Roles of *oipA* phase variation in virulence phenotype and association of *cag* pathogenicity island in *Helicobacter pylori* clinical isolates

Angulmaduwa Lokuruge Sacheera Kapruk Lakshan Angulmaduwa

The Graduate School

Yonsei University

Department of Applied Life Sciences

Roles of *oipA* phase variations in virulence phenotype and association of *cag* pathogenicity island in *Helicobacter pylori* clinical isolates

Directed by Professor Jeong-Heon Cha

The Doctoral Dissertation Submitted to
the Department of Applied Life Sciences,
the Graduate School of Yonsei University
in partial fulfilment of the requirements
for the degree of Doctor of Philosophy in Applied Life Science

Angulmaduwa Lokuruge Sacheera Kapruk Lakshan

Angulmaduwa

June 2024

This certifies that the Doctoral dissertation of

Angulmaduwa Lokuruge Sacheera Kapruk Lakshan

Angulmaduwa is approved.

Dissertation Supervisor: Jeong-Heon Cha

Yun-Jung Yoo

Sang-Sun Yoon

Ji-Hwan Ryu

Ki Woo Kim

The Graduate School, Yonsei University

June 2024

ACKNOWLEDGEMENT

First and foremost, I would like to express my deepest gratitude to my supervisor, Prof. Jeong-Heon Cha, for his invaluable guidance and unwavering support throughout my PhD journey. I am also immensely grateful to my committee members, Prof. Yun-Jung Yoo, Prof. Sang-Sun Yoon, Prof. Ji-Hwan Ryu, and Prof. Ki Woo Kim, for their time, insightful comments, and constructive suggestions on my dissertation. Special thanks to Prof. Yun-Jung Yoo for her kindness and support during my time here.

I would like to extend my appreciation to Prof. Eun-Jung Bak, Prof. Kavinda Tissera, Prof. Aeri Kim, and Teacher Yeong-Eui Jeong for their generous assistance, valuable advice on lab work and life, and unwavering support. Special thanks to Dr. Jing Lai, my dear senior, for teaching me many experimental skills and for her constant comfort and encouragement during difficult times. My gratitude also goes to Dr. Myeong-A Kim, Ji-Won Seo, Youn-Soo Lee, and Ashansa Ramanayake for their support, encouragement, and invaluable insights both in and out of the lab. I would like to acknowledge my other lab members: Min-Seo Kim, Jaeyun Nam, Dowon Seo, and Roshan Pradeep, for their assistance with my experiments whenever I needed it.

I am also deeply thankful to Prof. Sung-il Jang, Dr. Hanfu Su, and Prof. Yong-Joon Cho for their help with many of my experiments and analyses despite their extremely busy schedules. I would like to mention my dear professors from Choonhae College of Health Sciences: Prof. Kim Kum Ja, Prof. Kim Chan Hee, and Prof. Kim Sun-il, who helped me adjust to life in Korea and showed me the beauty of this country during my stay.

My heartfelt thanks go to my friends in Korea: Dr. Dinuka Adasooriya, Dr. Amanda Herath, Malith Ranathunga, Dylan Senith, Dylan de Silva, Niluka, Deepika, Ruuth, and Laksara from Sri Lanka, as well as Prof. Anish Adpaikar, Dr. Shu Jin Lee, Shin Ko, Prof. Uthkarsh Mangal, Chanshik Park, Akila Cooray, and Dayan Kim from Yonsei University College of Dentistry. Special thanks to my housemates and dorm friends Wei Ning Chiu, Hong Min, Soo-Seong Hyun, Srijita Das, and Doyong Kim. Their companionship and support helped me maintain a balanced social life during my stressful PhD journey.

I am deeply grateful to my Sri Lankan mentors, Prof. Rasika Illeperuma, Prof. Veranja Liyanpathirana, Prof. Nilanthi Dissanayake, Prof. Nimal Dharmasena, and Prof. Swarna Wimalasiri, who always cared about my studies and life and provided generous help and guidance throughout my academic journey. I would like to

mention my dearest friend Achini Poornima, Prof. Morris, Dr. Jeewanthi Ekanayake, and Thimira Amarasinghe, for being my unpaid therapists, taking on my psychological burdens, and helping me move forward.

I also appreciate my friends from Sri Lanka, especially from the Class of 2011 of the Faculty of Allied Health Sciences, University of Peradeniya: Isuru Ekanayke, Chamindi Sulakkhana, Irosha Subhasena, Melani Perera, Dilip Jayakody, Mekhala Nishshanka, Tharindu Batuwanthudawa, Eranga Udaya, Nuwan Bandaranayake, Pubudu Priyankara, Thulasika Oppilamani, Vinujaa Selvaraja, and many more, as well as friends from other batches: Ishan Dimuthu, Rakitha Jayatissa, and Nuwan. With their continuous encouragement and care, I was able to come to Korea and complete my PhD.

Finally, I would like to thank my family. I am truly grateful to my uncles, aunts, and cousins for their generous support and best wishes. I am extremely thankful to my dear Mom, my Brothers (Ayya, Rang Ayya, Chutta, and *Paty*), my Sisters (Loku Akka, Chutti Akka, and Rosi Akka), and my dearest father and younger brother who longed to see my dream come true. Despite all the hardships, they continually encouraged and helped me at every stage of my life, tenderly taking care of everything and providing me with the peace of mind to accomplish this

achievement.

තාත්තා නැති වූ දා පටන් වයස 16ක් වුවත් මහා පවුලක බර කරට ගත් ලොකු අයියාටත්, කුඩා කල පටන් දෙවෙනි අම්මා වූ ලොකු අක්කාටත්, ජීවිතය අමාරුම කාල වල දුක ඇහූ චුට්ටි අක්කාටත්,6 වසරේ පටන් බස් එකේ යන තැන සිට ජීවිතය කියාදුන් රෝසි අක්කාටත්, ලොකු අයියාට පවුල ගිලිහෙන මොහොතේ තම අධ්‍යාපනය කැප කර පවුලේ බර කරට ගත් රං අයියාටත්, මම නැතත් ගෙදර අඩුවක් නොදැනෙන්නට දුන්, කුඩා කල පටන් මගෙ නිවුන් සොයුරා මෙන් උන් චුට්ටා ටත්, අපි හැමෝටම ජීවිතයේ සහෝදර බැඳීම් වල වටින කම් මතක් කරදෙමින් යන්නට ගිය මගේ දයාබර චූටි මල්ලි පැටීටත්, අවසාන වශයෙන් ජීවිතයේ මහා අවු වැසි හෙන මැද, කඳු කඩා වැටෙන්දිත්, මහා කන්දක් සේ මුහුණ දෙමින් තනිව දරුවන් 8ක් ගොඩ නැගූ මගේ අම්මාට මගේ අධ්‍යාපනයේ අග්‍ර ඵලය, මාගේ ආචාර්ය උපාධියේ නිබන්දනය පුද කරමි!

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Roles of *oipA* phase variation in virulence phenotype and association of *cag* pathogenicity island in *Helicobacter pylori* clinical isolates

Angulmaduwa Lokuruge Sacheera Kapruk Lakshan Angulmaduwa

Department of Applied Life Science

The Graduate School,

Yonsei University

Directed by Professor Jeong-Heon Cha

The Outer Inflammatory Protein A (OipA) in *Helicobacter pylori* has been identified as a critical virulence factor that induces inflammation in the host's gastric mucosa. The expression of the *oipA* gene is regulated by phase variation, allowing clinical *H. pylori* isolates to switch between 'on' and 'off' states of *oipA* gene expression. This study aims to elucidate the impact of *oipA* phase variation on the virulence phenotype and its relationship with other essential virulence determinants in clinical *H. pylori* isolates.

We analyzed 18 *oipA* on and 22 *oipA* off *H. pylori* clinical isolates from the USA to assess their virulence phenotype. *oipA* on isolates consistently demonstrated high levels of interleukin-8 induction (16/18), gastrin expression (16/18), and cell elongation (16/18) in infected AGS cells, whereas *oipA* off isolates showed significantly lower levels of Interleukin-8 (1/22), gastrin expression (1/22), and cell elongation (1/22). Moreover, a majority of the *oipA* on isolates expressed CagA (16/18) and translocated the CagA into host cells (14/18), indicating a higher virulence potential. These phenotypic distinctions suggested a differential presence of *cag* pathogenicity island (*cag*PAI) between the two groups. Subsequent PCR analysis revealed that *oipA* on isolates predominantly (16/18) harbored *cag*PAI genes, while the majority of *oipA* off isolates (20/22) lacked. Comparative genomic analysis of selected 20 strains identified that *oipA* on isolates possessed intact

*cag*PAI and further harbored a unique set of 10 genes whereas, *oipA* off isolates were negative for *cag*PAI and further carried 28 distinct genes. To expand its association between *oipA* phase variation and *cag*PAI to diverse geographic regions, global genome analysis was performed using 613 *H. pylori* genome sequences available in the database and showed the strong association between *oipA* status and presence or absence of *cag*PAI in geographic regions of hpEurope, hpEastAsia, and hpAfrica1.

The results of this research emphasize the critical role of *oipA* in regulating the virulence of *Helicobacter pylori*. Specifically, this *oipA* phase variation influences bacterial virulence by modulating the presence or absence of the *cag*PAI. This regulation underscores a significant mechanistic pathway through which *oipA* phase variation can dictate the pathogenicity of *H. pylori*, thereby impacting host-pathogen interactions and influencing disease outcomes.

Key word: *Helicobacter pylori*, *oipA*, phase variation, *cag*PAI, pathogenes

I. INTRODUCTION

Helicobacter pylori (H. pylori) is a microaerophilic, gram-negative bacterium that colonizes the stomach lining of nearly 50% of the world's population[1–4]. While most *H. pylori* infections remain asymptomatic, some can lead to significant clinical diseases in some patients[5,6]. While, chronic gastritis, characterized by inflammation of the stomach lining, is the most common outcome[6], in some individuals, the persistent inflammation can progress to atrophic gastritis, leading to intestinal metaplasia, while significantly increasing the risk of dysplasia. Dysplasia represents a precancerous condition that can eventually evolve into gastric adenocarcinoma[5,7,8]. Gastric cancer (gastric adenocarcinoma) is the fifth most common cancer and the third leading cause of cancer-related deaths worldwide[9,10] with the global annual incidence rates between 15.6 and 18.1 per 100,000 individuals for men and between 6.7 and 7.8 per 100,000 individuals for women[11]. *H. pylori* infection is the major risk factor for gastric cancers, accounting for 75% of cases according to epidemiological studies[12]. *H. pylori* is classified as a Group I carcinogen by the International Agency for Research on Cancer (IARC)[8,13–16].

H. pylori is a bacterial species with an exceptional degree of genetic diversity and variability[17–21]. It was acquired by humans in Africa around 100 000 years ago and has accompanied modern humans during multiple ancient and more recent

migrations out of Africa[22–24]. In present times, the common and complex history of humans and *H. pylori* is now reflected by a highly distinctive phylogeographic population structure[18,25–27].

The prevalence of *H. pylori* infection and gastric cancer varies geographically, with East Asian countries such as China, Japan, and Korea showing higher rates[28,29]. This disparity can be explained by environmental factors, dietary habits, host genetic susceptibility, and most importantly differences in *H. pylori* genetics[5,30,31]. Also, the severity of the diseases caused by *H. pylori* is found to be associated with its polymorphisms in virulence factors, with most studied being *cytotoxin-associated gene A (cagA), vacuolating cytotoxin (vacA),* and *bab* paralogue[5,32–35].

cag Pathogenicity Island (_cag_PAI)

The *cag*PAI is a 40 kb genomic island acquired through horizontal gene transfer and contains approximately 27 genes including *cagA* which encodes CagA; the first identified bacterial protein involved in human cancer[36–38]. The presence of *cag*PAI or *cagA* is strongly correlated with increased virulence and more severe clinical outcomes in patients[39–41]. Also, presence or absence of *cag*PAI found to be vary among geographic regions[22,24]. *H. pylori* isolates worldwide are categorized in to about 6 major geographic populations and some more sub-populations based on

their genetic diversity which reflects the origin geographic region[18,24,42]. The hpEastAsia and hpAfrica1 geographic populations show a nearly 100% presence of *cag*PAI while the hpEurope population is a mix of both presence and absence[17,18].

The *cag*PAI encodes a type IV secretion system (T4SS) that translocate CagA toxin and other effector molecules into host cells[36,43–47]. Once translocated into the host cell, host cell kinases can phosphorylate CagA at conserved tyrosine residues located within the EPIYA (Glu-Pro-Ile-Tyr-Ala) motifs[34,48–51]. This phosphorylation alters multiple host signaling pathways, causing dysregulation of epithelial structure and integrity by inducing cytoskeletal rearrangement and increasing cellular mobility, resulting in the characteristic "hummingbird phenotype/Cell elongation". Not just phosphorylated CagA but the CagA that is translocated but not phosphorylated, also found to promote proliferation and inflammation by activating β-catenin and NF-κB signaling[52–56].

Other than injecting toxins to the host cells the T4SS of *H. pylori* facilitates direct interaction with gastric epithelial cells, initiating the secretion of various pro-inflammatory cytokines upon infection, including interleukin-8 (IL-8) which plays a central role in the recruitment of neutrophils and initiation of inflammatory responses[55–60]. IL-8 also can be stimulated through the activation of NF-κB and

MAPK signaling pathways by the CagA and other virulence factors delivered via the T4SS[56,61]. This interplay between *H. pylori* and IL-8 secretion is critical in its pathogenesis, contributing to chronic inflammation and the development of gastric diseases such as peptic ulcers and gastric cancer. Simultaneously, *H. pylori* infections were found to induce hypergastrinemia and decrease acid secretion via the T4SS and CagA[61].

***outer inflammatory protein A* (*oipA*)**

The outer membrane proteins (OMPs) are critical virulence factors in *H. pylori* pathogenicity. This includes AlpA, AlpB, BabA, HomB, HopZ, and SabA, which are associated with variable disease outcomes[53,62–67]. Mainly OMPs are found to facilitate bacterial adhesion to host cells and induce pro-inflammatory cytokine production, significantly impacting the infection's progression[63,64]. OMPs are encoded by a family of paralogous genes that account for 4% of the *H. pylori* genome and have been divided into 5 paralogous gene families according to their functions: Hop (outer membrane porins), Hor (Hop-related proteins), Hof (*H. pylori* OMP), Hom (*H. pylori* outer membrane), and Iron-regulated OMPs[68,69]. The expression of several hop genes (*oipA, sabA, sabB, babA, babB, and babC*) could be regulated by the slipped strand mispairing mechanism (SSM) according to the CT dinucleotide repeat number in their signal-peptide-coding region[33,70,71].

Outer inflammatory protein A (OipA), encoded by the *oipA* (*hp0638/hopH*) gene, has been identified as a factor in promoting IL-8 secretion from gastric epithelial cells and enhancing gastric inflammation[72,73]. The expression of *oipA* is phase-variable, regulated by slipped strand mispairing (SSM) according to the number of cytosine-thymidine (CT) dinucleotide repeats in its signal-peptide-coding region[33,70,71,74]. This phase variation can switch *oipA* 'on' or 'off,' which may impact the bacterium's ability to induce inflammatory responses[71].

Geographical differences in the prevalence and expression of *oipA* have been noted, with East Asian strains often carrying two *oipA* genes, unlike European or African counterparts[31,75]. This genetic diversity may contribute to the differences in rates of severe gastric diseases observed in these populations[31]. Epidemiological studies have shown significant associations between *oipA* phase variations and clinical outcomes such as duodenal ulcers and gastric cancer, although some studies report no such link, highlighting the need for further research into the role of *oipA* in *H. pylori* pathogenicity[76–79].

Another intriguing aspect of *oipA* is that the 'on' phase variation is found to have a linkage with some important virulent genes in *H. pylori* such as *cagA*, *babA2*, and

vacA genotype[80]. Understanding the roles of *oipA* phase variation and its association with *cag*PAI and other virulence factors is crucial in elucidating the pathogenic mechanisms of *H. pylori*. This knowledge can help develop targeted strategies for managing *H. pylori* infections and mitigating the risk of severe gastric diseases, particularly in high-risk populations. Thus, this study endeavors to elucidate the roles of *oipA* phase variations, in the context of *cag*PAI functional perturbations using clinical *H. pylori* isolates.

II. MATERIALS & METHODS

1. Bacterial strains and cultures

All 40 isolates were obtained from patients at Vanderbilt University Medical Center, Nashville TN, USA. Written informed consent was received from each patient, and the protocol was approved by the Vanderbilt University and the Nashville Department of Veterans Affairs Institutional Review Board (IRB# 5190). Gastric biopsies were harvested from individuals at the Veterans Affairs Medical Center in Nashville undergoing upper endoscopy, and used for bacterial culture. Isolates from biopsies were confirmed to be *H. pylori* by positive urease, catalase, and oxidase tests, and typical appearance on Gram stain. These strains are referred to as American *H. pylori* clinical isolates (AH). All strains used in this study are listed in Table 1.

H. pylori strains were grown on horse blood agar plates supplemented with Vancomycin (10 μg/ml), Polymixin-B (0.33 μg/ml), Trimethoprim (5 μg/ml), Amphotericin-B (8 μg/ml), and Beta-cyclodextrin (2 μg/ml) dissolved in DMSO (1%) for 2 to 3 days u to 2 passages from the frozen stock[21,81].

For liquid cultures, *H. pylori* strains were initially cultured in brucella broth (BD, Franklin Lakes, NJ) supplemented with 10% fetal bovine serum (FBS) (Gibco,

Grand Island, NY, USA) and 10 μg/ml vancomycin (Duchefa, Haarlem, Netherlands) for one day with shaking at 120 rpm, 37℃, and were inoculated into new media with an initial optical density of 0.05 at 600 nm. These cultures were grown under the same condition as an initial culture for 16-18 hours[21,81].

All *H. pylori* cultures were grown under microaerophilic conditions (5% O_2, 10% CO_2, 85% N_2) generated either by "Anaeropack Microaero gas generating system (Mitsubishi Gas Chemical, Tokyo, Japan)" or "Anoxomat gas evacuation and replacement system (Anoxomat, Advanced Instruments, Norwood, MA)".

Table 1. *H. pylori* strains used in this study

Name	Name
AH 868	AH 871
AH 869	AH 872
AH 870	AH 873
AH 875	AH 874
AH 876	AH 878
AH 877	AH 882
AH 879	AH 884
AH 880	AH 886
AH 881	AH 888
AH 883	AH 890
AH 885	AH 891
AH 887	AH 892
AH 889	AH J44
AH 893	AH 895
AH 894	AH J63
AH 897	AH 896
AH 898	AH J75
AH 900	AH 899
G27	AH 901
G27 Δ*cag*PAI	AH 904
G27 Δ*cagA*	AH J195
G27 Δ*oipA*	AH 908

2. Genotyping of the *oipA* phase variation

Bacterial genomic DNA was extracted using the Wizard Genomic DNA Purification Kit (Promega). The *oipA* gene was amplified by PCR using LA-f/LA-r and LA-f/LB-r. PCR products were purified and Sanger dideoxy DNA sequencing was performed at Cosmo Genetech Co., Ltd. (Seoul, South Korea using the primers *oipA*-F and *oipA*-R (Table 2). The resulting DNA sequences[65,75] were analyzed using CLC main workbench software.

3. Mammalian cell cultures

AGS cells (ATCC CRL-1739, human gastric adenocarcinoma epithelial cell line), were cultured in Dulbecco's Modified Eagle Medium (DMEM) (Gibco, Thermo Fisher Scientific, Waltham, MA) supplemented with 10% FBS, 100 U/ml penicillin, and 100 μg/ml streptomycin (Gibco) and maintained at 37°C in a water-saturated 5% CO_2 air atmosphere.

4. Luciferase Assay for Gastrin Promoter Activity

G240-Luc cells were plated in 12-well plates at a density of 1 × 105 cells per well in 1 ml of DMEM supplemented with 10% FBS. When the cells reached 80% confluence, cells were infected with H. pylori. Five hours after stimulation, gastrin promoter activity was measured using the luciferase assay as described previously[61].

Briefly, cells were washed with PBS and lysed with passive lysis buffer (Promega). Cells were scraped, and centrifuged at 12,000 × g for 2 min at 4℃. Luciferase activities were measured with a luciferase assay system (Promega) and a Centro XS3 microplate luminometer LB 960 (Berthold Technologies). The total protein in cell lysates was determined by the bicinchoninic acid assay (Pierce Biotechnology) and was used for normalization.

5. IL-8 secretion assay

AGS cells were infected by *H. pylori* strains as previously described in the cell elongation assay. Cell culture supernatant was collected at 6 hours' post-infection and was used for IL-8 enzyme-linked immunosorbent assay (ELISA). Assays were performed using human IL-8 ELISA Max Deluxe (BioLegend, San Diego, CA) following the manufacturer's instructions, and absorbance was measured using an Epoch microplate spectrophotometer (BioTek, Winooski, VT). Data were presented as means SD of results from 3 independent experiments.

6. Cell elongation assay

2×10^5 AGS cells per well were seeded onto 12-well cell culture plates and incubated for 1 day at 37℃, when the cell confluence reached 60-70%, cells were washed with phosphate-buffered saline (PBS, pH 7.4) and the medium was changed

to 1 mL of plain DMEM media at 2 h before the infection. Liquid cultures of *H. pylori* were centrifuged and the bacteria were suspended in plain DMEM, and AGS cells were infected at a multiplicity of infection (MOI) of 100. At 6 hours' post-infection, cells were washed with PBS and fixed with 4% paraformaldehyde. Images of the cells were taken under 200 × magnification.

7. Immunoblot assay

Expression of OipA, CagA, phosphorylated CagA, UreA, and GAPDH were detected by immunoblot. To prepare lysates of infected cells, AGS cells were seeded onto 6-well cell culture plates at a density of 4×10^5 cells per well and were then incubated for 1 day at 37℃. At 2 hours before the infection, cells were washed with PBS and the medium was changed to 2 mL of plain DMEM. Liquid cultures of *H. pylori* were centrifuged and the bacteria were suspended in plain DMEM, and AGS cells were infected at an MOI of 100. At 6 hours post-infection, cells were washed with PBS and then lysed with 120 μl of cell lysis buffer supplemented with protease inhibitor cocktail. Protein concentrations of lysates were measured using Pierce bicinchoninic acid (BCA) protein assay reagent (Thermo Fisher Scientific, Waltham, MA). 20 μg of total protein for each sample were electrophoresed in a 10% polyacrylamide gel and then transferred to a PVDF membrane (Millipore, Billerica, MA).

The expression OipA was detected using rabbit polyclonal anti-OipA antibody ("NYYSDDYGDKLDYK" generously provided by Professor Steffen Backert [73]). Primary antibody was diluted 1: 5,000 in 3% BSA-TBST. Membranes were further probed with goat anti-rabbit IgG-HRP (Santa Cruz Biotechnology) diluted 1:5000 in 3% skim milk TBST, and subsequently visualized by chemiluminescence.

Phosphorylated CagA expression was detected by using mouse monoclonal anti-phosphotyrosine antibody pY99 ((Santa Cruz Biotechnology, Santa Cruz, CA) diluted 1:10000 in 3% BCA in TBST (20 mM Tris-HCl, pH 7.5, 150 mM NaCl, 0.05% Tween 20, 3% bovine serum albumin) as primary antibody and using HRP-conjugated goat anti-mouse IgG (Santa Cruz Biotechnology, Santa Cruz, CA) diluted 1:5000 in 3% skim milk TBST as the secondary antibody, bounded protein was detected with chemiluminescence using ECL reagents (GE Healthcare, Buckinghamshire, UK).

The expressions of total CagA, GAPDH, and UreA were detected using rabbit polyclonal anti-CagA antibody b-300 (Santa Cruz Biotechnology, Dallas, TX), polyclonal antibody (sp56) rabbit polyclonal anti-GAPDH antibody (Koma Biotech, Seoul, South Korea), and rabbit polyclonal anti-UreA antibody b-234 (Santa Cruz Biotechnology), respectively. Primary antibodies were diluted 1: 5,000 in 3% BSA-

TBST. These membranes were further probed with goat anti-rabbit IgG-HRP (Santa Cruz Biotechnology) diluted 1:5000 in 3% skim milk TBST, and subsequently visualized by chemiluminescence.

8. Field Emission Scanning Electron Microscopy (FESEM)

High-resolution field emission scanning electron microscopy (FESEM) of H. pylori in contact with AGS cells. The co-culture of AGS cells and H. pylori was prepared as described above. Briefly, 1 × 105 AGS cells/well were seeded onto the 2% gelatin-coated coverslips in DMEM supplemented with heat-inactivated 10% fetal bovine serum and 1% penicillin/streptomycin, and grew for 24 h until cell confluence reaches 60–70%. After 2 h for serum starvation in plain DMEM, AGS cells were infected by H. pylori at an MOI of 100:1 for 4 h at 37℃ in the presence of 5% CO2. Cells were fixed with 2.0% paraformaldehyde, 2.5% glutaraldehyde in PBS. Coverslips were washed with sodium cacodylate buffer and fixed with 1% osmium tetroxide at room temperature for 2 hours. After washing with sodium cacodylate buffer, the coverslips were dehydrated with sequential washes of increasing concentrations of ethanol. Samples were then dried at the critical point, mounted onto sample stubs, grounded with a thin strip of silver paint at the sample edge, and sputter-coated with gold before viewing with a Zeiss Supra 35 V FEG scanning electron microscope.

9. *cag*PAI region genotyping

Bacterial genomic DNA was extracted using the Wizard Genomic DNA Purification Kit (Promega). A PCR-based method was designed to identify the presence of *cag*PAI region. 7 sets of PCR were performed for each sample that covered the 40 Kb *cag*PAI region using the primers described in Figure 9 and Table 2. PCR products were resolved on a 1% agarose gel for 40 minutes at 120 V. Gel pictures were taken and analyzed for the presence or absence of each band.

Table 2. Primers used in this study

Region	Primer name	Length	Sequence
***cag*PAI**	Fg1-F	19	GCT GTA TTC TTG CGC CAA T
	Fg1-R	22	CAA CTA TTT GAA CAA CCT TGT G
	Fg2-F	23	CAG ATA AGA AGC CAC TAG GTC TG
	Fg2-R	21	AGC AAT GAA CAG ATT ATC AAC
	Fg3-F	24	GTA GTA ATT GTA GTT TCT AGG CAC
	Fg3-R	21	CCC ACC CAT ACA CAA TCC TAA
	Fg4-F	22	CAC CTA GCA ACT CAC AGA GCA A
	Fg4-R	22	CAT AGG CAT AAG GGT TAG GAA G
	Fg5-F	21	GAA GAA GTG GCT GCA AAA GAA
	Fg5-R	21	ATC AAT GCG ATG TGG GGC ATT
	Fg6-F	24	CTA ATA TAC CGC TCA TCA TTT CAA
	Fg6-R	26	TTC GTG ATT GTC AAT AGA GTC CTT AA
	Fg7-F	20	CCG CTT CAC ATG TAA TCG TA
	Fg7-R	20	TCT AGC TTT AGA AGA GAT GC
oipA	oipA-F	20	GGA CTA ACG ATA AGC GAG CG
	oipA-R	20	CCA ATC ACA AGC CCT GAA GA

10. Whole Genome Sequencing, assembly, and annotation

Selected strains were further cultured for whole genome sequencing. Bacterial genomic DNA was extracted using the Wizard Genomic DNA Purification Kit (Promega) and sent to Chun-Lab Inc. Korea for sequencing. Whole genomes of the 24 isolates were determined using third-generation (Pac-Bio SMRT) sequencing technology, as previously described[20]. The raw data were assembled with the SMRT Analysis Software v2.3.0 with the HGAP2 protocol and default parameters, and then processed for annotation by the NCBI Prokaryotic Genome Annotation Pipeline version 4.3[82].

11. Phylogenetic analysis

Phylogenetic analysis based on core genome pairwise distance was performed using the Efficient Database framework for comparative Genome Analyses using BLAST score Ratios (EDGAR 2.0)37. Briefly, 927 orthologous genes shared among all 20 *H. pylori* genomes were identified by bidirectional BLASTn38. Each of the orthologous gene was aligned with MUSCLE39. The alignments were then trimmed and concatenated to a single 855,682-bp alignment. The concatenated alignment was further used to compute a pairwise genome distance matrix; a phylogenetic tree was then constructed based on this distance matrix using a Neighbor-joining method as implemented in PHYLIP40. The Neighbor-joining tree

was visualized using Mega 7.041.

Phylogenetic analysis based on chromosomal inversion was computed using progressive Mauve and MGR42,43; all plasmids were first removed from the genome sequences. Briefly, the completed chromosomal sequences of 20 strains were aligned with progressive Mauve using the default parameters; these resulted in 107 and 80 conserved local co-linear blocks (LCBs), respectively. The relative position and orientation of the LCBs were output as signed permutations. Further, the permutations were input to MGR to compute phylogeny trees based on reversals with the uni-chromosomal circular mode. MGR generates a tree such that the sum of the rearrangements is minimized over all of the edges of the tree. Finally, the tree was midpoint-rooted and the tree topology was visualized using Mega 7.0.

12. Statistical analysis

Statistical analysis was performed using the IBM SPSS Statistics 23 program (IBM, Armonk, NY) and Statistical R Core Team (2021) (R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/). The Fisher exact test, Pearson's Chi-squared test with Yates' continuity correction, or Pearson's Chi-squared test with Benjamini-Hochberg correction were used to analyze the association between populations of

H. pylori clinical isolates and *oipA* phase variation status and the presence or absence of *cag*PAI. A *p*-value less than 0.05 was considered to be statistically significant.

III. RESULTS:

PART I: *oipA* phase variation and associated virulence phenotypes in clinical *H. pylori* isolates

1. *oipA* phase variation and OipA protein expression is co-related

We randomly selected 40 *Helicobacter pylori* isolates from patients in the United States[65,75], focusing on those that exhibited both 'on' and 'off' phase variations in the *oipA* gene. Specifically, 18 isolates with the *oipA* 'on' phase (Figure 1), while 22 isolates with the 'off' phase (Figure 2). These samples were obtained from patients with various gastric diseases, primarily gastritis, with 2 samples originating from patients with cancerous conditions (Table 3).

Table 3. *oipA* phase variations and associated disease conditions of clinical isolates.

oipA phase variation	Disease condition	Isolates count
on	Cancer	1
	Gastric ulcer	2
	Duodenal Ulcers	6
	Esophagitis	4
	Gastritis	5
off	Barrett's Esophagus	1
	Gastric ulcer	3
	Duodenal Ulcers	6
	Esophagitis	1
	Duodenitis	4
	Gastritis	7

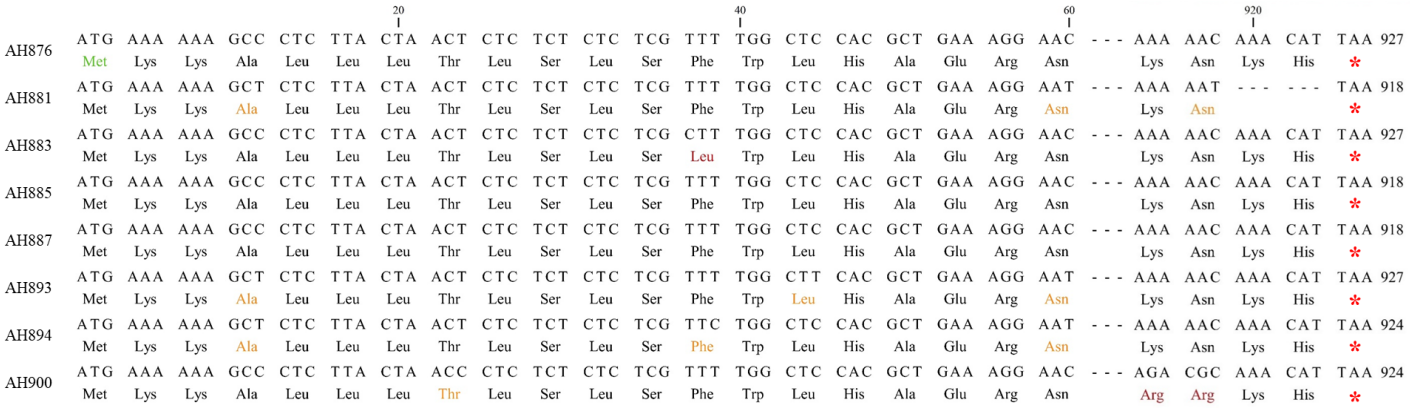


Figure 1. *oipA* gene sequence and its amino acid sequence of clinical isolates with *oipA* 'on' phase variation.

The expression of the *oipA* gene is regulated by slipped strand mispairing based on CT dinucleotide repeats. When the phase is 'on', the *oipA* gene expresses the OipA protein, which consists of approximately 307 amino acids.

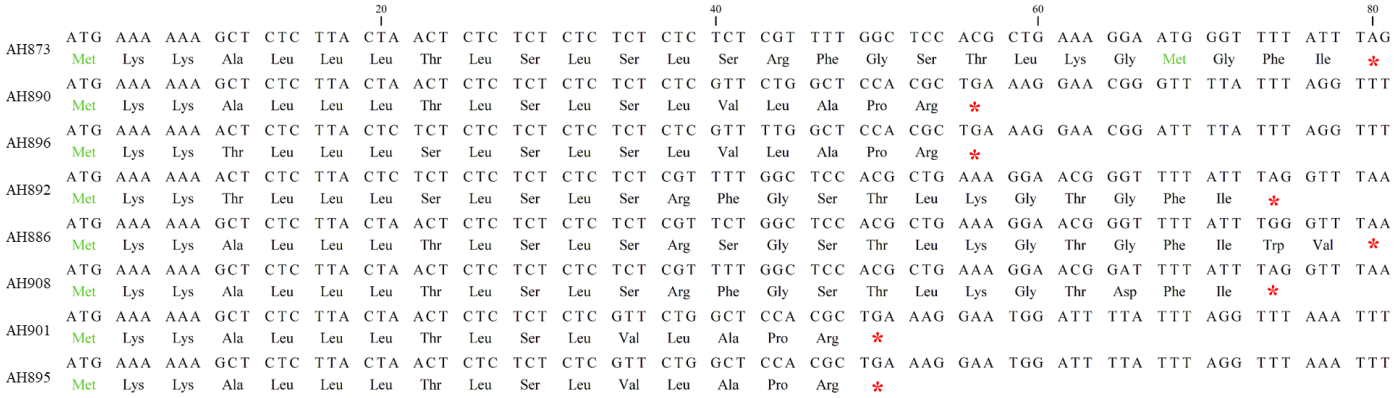


Figure 2. *oipA* gene sequence and its amino acid sequence of clinical isolates with *oipA* 'off' phase variation.

The expression of the *oipA* gene is regulated by slipped strand mispairing based on CT dinucleotide repeats. When the phase is 'off', the *oipA* gene is unlikely to express the OipA protein. In this 'off' phase, in silico translation results in a truncated protein sequence of approximately 16 amino acids due to a frameshift caused by the slipped strand mispairing of the CT repeats.

Next, we confirmed whether the *oipA* phase variation correlates with OipA protein expression in vitro when *H. pylori* is stimulated with AGS cells. We infected AGS cells for 6 hours, collected the cell lysates, and performed immunoblotting to detect the OipA protein. Consistent with the PCR and sequencing results, all isolates with the *oipA* gene in the 'on' phase displayed a positive band for the OipA protein in the immunoblot. Conversely, we did not detect the OipA protein in isolates where the oipA gene was in the 'off' phase (Figure 3).

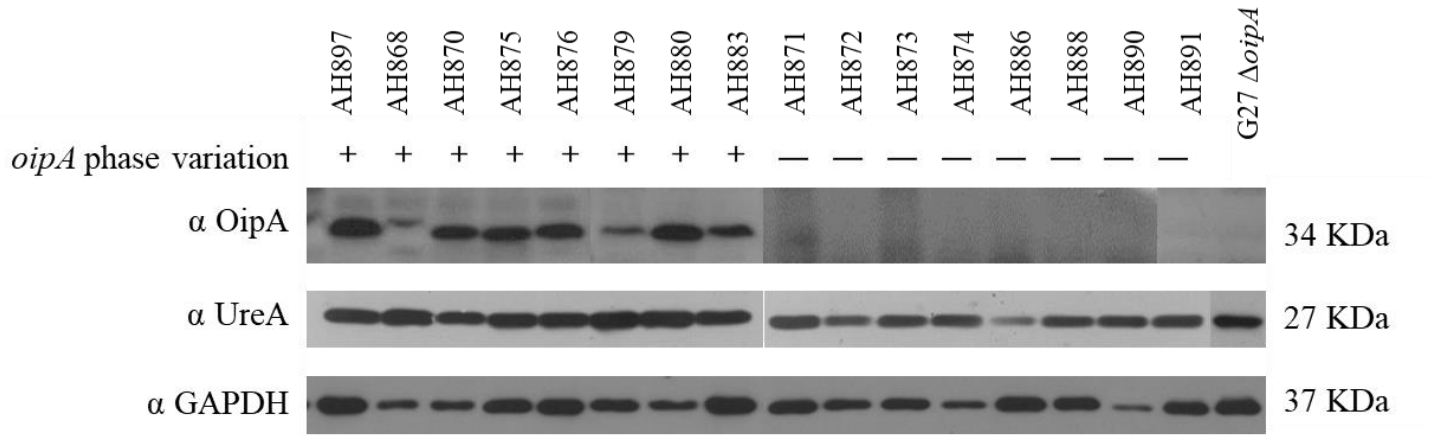


Figure 3. Immunoblot detection of OipA protein expression in *H. pylori* clinical isolates.

AGS cells were infected with all *H. pylori* isolates with *oipA* phase variation 'on' (+) or 'off' (-) for 6 hours. Cell lysates were then subjected to immunoblotting using an anti-OipA antibody. UreA was used as the *H. pylori* control, and GAPDH was used as the AGS cell control.

2. Majority of *H. pylori* clinical isolates carrying *oipA* 'on' phase induces IL-8 in AGS cells

Next, given its known association with inflammation, we investigated how *oipA* phase variation affects interleukin-8 (IL-8) induction. *H. pylori* isolates were co-cultured with AGS cells for 6 hours, after which cell culture supernatants were collected to measure IL-8 induction. The results showed that 16 out of 18 (88.9%) *oipA* 'on' isolates induced IL-8 similar to the positive control, whereas only 2 out of 22 (9.0%) *oipA* 'off' isolates were able to induce IL-8 (Figure 4). These patterns between the *oipA* 'on' and *oipA* 'off' isolates suggest a potential association between *oipA* phase variation and the ability to induce IL-8 in *H. pylori* clinical isolates.

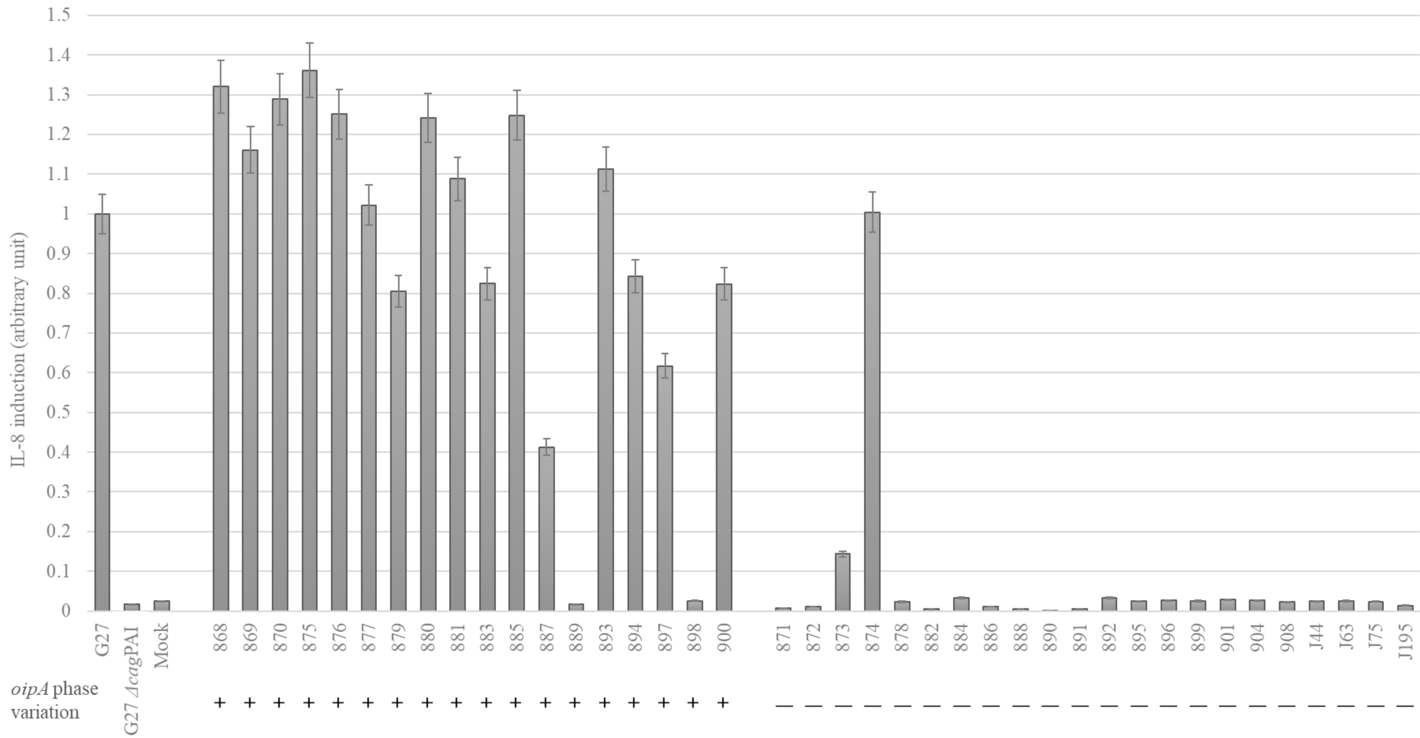


Figure 4. IL-8 Induction by *H. pylori* clinical Isolates with *oipA* phase variations 'on' (+) or 'off' (-).

Normalized levels of IL-8 induction in AGS cells after *H. pylori* infection. AGS cells were infected with clinical isolates for 6 hours, and cell culture supernatants were analyzed for IL-8 induction (pg/ml) using ELISA. Values were normalized using G27 WT as the positive control, set to 1.

3. *oipA* phase variation of clinical isolates affects the gastrin expression in G240-Luc cells.

To evaluate the impact of *oipA* phase variation on gastrin expression, *H. pylori* clinical isolates were co-cultured with G240-Luc cells containing a gastrin promoter-luciferase fusion. After 6 hours, cells were washed with PBS, lysed, and samples were collected to conduct the Luciferase assay. The results revealed that 16 out of 18 (88.9%) *oipA* 'on' samples induced a positive gastrin expression response, whereas only 2 out of 22 (9.0%) *oipA* 'off' samples exhibited a positive gastrin expression response (pg/ml) (Figure 5). These findings suggest that *oipA* phase variation may play a role in modulating gastrin expression.

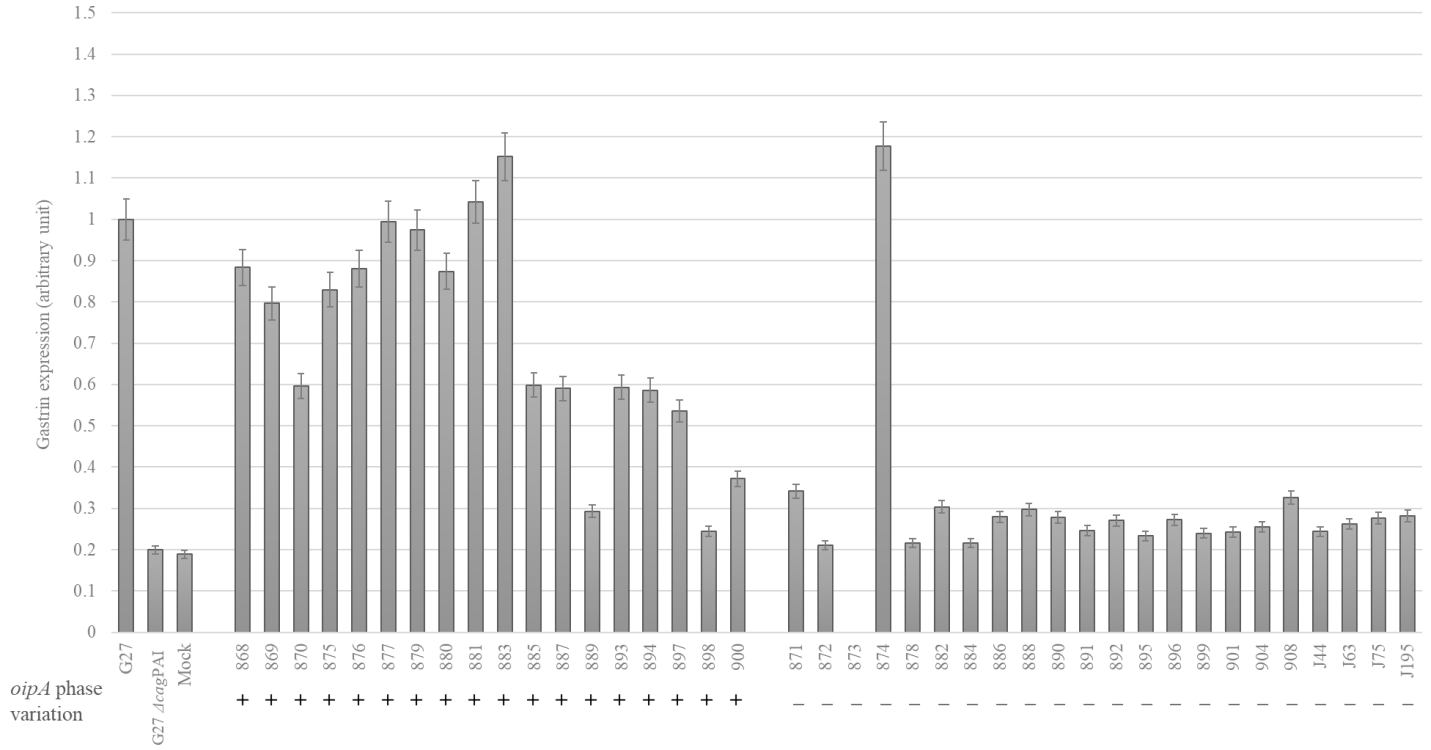


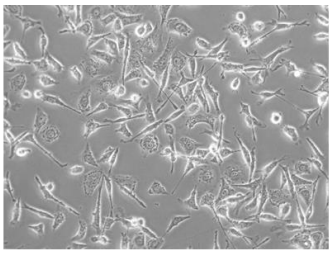
Figure 5. Gastrin expression by *H. pylori* clinical Isolates with *oipA* phase variations ʻon' (+) or ʻoff' (-).

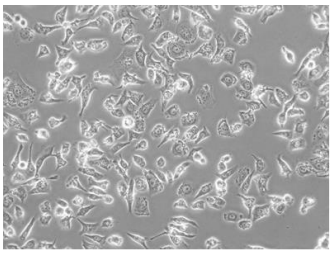
Normalized levels of Gastrin expression in G240-Luc cells cells after *H. pylori* infection. G240-Luc cells were infected with clinical isolates for 6 hours, and cell lysates were analyzed for Gastrin expression using Luciferase assay. Values were normalized using G27 WT as the positive control, set to 1.

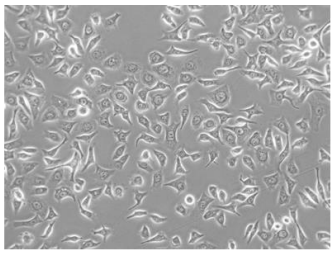
4. *oipA* phase variation of clinical isolates may associated with CagA translocation and phosphorylation.

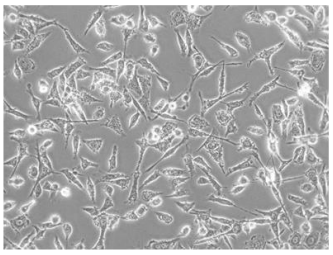
Since both IL-8 induction and gastrin expression induction by *H. pylori* are associated with the presence of *cag*PAI, we next wanted to see if these clinical *H. pylori* isolates show a correlation with the Cell elongation phenotype/ hummingbird phenotype which is a determinant of translocation and phosphorylation of CagA toxin to the AGS cells.

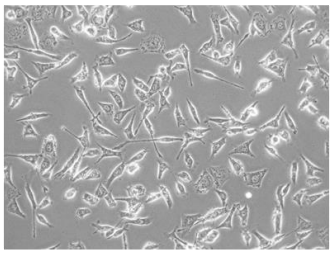
We infected AGS cells with H. pylori isolates for 6 hours, and the cells were fixed with 4% PFA. Images of the cells were captured under ×200 magnification for further analysis. 12 out of 18 (66.6%) *oipA* 'on' isolates induced the cell elongation/hummingbird phenotype. In contrast, only 1 out of 22 (4.5%) *oipA* 'off' isolates induced the cell elongation/hummingbird phenotype (Figure 6). These results suggest that oipA phase variation in clinical isolates may play a role in inducing the cell elongation/hummingbird phenotype underscoring the potential role of *oipA* phase variations in influencing the virulence characteristics of *H. pylori*, affecting its interaction with host cells.

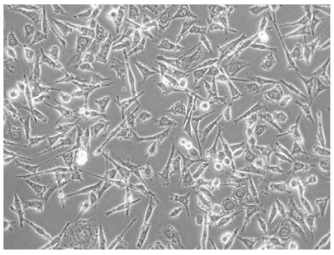

G27 WT

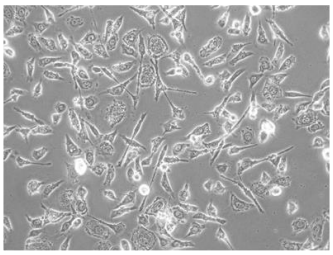

G27 *Δcag*PAI

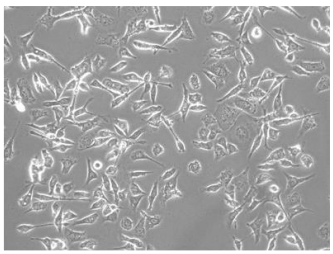

MOCK

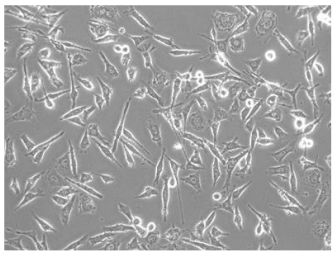

AH868 (+)

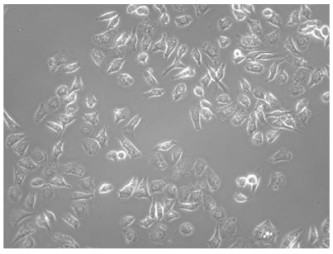

AH870 (+)

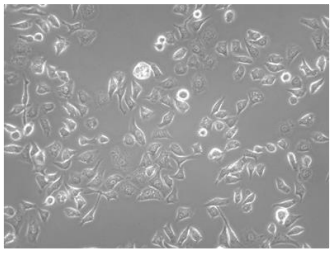

AH875 (+)

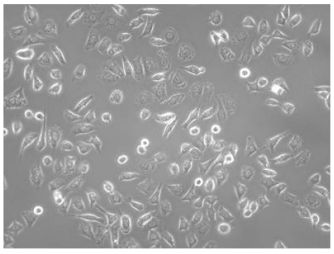

AH876 (+)

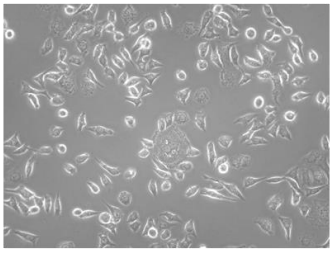

AH883 (+)

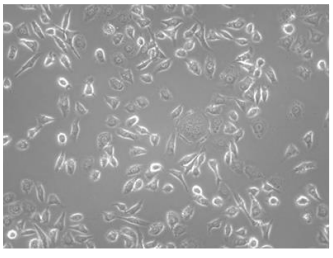

AH893 (+)


AH872 (-)


AH878 (-)


AH884 (-)


AH886 (-)


AH888 (-)

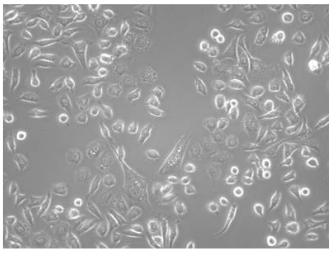

AH891 (-)

Figure 6. Cell elongation in AGS cells induced by *H. pylori* clinical isolates.

AGS cells were infected with *H. pylori* clinical isolates that possess *oipA* phase variations 'on' (+) or 'off' (-) at a multiplicity of infection (MOI) of 100 for 6 hours. After infection, cells were washed and fixed with 4% paraformaldehyde (PFA), and micrographs were obtained under 200X magnification. *H. pylori* strain G27 WT was used as the positive control, and G27*Δcag*PAI as the negative control.

Subsequently, we performed immunoblot analysis to assess CagA expression, translocation into AGS cells, and the tyrosine phosphorylation of CagA for each isolate. AGS cells were infected with *H. pylori* isolates for 6 hours, followed by cell lysis and isolation of proteins for immunoblot analysis. Among the *oipA* 'on' isolates, 16 out of 18 (94.1%) were positive for total CagA expression (α CagA), whereas, among the *oipA* 'off' isolates, only 2 out of 22 (9.5%) were positive (Figure 7). CagA translocation to AGS cells via the Type 4 secretion system (T4SS) was determined by detecting phosphorylated CagA (α PY99), with 14 out of 18 (77.7%) *oipA* 'on' isolates being positive, compared to only 2 out of 22 (9.5%) *oipA* 'off' isolates (Figure 7). These results indicate that *oipA* phase variation in *H. pylori* clinical isolates affects both CagA expression and its translocation/phosphorylation in AGS cells, highlighting the role of *oipA* in modulating *cag*PAI-related virulence mechanisms.

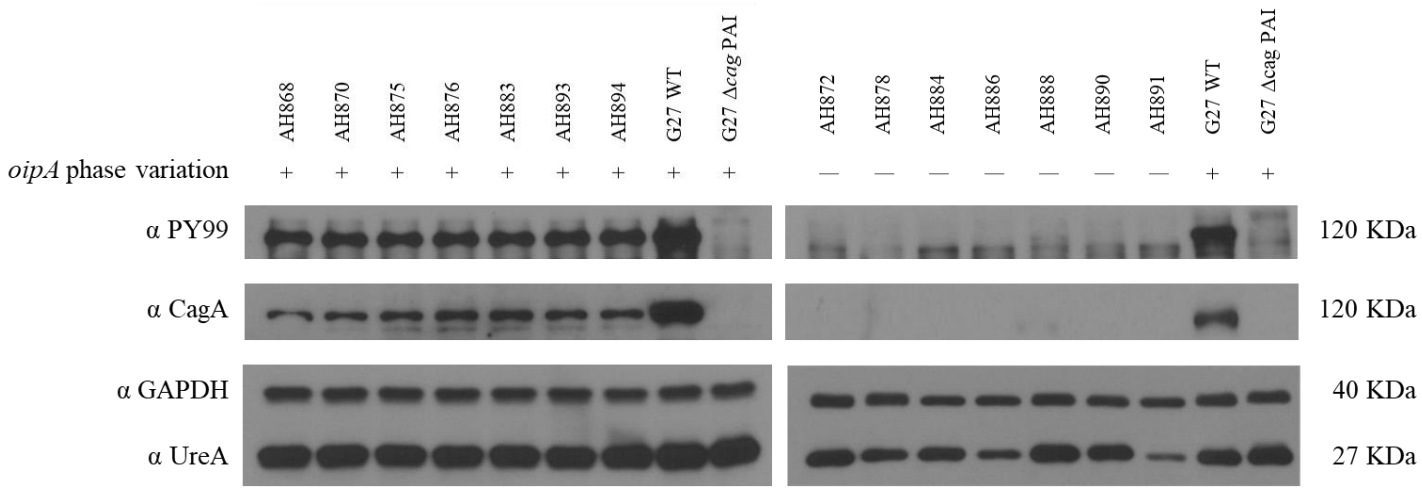


Figure 7. Total CagA Expression and Translocation/Phosphorylation by Immunoblot.

AGS cells were infected with *H. pylori* clinical isolates that possess *oipA* phase variations 'on' (+) or 'off' (-) at a multiplicity of infection (MOI) of 100 for 6 hours. After infection, cells were washed, lysed and protein were isolated for immunoblot. Total CagA expression was detected by α CagA antibody and tyrosine phosphorylation of translocated CagA was detected by α PY99 antibody. UreA and GAPDH were used as *H. pylori* control and AGS cells control respectively.

5. Field Emission Scanning Electron Microscopy to visualize the T4SS

Given the strong presentation of positive phenotypes related to the *cag*PAI in *oipA* 'on' isolates and the strong presentation of negative phenotypes related to the *cag*PAI in *oipA* 'off' isolates, we decided to utilize Field Emission Scanning Electron Microscopy (FESEM) to visualize the T4SS. *H. pylori* isolates were co-cultured with AGS cells and used for the FESEM imaging. Unlike the well-known positive control used G27 WT, we could not clearly identify the T4SS structure under FESEM in clinical isolates.

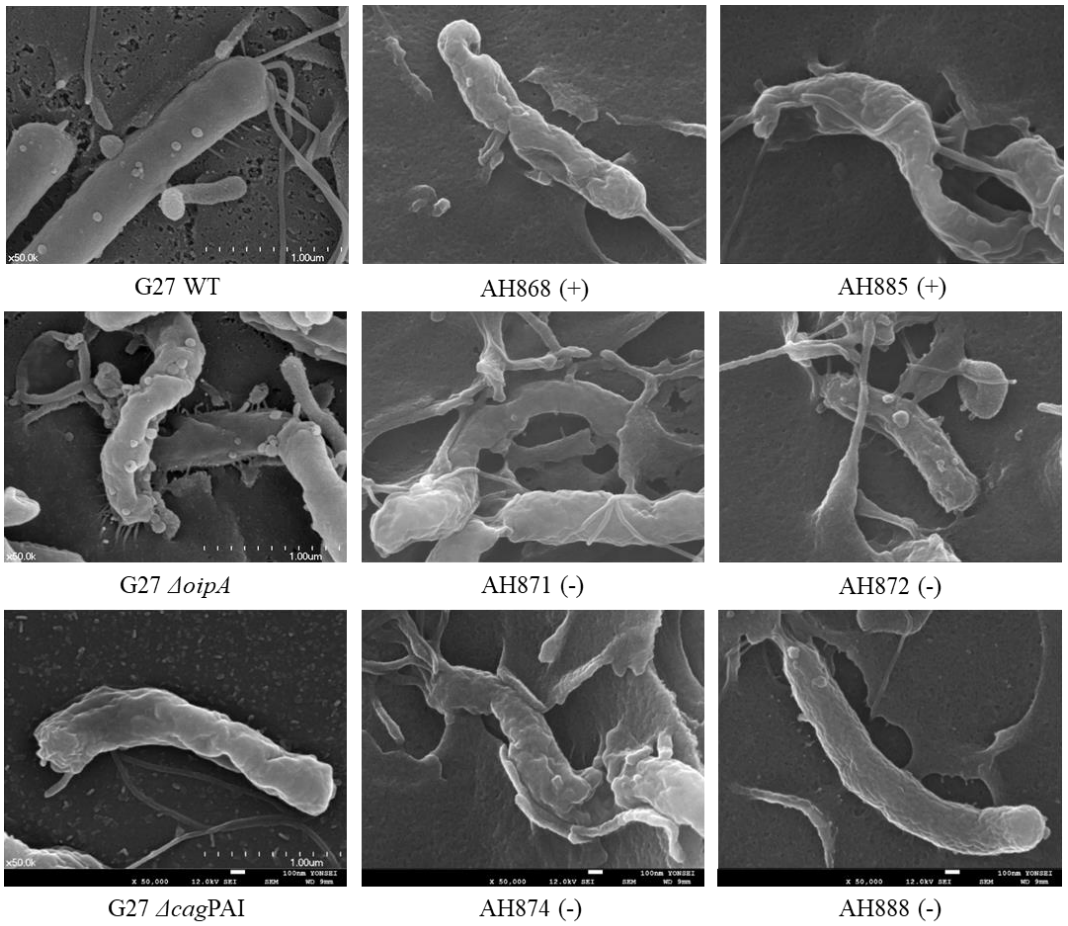


Figure 8. Field Emission Scanning Electron Microscopy (FESEM) to visualize the T4SS among *oipA* 'on' (+) vs *oipA* 'off' (-) clinical isolates.

G27 WT was used as the positive control and G27 Δ*cag*PAI was used as the negative control. Also G27 Δ*oipA* was used as a control for *oipA*.

6. *oipA* phase variations may be associated with the virulence phenotypes of *cag*PAI-T4SS

In aggregate, most of the *oipA* 'on' clinical isolates displayed phenotypic similarities to *cag*PAI-positive strains for IL-8 induction, Gastrin expression, cell elongation, CagA expression, and CagA translocation/ tyrosine phosphorylation (Table 4). Conversely, *oipA* phase variation 'off' clinical isolates displayed phenotypic similarities to *cag*PAI-negative strains (Table 5). These findings suggest that *oipA* phase variations in clinical isolates may be associated with the presence or absence of the *cag*PAI in the genome.

Table 4. Summary of Virulence phenotypes of *oipA* phase variation 'on' (+) *H. pylori* clinical isolates

Name	*oipA* phase variation	OipA expression	Total CagA	pCagA	Cell elongation	Gastrin expression	IL-8 Induction
AH868	+	+	+	+	+	+	+
AH869	+	+	+	+	+	+	+
AH870	+	+	+	+	+	+	+
AH875	+	+	+	+	+	+	+
AH876	+	+	+	+	+	+	+
AH877	+	+	+	+	+	+	+
AH879	+	+	+	+	-	+	+
AH880	+	+	+	+	+	+	+
AH881	+	+	+	-	-	+	+
AH883	+	+	+	+	+	+	+
AH885	+	+	+	+	+	+	+
AH887	+	+	+	+	-	+	+
AH889	+	+	-	-	-	-	-
AH893	+	+	+	+	+	+	+
AH894	+	+	+	+	+	+	+
AH897	+	+	+	-	-	+	+
AH898	+	+	-	-	-	-	-
AH900	+	+	+	+	+	+	+
	18/18	**18/18**	**16/18**	**14/18**	**12/18**	**16/18**	**16/18**

Table 5. Summary of Virulence phenotypes of *oipA* phase variation 'off' (-) *H. pylori* clinical isolates

	oipA phase variation	OipA expression	Total CagA	pCagA	Cell elongation	Gastrin expression	IL-8 Induction
AH871	-	-	-	-	-	-	-
AH872	-	-	-	-	-	-	-
AH873	-	-	+	+	-	+	+
AH874	-	-	+	+	+	+	+
AH878	-	-	-	-	-	-	-
AH882	-	-	-	-	-	-	-
AH884	-	-	-	-	-	-	-
AH886	-	-	-	-	-	-	-
AH888	-	-	-	-	-	-	-
AH890	-	-	-	-	-	-	-
AH891	-	-	-	-	-	-	-
AH892	-	-	-	-	-	-	-
AH1303	-	-	-	-	-	-	-
AH895	-	-	-	-	-	-	-
AH1304	-	-	-	-	-	-	-
AH896	-	-	-	-	-	-	-
AH1305	-	-	-	-	-	-	-
AH899	-	-	-	-	-	-	-
AH901	-	-	-	-	-	-	-
AH904	-	-	-	-	-	-	-
AH1309	-	-	-	-	-	-	-
AH908	-	-	-	-	-	-	-
	0/22	**0/22**	**2/21**	**2/21**	**1/22**	**2/21**	**2/22**

7. The phase variations of *oipA* suggest a strong correlation with presence or absence of *cag*PAI in *H. pylori* clinical isolates.

The strong trend in the phenotype of *oipA* on vs *oipA* off isolates which mirrored the phenotype of *cag*PAI prompted us to investigate the presence or absence of the *cag*PAI in the genome of these clinical isolates. We used a PCR-based method to analyze the *cag*PAI region by designing 7 sets of primers using universal sequences to amplify 7 separate overlapping sections, thereby completely covering the 40 kb cagPAI (Figure 9).

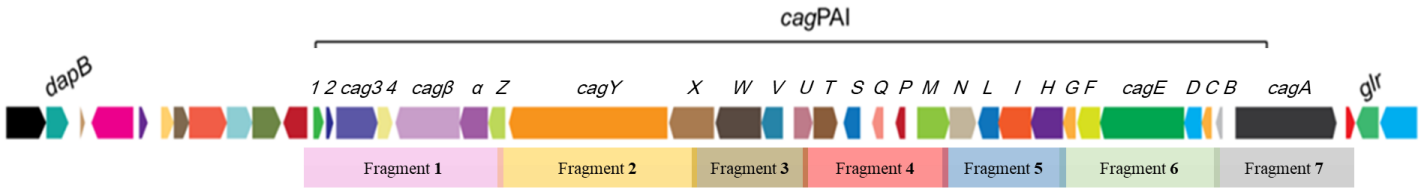


Figure 9. PCR screening of the *cag*PAI region.

PCR based technique was employed to analyze the *cag*PAI by amplifying 7 separate sections of *cag*PAI each overlapping another thus completely covering the 40Kb *cag*PAI.

1. *HP0519-cagZ* (8~ Kbp), **2.** *cagZ-cagX* (7~ Kbp), **3.** *cagX-cagU* (4~ Kbp), **4.** *cagU-cagM* (6~ Kbp),

5. *cagM-cagH* (5.5~ Kbp), **6.** *cagH-cagD* (5.2~ Kbp), and **7.** *cagD-mur1* (7~ Kbp)

(Figure is adapted from the Doctoral Dissertation of Hanfu Su 2017)

The majority of *oipA* 'on' isolates exhibited bands in almost all 7 PCRs (Table 6), indicating the presence of the complete *cag*PAI region. Conversely, the majority of *oipA* 'off' isolates either showed no bands in any of the 7 PCRs or only in some (Table 7), suggesting absence or partial presence of *cag*PAI thus being nonfunctional. These findings highlight a strong correlation between *oipA* phase variation and the *cag*PAI in these *H. pylori* clinical isolates, underscoring the potential utility of *oipA* phase variation in modulating *cag*PAI and its association in virulence of *H. pylori.*

Table 6. PCR screening of the *cag*PAI region in *oipA* 'on' (+) clinical isolates.

Name	*oipA* phase variation	Fragment 1	Fragment 2	Fragment 3	Fragment 4	Fragment 5	Fragment 6	Fragment 7	
AH868	+	+	+	+	+	+	+	+	7/7
AH869	+	+	+	+	+	+	+	+	7/7
AH870	+	-	+	+	+	+	+	+	6/7
AH875	+	+	+	+	+	+	+	+	7/7
AH876	+	+	+	+	+	+	+	+	7/7
AH877	+	-	+	+	+	+	-	-	4/7
AH879	+	-	+	-	-	-	+	+	3/7
AH880	+	+	+	+	+	+	+	+	7/7
AH881	+	+	+	+	+	+	+	+	7/7
AH883	+	+	+	+	+	+	+	+	7/7
AH885	+	+	+	+	+	+	+	+	7/7
AH887	+	+	+	+	+	+	+	+	7/7
AH889	+	+	-	-	-	-	-	-	1/7
AH893	+	+	+	+	+	+	+	+	7/7
AH894	+	+	-	+	+	+	+	+	6/7
AH897	+	+	+	+	+	+	+	+	7/7
AH898	+	+	-	-	-	-	-	-	1/7
AH900	+	+	+	+	+	+	+	+	7/7

Table 7. PCR screening of the *cag*PAI region in *oipA* 'off' (-) clinical isolates.

Name	*oipA* phase variation	Fragment 1	Fragment 2	Fragment 3	Fragment 4	Fragment 5	Fragment 6	Fragment 7	
AH871	-	-	-	-	-	-	-	-	0/7
AH872	-	-	-	-	-	-	-	-	0/7
AH873	-	-	+	+	+	+	-	-	4/7
AH874	-	+	+	+	+	+	+	+	7/7
AH878	-	-	+	+	-	-	-	+	2/7
AH882	-	-	+	+	+	-	-	-	3/7
AH884	-	-	-	-	-	-	-	-	0/7
AH886	-	-	-	-	-	-	-	-	0/7
AH888	-	+	-	+	-	-	-	-	2/7
AH890	-	-	-	-	-	-	-	-	0/7
AH891	-	-	-	-	-	-	-	-	0/7
AH892	-	-	-	-	-	-	-	-	0/7
AH1303	-	-	-	+	-	-	-	-	1/7
AH895	-	-	-	-	-	-	-	-	0/7
AH1304	-	-	-	-	-	-	-	-	0/7
AH896	-	-	-	-	-	-	-	-	0/7
AH1305	-	-	-	-	-	-	-	-	0/7
AH899	-	-	-	-	-	-	-	-	0/7
AH901	-	-	-	-	-	-	-	-	0/7
AH904	-	-	-	-	-	-	-	-	0/7
AH1309	-	+	-	-	-	-	-	-	1/7
AH908	-	-	-	-	-	-	-	-	0/7

8. Association of *oipA* phase variations and *cag*PAI.

Establishing that the *oipA* status may associated with the presence or absence of the *cag*PAI, we next wanted to explore the differences between *oipA* on *H. pylori* and *oipA* off *H. pylori* at the genomic level. Thus, we sought to define the complete genome sequence from a subset of *oipA* 'on' an *oipA* 'off' clinical isolates. Out of all 40 clinical isolates studied, we selected 20 representative clinical isolates from each *oipA* 'on' (10) and *oipA* 'off' (10) group, ensuring that *oipA* phase variation 'on' isolates exhibited a WT-like virulence phenotype with intact *cag*PAI, while *oipA* phase variation 'off' isolates displayed otherwise. Also we selected 2 strains from each group that displayed a notable contradictory phenotype for whole genome sequencing.

After establishing the complete genome sequences first, we confirmed that the 10 *oipA* 'on' clinical isolates carried the almost all 27~ genes of cagPAI while none of the *oipA* 'off' clinical isolates carried any *cag*PAI genes. This further confirmed that virulence phenotype exhibited by the *oipA* phase variation 'off' isolates and *oipA* phase variation 'off' isolates are indeed associated with the virulence phenotype of *cag*PAI and are associated with the presence or absence of the *cag*PAI (Table 8).

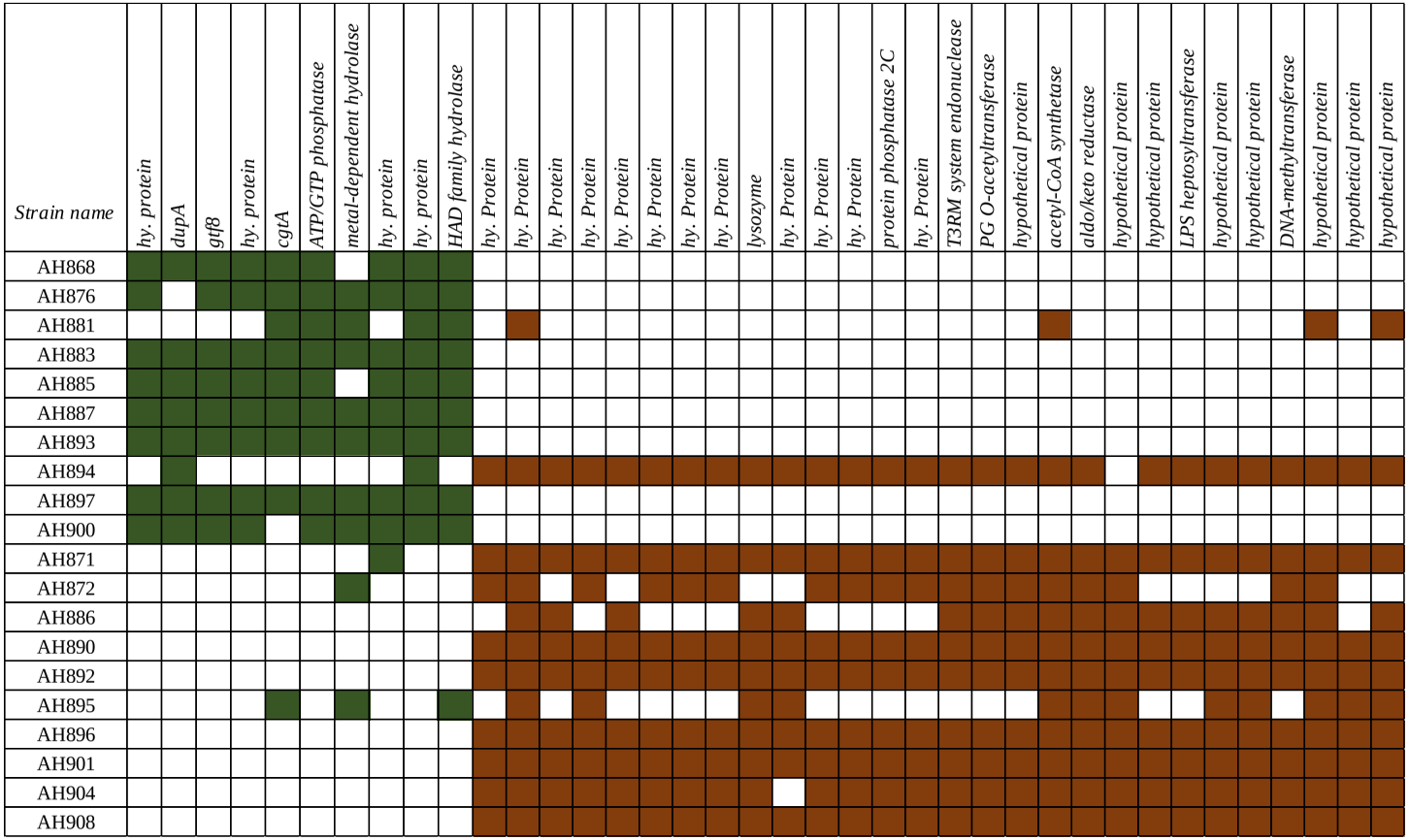
Table 8. Presence or absence of *cag*PAI genes in selected clinical isolates with *oipA* 'on' (+) or *oipA* 'off' (-) phase variations.

Strain name	*oipA* phase variation	*cag1*	*cag2*	*cag3*	*cag4*	*cagβ*	*cagα*	*cagZ*	*cagY*	*cagX*	*cagW*	*cagV*	*cagU*	*cagT*	*cagS*	*cagQ*	*cagM*	*cagN*	*cagL*	*cagI*	*cagH*	*cagG*	*cagF*	*cagE*	*cagD*	*cagC*	*cagB*	*cagA*
AH868	+	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
AH876	+	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
AH881	+	■	□	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
AH883	+	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
AH885	+	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
AH887	+	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
AH893	+	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
AH894	+	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	□	■
AH897	+	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	□	■
AH900	+	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
AH871	-																											
AH872	-																											
AH886	-																											
AH890	-																											
AH892	-																											
AH895	-																											
AH896	-																											
AH901	-																											
AH904	-																											
AH908	-																											

9. Comparative analysis revealed other genes that may be involved in the virulence phenotype mediated via *cag*PAI and *oipA* phase variations.

We next wanted to find out what other differences we could see in genomes of this *oipA* 'on' phase variation and *oipA* 'off' phase variation clinical isolates contributing to these distinct phenotypes. We performed comparative genome analysis to see what other genes that might come up with *oipA* 'on' phase variation and the presence of *cag*PAI or *oipA* 'off' phase variation and absence of *cag*PAI. Surprisingly, we found that 10 genes show up that are at least 80% associated with the *oipA* 'on' phase variation and the presence of *cag*PAI (Table 9). These 10 genes suggest that the virulence phenotype of *cag*PAI and its association with *oipA* 'on' phase may be rather complex and these mentioned 10 genes also may contribute to the phenotype. Similarly, we found that there is another set of 28 genes that is at least 80% associated with the absence of *cag*PAI and *oipA* 'off' phase variation (Table 9). Compared to *cag*PAI positive and *oipA* 'on' clinical isolates, *cag*PAI negative and *oipA* 'off' isolates possess many more genes which leads us to speculate that *cag*PAI negative isolates may have their own unique way of modulating virulence in the absence of *cag*PAI.

Table 9. Comparative analysis revealed other genes that are associated with *oipA* phase variation and presence or absence of *cag*PAI.



10. Outlier clinical isolates may lead us to understand the role of *oipA* phase variation on modulating *cag*PAI virulence.

We found 2 specific strains that displayed the opposite characteristics compared to the rest of the strains of *oipA* 'on' or *oipA* 'off', B221 with *oipA* phase variation 'on' and B129 with *oipA* phase variation 'off' (Table 3, 4, 5, & 6). We wanted to understand if we could use these 2 *H. pylori* strains to understand the role of *oipA* phase variation on the acquisition and excision of *cag*PAI. Thus, we performed whole genome sequencing of these two strains. Interestingly we confirmed that B129 possesses the complete *cag*PAI while B221 does not (Table 10).

Table 10. Association of *cag*PAI genes in B221 and B129

Strain name		AH889	AH874
oipA phase variation		on	off
*cag*PAI genes	*cag1*		
	cag2		
	cag3		
	cag4		
	cagβ		
	cagα		
	cagZ		
	cagY		
	cagX		
	cagW		
	cagV		
	cagU		
	cagT		
	cagS		
	cagQ		
	cagM		
	cagN		
	cagL		
	cagI		
	cagH		
	cagG		
	cagF		
	cagE		
	cagD		
	cagC		
	cagB		
	cagA		

We next aimed to identify additional genomic differences in these two isolates, similarly, we conducted a comparative genome analysis to pinpoint other genes associated with the presence of *cag*PAI in *oipA* 'on' phase variants (10 genes), and the absence of *cag*PAI in *oipA* 'off' phase variants (28 genes).

Surprisingly, our analysis revealed 8 out 10 genes that were associated with the *oipA* 'on' phase variation and the presence of *cag*PAI were present in the B129 isolate (Table 11.A). Also, our analysis revealed 18 out of 28 genes that were associated with the absence of *cag*PAI and the *oipA* 'off' phase variation was present in the B221 isolate (Table 11.A). These findings suggest that these associated genes may be closely associated with the presence or absence of *cag*PAI and this might help us to understand the roles of *oipA* phase variations in this.

Table 11. A. Association of other genes in B221 and B129.

Strain name		AH889	AH874
oipA phase variation		on	off
genes associated with the presence of *cag*PAI	*hy. protein*		
	dupA		
	gtf8		
	hy. protein		
	cgtA		
	ATP/GTP phosphatase		
	metal-dependent hydrolase		
	hy. protein		
	hy. protein		
	HAD family hydrolase		

Table 11. B. Association of other genes in B221 and B129

Strain name		AH889	AH874
oipA phase variation		on	off
genes associated with the absennce of cagPAI	*hy. Protein*		
	hy. Protein		
	hy. Protein		
	hy. Protein		
	hy. Protein		
	hy. Protein		
	hy. Protein		
	hy. Protein		
	lysozyme		
	hy. Protein		
	hy. Protein		
	hy. Protein		
	protein phosphatase 2C		
	hy. Protein		
	type III RM system end.		
	pg O-acetyltransferase		
	hypothetical protein		
	acetyl-CoA synthetase		
	aldo/keto reductase		
	hypothetical protein		
	hypothetical protein		
	LPS heptosyltransferase		
	hypothetical protein		
	hypothetical protein		
	DNA-MT		
	hypothetical protein		
	hypothetical protein		
	hypothetical protein		

11. Phylogenetic analysis of clinical *H. pylori* isolates

To understand the evolutionary relationships, in the dynamics of the *oipA* phase variations, and presence or absence of *cag*PAI, we generated a Maximum-Likelihood (ML) Genome Tree (Figure 10). The tree is divided into three main clades (or groups), each representing a different subset of the isolates. The clades show the evolutionary relationships among the isolates, with closer branches indicating more closely related isolates.

The first clade consists of AH894, AH900, AH876, AH885, AH883, AH887, AH893, AH868, and AH897 where all of which are *oipA* 'on' and *cag*PAI-positive, thus being the virulent clade. The strong bootstrap values (82%-99%) in the virulent clade indicate robust and reliable evolutionary relationships among these *cag*PAI-positive isolates. This suggests a common evolutionary path characterized by the retention of virulence factors.

The second (Mixed virulent clade) and third (nonvirulent clade) clades consist of AH886, AH904, AH895, AH872, 26695, AH881 and AH892, AH896, AH901, AH871, AH890, AH908 respectively. Surprisingly, AH881 (*oipA* 'on' with *cag*PAI present) is positioned closely to *oipA* off and *cag*PAI absent isolates. This close positioning may suggest a recent divergence or potential horizontal gene transfer

event where AH881 might have acquired *cag*PAI or maintained it while others may have lost it indicating selective pressure to keep these traits for pathogenicity.

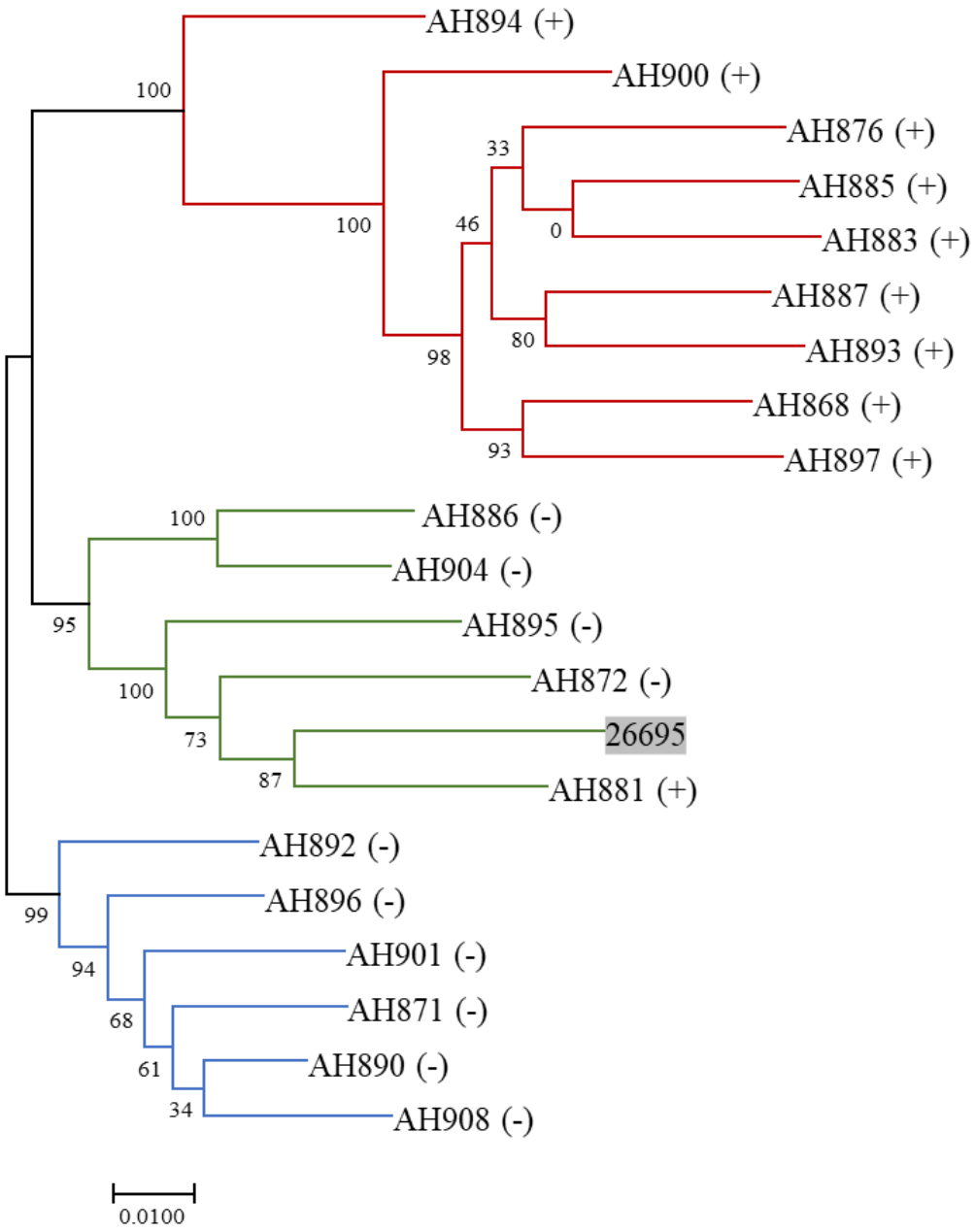


Figure 10. Maximum-Likelihood Genome Tree of 20 clinical isolates

***oipA* on (+) or *oipA* 'off' (-)**

Distance calculation method: Jukes-Cantor

ML Model: Generalized Time-Reversible, CAT approximation with 20 rate categories.

Bootstrap Values are indicated at the nodes (branch points) of the tree. High bootstrap values (generally above 70%) suggest that the grouping is well-supported.

The lengths of the branches represent the amount of genetic change. Longer branches indicate more significant evolutionary change.

The structure (topology) of the tree shows which strains are more closely related to each other.

PART II: Association of *oipA* phase variation and presence or absence of *cag*PAI in *H. pylori:* an analysis across geographic regions

1. *H. pylori* genomes preparation and trimming.

To further understand and confirm this association, we used available genome data of *H. pylori* isolates from worldwide. We had to exclude certain genome samples due to those samples being unusable for our analysis (Figure 11).

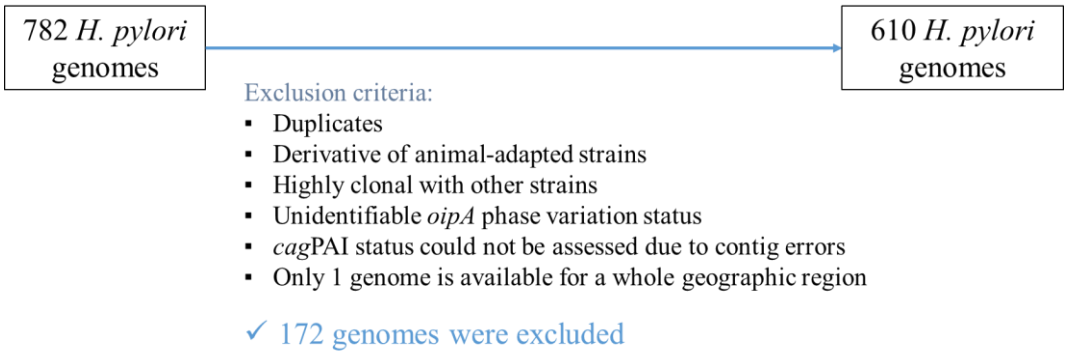


Figure 11. *H. pylori* genomes preparation and trimming.

Out of original collection of 782 genomes, we excluded 172 genomes due to lack of usability for our specific objectives of the analysis.

2. Distribution of Genomes across different geographic populations.

After the exclusion of 172 genomes, the remaining 610 genomes were spread across 5 distinct geographic regions (Figure 12). The hpEastAsia geographic population had the highest number of genomes (247 genomes), while the hpAfrica2 geographic population had the lowest number of genomes (14 genomes).

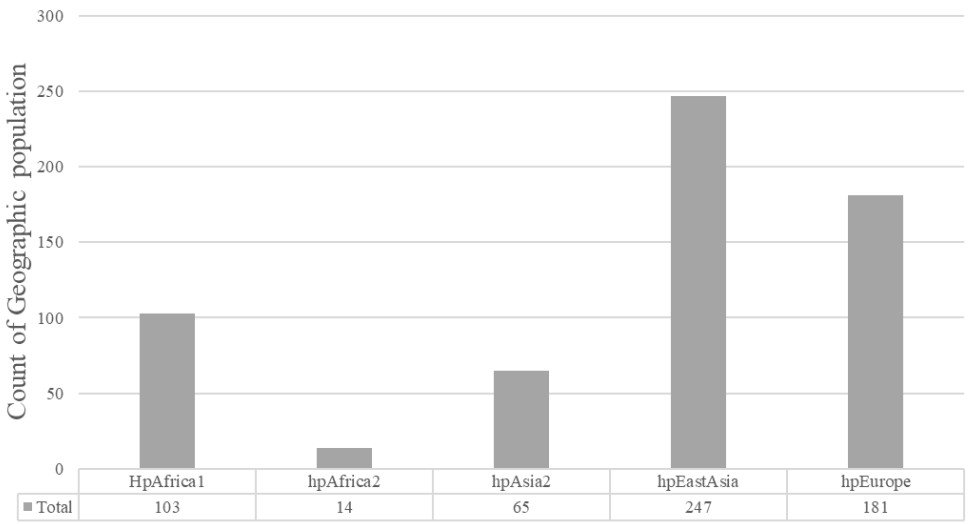


Figure 12. Counts of *H. pylori* isolates from each of the 5 geographic populations.

3. Distribution of *oipA* phase variation among geographic regions

The distribution of *oipA* phase variations varied among each geographic population. hpEurope displayed a much similar distribution for both *oipA* phase variations while, other geographic populations tend to be polarized towards one *oipA* phase variation, specifically *oipA* 'on' (Figure 13).

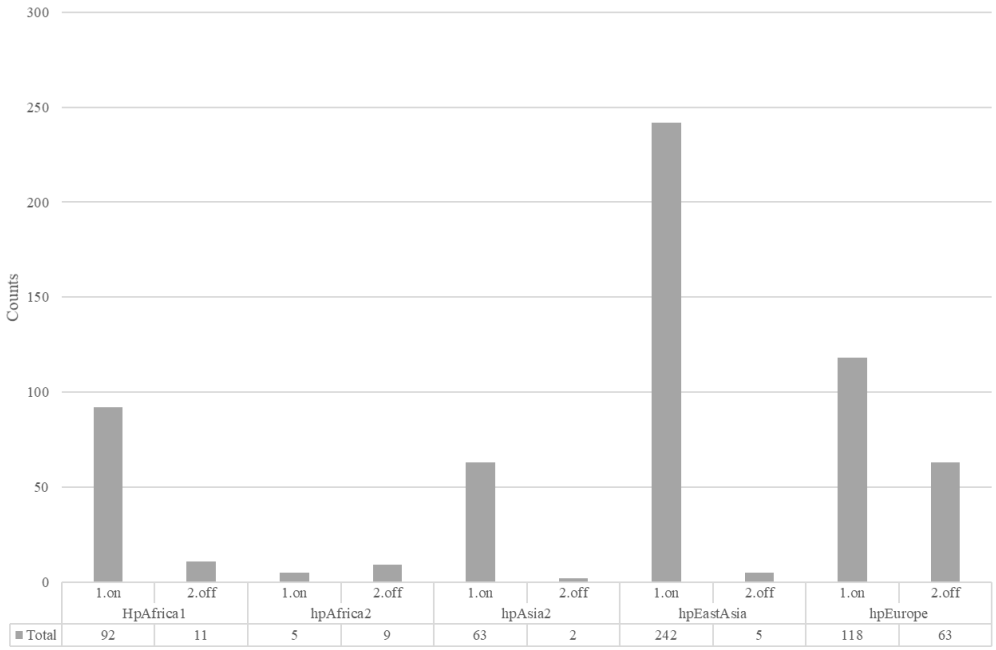


Figure 13. Counts of *oipA* phase variations among each of the 5 geographic regions.

4. Distribution of presence or absence of *cag*PAI among geographic regions

The distribution of presence or absence of *cag*PAI varied among each geographic population. hpEurope displayed a much similar distribution for both presence or absence of *cag*PAI while, other geographic populations tend to be polarized towards either the present or absent *cag*PAI, mainly being *cag*PAI present (Figure 14).

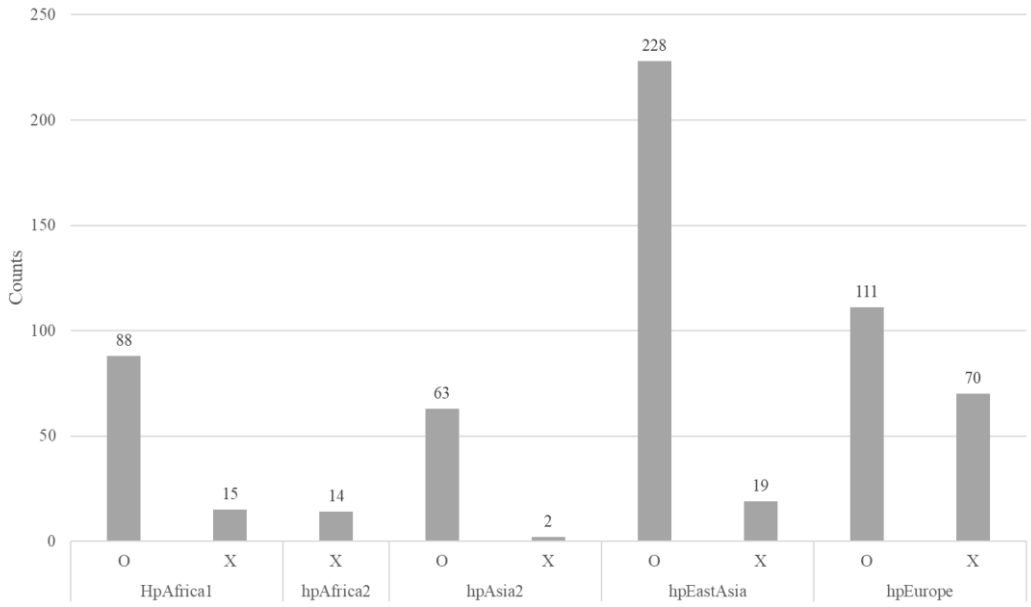


Figure 14. Counts of presence (o) or absence (x) of *cag*PAI among each of the 5 geographic regions.

5. Association of *oipA* phase variation and presence or absence of *cag*PAI among geographic regions

We first categorized the whole pool in to 4 categories based on *oipA* phase variation and the presence or absence of *cag*PAI (Figure 15). Also, these counts were categorized according to their geographic populations (Figure 16). Our analysis revealed regional associations between *oipA* phase variation 'on' status and the presence of intact *cag*PAI and *oipA* phase variation 'off' status and the absence of *cag*PAI, specifically highlighting the complex host-pathogen interactions in different geographic regions. Notably, a distinct association was observed in the HpEurope region where both *oipA* 'on' and *oipA* 'off' phase variations were commonly found (Table 12).

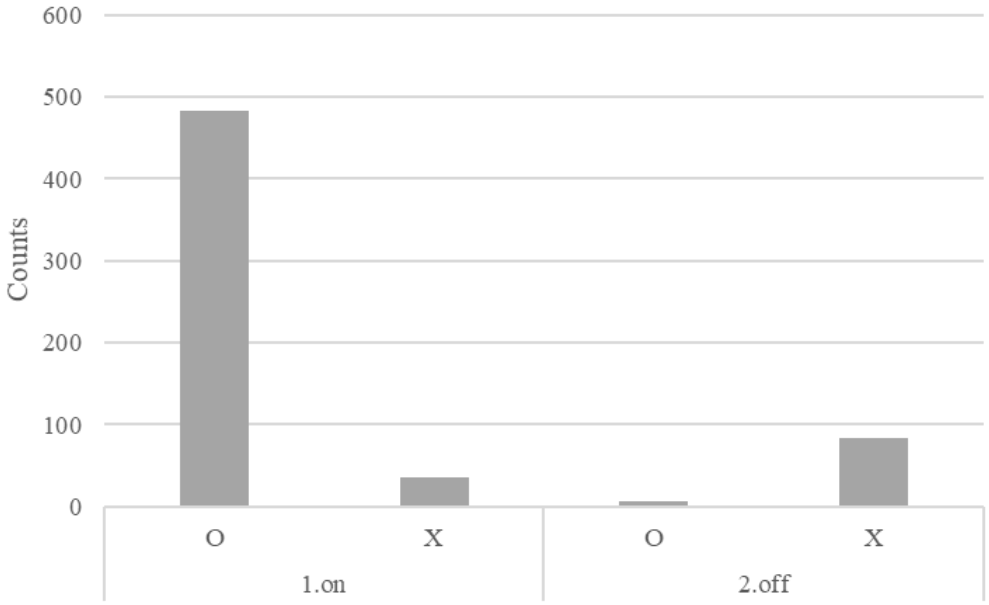


Figure 15. Distribution of presence or absence of *cag*PAI based on *oipA* phase variation.

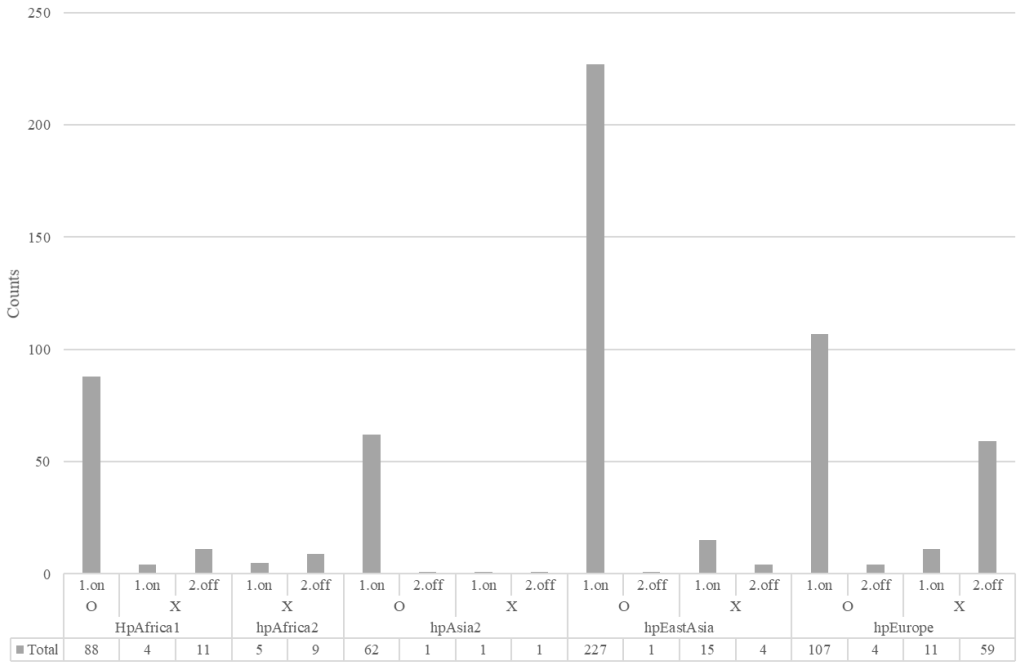


Figure 16. Counts of *oipA* phase variations along with presence (o) or absence (x) of *cag*PAI among each of the 5 geographic regions.

Table 12. Association between *oipA* phase variation and presence or absence of *cag*PAI.

Geographic population	presence (o) or absence (x) of *cag*PAI	*oipA* phase variation		Strain number	p-value
		off	on		
hpAfrica1	o	0	88	103	4.21E-11
	x	11	4		
hpAfrica2	o	0	0	14	1
	x	9	5		
hpEastAsia	o	2	226	247	9.42E-05
	x	4	15		
hpAsia2	o	1	62	65	0.06106
	x	1	1		
hpEurope	o	4	107	181	< 2.2E-16
	x	58	12		
Total	o	7	483	610	< 2.2E-16
	x	83	37		

IV. DISCUSSION

The study aimed to elucidate the genomic and phenotypic differences associated with *oipA* phase variations in clinical *Helicobacter pylori* isolates, and how these variations correlate with the major virulent factor: *cag* pathogenicity island (*cag*PAI). Our findings indicate a significant relationship between *oipA* phase variations and the virulence phenotypes of *H. pylori*, with distinct associations observed between the *oipA* 'on' and 'off' phases and the presence or absence of *cag*PAI.

Our investigation into *oipA* phase variation demonstrated that the Clinical isolates exhibiting the *oipA* 'on' phase not only expressed the OipA protein[74] but also induced higher levels of interleukin-8 (IL-8) and Gastrin in an in-vitro setting. These isolates also exhibited the cell elongation/hummingbird phenotype, a hallmark of CagA translocation and phosphorylation. In contrast, *oipA* 'off' isolates showed minimal to no IL-8 induction and gastrin expression and rarely induced the cell elongation phenotype. There were other studies conducted also acknowledge the role of *oipA* in relates to these phenotypes[72]. Together, these findings underscore the role of *oipA* phase variations in modulating the virulence characteristics of *H. pylori* which may associated with the *cag*PAI.

The strong correlation between *oipA* 'on' isolates and the presence of intact *cag*PAI was further supported by PCR and sequencing analyses. The majority of *oipA* 'on' isolates possessed intact *cag*PAI, while *oipA* 'off' isolates often lacked the entire *cag*PAI or had possibly nonfunctional remnants similar with some other previous studies[73,80,83]. This suggests that the virulence phenotype associated with *oipA* 'on' isolates may be largely driven by the presence of *cag*PAI, whereas *oipA* 'off' isolates may utilize alternative mechanisms to modulate virulence.

The comparative genome analysis revealed additional genes associated with the presence or absence of *cag*PAI in relation to *oipA* phase variation. Ten genes were identified that are at least 80% associated with the *oipA* 'on' phase and the presence of *cag*PAI, suggesting a complex interplay in the virulence phenotype. Conversely, 28 genes were associated with the absence of *cag*PAI and *oipA* 'off' phase, indicating that these isolates might have unique virulence modulation mechanisms independent of *cag*PAI. Also, it is worth noting that *cag*PAI negative strains possess higher number of unique genes compared to that of *cag*PAI positive strains. This lead us to suggest that these strains may use these 28 genes for modulations of virulence in the absence of *cag*PAI. There are reports discuss about the association of known virulent genes such as *vacA* with *cag*PAI, but there were no enough studies discuss about these specific genes and their roles in modulating the

virulence of *H. pylori.*

Interestingly, two outlier strains, B221 and B129, which did not follow the typical association patterns, provided insights into the genomic influences on virulence. B129, despite being *oipA* 'off', carried the complete *cag*PAI, while B221, with an *oipA* 'on' phase, lacked the entire *cag*PAI. Comparative genome analysis of these isolates revealed that the presence of specific genes could override the typical *oipA*-*cag*PAI association, suggesting potential horizontal gene transfer events or recent evolutionary divergence.

The phylogenetic analysis provided further insights into the evolutionary relationships among the isolates. The clustering of virulent *oipA* 'on' and *cag*PAI-positive isolates into a distinct clade with common evolutionary path characterized by the retention of virulence factors. This evolutionary insight underscores the importance of both *oipA* phase variation and *cag*PAI in the pathogenicity of *H. pylori.* It will be interesting to see if *oipA* phase variations played a role in acquisition or excision of *cag*PAI, along the journey of *H. pylori* evolution.

Analyzing the geographic distribution of *oipA* phase variations and *cag*PAI revealed regional differences. While some regions, such as hpEurope, displayed a balanced

distribution of *oipA* phase variations and *cag*PAI, others showed a distinct polarization towards either the 'on' phase and *cag*PAI presence or the 'off' phase and *cag*PAI absence. These findings were well supported by another recent study[84]. Most importantly, these regional differences highlight the complex host-pathogen interactions and suggest that local evolutionary pressures may influence the prevalence of these virulence factors. Again, further analyses and studies will help us to understand if *oipA* as a phase variable gene and as an Outer membrane protein played a role in regulating the *cag*PAI in a mechanistic pathway.

This study demonstrates a strong correlation between *oipA* phase variations and the *cag*PAI in *H. pylori* clinical isolates, significantly influencing their virulence phenotypes. The identification of additional genes associated with these variations suggests a complex genetic interplay that modulates the pathogenicity of *H. pylori.* Understanding these relationships enhances our comprehension of *H. pylori* virulence and could inform targeted therapeutic strategies and diagnostic tools to manage *H. pylori*-related diseases. Future research should focus on the functional characterization of the identified genes and their role in *H. pylori* pathogenicity, as well as the evolutionary mechanisms driving these associations across different geographic regions.

V. APPENDIX

Supplementary Table 1. Selected *H. pylori* genomes from different geographic populations

No	strain name	population	*oipA* phase variation	presence or absence of *cag*PAI
1	26695	hpEurope	on	O
2	J99	HpAfrica1	on	O
3	51	hpEastAsia	on	O
4	HPAG1	hpEurope	on	O
5	Shi470	hpEastAsia	on	O
6	G27	hpEurope	on	O
7	P12	hpEurope	on	O
8	52	hpEastAsia	on	O
9	B38	hpEurope	off	X
10	908	hpAfrica1	on	O
11	SJM180	hpEurope	off	O
12	PeCan4	hpEastAsia	on	O
13	Cuz20	hpEastAsia	on	O
14	35A	hpEastAsia	on	O
15	India7	hpAsia2	on	O
16	83	hpEastAsia	on	O
17	Puno120	hpEastAsia	on	O
18	Puno135	hpEastAsia	on	O
19	ELS37	hpEurope	on	O

20	HUP-B14	hpEurope	off	O
21	F16	hpEastAsia	on	O
22	F30	hpEastAsia	on	O
23	F32	hpEastAsia	on	O
24	F57	hpEastAsia	on	O
25	Shi417	hpEastAsia	on	O
26	Shi169	hpEastAsia	on	O
27	Shi112	hpEastAsia	on	O
28	PeCan18	hpAfrica1	on	O
29	OK113	hpEastAsia	on	O
30	OK310	hpEastAsia	on	O
31	UM032	hpEastAsia	on	O
32	UM037	hpEurope	on	O
33	UM066	hpEastAsia	on	O
34	BM012A	hpEurope	on	O
35	oki102	hpAsia2	on	O
36	oki112	hpAsia2	on	O
37	oki128	hpEastAsia	off	X
38	oki154	hpEastAsia	on	X
39	oki422	hpAsia2	on	O
40	oki673	hpEastAsia	off	X
41	oki828	hpEastAsia	on	X
42	oki898	hpAsia2	on	O
43	J166	hpEurope	on	O
44	BM013A	hpEurope	off	X

45	Hp238	hpEastAsia	on	O
46	NY40	hpEurope	on	O
47	29CaP	hpEurope	off	X
48	7C	hpEurope	off	X
49	L7	hpEastAsia	on	O
50	DU15	hpEastAsia	on	O
51	CC33C	hpAfrica1	on	O
52	K26A1	hpAfrica2	off	X
53	PMSS1	hpEurope	on	O
54	G272	hpAsia2	on	O
55	HPJP26	hpEastAsia	on	O
56	B128_1	hpEurope	off	O
57	BCM-300	hpAfrica1	on	O
58	MKM5	hpEastAsia	on	O
59	F28	hpEastAsia	on	O
60	F38	hpEastAsia	on	O
61	F63	hpEastAsia	on	O
62	F51	hpEastAsia	on	X
63	F78	hpEastAsia	on	O
64	F13	hpEastAsia	on	O
65	F17	hpEastAsia	on	O
66	F18	hpEastAsia	on	O
67	F209	hpEastAsia	on	O
68	F20	hpEastAsia	on	O
69	F210	hpEastAsia	on	O

70	F211	hpEastAsia	on	O
71	F21	hpEastAsia	on	O
72	F23	hpEastAsia	on	O
73	F24	hpEastAsia	on	O
74	F55	hpEastAsia	on	O
75	F67	hpEastAsia	on	O
76	F70	hpEastAsia	on	O
77	F72	hpEastAsia	on	O
78	F75	hpEastAsia	on	O
79	F90	hpEastAsia	on	O
80	F94	hpEastAsia	on	O
81	MKF10	hpEastAsia	on	O
82	MKF3	hpEastAsia	on	O
83	MKF8	hpEastAsia	on	O
84	MKM1	hpEastAsia	on	O
85	MKM6	hpEastAsia	on	O
86	HLJHP256	hpEastAsia	on	O
87	HLJHP271	hpEastAsia	on	O
88	HLJHP253	hpEastAsia	on	O
89	HLJHP193	hpEastAsia	on	O
90	GAM103Bi	hpAfrica1	on	O
91	GAM105Ai	hpAfrica1	on	O
92	GAM114Ai	hpAfrica1	on	O
93	GAM201Ai	hpAfrica1	off	X
94	GAM119Bi	hpAfrica1	on	O

95	GAM121Aii	hpAfrica1	on	O
96	GAM210Bi	hpAfrica1	on	O
97	GAM245Ai	hpAfrica1	on	O
98	GAM246Ai	hpAfrica1	on	O
99	GAM249T	hpAfrica1	on	O
100	GAM254Ai	hpAfrica1	on	O
101	GAM263BFi	hpAfrica1	on	O
102	GAM264Ai	hpAfrica1	off	X
103	GAM270ASi	hpAfrica1	on	O
104	GAM42Ai	HpAfrica1	on	O
105	GAM83T	HpAfrica1	off	X
106	GAM93Bi	hpAfrica1	on	O
107	GAMchJs106B	HpAfrica1	off	X
108	HP116Bi	hpAfrica1	on	O
109	HP250AFiV	HpAfrica1	on	O
110	HP260BFii	HpAfrica1	off	X
111	GAMchJs117Ai	HpAfrica1	off	X
112	GAMchJs124i	HpAfrica1	off	X
113	Hp-H-41	HpAfrica1	on	O
114	Hp-A-8	HpAfrica1	on	O
115	Hp-H-36	HpAfrica1	on	O
116	Hp-H-30	HpAfrica1	on	O
117	Hp-H-29	HpAfrica1	on	O
118	Hp-H-42	HpAfrica1	on	O

119	Hp-H-44	HpAfrica1	on	O
120	Hp-H-45	HpEurope	on	O
121	Hp-H-43	HpEurope	off	X
122	Hp-A-6	HpAfrica1	on	O
123	Hp-A-14	HpEurope	off	X
124	Hp-A-26	HpEurope	on	X
125	Hp-A-27	HpEurope	on	O
126	Hp-H-3	HpAfrica1	on	O
127	Hp-H-9	HpEurope	on	O
128	Hp-H-6	HpAfrica1	on	O
129	Hp-H-19	HpAfrica1	on	O
130	Hp-H-4	HpAfrica1	on	O
131	Hp-H-23	HpAfrica1	on	O
132	Hp-H-21	HpAfrica1	on	X
133	Hp-P-11	HpAfrica1	on	O
134	Hp-H-34	HpAfrica1	on	X
135	Hp-P-15	HpEurope	off	X
136	Hp-P-16	HpEurope	off	X
137	Hp-P-23	HpEurope	off	X
138	Hp-P-74	HpEurope	on	X
139	Hp-H-5b	HpAfrica1	on	O
140	Hp-P-13b	HpAfrica1	on	O
141	Hp-H-11	HpEurope	on	O
142	Hp-H-18	HpAfrica1	on	O
143	NQ4161	hpEastAsia	on	X

144	NQ4110	HpEurope	off	X
145	NQ4099	HpEurope	on	O
146	NQ4076	HpEurope	on	O
147	NQ4053	HpEurope	off	X
148	NQ4044	HpEurope	on	O
149	CPY6311	hpEastAsia	on	O
150	CPY6271	hpEastAsia	on	O
151	CPY6261	hpEastAsia	on	O
152	CPY6081	hpEastAsia	on	O
153	CPY3281	hpEastAsia	on	O
154	Hp-P-8b	HpAfrica1	on	X
155	Hp-P-1	HpAfrica1	on	O
156	Hp-P-2	HpAfrica1	on	O
157	NQ4200	HpEurope	on	O
158	NQ4228	HpEurope	on	O
159	NQ4216	HpEurope	on	O
160	Hp-A-5	HpAfrica1	on	O
161	Hp-A-9	HpEurope	on	O
162	Hp-H-24	HpAfrica1	on	O
163	Hp-H-27	HpEurope	off	X
164	Hp-H-28	HpEurope	on	O
165	Hp-A-16	HpAfrica1	on	O
166	CPY1962	hpEastAsia	on	O
167	CPY1313	hpEastAsia	on	O
168	CPY1124	hpEastAsia	on	O

169	Hp-H-16	HpAfrica1	on	O
170	Hp-A-17	HpAfrica1	off	X
171	Hp-A-20	HpAfrica1	on	O
172	Hp-A-4	HpAfrica1	on	O
173	Hp-H-10	HpAfrica1	on	O
174	Hp-P-4c	HpAfrica1	on	O
175	Hp-P-26	HpAfrica1	on	O
176	Hp-P-30	HpEurope	off	X
177	Hp-P-41	HpAfrica1	on	O
178	Hp-P-62	HpAfrica1	on	O
179	Hp-P-3b	HpAfrica1	on	O
180	BCS100H1	hpEastAsia	on	X
181	NQ1671	HpEurope	on	O
182	NQ4191	HpEurope	on	O
183	R030b	HpEurope	on	O
184	R036d	HpEurope	on	O
185	R037c	HpEurope	on	X
186	R038b	HpEurope	off	X
187	R046Wa	HpEurope	on	X
188	R055a	HpEurope	on	O
189	R056a	HpEurope	on	O
190	R32b	HpEurope	on	O
191	R018c	HpEurope	on	O
192	A45	hpEastAsia	off	X
193	UMB_G1	HpEurope	off	X

194	CCHI-33	HpAfrica1	on	O
195	CPY1662	hpEastAsia	on	O
196	Hp-A-11	HpEurope	on	O
197	HPARG63	HpEurope	on	O
198	UM111	hpEastAsia	on	O
199	UM023	hpEastAsia	on	O
200	UM038	hpEastAsia	on	O
201	UM065	hpEastAsia	on	O
202	UM084	hpAsia2	on	O
203	UM114	hpAsia2	on	O
204	UM085	hpEastAsia	on	O
205	UM067	hpAsia2	on	O
206	SA302C	hpEurope	off	X
207	SA40A	HpAfrica2	off	X
208	SA156A	HpAfrica1	on	O
209	SA37A	HpEurope	off	X
210	SA171A	HpEurope	off	X
211	SA163C	HpAfrica1	on	O
212	SA175C	HpAfrica2	off	X
213	SA166A	hpAfrica2	off	X
214	SA47A	HpAfrica2	on	X
215	SA34A	HpAfrica2	on	X
216	SA222C	hpAsia2	on	X
217	SA210C	HpAfrica1	on	O
218	SA253C	hpAfrica2	off	X

219	SA31C	HpAfrica1	on	O
220	SA172A	HpEurope	off	X
221	SA233C	hpAfrica2	on	X
222	SA171C	hpEurope	off	X
223	SA300C	HpAfrica1	on	O
224	SA221C	hpEurope	off	X
225	SA175A	hpAfrica2	on	X
226	SA165A	HpEurope	off	X
227	SA160C	hpAfrica2	off	X
228	SA45C	hpAfrica1	on	O
229	SA194A	HpAfrica2	on	X
230	SA251A	HpAfrica2	off	X
231	SA146A	HpAfrica1	on	O
232	HLJ039	hpEastAsia	on	O
233	wls-5-13	hpEastAsia	off	O
234	wls-5-18	hpEastAsia	on	O
235	SA35C	hpAfrica1	on	O
236	Manado-1	hpEastAsia	on	O
237	H3018	HpEurope	on	O
238	UM087	hpAsia2	on	O
239	UM045	HpEurope	on	O
240	UM139	hpAsia2	on	O
241	UM119	hpEastAsia	on	O
242	UM158	hpAsia2	on	O
243	UM131	hpAsia2	on	O

244	UM122	hpAsia2	on	O
245	UM209	hpAsia2	off	X
246	UM228	hpAsia2	on	O
247	UM246	hpEastAsia	on	O
248	UM291	hpEastAsia	on	O
249	UM211	hpAsia2	on	O
250	UM352	hpEastAsia	on	O
251	UM370	hpEastAsia	on	O
252	UM411	hpAsia2	on	O
253	UM196	hpEastAsia	on	O
254	UM152	hpAsia2	on	O
255	UM165	hpAsia2	on	O
256	UM300	hpAsia2	on	O
257	UM408	hpAsia2	on	O
258	UM163	hpEastAsia	on	O
259	UM171S	hpEastAsia	on	X
260	UM233S	hpEastAsia	on	O
261	UM400AS	HpAsia2	on	O
262	UM303S	hpEastAsia	on	O
263	UM443S	hpEastAsia	on	O
264	241	hpEastAsia	on	O
265	428	hpEastAsia	on	O
266	178	hpEastAsia	on	O
267	132	hpAsia2	on	O
268	H9	hpEastAsia	on	O

269	S380A	hpEastAsia	on	O
270	H30	hpEastAsia	on	O
271	1177	hpEastAsia	on	O
272	S468A	hpEastAsia	on	O
273	C664	hpEastAsia	on	O
274	C333	hpEastAsia	on	O
275	26083	HpEurope	on	O
276	22389	HpEurope	off	X
277	22352	HpEurope	on	O
278	22347	HpEurope	on	O
279	22278	HpEurope	on	O
280	2029	HpEurope	on	O
281	3046	HpEurope	on	O
282	3076	HpEurope	off	X
283	22312	HpEurope	on	O
284	3053	HpEurope	off	X
285	A037	HpEurope	on	O
286	26084	hpEastAsia	on	O
287	26024	HpEurope	on	O
288	24013	HpEurope	off	X
289	22402	HpEurope	on	O
290	22386	HpAfrica1	off	X
291	22366	HpEurope	on	O
292	22367	HpEurope	on	O
293	22362	hpEastAsia	on	O

294	22331	HpEurope	on	O
295	22151	HpEurope	on	O
296	22095	HpAfrica1	on	O
297	22023	HpEurope	off	X
298	3136	HpEurope	off	X
299	2015	HpEurope	on	X
300	1102	HpEurope	on	O
301	3125	HpEurope	on	O
302	22003	HpEurope	on	O
303	3118	HpEurope	on	O
304	3120	HpEurope	on	O
305	KH43	hpAsia2	on	O
306	KH44	hpEurope	on	O
307	KH45	hpAsia2	on	O
308	KH41	hpEurope	on	O
309	KH39	hpAsia2	on	O
310	KH28	hpEurope	on	O
311	KH25	hpAsia2	on	O
312	KH36	hpEurope	on	O
313	KH27	hpEurope	on	O
314	KH20	hpEurope	on	O
315	KH21	hpEurope	on	O
316	KH16	hpEurope	off	X
317	KH11	hpAsia2	on	O
318	KH10	hpAsia2	on	O

319	KH4	hpAsia2	on	O
320	KH3	hpEurope	on	X
321	KH40	hpEurope	off	X
322	KH37	hpAsia2	on	O
323	KH35	hpEurope	on	O
324	KH34	hpAsia2	on	O
325	KH33	hpEurope	on	O
326	KH31	hpAsia2	on	O
327	KH32	hpAsia2	on	O
328	KH30	hpEurope	on	O
329	KH22	hpAsia2	on	O
330	KH19	hpEurope	off	X
331	KH17	hpAsia2	on	O
332	KH15	hpAsia2	on	O
333	KH14	hpAsia2	on	O
334	KH12	hpEurope	on	X
335	KH9	hpEurope	on	O
336	KH8	hpEurope	off	X
337	KH7	hpAsia2	on	O
338	KH6	hpEurope	off	X
339	KH2	hpEurope	off	X
340	KH1	hpEurope	on	O
341	Feb-55	HpEurope	on	O
342	456	hpEurope	on	O
343	50	HpEurope	on	O

344	30950	hpEurope	on	O
345	565-99	hpEurope	off	X
346	638	hpEurope	on	X
347	36166	hpEurope	off	X
348	3699	HpEurope	on	O
349	GC54-HL	hpEurope	on	O
350	B30	HpEurope	off	X
351	B40	HpEurope	off	X
352	B25	hpEurope	off	X
353	B508A-T4	HpEurope	off	X
354	NCTC13207	HpEurope	off	X
355	NCTC13094	HpEurope	on	O
356	YN1-91	hpEastAsia	on	O
357	YN4-84	hpEastAsia	on	O
358	207-99	hpEurope	off	X
359	655-99	hpEurope	off	X
360	1786-05	hpEurope	off	X
361	UM229S	hpEastAsia	on	O
362	UM137S	hpEastAsia	on	O
363	PZ5019_3A3	HpAfrica1	on	O
364	SV397_2	HpEurope	on	O
365	SV449_1	HpEurope	on	O
366	SV380_1	HpAfrica1	on	O
367	SV355_2	HpEurope	on	O
368	PZ5016_3A3	HpAfrica1	on	O

369	XZ274	hpEastAsia	on	O
370	SouthAfrica20	hpAfrica2	off	X
371	Hp-P-25	HpAfrica1	on	O
372	CG-IMSS-2012	HpEurope	on	O
373	v225d	hpEastAsia	on	O
374	Gambia94-24	hpAfrica1	on	O
375	Lithuania75	hpEurope	on	O
376	SouthAfrica7	hpAfrica2	off	X
377	Santal49	hpAsia2	on	O
378	Aklavik117	hpEastAsia	on	X
379	Aklavik86	hpEastAsia	on	X
380	NCTC-11637	HpEurope	on	O
381	BJWJ10	hpEastAsia	on	O
382	BJWJ17	hpEastAsia	on	O
383	BJXHBAO	hpEastAsia	on	O
384	BJXHWU	hpEastAsia	on	O
385	DU1829	hpEastAsia	on	O
386	DU1838	hpEastAsia	on	O
387	DU1867	hpEastAsia	on	O
388	DU1888	hpEastAsia	on	O
389	F216	hpEastAsia	on	O
390	F219	hpEastAsia	on	O
391	F221	hpEastAsia	on	O
392	F225	hpEastAsia	on	O
393	F226	hpEastAsia	on	O

394	F227	hpEastAsia	on	O
395	F228	hpEastAsia	on	O
396	F232	hpEastAsia	on	O
397	F233	hpEastAsia	on	O
398	F234	hpEastAsia	on	O
399	F235	hpEastAsia	on	O
400	F93	hpEastAsia	on	O
401	GU1861	hpEastAsia	on	O
402	GU1862	hpEastAsia	on	O
403	GU1871	hpEastAsia	on	O
404	GU1882	hpEastAsia	on	O
405	GU1883	hpEastAsia	on	O
406	HLJ011	hpEastAsia	on	O
407	HLJ014	hpEastAsia	on	O
408	HLJ022	hpEastAsia	on	O
409	HLJ030	hpEastAsia	on	O
410	HLJ15A	hpEastAsia	on	O
411	HLJ201	hpEastAsia	on	O
412	HLJ215	hpEastAsia	on	O
413	HLJ220	hpEastAsia	on	O
414	HLJ226	hpEastAsia	on	O
415	HLJ229	hpEastAsia	on	O
416	HLJ259	hpEastAsia	on	O
417	HLJ262	hpEastAsia	on	O
418	HLJ270	hpEastAsia	on	X

419	HP32	hpEastAsia	on	O
420	HP37	hpEastAsia	on	O
421	HP76	hpEastAsia	on	O
422	HP85	hpEastAsia	on	O
423	HP97	hpEastAsia	on	O
424	HZT090500664	hpEastAsia	on	O
425	HZT090501406	hpEastAsia	on	O
426	HZT090600333	hpEastAsia	on	O
427	HZT090600337	hpEastAsia	on	O
428	HZT342	hpEastAsia	on	O
429	K111	hpEastAsia	on	O
430	K115	hpEastAsia	on	O
431	K131	hpEastAsia	on	O
432	K154	hpEastAsia	on	O
433	K16	hpEastAsia	on	O
434	K172	hpEastAsia	on	O
435	K262	hpAfrica1	on	O
436	K27	hpEastAsia	on	O
437	K29	hpEastAsia	on	X
438	K46	hpEastAsia	on	O
439	K47	hpEastAsia	on	O
440	K60	hpEastAsia	on	O
441	K64	hpEastAsia	on	O
442	K74	hpEastAsia	on	O

443	K80	hpEastAsia	on	O
444	K93	hpEastAsia	on	O
445	M1303	hpEastAsia	on	O
446	M3021	hpEastAsia	on	O
447	M3022	hpEastAsia	on	O
448	M3029	hpEastAsia	on	O
449	M374	hpEastAsia	on	O
450	M401	hpEastAsia	on	O
451	M523	hpEastAsia	on	X
452	M588	hpEastAsia	on	O
453	M628	hpEastAsia	on	O
454	M670	hpEastAsia	on	X
455	M712	hpEastAsia	on	O
456	M736	hpEastAsia	on	O
457	MALT1706	hpEastAsia	on	O
458	MALT1726	hpEastAsia	on	O
459	MALT1785	hpEastAsia	on	X
460	oi10-358	hpEastAsia	on	O
461	OK103	hpEastAsia	on	O
462	OK130	hpEastAsia	on	O
463	OK134	hpEastAsia	on	O
464	OK136	hpEastAsia	off	X
465	OK144	hpAsia2	on	O
466	OK179	hpAsia2	on	O
467	OK180	hpAsia2	on	O

468	OK185	hpAsia2	off	O
469	OK188	hpEastAsia	on	O
470	OK195	hpEastAsia	on	O
471	OK206	hpEastAsia	on	O
472	OK212	hpEastAsia	on	O
473	OK301	hpEastAsia	on	O
474	OK302	hpEastAsia	on	O
475	OK303	hpEastAsia	on	O
476	OK304	hpEastAsia	on	O
477	OK305	hpEastAsia	on	O
478	OK306	hpEastAsia	on	O
479	OK307	hpEastAsia	on	O
480	OK308	hpEastAsia	on	O
481	OK311	hpEastAsia	on	O
482	OK312	hpEastAsia	on	O
483	OK313	hpEastAsia	on	O
484	OK314	hpEastAsia	on	O
485	OK316	hpEastAsia	on	O
486	OK317	hpEurope	on	O
487	OK318	hpEastAsia	on	O
488	OK99	hpEastAsia	on	O
489	P164	hpEastAsia	on	O
490	P200	hpEastAsia	on	O
491	P214	hpEastAsia	on	O
492	PIK29	hpEastAsia	on	O

493	PIK93	hpEastAsia	on	O
494	XA1	hpEurope	on	O
495	XA6	hpEastAsia	on	O
496	YN1-100	hpEastAsia	on	O
497	YN1-101	hpEastAsia	on	O
498	YN1-136	hpEastAsia	on	O
499	YN1-139	hpEastAsia	on	O
500	YN1-92	hpEastAsia	on	O
501	YN1-99	hpEastAsia	on	O
502	YN3-2	hpEastAsia	on	O
503	YN3-21	hpEastAsia	on	O
504	YN3-34	hpEastAsia	on	O
505	YN3-36	hpEastAsia	on	O
506	YN3-77	hpEastAsia	on	O
507	YN4-105	hpEastAsia	on	O
508	YN4-109	hpEastAsia	on	O
509	YN4-11	hpEastAsia	on	O
510	YN4-125	hpEastAsia	on	O
511	YN4-134	hpEastAsia	on	O
512	YN4-58	hpEastAsia	on	X
513	YN4-62	hpEastAsia	on	O
514	YN4-83	hpEastAsia	on	O
515	YTC-5	hpEastAsia	on	O
516	ZHOU183	hpEastAsia	on	O
517	AH871	hpEurope	off	X

518	AH876	hpAfrica1	on	O
519	AH881	hpEurope	on	O
520	AH885	hpAfrica1	on	O
521	AH886	hpEurope	off	X
522	AH887	hpAfrica1	on	O
523	AH888	hpAfrica1	off	X
524	AH889	hpEurope	on	X
525	AH893	hpAfrica1	on	O
526	AH895	hpEurope	off	X
527	AH898	hpAfrica1	on	X
528	AH904	hpEurope	off	X
529	3800	hpEurope	off	X
530	BON254	hpEurope	off	X
531	Mandalay03	hpAsia2	on	O
532	Mandalay13	hpEurope	on	O
533	Mandalay30	hpAsia2	on	O
534	Mandalay38	hpEastAsia	on	O
535	Mandalay46	hpEurope	on	O
536	Mandalay60	hpAsia2	on	O
537	Myanmar51	hpAsia2	on	O
538	Myanmar52	hpAsia2	on	O
539	Myanmar66	hpAsia2	on	O
540	NP04	hpAsia2	on	O
541	NP05	hpAsia2	on	O
542	NP05-105	hpAsia2	on	O

543	NP05-107	hpAsia2	on	O
544	NP05-112	hpEurope	on	O
545	NP05-121	hpAsia2	on	O
546	NP05-124	hpAsia2	on	O
547	NP05-227	hpEurope	off	O
548	NP05-250	hpAsia2	on	O
549	NP05-261	hpEurope	on	O
550	NP05-266	hpEurope	on	O
551	NP05-272	hpEurope	on	O
552	NP05-278	hpAsia2	on	O
553	NP05-282	hpEurope	on	O
554	SSR43	hpEurope	off	X
555	Yangon132	hpEurope	on	O
556	Yangon142	hpAsia2	on	O
557	Yangon159	hpEastAsia	on	O
558	Yangon173	hpEastAsia	on	O
559	Yangon179	hpAsia2	on	O
560	Yangon188	hpAsia2	on	O
561	Yangon202	hpAsia2	on	O
562	Yangon222	hpEastAsia	on	O
563	Yangon233	hpAsia2	on	O
564	Yangon244	hpEastAsia	on	O
565	444	hpEurope	on	O
566	21580	hpEurope	on	O
567	Nic01-A	hpEurope	on	O

568	Nic03-C	hpAfrica1	on	O
569	Nic04-A	hpEurope	on	O
570	Nic05-C	hpEurope	on	O
571	Nic06-A	hpEurope	on	O
572	Nic07-C	hpAfrica1	on	O
573	Nic08-C	hpEurope	off	X
574	Nic09-C	hpEurope	off	X
575	Nic10-C	hpEurope	on	O
576	Nic11-A	hpAfrica1	on	O
577	Nic12-C	hpEurope	on	O
578	Nic13-C	hpAfrica1	on	O
579	Nic14-A	hpAfrica1	on	O
580	Nic14-C	hpEurope	on	X
581	Nic15-C	hpEastAsia	on	X
582	Nic16-C	hpAfrica1	on	O
583	Nic17-A	hpEurope	on	O
584	Nic18-C	hpAfrica1	on	O
585	Nic19-C	hpAfrica1	on	O
586	Nic20-A	hpEurope	on	X
587	Nic20-C	hpAfrica1	on	O
588	Nic23-A	hpEurope	on	O
589	Nic25-A	hpAfrica1	on	O
590	Nic26-A	hpAfrica1	on	O
591	Nic28-A	hpAfrica1	on	O
592	Nic29-A	hpAfrica1	on	O

593	Nic30-A	hpEurope	on	O
594	Nic32-A	hpAfrica1	off	X
595	AH868	hpAfrica1	on	O
596	AH872	hpEurope	off	X
597	AH883	hpAfrica1	on	O
598	AH890	hpEurope	off	X
599	AH892	hpEurope	off	X
600	AH894	hpEurope	on	O
601	AH896	hpEurope	off	X
602	AH900	hpAfrica1	on	O
603	AH901	hpEurope	off	X
604	AH908	hpEurope	off	X
605	B125A-2	hpEurope	on	O
606	B130A	hpEurope	on	O
607	B136A	hpEurope	on	O
608	B140	hpEurope	on	O
609	B147-2	hpEurope	on	O
610	J182	hpEurope	on	O

VI. REFERENCES

1. Chen YC, Malfertheiner P, Yu HT, et al. Global Prevalence of *Helicobacter pylori* Infection and Incidence of Gastric Cancer Between 1980 and 2022. *Gastroenterology*. 2024;166(4):605-619. doi:10.1053/j.gastro.2023.12.022

2. Marshall BJ, Warren JR. Unidentified curved bacilli in the stomach of patients with gastritis and peptic ulceration. *The Lancet*. 1984;323(8390). doi:10.1016/S0140-6736(84)91816-6

3. Warren JR, Marshall B. Unidentified curved bacilli on gastric epithelium in active chronic gastritis. *Lancet*. 1983;1(8336):1273-1275.

4. Dubois A. Spiral Bacteria in the Human Stomach: The Gastric *Helicobacters*. *Emerg Infect Dis*. 1995;1(3):79-85. doi:10.3201/eid0103.950302

5. Polk DB, Peek RM. *Helicobacter pylori*: gastric cancer and beyond. *Nat Rev Cancer*. 2010;10(6):403-414. doi:10.1038/nrc2857

6. Taylor DN, Blaser MJ. The Epidemiology of *Helicobacter pylori* Infection. *Epidemiologic Reviews*. 1991;13(1):42-59. doi:10.1093/oxfordjournals.epirev.a036078

7. Correa P. Human gastric carcinogenesis: a multistep and multifactorial process--First American Cancer Society Award Lecture on Cancer Epidemiology and Prevention. *Cancer Res*. 1992;52(24):6735-6740.

8. Moss SF. The Clinical Evidence Linking *Helicobacter pylori* to Gastric Cancer. *CMGH*. 2017;3(2). doi:10.1016/j.jcmgh.2016.12.001

9. Ferlay J, Soerjomataram I, Dikshit R, et al. Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. *International Journal of Cancer*. 2015;136(5). doi:10.1002/ijc.29210

10. Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians*. 2021;71(3). doi:10.3322/caac.21660

11. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. *CA A Cancer J Clinicians*. 2015;65(2):87-108. doi:10.3322/caac.21262

12. Parkin DM. The global health burden of infection-associated cancers in the year 2002. *Intl Journal of Cancer*. 2006;118(12):3030-3044. doi:10.1002/ijc.21731

13. Forman D, Coleman M, De Backer A, et al. Epidemiology of, and risk factors for, *Helicobacter pylori* infection among 3194 asymptomatic subjects in 17 populations. *Gut*. 1993;34(12). doi:10.1136/gut.34.12.1672

14. Forman D, Coleman M, De Backer G, et al. An international association between *Helicobacter pylori* infection and gastric cancer. *The Lancet*. 1993;341(8857). doi:10.1016/0140-6736(93)90938-D

15. Blaser MJ. *Helicobacter pylori* and gastric diseases. *BMJ*. 1998;316(7143). doi:10.1136/bmj.316.7143.1507

16. Schistosomes, liver flukes and *Helicobacter pylori*. IARC Working Group on the Evaluation of Carcinogenic Risks to Humans. Lyon, 7-14 June 1994. In: *IARC*. Vol 61. ; 1994.

17. Olbermann P, Josenhans C, Moodley Y, et al. A Global Overview of the Genetic and Functional Diversity in the *Helicobacter pylori cag* Pathogenicity Island. Malik HS, ed. *PLoS Genet*. 2010;6(8):e1001069. doi:10.1371/journal.pgen.1001069

18. Thorell K, Muñoz-Ramírez ZY, Wang D, et al. The *Helicobacter pylori* Genome Project: insights into H. pylori population structure from analysis of a worldwide collection of complete genomes. *Nat Commun*. 2023;14(1):8184. doi:10.1038/s41467-023-43562-y

19. Kawai M, Furuta Y, Yahara K, et al. Evolution in an oncogenic bacterial species with extreme genome plasticity: *Helicobacter pylori* East Asian genomes. *BMC Microbiology*. 2011;11. doi:10.1186/1471-2180-11-104

20. Su H, Tissera K, Jang S, et al. Evolutionary mechanism leading to the multi-*cagA* genotype in *Helicobacter pylori*. *Sci Rep*. 2019;9(1):11203. doi:10.1038/s41598-019-47240-2

21. Jang S, Su H, Blum FC, et al. Dynamic expansion and contraction of *cagA* copy number in *Helicobacter pylori* impact development of gastric disease. *mBio*. 2017;8(1). doi:10.1128/mbio.01779-16

22. Yamaoka Y. *Helicobacter pylori* typing as a tool for tracking human migration. *Clinical Microbiology and Infection*. 2009;15(9). doi:10.1111/j.1469-0691.2009.02967.x

23. Muñoz-Ramirez ZY, Pascoe B, Mendez-Tenorio A, et al. A 500-year tale of co-evolution, adaptation, and virulence: *Helicobacter pylori* in the Americas. *The ISME Journal*. 2021;15(1):78-92. doi:10.1038/s41396-020-00758-0

24. Falush D, Wirth T, Linz B, et al. Traces of Human Migrations in *Helicobacter pylori* Populations. *Science*. 2003;299(5612):1582-1585. doi:10.1126/science.1080857

25. Falush D, Stephens M, Pritchard JK. Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics*. 2003;164(4). doi:10.1093/genetics/164.4.1567

26. Maiden MCJ, Bygraves JA, Feil E, et al. Multilocus sequence typing: A portable approach to the identification of clones within populations of pathogenic microorganisms. *Proceedings of the National Academy of Sciences of the United States of America*. 1998;95(6). doi:10.1073/pnas.95.6.3140

27. Atherton JC, Blaser MJ. Coadaptation of *Helicobacter pylori* and humans: Ancient history, modern implications. *Journal of Clinical Investigation*. 2009;119(9). doi:10.1172/JCI38605

28. Kim Y, Park J, Nam BH, Ki M. Stomach cancer incidence rates among Americans, Asian Americans and Native Asians from 1988 to 2011. *Epidemiol Health*. 2015;37:e2015006. doi:10.4178/epih/e2015006

29. Miwa H, Go MF, Sato N. H. pylori and gastric cancer: the Asian enigma. *Am J Gastroenterology*. 2002;97(5):1106-1112. doi:10.1111/j.1572-0241.2002.05663.x

30. Kodaman N, Pazos A, Schneider BG, et al. Human and *Helicobacter pylori* coevolution shapes the risk of gastric disease. *Proc Natl Acad Sci USA*. 2014;111(4):1455-1460. doi:10.1073/pnas.1318093111

31. Yamaoka Y, Kato M, Asaka M. Geographic Differences in Gastric Cancer Incidence Can be Explained by Differences between *Helicobacter pylori* Strains. *Intern Med*. 2008;47(12):1077-1083. doi:10.2169/internalmedicine.47.0975

32. Xiang Z, Censini S, Bayeli PF, et al. Analysis of expression of CagA and VacA virulence factors in 43 strains of *Helicobacter pylori* reveals that clinical isolates can be divided into two major types and that CagA is not necessary for expression of the vacuolating cytotoxin. *Infection and Immunity*. 1995;63(1). doi:10.1128/iai.63.1.94-98.1995

33. Ando T, Peek RM, Pride D, et al. Polymorphisms of *Helicobacter pylori* HP0638 reflect geographic origin and correlate with *cagA* status. *Journal of Clinical Microbiology*. 2002;40(1). doi:10.1128/JCM.40.1.239-246.2002

34. Jones KR, Joo YM, Jang S, et al. Polymorphism in the *cagA* EPIYA motif impacts development of gastric cancer. *Journal of Clinical Microbiology*. 2009;47(4). doi:10.1128/JCM.02330-08

35. Malfertheiner P, Camargo MC, El-Omar E, et al. *Helicobacter pylori* infection. *Nat Rev Dis Primers*. 2023;9(1):19. doi:10.1038/s41572-023-00431-8

36. Backert S, Tegtmeyer N, Fischer W. Composition, structure and function of the *Helicobacter pylori cag* pathogenicity island encoded type IV secretion system. *Future Microbiology*. 2015;10(6). doi:10.2217/fmb.15.32

37. Censini S, Lange C, Xiang Z, et al. *cag*, a pathogenicity island of *Helicobacter pylori*, encodes type I-specific and disease-associated virulence factors. *Proceedings of the National Academy of Sciences of the United States of America*. 1996;93(25). doi:10.1073/pnas.93.25.14648

38. Gordon D. CagA protein from *Helicobacter pylori* is a Trojan horse to epithelial cells. *Gastroenterology*. 2000;118(5):817. doi:10.1016/S0016-5085(00)70161-2

39. Blaser MJ, Perez-Perez GI, Kleanthous H, et al. Infection with *Helicobacter pylori* strains possessing *cagA* is associated with an increased risk of developing adenocarcinoma of the stomach. *Cancer Res*. 1995;55(10):2111-2115.

40. Enroth H, Kraaz W, Engstrand L, Nyrén O, Rohan T. *Helicobacter pylori* strain types and risk of gastric cancer: a case-control study. *Cancer Epidemiol Biomarkers Prev*. 2000;9(9):981-985.

41. Tran SC, Bryant KN, Cover TL. The *Helicobacter pylori cag* pathogenicity island as a determinant of gastric cancer risk. *Gut Microbes*. 2024;16(1):2314201. doi:10.1080/19490976.2024.2314201

42. Rosenberg NA, Pritchard JK, Weber JL, et al. Genetic Structure of Human Populations. *Science*. 2002;298(5602):2381-2385. doi:10.1126/science.1078311

43. Backert S, Ziska E, Brinkmann V, et al. Translocation of the *Helicobacter pylori* CagA protein in gastric epithelial cells by a type IV secretion apparatus. *Cellular Microbiology*. 2000;2(2). doi:10.1046/j.1462-5822.2000.00043.x

44. Kutter S, Buhrdorf R, Haas J, Schneider-Brachert W, Haas R, Fischer W. Protein subassemblies of the *Helicobacter pylori cag* type IV secretion system

revealed by localization and interaction studies. *Journal of Bacteriology*. 2008;190(6). doi:10.1128/JB.01341-07

45. Johnson EM, Gaddy JA, Voss BJ, Hennig EE, Covera TL. Genes required for assembly of pili associated with the *Helicobacter pylori cag* type IV secretion system. *Infection and Immunity*. 2014;82(8). doi:10.1128/IAI.01640-14

46. Frick-Cheng AE, Pyburn TM, Voss BJ, McDonald WH, Ohi MD, Cover TL. Molecular and structural analysis of the *Helicobacter pylori cag* type IV secretion system core complex. *mBio*. 2016;7(1). doi:10.1128/mBio.02001-15

47. Chang YW, Shaffer CL, Rettberg LA, Ghosal D, Jensen GJ. In Vivo Structures of the *Helicobacter pylori cag* Type IV Secretion System. *Cell Reports*. 2018;23(3). doi:10.1016/j.celrep.2018.03.085

48. Backert S, Moese S, Selbach M, Brinkmann V, Meyer TF. Phosphorylation of tyrosine 972 of the *Helicobacter pylori* CagA protein is essential for induction of a scattering phenotype in gastric epithelial cells. *Molecular Microbiology*. 2001;42(3). doi:10.1046/j.1365-2958.2001.02649.x

49. Ren S, Higashi H, Lu H, Azuma T, Hatakeyama M. Structural basis and functional consequence of *Helicobacter pylori cagA* multimerization in cells. *Journal of Biological Chemistry*. 2006;281(43). doi:10.1074/jbc.M606172200

50. Higashi H, Tsutsumi R, Fujita A, et al. Biological activity of the *Helicobacter pylori* virulence factor CagA is determined by variation in the tyrosine phosphorylation sites. *Proceedings of the National Academy of Sciences of the United States of America*. 2002;99(22). doi:10.1073/pnas.222375399

51. Stein M, Rappuoli R, Covacci A. Tyrosine phosphorylation of the *Helicobacter pylori* CagA antigen after cag-driven host cell translocation. *Proceedings of the National Academy of Sciences of the United States of America*. 2000;97(3). doi:10.1073/pnas.97.3.1263

52. Segal ED, Cha J, Lo J, Falkow S, Tompkins LS. Altered states: Involvement of phosphorylated CagA in the induction of host cellular growth changes by *Helicobacter pylori*. *Proceedings of the National Academy of Sciences of the United States of America*. 1999;96(25). doi:10.1073/pnas.96.25.14559

53. Guruge JL, Falk PG, Lorenz RG, et al. Epithelial attachment alters the outcome of *Helicobacter pylori* infection. *Proceedings of the National Academy of Sciences of the United States of America*. 1998;95(7). doi:10.1073/pnas.95.7.3925

54. Keates S, Hitti YS, Upton M, Kelly CP. *Helicobacter pylori* infection activates NF-κB in gastric epithelial cells. *Gastroenterology*. 1997;113(4). doi:10.1053/gast.1997.v113.pm9322504

55. Sokolova O, Borgmann M, Rieke C, Schweitzer K, Rothkötter HJ, Naumann M. *Helicobacter pylori* induces type 4 secretion system-dependent, but CagA-independent activation of IκBs and NF-κB/RelA at early time points. *International Journal of Medical Microbiology*. 2013;303(8). doi:10.1016/j.ijmm.2013.07.008

56. Seo JH, Lim JW, Kim H, Kim KH. *Helicobacter pylori* in a Korean isolate activates mitogen-activated protein kinases, AP-1, and NF-κB and induces chemokine expression in gastric epithelial AGS cells. *Laboratory Investigation*. 2004;84(1):49-62. doi:10.1038/labinvest.3700010

57. Crabtree JE, Covacci A, Farmery SM, et al. *Helicobacter pylori* induced interleukin-8 expression in gastric epithelial cells is associated with CagA positive phenotype. *Journal of Clinical Pathology*. 1995;48(1). doi:10.1136/jcp.48.1.41

58. Lee KE, Khoi PN, Xia Y, et al. *Helicobacter pylori* and interleukin-8 in gastric cancer. *World Journal of Gastroenterology*. 2013;19(45). doi:10.3748/wjg.v19.i45.8192

59. Odenbreit S, Kavermann H, Püls J, Haas R. CagA tyrosine phosphorylation and interleukin-8 induction by *Helicobacter pylori* are independent from AlpAB, HopZ and Bab group outer membrane proteins. *International Journal of Medical Microbiology*. 2002;292(3-4). doi:10.1078/1438-4221-00205

60. Yamaoka Y, Kudo T, Lu H, Casola A, Brasier AR, Graham DY. Role of Interferon-Stimulated Responsive Element-like Element in Interleukin-8 Promoter in *Helicobacter pylori* Infection. *Gastroenterology*. 2004;126(4). doi:10.1053/j.gastro.2003.12.048

61. Gunawardhana N, Jang S, Choi YH, et al. *Helicobacter pylori*-Induced HB-EGF Upregulates Gastrin Expression via the EGF Receptor, C-Raf, Mek1, and Erk2 in the MAPK Pathway. *Front Cell Infect Microbiol*. 2018;7:541. doi:10.3389/fcimb.2017.00541

62. Alm RA, Bina J, Andrews BM, Doig P, Hancock REW, Trust TJ. Comparative genomics of *Helicobacter pylori*: Analysis of the outer membrane protein families. *Infection and Immunity*. 2000;68(7). doi:10.1128/IAI.68.7.4155-4168.2000

63. Cover TL. Role of *Helicobacter pylori* outer membrane proteins in gastroduodenal disease. *Journal of Infectious Diseases*. 2006;194(10). doi:10.1086/508432

64. Matsuo Y, Kido Y, Yamaoka Y. *Helicobacter pylori* outer membrane protein-related pathogenesis. *Toxins*. 2017;9(3). doi:10.3390/toxins9030101

65. Kim A, Servetas SL, Kang J, et al. *Helicobacter pylori* bab Paralog Distribution and Association with *cagA*, *vacA*, and *homA*/B Genotypes in American and South Korean Clinical Isolates. Hozbor DF, ed. *PLoS ONE*. 2015;10(8):e0137078. doi:10.1371/journal.pone.0137078

66. Bartpho TS, Wattanawongdon W, Tongtawee T, Paoin C, Kangwantas K, Dechsukhum C. Precancerous Gastric Lesions with *Helicobacter pylori vacA* $^{+}$ / *babA* 2 $^{+}$ / *oipA* $^{+}$ Genotype Increase the Risk of Gastric Cancer. *BioMed Research International*. 2020;2020:1-8. doi:10.1155/2020/7243029

67. Borén T, Falk P, Roth KA, Larson G, Normark S. Attachment of *Helicobacter pylori* to Human Gastric Epithelium Mediated by Blood Group Antigens. *Science*. 1993;262(5141):1892-1895. doi:10.1126/science.8018146

68. Yamaoka Y, Ojo O, Fujimoto S, et al. *Helicobacter pylori* outer membrane proteins and gastroduodenal disease. *Gut*. 2006;55(6). doi:10.1136/gut.2005.083014

69. Yamaoka Y, Kita M, Kodama T, et al. *Helicobacter pylori* infection in mice: Role of outer membrane proteins in colonization and inflammation. *Gastroenterology*. 2002;123(6). doi:10.1053/gast.2002.37074

70. Sallas ML, dos Santos MP, Orcini WA, et al. Status (on/off) of *oipA* gene: their associations with gastritis and gastric cancer and geographic origins. *Archives of Microbiology*. 2019;201(1). doi:10.1007/s00203-018-1580-5

71. Oktem-Okullu S, Karaman T, Akcelik-Deveci S, et al. Effect of the switch status of *Helicobacter pylori* outer inflammatory protein A on gastric diseases. *AMB Expr*. 2023;13(1):109. doi:10.1186/s13568-023-01621-z

72. Yamaoka Y, Kwon DH, Graham DY. A Mr 34,000 proinflammatory outer membrane protein (*oipA*) of *Helicobacter pylori*. *Proceedings of the National Academy of Sciences of the United States of America*. 2000;97(13). doi:10.1073/pnas.130079797

73. Dossumbekova A, Prinz C, Mages J, et al. *Helicobacter pylori* HopH (OipA) and bacterial pathogenicity: Genetic and functional genomic analysis of hopH gene polymorphisms. *Journal of Infectious Diseases*. 2006;194(10). doi:10.1086/508426

74. Kudo T, Nurgalieva ZZ, Conner ME, et al. Correlation between *Helicobacter pylori* OipA Protein Expression and *oipA* Gene Switch Status. *Journal of Clinical Microbiology*. 2004;42(5). doi:10.1128/JCM.42.5.2279-2281.2004

75. Kim A, Lai J, Merrell DS, Kim JH, Su H, Cha JH. Geographic diversity in *Helicobacter pylori oipA* genotype between Korean and United States isolates. *J Microbiol*. 2021;59(12):1125-1132. doi:10.1007/s12275-021-1450-8

76. Souod N, Sarshar M, Dabiri H, et al. The study of the *oipA* and dupA genes in *Helicobacter pylori* strains and their relationship with different gastroduodenal diseases. *Gastroenterology and Hepatology from Bed to Bench*. 2015;8.

77. Liu J, He C, Chen M, Wang Z, Xing C, Yuan Y. Association of presence/absence and on/off patterns of *Helicobacter pylori oipA* gene with peptic ulcer disease and gastric cancer risks: A meta-analysis. *BMC Infectious Diseases*. 2013;13(1). doi:10.1186/1471-2334-13-555

78. Yamaoka Y, Kikuchi S, ElZimaity HMT, Gutierrez O, Osato MS, Graham DY. Importance of *Helicobacter pylori oipA* in clinical presentation, gastric inflammation, and mucosal interleukin 8 production. *Gastroenterology*. 2002;123(2). doi:10.1053/gast.2002.34781

79. Fock KM, Ang TL. Epidemiology of *Helicobacter pylori* infection and gastric cancer in Asia. *Journal of Gastroenterology and Hepatology (Australia)*. 2010;25(3). doi:10.1111/j.1440-1746.2009.06188.x

80. Farzi N, Yadegar A, Aghdaei HA, Yamaoka Y, Zali MR. Genetic diversity and functional analysis of *oipA* gene in association with other virulence factors among *Helicobacter pylori* isolates from Iranian patients with different gastric diseases. *Infection, Genetics and Evolution*. 2018;60. doi:10.1016/j.meegid.2018.02.017

81. Carpenter BM, McDaniel TK, Whitmire JM, et al. Expanding the *Helicobacter pylori* Genetic Toolbox: Modification of an Endogenous Plasmid for Use as a Transcriptional Reporter and Complementation Vector. *Appl Environ Microbiol*. 2007;73(23):7506-7514. doi:10.1128/AEM.01084-07

82. Haft DH, DiCuccio M, Badretdin A, et al. RefSeq: an update on prokaryotic genome annotation and curation. *Nucleic Acids Research*. 2018;46(D1):D851-D860. doi:10.1093/nar/gkx1068

83. Yanovich OO, Titov LP, Doroshko MV. Genetic Diversity and Functional Status of the *oipA* Gene of *Helicobacter pylori*. Its Association with Genes of the Pathogenicity Island in Patients with Gastroduodenal Diseases in the Republic of Belarus. *Mol Genet Microbiol Virol*. 2022;37(4):173-178. doi:10.3103/S0891416822040097

84. Suerbaum S, Ailloud F. Genome and population dynamics during chronic infection with *Helicobacter pylori*. *Current Opinion in Immunology*. 2023;82:102304. doi:10.1016/j.coi.2023.102304

VII. ABSTRACT (KOREAN)

H. pylori 임상 균주에서 독성 표현형 및 *cag*PAI 연관성에서 *oipA* 위상 변이의 역할

지도교수 차정헌

연세대학교 대학원 응용생명과학과

Angulmaduwa Lokuruge Sacheera Kapruk Lakshan

Angulmaduwa

*H. pylori*의 Outer Inflammatory Protein A(OipA)는 숙주의 위 점막에서 염증을 유발하는 중요한 독성 인자로 확인되었습니다. *oipA* 유전자의 발현은 위상 변이에 의해 조절되어 임상 *H. pylori* 분리체가 *oipA* 유전자 발현의 '켜짐' 상태와 '꺼짐' 상태 사이를 전환할 수 있도록 합니다. 이 연구는 독성 표현형에 대한 *oipA* 위상 변이의 영향과 임상 *H. pylori* 분리체에서 다른 필수 독성 결정 요인과의 관계를 밝히는 것을 목표로 합니다.

우리는 미국에서 18개의 *oipA* 'on'과 22개의 *oipA* 'off'- *H. pylori* 임상 분리체를 분석하여 독성 표현형을 평가했습니다. *oipA* 'on' 분리체가 감염된 AGS 세포에서 IL-8 induction (16/18), gastrin expression (16/18) 및 cell elongation (16/18)의 높은 수준을 일관되게 나타냈습니다. *oipA* off 분리체는 상당히 낮은 수준의 IL-8 (1/22), gastrin expression (1/22) 및 cell elongation (1/22)을 나타냈습니다. 게다가, *oipA* 'on' 분리체의 대다수는 CagA(16/18)를 발현하고 CagA를 숙주 세포(14/18)로 전위시켰는데, 이는 더 높은 독성 가능성을 나타냅니다. 이러한 표현형 차이는 두 그룹 사이에 *cag*PAI가 다르게 존재함을 시사했습니다. 후속 PCR 분석 결과, *oipA* on 분리체에는 주로(16/18) *cag*PAI 유전자가 포함되어 있는 반면, 대부분의 *oipA* off 분리체(20/22)에는 부족한 것으로 나타났습니다. 선택된 20종의 비교 게놈 분석을 통해 *oipA* on 분리체는 온전한

*cag*PAI를 보유하고 추가로 10개의 유전자로 구성된 고유한 세트를 보유하고 있는 반면, *oipA* off 분리체는 cagPAI에 대해 음성이었고 추가로 28개의 별개의 유전자를 운반하는 것으로 확인되었습니다. *oipA* 위상 변이와 *cag*PAI 사이의 연관성을 다양한 지리적 영역으로 확장하기 위해 데이터베이스에서 사용 가능한 613개의 *H. pylori* 게놈 서열을 사용하여 전체 게놈 분석을 수행했습니다. 이는 hpEurope, hpEastAsia 및 hpAfrica1의 지리적 지역에서 *oipA* 위상 변이 상태와 *cag*PAI의 유무 사이의 강한 연관성을 보여주었습니다.

이 연구의 결과는 *H. pylori*의 독성을 조절하는 데 있어 *oipA*의 중요한 역할을 강조합니다. 특히, 이 *oipA* 위상 변이는 *cag*PAI의 존재 또는 부재를 조절하여 세균 독성에 영향을 미칩니다. 이 조절은 *oipA* 위상 변이가 *H. pylori*의 병원성을 지시하여 숙주-병원체 상호 작용에 영향을 미치고 질병 결과에 영향을 미칠 수 있는 중요한 기계론적 경로를 강조합니다.