic Acids Res. 41(D1):D348-D352. https://doi.org/10.1093/nar/gks1243

Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, Geer RC, He J, Gwadz M, Hurwitz DI, Lanczycki CJ, et al. 2015. CDD: NCBI's Conserved Domain Database. Nucleic Acids Res. 43(D1):D222-D226. https://doi.org/10.1093/nar/gku1221

Marchler-Bauer A, Bo Y, Han L, He J, Lanczycki CJ, Lu S, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, Gwadz M, et al. 2017. CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. Nucleic Acids Res. 45(D1):D200−D203. https://doi.org/10.1093/nar/gkw1129

Olivera-Severo D, Uberti AF, Marques MS, Pinto MT, Gomez-Lazaro M, Figueiredo C, Leite M, Carlini CR. 2017. A new role for *Helicobacter pylori* urease: contributions to angiogenesis. Front Microbiol. 8:1883. https://doi.org/10.3389/fmicb.2017.01883

Qumar S, Nguyen TH, Nahar S, Sarker N, Baker S, Bulach D, Ahmed N, Rahman M. 2021. A comparative whole genome analysis of *Helicobacter pylori* from a human dense South Asian setting. *Helicobacter*. 26(1):e12766. https://doi.org/10.1111/hel.12766

Šterbenc A, Jarc E, Poljak M, Homan M. 2019. *Helicobacter pylori* virulence genes. World J Gastroentero. 25(33):4870-4884. https://doi.org/10.3748/wjg.v25.i33.4870

Tamura K, Nei M. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol. 10(3):512-526. https://doi.org/10.1093/oxfordjournals.molbev.a040023

van Duynhoven YTHP, de Jonge R. 2001. Transmission of *Helicobacter pylori*: a role for food? B World Health Organ. 79(5):455-460. https://iris.who.int/handle/10665/268330