Roles of *oipA* in Type IV secretion system of *Helicobacter pylori*

Jing Lai

The Graduate School

Yonsei University

Department of Applied Life Science

Roles of *oipA* in Type IV secretion system of *Helicobacter pylori*

Directed by Professor Jeong-Heon Cha

The Doctoral Dissertation Submitted to

the Department of Applied Life Science,

the Graduate School of Yonsei University

in partial fulfillment of the requirements

for the degree of Doctor of Philosophy in Applied Life Science

Jing Lai

June 2021

This certifies that the Doctoral dissertation

of Jing Lai is approved.

Thesis Supervisor: Jeong-Heon Cha

Yun-Jung Yoo

Sang-Sun Yoon

Ji-Hwan Ryu

Ki Woo Kim

The Graduate School

Yonsei University

June 2021

ACKNOWLEDGMENT

First and foremost, I would like to thank my supervisor, Prof. Jeong-Heon Cha, who provided me this precious opportunity and constant support during my Ph.D. and acted as my role model not only in academia but also in life. Without his guidance and support, I could never have passed this millstone and experienced such tremendous happiness in my life.

I'm deeply indebted to my master supervisor, Prof. Gang Luo, an outstanding dentist, who encouraged me to pursue the Ph.D. degree, and always be extraordinarily generous and kind whenever I need some help from him. I am also deeply grateful to Prof. Zheng-guo Piao and Prof. Xin-hong Wang, who always care about my study and life, and supply generous help and support.

I would like to express my deepest appreciation to my committee, Prof. Yun-Jung Yoo, Prof. Sang-Sun Yoon, Prof. Ji-Hwan Ryu, and Prof. Ki Woo Kim for sparing the time regardless of their busy schedule to offer me valuable advice. Special thanks to Prof. Yun-Jung Yoo, who always be nice and considerate to me. I am also thankful to Prof. Ki Woo Kim for his careful inspection to my thesis and always giving me warm greetings: "Hi, how are you?" whenever I meet him. I am also grateful to Prof. Eun-Jung Bak for her kindness during my life in Korea.

I would like to thank Prof. Sungil Jang who gave me invaluable instructions and support on research, especially for the guidance for genome data analysis and study of R studio. He is a great senior and scientist to look up to and learn from. Many appreciations should go to Dr. Aeryun Kim, my dear senior, who taught me lots of experimental skills and always comfort and encourage me whenever I experienced frustrations during the period. Also, I express my gratitude to Dr. Yunhui Choi, a great senior with a sound knowledge of research and always be willing to teach us. I am also genuinely grateful to Dr. Ponsuge Kavinda Sudharaka Tissera, the first senior I met in Korea, who taught me the most basic skills and always be kind enough for giving unselfish assistance both mentally and physically. I would like to express my sincere appreciation to Dr. Aeri Kim, for giving me valuable advice about lab work and life, teaching me how to conduct laser scanning confocal microscope, and appreciate her support which provides without any hesitation. Also, I am grateful for the support from our teaching assistance teachers Yeong-eui Jeon and Ji-Won Seo, without their supports and help in Official affairs and also in daily life, I could not have realized my goal to become a Ph.D.

I have to express my heartfelt gratitude to my best friend in Korea, Myeong-a Kim, I am lucky enough to have her in the lab with me during these four years. She is always being supportive and considerate, always be ready to hug me, and share happiness and depression.

Also, I want to thank all the other lab members for their support and encouragement. Special thanks to Sacheera Angulmaduwa, who always be friendly and ready to help me; Youn-Soo Lee, the sweet and youngest lab member, who always be considerate; and

Ashansa Ramanayake, who always be caring, and cheer me up when I had a hard time.

In addition, I take this opportunity to express thanks to my dear friends who I met in Korea. Firstly, I want to deeply thank Bingjun Jin and Dongzi Chen, the lovely and sincere couple who offered tremendous help and support to me, especially for creating the chance for me and my boyfriend to know each other and build a relationship. In addition, with the accompany and support of my great friends, Jie Jin, Jiayi Li, and Haoyang Zhang, I could have enjoyed the leisure time from the laborious lab life, without them, I could not maintain my mental health during such a stressful Ph.D. life in a foreign country. I also would like to thank Lei Wu, Qing Ge, Kelun Zhang, and Yaru Bao, who gave me great help and encouragement, it's much easier for me to adapt the life in Korea. I am also grateful to my friends and colleagues in China, Zhongqun miao, Xin He, Yinshen Yang, Moyang Yang, Muming Hou, Kai Yang, Andy Shi, with their continuous encouragement and care, I could have come to Korea and enjoyed this Ph.D. journey.

Finally, I would like to thank my family especially my father who continually encouraged and helped me at every stage of my life and longed to see my dream come true. Many thanks need to give to my dear aunt Limin Zhou, for tenderly accompanying and taking care of my father as his life partner, and providing me a comfortable mind to accomplish this achievement. I deeply missed my dear mother and grandparents, it's such a great pity that they could not witness this joyful moment in my life. I am truly grateful to my uncles who gave me generous mental and financial support. I also would like to thank my dear sisters for the assistance and psychological support they provided. Last but not

least, I would like to thank my beloved boyfriend, Lei Liu, who always tries to cheer me up when I am down, always have faith in me, and always stands by me unconditionally. Thank you so much for your extraordinary support, encouragement, and love, I was able to accomplish this important task in my life.

CONTENTS

CONTENTS vii

LIST OF FIGURES x

LIST OF TABLES xii

ABSTRACT(IN ENGLISH) xiii

I. INTRODUCTION 2

II. MATERIALS AND METHODS 9

1. World-wide *Helicobacter pylori* genome analysis 9

2. Population assignment using MLST 9

3. Bacterial strains and cultures 10

4. AGS cell culture 13

5. *oipA* full sequence and alignment 13

6. *oipA* genotyping of AH868 single colonies by colony PCR 17

7. Generation of *oipA* isogenic mutant strains 18

8. Generation of *oipA* complementation mutants.....20

9. Cell elongation assay.....21

10. IL-8 secretion assay.....21

11. Bacterial adherence assay.....22

12. Immunoblot assay.....23

13. DNA sequencing.....25

14. Statistical analysis.....25

III. RESULTS.....26

1. OipA alignment within and between Korean and US *H. pylori*.....26

2. *H. pylori* population assignment.....31

3. Association between *oipA* copy number, *oipA* status or CT repeat number and different *H. pylori* populations.....34

4. Nature *oipA* phase variation in AH868 didn't affect *H. pylori* virulent phenotypes.....38

5. K74 *oipA* isogenic mutants induced wild-type level of IL-8 and cell elongation in AGS cells.....41

6. IL-8 and cell elongation phenotypes induced by G27 Δ*oipA* mutants.....45

7. 9.6% of G27 *oipA* off mutants showed the defective ability of cell elongation induction in AGS cells. 49

8. Complementation of G27 *oipA* off #3 with *oipA* "on" gene restored the capacity of cell elongation induction at a rate of 12.5%.....53

9. Complementation of G27 *oipA* off #3 with *oipA* "off" gene failed to restore the defective capacity of cell elongation induction.....56

10. Variation of cell elongation induction abilities in G27 *oipA* off isogenic mutants were associated with different level of CagA phosphorylation.....59

11. Variation of cell elongation induction abilities in G27 *oipA* off isogenic mutants were not caused by the adherence ability defect 63

IV. DISCUSSION.....65

V. REFERENCES.....72

ABSTRACT(IN KOREAN).....83

LIST OF FIGURES

Figure 1. *oipA* copy number, status, CT-repeat number varies between US and Korean *H. pylori* populations. 8

Figure 2. Schematic of hypothetical steps for *oipA* duplication through the DDAI process and primer design strategy for *oipA* full sequence. ... 16

Figure 3. Alignment of Korean OipA1(A), Korean OipA2(B), and US OipA(C). 28

Figure 4. Alignment of the consensus between Korean OipA1 and OipA2(A), Korean OipA , and US OipA(B). 30

Figure 5. World-wide *H. pylori* genome data included in this study. 33

Figure 6.*oipA* Sequence of AH868 wild type and AH868/CT^{11} at 5' region. 39

Figure 7. *oipA* nature phase variation in AH868 didn't affect the *H. pylori* virulent phenotypes. 40

Figure 8. Schematic of K74 Δ*oipA* isogenic mutants. 42

Figure 9. Cell elongation and IL-8 secretion in AGS cells infected by K74 *oipA* isogenic mutants. 44

Figure 10. Cell elongation and IL-8 secretion in AGS cells infected by G27 Δ*oipA* isogenic mutants. 46

Figure 11. Cell elongation and IL-8 secretion in AGS cells infected by G27 Δ*oipA* complementation mutants. 48

Figure 12. Schematic of G27 *oipA* off isogenic mutant. 50

Figure 13. *oipA* off in G27 didn't affect the phenotype of IL-8 induction. 51

Figure 14. Morphology of AGS cells infected with G27 *H. pylori oipA* off isogenic mutants. 53

Figure 15. Morphology of AGS cells infected with G27 $oipA^{off\text{-}\#3}/oipA^{on}$ isogenic mutants. 55

Figure 16. Morphology of AGS cells infected with G27 $oipA^{off\text{-}\#3}/oipA^{off}$ complement isogenic mutants. 58

Figure 17. Relative levels of p-CagA protein in G27 *oipA* off mutants. ..60

Figure 18. Relative levels of p-CagA protein in G27 $oipA^{off\text{-}\#3}/oipA^{on}$ mutants. 62

Figure 19. Bacterial adherence assay. 64

LIST OF TABLES

Table 1. *E. coli* and *H. pylori* strains used in this study ·····12

Table 2. Primer sequences used for this study ·····15

Table 3. Multilocus sequence types of *Helicobacter pylori* according to geographical area ·····32

Table 4. Frequency of *oipA* genotypes with *H. pylori* populations from worldwide *H. pylori* population ·····36

Table 5. Significant associations of population types with a variation of *oipA* genotypes ·····37

Table 6. CT-repeat number of *oipA* in single colonies of AH868 ·····38

ABSTRACT

Roles of *oipA* in Type IV secretion system of *Helicobacter pylori*

Jing Lai

Department of Applied Life Science

The Graduate School, Yonsei University

(Directed by Professor Jeong-Heon Cha)

Helicobacter pylori (*H. pylori*) is a gram-negative bacterial pathogen that is a major factor to cause gastric diseases including gastric cancer. Among virulent factors in *H. pylori*, outer inflammatory protein A (OipA) encoded by *oipA* gene (*HP0638*/*hopH*) is reported to be important for *H. pylori* pathogenesis.

Expression of OipA is regulated by a slipped strand mispairing (SSM)-mediated phase variation based on Cytosine-Thymine (CT) dinucleotide repeat number in the 5′ region of *oipA* gene; "on" phase variation expresses a full length (usually 305 amino acids) of OipA which is functional, and "off" phase variation shifts a reading frame to express a truncated nonfunctional OipA about 24 amino acids. Intriguingly, previous unpublished data in our lab using 233 Korean and 89 United States (US) *H. pylori* clinical isolates suggested diverse *oipA* genotypes such as the gene copy number and phase variation and CT-repeat number among different *H. pylori* populations. Moreover, the role of *oipA* genotypes in pathogenesis remains to be discovered.

This study aimed to identify the *oipA* genotypes in worldwide populations by comparative genome analysis using 613 *H. pylori* genome data and to investigate an *oipA* function in the virulence of cell elongation and IL-8 induction. First, the OipA consensus amino acid sequences of US or Korean *H. pylori* populations were determined by sequencing 43 Korean and 19 US strains, and it showed that *oipA* genes were highly conserved in two populations. For the comparative genome analysis, the worldwide *H. pylori* strains were grouped into seven population types. It showed that the *oipA* copy number, *oipA* "on" and "off" status, or CT-repeat number were associated significantly

among *H. pylori* populations. Remarkably, more strains in hpEastAsia population (154/248, 62.1%) possess two *oipA* genes in a statistically significant association versus 4/103 (3.9%) of hpAfrica1 ($p = 0.0000$), 0/14 (0%) of hpAfrica2 ($p = 0.0001$), 1/54 (1.5%) of hpAsia2 ($p = 0.0000$) or 10/181(5.5%) of hpEurope ($p = 0.0000$) strains with two *oipA* genes. In CT-repeat number, there are significant associations of hpEastAsia *oipA* with less than 3 CT repeats (358/399,89.7%) when compared with hpAfrica1(1/105, 1%), hpAfrica2 (0/12, 0%), hpAsia2 (25/66, 37.9%) or hpEurope (6/189, 3.2%) (all $p = 0.0000$); moreover, significant associations of hpAsia2 with less than 3 CT repeats (25/66, 37.9%) were also found when compared with hpAfrica1 (1/105, 1%) or hpEurope (6/189, 3.2%) (both $p = 0.0000$). Our data suggested that different populations may have a different strategy to regulate the *oipA* expression. The great diversity in the *oipA* gene number, CT-repeat number, and *oipA* status among *H. pylori* populations may also confer a benefit in the adaptation to the diverse hosts since many genes of outer membrane protein in *H. pylori* are involved in the adaptation by highly dynamic changes of the gene expression.

Next, *H. pylori* strains with naturally occurred *oipA* on and off phase variation and genetically manipulated *oipA* were characterized to understand the role of *oipA* in virulence. We identified a colony with 11 CT-repeats *oipA* "off" phase variation naturally occurred from US AH868 *H. pylori* originally with 12 CT-repeats *oipA* "on". Cell elongation and IL-8 secretion assays were performed using two different clones with the *oipA* on and off phase variation in AH868, which showed no change in cell elongation and IL-8 induction. In the case of Korean strain K74, 18 transformants of Δ*oipA* isogenic

mutants were screened to identify the double deletion, Δ*oipA1*/Δ*oipA2* at two different loci. Only one double Δ*oipA* transformant, K74Δ*oipA1* /Δ*oipA2,* and 17 single Δ*oipA* transformants, K74Δ*oipA1* /*oipA2 or oipA1* /Δ*oipA2,* were generated, which all 18 transformants showed the similar induction level of cell elongation and IL-8 secretion to K74 wild type. In a hpEurope strain G27, three Δ*oipA* transformants by genetic manipulation were selected and characterized for cell elongation and IL-8 secretion. Unexpectedly one G27 Δ*oipA* transformant showed defected virulence phenotypes of cell elongation and IL-8 secretion while two G27Δ*oipA* transformants showed similar induction levels of cell elongation and IL-8 secretion to G27 wild type. However, the complement of G27Δ*oipA* transformant with G27 *oipA* didn't recover the ability to induce cell elongation and IL-8 secretion, suggesting that the defected virulence phenotypes occurred by secondary mutation other than Δ*oipA*.

Since the results of G27 have shown inconsistent Δ*oipA* phenotypes of cell elongation and IL-8 secretion, a different approach was employed to imitate natural *oipA* off phase variation and thus to investigate the role of *oipA* phase variation in the virulence phenotypes. We generated *oipA* "off" phase variation in G27 *H. pylori* by transforming an *oipA*^{off} construct by homologous recombination and the transformants became permanently *oipA* "off". To avoid bias for the previous *oipA* investigation, a large number of the G27 *oipA*^{off} transformants were selected and analyzed for the virulence phenotypes. Three out of 31 G27 *oipA*^{off} transformants (9.6%) showed a significant decrease in cell elongation but

a similar level of IL-8 secretion to G27 wild type while the rest of G27 *oipA*^{off} transformants (28/31, 90.4%) showed the wild-type induction levels for cell elongation and IL-8 secretion. Since 9.6% of transformants showed a defect in cell elongation induction, the unforeseen secondary mutation was suspected to be responsible for the defect. Thus, to determine whether the defect was due to *oipA* "off" phase variation, G27 *oipA*$^{off\#3}$ was complemented by G27 wild-type *oipA* "on", resulting in G27 *oipA*$^{off\#3}$/*oipA*^{on}. From the analysis of a large number of the transformants, four out of 32 transformants (12.5%) showed the restoration of the capability to induce cell elongation while the 87.5% transformants remained a decrease in cell elongation. As a control, the *oipA* "off" construct in G27 *oipA*$^{off\#3}$ was replaced by the *oipA* "off" without *kan-sacB* cassette and none of the 32 transformants showed the restoration. Further Western blot analysis suggested that the association between *oipA* phase variation and cell elongation phenotype is correlated with CagA phosphorylation. Thus, it strongly suggests that *oipA* phase variation is associated with the virulence phenotype of cell elongation, probably in an indirect way.

Since *oipA* is associated with the virulence phenotype of cell elongation by differential CagA phosphorylation which plays a critical role in the development of gastric cancer, it is tempting to speculate that these different *oipA* genotypes among the populations may contribute to the distinct gastric cancer rates.

Key word: *Helicobacter pylori*, *oipA*, phase variation, pathogenesis

Roles of *oipA* in Type IV secretion system of *Helicobacter pylori*

Jing Lai

Department of Applied Life Science

The Graduate School, Yonsei University

(Directed by Professor Jeong-Heon Cha)

I. INTRODUCTION

Gastric cancer is the fifth most common cancer worldwide[1, 2], the epidemiology of gastric cancer has substantial geographical heterogeneity, part of this geographical variation correlated with *H. pylori* infection rates as well as *H. pylori* genetic diversity across populations[3, 4]. *Helicobacter pylori* is a microaerophilic, gram-negative bacterium with spiral shape bacterium discovered in the epithelial lining of the stomach[5], based on several clinico-epidemiological studies which showed the close relationship between *H. pylori* infection and gastric cancer, the international agency for research on cancer[6-9], World Health Organization (IARC/WHO) identified *H. pylori* as a group I carcinogen in 1994[10].

The development of *H. pylori*-associated gastric diseases is thought to be related to various bacterial virulent factors along with the host genetics and environmental influences[11]. One of the well-studied virulent factors is *cag* pathogenicity island (*cag*PAI), which is a ~40 kb DNA insertion element that was likely from horizontal acquisition[12-14]. It contains 37-42 *cag* genes, most of the proteins encoded by these genes are involved in the assembly of type IV secretion system (T4SS)[15-17]. T4SS is a syringe-like structure that could translocate bacterial macromolecules, including peptidoglycan[18] and the cytotoxin-associated gene A (CagA) into host cells[17-22].

CagA is the first identified bacterial protein involved in human cancer. Once be translocated into the host cell by T4SS, CagA could be phosphorylated by host cell kinases at a conserved tyrosine residue located within the EPIYA (Glu-Pro-Ile-Tyr-Ala) motif[23-27],

and alter multiple host signaling pathways. Phosphorylated CagA causes dysregulation of epithelial structure and integrity, by inducing cytoskeletal rearrangement, increased cellular mobility, and elongated cell shape referred to as the "hummingbird phenotype"[28-31]. Recently, the nonphosphorylated CagA was proved to be involved in promoting proliferation and inflammation by the activation of β-catenin and NF-κB signaling[18].

Despite *cag* genes, another large family of outer membrane proteins (OMPs) has been reported to be associated with the outcome of *H. pylori*-related gastric diseases since they play a vital role in facilitating the colonization of *H. pylori* to host cells at the primary stage[32-34]. The OMPs are encoded by a family of paralogous genes which 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[35]. 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[35].

HopH, one of the OMPs belongs to Hop family, encoded by *hopH* (*hp*0638/*oipA*), has been designated as outer membrane inflammatory protein A (OipA) because of its role in promoting secretion of interleukin (IL)-8 from gastric epithelial cells *in vitro* and enhanced gastric inflammation *in vivo*[36]. Its expression can be regulated by the SSM according to the CT nucleotide repeat number in their signal-peptide-coding region. The *oipA* 'on' phase variation expresses a full-length OipA, typically 305 amino acids; 'off' phase variation

shifts the reading frame to produce only approximately 24 amino acids. The genetic diversity of *oipA* genotypes such as *oipA* "on" and "off" status, CT repeat patterns among different *H. pylori* populations were reported by several studies[37-39]. Besides, a second *oipA* gene was found in some East-Asian *H. pylori* strains while in none of Western *H. pylori* strains[40, 41]. The duplication of *oipA* gene can be explained by the theory of DNA duplication associated with inversion (DDAI)[40], and DDAI could be followed by another inversion through homologous recombination between the duplicated regions. In Korean strains which carrying two *oipA* genes, there are two possible gene rearrangements adjacent to each *oipA* gene (Figure 2). In our previous study, we screened the *oipA* status and CT repeat number in 89 US *H. pylori* and 233 Korean strains, found that none of the US *H. pylori* have two *oipA* genes, while except 4 strains, all Korean strains have two *oipA* genes in their genome (Figure 1), this result confirmed the dramatic *oipA* genotype variability between different *H. pylori* populations.

The involvement of functional OipA as adhesion in facilitating *H. pylori* colonization to gastric mucosa has been reported in some animal studies. In one study that used Mongolian gerbils, the TN2 *oipA* isogenic mutant failed to establish colonization in gastric mucosa[42]. Also, in another study using mice as experiment animals, the CPY2052 *oipA* isogenic mutant showed a significantly lower level of colonization than the wild-type[43].

However, the role of *oipA* in *H. pylori*-related inflammation is still under controversy since its identification in 2001[36]. The chronic infection of *H. pylori* is characterized by the inflammation of gastric epithelium and is considered a major cause of gastric diseases. *H.*

pylori infection can induce the expression of a number of the genes in host cells which involved in inflammation including IL-8[44-46]. Thus, the secretion amount of IL-8 by host cells can be used as a determinant of inflammation level. *H. pylori* can induce the IL-8 secretion of gastric epithelium cells by activating NF-κB pathway, and *H. pylori*-induced IL-8 induction is reported to be T4SS dependent at early-phase while CagA dependent at the late-phase of infection[47].

Among some epidemiology studies, a significant association between *oipA* status and some clinical gastric disease outcomes such as duodenal ulcer and peptic ulcer diseases has been found[36, 48-50], while other researchers found no link between *oipA* status with gastric diseases[51-53]; Another intriguing aspect of *oipA* is that its "on" status is found to have linkage with some important virulent genes in *H. pylori* such as *cag*PAI, *cagA*, *babA*2, and *vacA* phenotype[37, 39, 48, 53, 54].

For *in vitro* study, the controversy mainly comes from the role of *oipA* in inducing IL-8 secretion in host cells. Some researchers found that the *oipA* isogenic mutants showed a significant decrease in the ability to induce IL-8 secretion than their parent strains[36], while some other studies got no significant results regarding the inflammation induction role of *oipA*[37, 42, 55]. The inconsistent results were assumed to be related to the strain specificity, and none of the studies showed positive result confirmed the role of *oipA* by complementation trials.

Thus, we carefully revisited this issue, this study aims to characterize the *oipA* genotype to worldwide populations by comparative genome analysis using 613 *H. pylori* genome data and to explore its role in virulence phenotypes such as cell elongation and IL-8 induction in *H. pylori* strains.

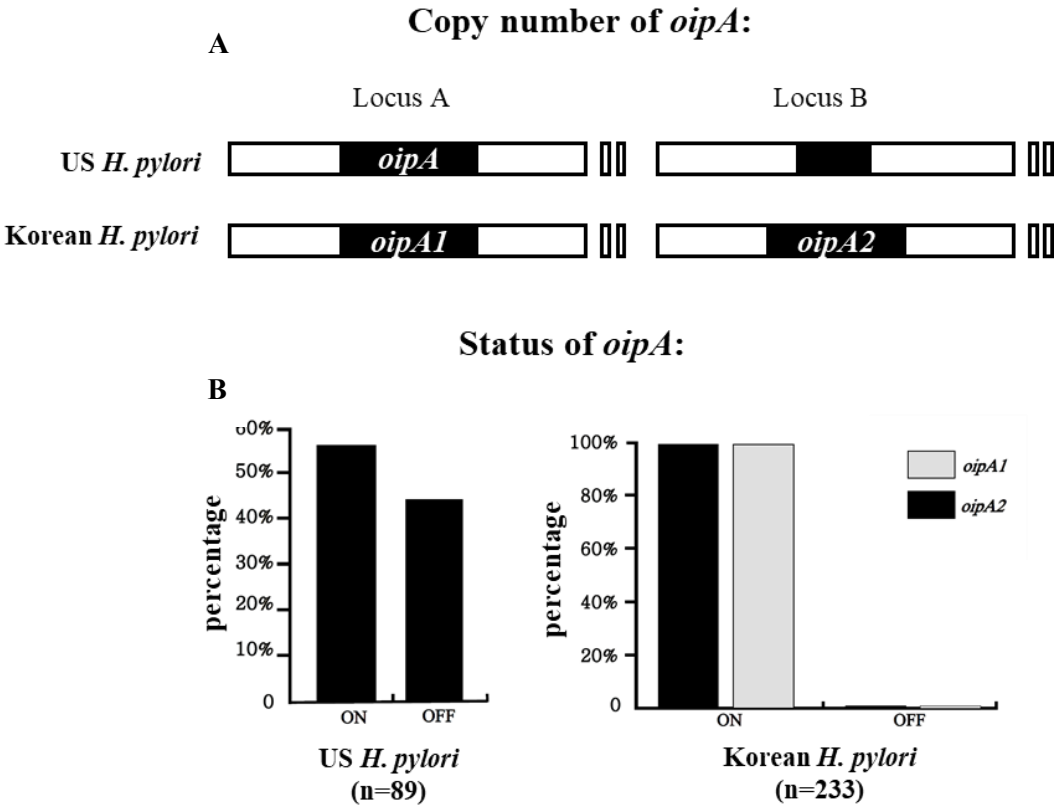


Figure 1. *oipA* copy number, status, CT-repeat number varies between US and Korean *H. pylori* populations.

We screened 89 United States clinical *H. pylori* strains and 233 Korean clinical *H. pylori* strains, the results showed that (A) only one *oipA* gene could be found among US clinical *H. pylori* strains, while most of Korean *H. pylori* have 2 copies of *oipA* and located at two different loci, Locus A and Locus B, respectively; (B) only 50 (56%) of US strains have *oipA* "on" status, while 99% of *oipA*1 and *oipA*2 of Korean strains have "on" status; (C. most of US *H. pylori oipA* contain more than 6 CT-repeats, while the *oipA* contains less than 3 CT-repeats are dominant among Korean *H. pylori.*

II. MATERIALS AND METHODS

1. Worldwide *Helicobacter pylori* genome analysis

A total of 782 *Helicobacter pylori* genomes were collected for this analysis. After excluded 169 of the genome which has duplicates, or from the derivative of an animal adapted strains, or population unidentifiable, or highly clonal with others according to multilocus sequence typing (MLST) analysis, or *oipA* CT-repeat number unidentifiable. Finally, a total of 613 *H. pylori* genome data collected from 27 different countries were included in this final genome comparative analysis.

2. Population assignment using MLST

Population assignments of global *H. pylori* genome samples were conducted with multilocus sequence typing (MLST) method using 7 housekeeping genes (*atpA*, *efp*, *mutY*, *ppa*, *trpC*, *ureI*, and *yphC*) as described previously[56-58]. Strain population assignment was performed using the "no admixture model" of STRUCTURE, the "admixture model" was also applied to verify the existence of admixture of distinct populations as previously described[59]. Phylogenetic tree reconstitution was made using the neighbor-joining algorithm implemented in MEGA 6.0 software.

3. Bacterial strains and cultures

All strains used in this study are listed in Table 1 and were cultured as previously described[60]. Simply, all *H. pylori* strains were grown on horse blood agar plates supplemented with antibiotics.

For cultures of *oipA* isogenic mutant strains that contained the *cat* cassette or *kan-sac*B cassette, chloramphenicol or kanamycin disulfate was added to the horse blood agar plates or liquid culture medium at a concentration of 25 μg/ml; 5% sucrose was added to the horse blood agar plates for selections and cultures of *oipA* isogenic complementation strains after losing the *kan-sac*B cassette.

As for AGS cell infection to measure cell elongation, IL-8 induction, and immunoblot assays, *H. pylori* strains were prepared as previously described[27, 60]. Briefly, *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°C, and were then inoculated into new media with an initial optical density of 0.05 at 600 nm. These cultures were grown under the same condition with initial culture for 16-18 h before infection.

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. *E. coli* and *H. pylori* strains used in this study

Strain	Relevant characteristic(s)	Selection marker(s)[α]
E. coli		
DH5α Δ*oipA*^{K74}	*E. coli* DH5α with plasmid Δ*oipA*^{K74}::*cat*	Amp, Cm
DH5α Δ*oipA*^{G27}	*E. coli* DH5α with plasmid Δ*oipA*^{G27}:: *kan-sac*B	Amp, Km, Sucrose^{S}
DH5α *oipA* off ^{G27}	*E. coli* DH5α with plasmid *oipA* off ^{G27}::*kan-sac*B	Amp, Km, Sucrose^{S}
DH5α *oipA* ^{on}	*E. coli* DH5α with plasmid *oipA* ^{on}	Amp
DH5α *oipA* ^{off}	*E. coli* DH5α with plasmid *oipA* ^{off}	Amp
H. pylori		
Wild-type		
G27	Wild-type	
K74	Wild-type	
Deletion mutants		
G27Δ*cag*PAI	G27 Δ*cag*PAI::*aphA*	Km
G27Δ*cag*A	G27Δ*cag*A::*cat*	Cm
K74Δ*cag*L	K74Δ*cag*L::*kan-sac*B	
oipA isogenic mutants		
G27Δ*oipA*	G27 Δ*oipA*::*kan-sac*B	Km, Sucrose^{S}
K74Δ*oipA*1	K74Δ*oipA*1::*cat*	Cm
K74Δ*oipA*2	K74Δ*oipA*2::*cat*	Cm
K74Δ*oipA*1/2	K74Δ*oipA*1/2::*cat*	Cm
G27 *oipA* off	G27 *oipA* off::*kan-sac*B	Km, Sucrose^{S}
oipA complementation mutants		
G27 Δ*cag*A/*oipA*		
G27 *oipA* off #3/*oipA*^{on}	G27 *oipA* off #3 complemented with *oipA* from G27	Sucrose^{R}
G27 *oipA* off #3/*oipA*^{off}	G27 *oipA* off #3 complemented with *oipA* from AH882	Sucrose^{R}

[α] Amp, ampicillin resistant; Cm, chloramphenicol resistant; Km, kanamycin resistant; Sucrose^{R}, sucrose resistant; Sucrose^{S}, sucrose sensitive

4. AGS cell culture

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.

5. *oipA* full sequence and alignment

Most Korean clinical strains carry two *oipA* genes, there are two possible chromosomal gene rearrangements between two *oipA* genes. Thus, we designed two primer sets to amplify each *oipA* gene for a PCR-based method, which are listed in Table 2. For *oipA*1, the primer sets of LA-f/LA-r and LA-f/LB-r were designed to amplify *oipA*1 full length and the primer sets of LB-f/LB-r and LB-f/LA-r were designed to amplify *oipA*2, respectively. The aligning sites of all primers are located on the outsides of *oipA* gene (figure. 2). The genomic DNA of 43 Korean *H. pylori* were used as templates, the PCR reaction was performed in a thermocycler (SimpliAmpTM Thermal Cycler, Thermo Fisher Scientific, Waltham, MA, USA) using the following cycling parameters: 94°C (5 min); 32 cycles of 94°C (30 s), 54°C (for *oipA*1) (30 s) or 50°C (for *oipA*2) (30 s), 72°C (2 min); and a final extension at 72°C (10 min) PCR products were purified with MEGAquick-

spinTM plus kit (iNtRON Biotechnology, Inc.) according to manufacturer's instruction and sequenced at both directions using primers of *oipA*-F and *oipA*-R. Besides, US *oipA* sequences from 19 strains were extracted from genome sequences.

The predicted amino acid sequence of OipA was generated by Vector NTI version 9.1 (Invitrogen, Carlsbad, CA). The alignments of the predicted amino acid sequences of Korean OipA1, Korean OipA2, and US OipA were generated using WebLogo generator as previously described. Consensus and multiple sequence alignments of OipA were generated from 43 Korean and 19 US clinical *H. pylori* isolates which are all *oipA* "on" strains using CLC main work bench 20.0.4. Alignments were shaded using Boxshade version 3.3.1 with similar amino acids shaded in gray, identical amino acids shaded in black, while different amino acids with no shade.

Table 2. Primer sequences used for this study

Primer name	Primer sequence (5'→3')
LA-f	AGC GTA AGG GAA GTA TCC AG
LB-f	AAC TGC TCA ACA TCG CTT CT
LA-r	AGA ATT TCA AAT GGC TAA AG
LB-r	GCT TGT AAA AGC ATG CGG AT
oipA-F[c]	GGA CTA ACG ATA AGC GAG CG
oipA-R[c]	CCA ATC ACA AGC CCT GAA GA
oipA-FWFp[a]	GTT TTG GCT CCA CGC TGA AAG G
oipA-FWRp[a]	CGT TCA ACC CCG GGA GGC TCG AGA ATC CGT ATT GAA ACC ATA GGG[d]
oipA-RWFp[a]	CAA TAC GGA TTC TCG AGC CTC CCG GGG TTV AAC GAA GGG TAA AAG GGC[d]
oipA-RWRp[a,b]	AGT TAA GAT AAA CGC TCA CCA C
oipA-off-FWFp[b]	TAG GCA CAT TCG CCC CAC AA
oipA-off-FWRp[b]	CGT TCA ACC CCG GGA GGC TCG AGC ACC GAT GCT ACC TTG TCC T[d]
oipA-off-RWFp[b]	CAT CGG TGC TCG AGC CTC CCG GGG TTG AAC GAA GGG TAA AAG GGC[d]
OipA FF[b]	CCA TGA AAA AAG TCC TCT TAC TAA C
OipA FR[b]	GAG TTA GTA AGA GGA CTT TTT TCA TGG
Kanamycin-R	TGC CTT GCC CCC TGA AAT CC
Cat-R	TCA TGC CGT TTG TGA CGG CT
oipA CT-R	TCC ACC ATT TCC TGC GAA TC

[a] primers were used for deleted *oipA* mutant
[b] primers were used for *oipA* off mutant
[c] primers were used for *oipA* sequencing
[d] Underlined CCCGGG indicates *sma*I, and underlined CTCGAG indicates *xho*I enzyme site

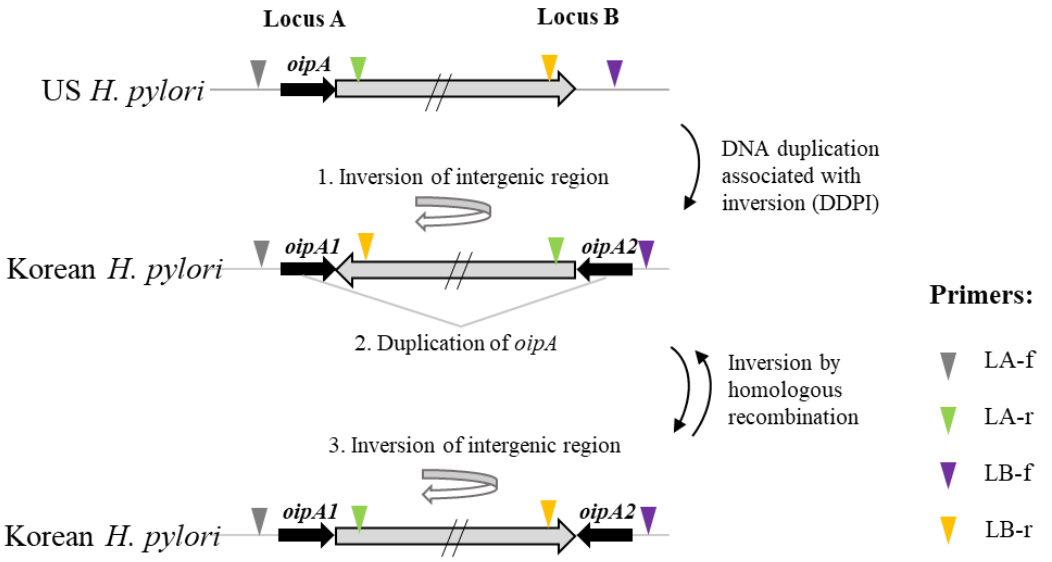


Figure 2. Schematic of hypothetical steps for *oipA* duplication through the DDAI process and primer design strategy for *oipA* full sequence.

The duplication of *oipA* gene in the genome can be explained by the theory of DNA duplication associated with inversion (DDAI)[40], and DDAI could be followed by another inversion through homologous recombination between the duplicated regions. Thus, in Korean strains which carrying two *oipA* genes, there are two possible gene rearrangements adjacent to each *oipA* gene. For *oipA*1, the primer sets of LA-f/LA-r and LA-f/LB-r were designed to amplify *oipA*1 full length and the primer sets of LB-f/LB-r and LB-f/LA-r were designed to amplify *oipA*2, respectively.

6. *oipA* genotyping of AH868 single colonies by colony PCR

Different bacterial volumes (1 μl, 10 μl, 100 μl) of serial dilution were taken from liquid culture of AH868 in Brucella Broth and were plated onto horse blood agar plates by shaking of glass beads. Plates were incubated for 3-4 days under microaerobic conditions until single colonies appeared. Single colonies were streaked onto a second blood agar plate for making further frozen stocks.

Half of each colony was scraped with the pipette tip and was dispersed into a PCR tube containing 20 μl of distilled water, PCR tubes were incubated at 99°C for 3 min in SimpliAmp Thermal Cycler (Applied Biosystems™, USA). The supernatant of the distilled water after centrifugation was used as the template of colony PCR. Primers LA-f and OipA-CT-R were used to amplify the 5' region of *oipA* which resulting in a 934-bp amplicon. The PCR was carried out as follows: a cycle at 94°C for 4 min; 30 cycles of 94°C for 30 s, 52°C for 30 s, and 72°C for 60 s; a final elongation step at 72°C for 10 min. PCR products were sequenced using oipA-CT-R primer and was analyzed for examination of CT repeat number.

7. Generation of *oipA* isogenic mutant strains

G27 Δ*oipA*, K74Δ*oipA,* and G27 *oipA* off isogenic mutants were constructed, respectively (Table 1). For G27 Δ*oipA* or K74Δ*oipA* isogenic mutants, the front and rear part of the respective deletion region of *oipA* in G27 or K74 were PCR amplified using primers (Table 2) with *Xho*I and *Sma*I restriction sites, the front and rear part were stitched to each other with splicing by overhang extension (SOE) PCR. PCR fragment was purified and cloned into pGEMT®-T Easy Vector Systems (Promega) following the manufacturers' instructions. A *kan-sac*B cassette which contains a kanamycin resistance gene (*aphA*) and sucrose sensitive gene (*sacB*) was inserted into *Xho*I-*Sma*I site of insertion part, yielding DH5α Δ$oipA^{G27}$; a chloramphenicol resistance gene (*cat*) was inserted into *Xho*I-*Sma*I site of insertion part, yielding DH5α Δ$oipA^{K74}$. The resulting plasmids were transformed into *E. coli* DH5α (RBCBioscience) respectively, according to the manufacturers' instructions, and transformants were grown overnight on Luria-Bertani (LB) plates containing kanamycin disulfate (25 mg/ml) or chloramphenicol (25 mg/ml). Resistant colonies were inoculated to selective LB broth, and plasmids from the resulting culture were purified with LaboPassTM Plasmid Mini kit (LaboPass). Plasmids were digested with *Xho*I and *Sma*I, and sequenced for confirmation of correct construction before the nature transformation of *H. pylori*.

G27 *oipA* off isogenic mutants were created similarly, but with two steps for the front part: firstly, DNA fragment F1 from the 5' part of US clinical isolates AH882 were amplified, which has *oipA* "off" status with 7 CT-repeats, a premature stop codon appears after 24 amino acids from the start codon. Additionally, the fragment F2 which contains

the *oipA* sequence downstream of CT-repeat part of G27 was amplified by primer sets (table2) respectively, and F1 and F2 were fused to obtain the front part by SOE PCR. The rear part was amplified with primers using the genomic DNA of G27 as a template, subsequently, the purified front and rear parts for the *oipA* off constructs were stitched to each other by SOEing PCR. PCR fragment was purified and then TA cloned into pGEMT®-T Easy Vector Systems (Promega). A *kan-sac*B cassette which contains a kanamycin resistance gene (*aphA*) and sucrose sensitive gene (*sacB*) was inserted into *Xho*I-*Sma*I site of the insertion part, yielding DH5α *oipA* off ^{G27} and the resulting plasmid was transformed into *E. coli* DH5α (RBCBioscience), and transformants were grown overnight on Luria-Bertani (LB) plates containing kanamycin (50 mg/ml). Plasmids were double digested with restriction enzymes *Xho*I and *Sma*I to confirm the correct size of the insert part and sequenced for confirmation of correct construction before the natural transformation of *H. pylori*.

After making these *oipA* constructs, they were transformed into the K74 WT and G27 WT by nature transformation. Transformation of *H. pylori* with the plasmid containing the *kan-sac*B or *cat* cassette and flanking *oipA* sequences, with selection on kanamycin (25 mg/ml) or chloramphenicol (25 mg/ml) respectively, resulted in the replacement of 455 to 500 bp of *oipA* in Δ*oipA* mutants and the replacement of 125 to 499 bp of *oipA* in *oipA* off mutants with cassette, respectively. Primer sets of LA-f and cat-R/Kanamycin-R or LB-f and cat-R are used to detect whether the antibiotic-resistant gene had been inserted into the desired locus. Especially, *cat* cassette was inserted to each locus of K74, thus we got three

kinds of K74 *oipA* isogenic mutants, which are either *oipA*1 or *oipA*2 knockout, or double *oipA* knock out mutants, they were designated as K74 Δ*oipA*1, K74 Δ*oipA*2, and K74 Δ*oipA*1/2, respectively.

8. Generation of *oipA* complementation mutants

oipA complementation mutants were constructed using countraselectable sucrose susceptibility method. Two kinds of *oipA* complementation constructs were constructed, $oipA^{on}$ and $oipA^{off}$. Simply, the fragment for *oipA* complementation from G27 ("*oipA*" on) or AH882 ("*oipA*" off) was amplified with primer *oipA*-FW-Fp and *oipA*-RW-Rp by PCR, and the product was purified and then TA cloned into pGEMT®-T Easy Vector Systems (Promega), yielding $oipA^{on}$ or $oipA^{off}$, respectively. The resulting plasmids were transformed into *E. coli* DH5α (RBCBioscience) according to the manufacturers' instructions, and transformants were grown overnight on Luria-Bertani (LB) plates containing ampicillin (100 mg/ml).

After confirming the sequence, construct $oipA^{on}$ was used to reinsert the G27 wild-type *oipA* into G27 Δ*oipA* #3 or G27 *oipA* off #3, resulting in the G27 Δ*oipA* /*oipA* or G27 *oipA* $off^{-\#3}$/$oipA^{on}$ transformants; $oipA^{off}$ was used to reinsert the AH882 wild type *oipA* into G27 *oipA* off #3, resulting in the G27 *oipA* $off^{-\#3}$/$oipA^{off}$ transformants. Sucrose-resistant colonies were fully sequenced at the *oipA* locus to confirm that they had undergone the desired *oipA* complementation.

9. 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°C, when the cell confluency 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 suspended in plain DMEM, and AGS cells were infected at a multiplicity of infection (MOI) of 100. At 6 h (for G27 and its isogenic mutants) or 8 h (for K74 and its isogenic mutants) post-infection, cells were fixed with 4% paraformaldehyde. Images of the cells were taken under 200 × or 400 × magnifications, and cell elongation was calculated as previously described[61]. Briefly, one hundred cells were randomly selected from each well, and cell elongation was calculated by dividing the length of the longest protrusion of a cell by the breadth of the cell. The values corresponding to the elongation of cells induced by each *H. pylori* mutant strain are presented as box plots generated using RStudio Team (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL http://www.rstudio.com/.

10. IL-8 secretion assay

AGS cells were infected by *H. pylori* strains as previously described in cell elongation assay. 0.8 ml of cell culture medium was taken at 6 h (for G27 and its isogenic mutants) or 8 h (for K74 and its isogenic mutants) post-infection before fixation 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 replicates.

11. Bacterial adherence assay

The adherence assay was conducted as previously described[62]. Briefly, 2×10^5 AGS cells were seeded onto a 24-well cell culture plate in DMEM supplemented with 10% FBS and then incubated for 2 days until they formed the confluent monolayers. *H. pylori* strains (about 10^8 bacterial/ mL) were added into each well and incubated for 1 min. After incubation, the plate was washed three times with fresh DMEM to remove non-adherent bacteria, afterward, monolayers were incubated with 120 μl DMEM containing 0.5% saponin (saponin rapidly lyses AGS cells but not the bacteria) for 3 min, subsequently, cells were harvested by scraping the monolayers off from the wells. The samples were then serially diluted and plated for CFU counts to determine the number of adherent *H. pylori*.

12. Immunoblot assay

Expression of phosphorylated CagA, CagA, OipA, GAPDH, and UreA 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°C. At 2 h before infection, cells were washed with PBS and the medium was changed to 2 ml of plain DMEM. Liquid cultures of *H. pylori* were suspended in plain DMEM, and AGS cells were infected at an MOI of 100. At 6 h 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 bacterial lysates and infected cell 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).

Phosphorylated CagA expression was detected by using mouse monoclonal anti-phosphotyrosine antibody pY99 ((Santa Cruz Biotechnology, Santa Cruz, CA) diluted 1:10,000 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 H RP-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 expression of total CagA, OipA, GAPDH, and UreA were detected by using rabbit polyclonal anti-CagA antibody b-300 (Santa Cruz Biotechnology, Dallas, TX), polyclonal antibody (sp56) kindly provided by Yoshio Yamaoka, rabbit polyclonal anti-GAPDH antibody (Koma Biotech, Seoul, South Korea), and rabbit polyclonal anti-UreA antibody b-234 (Santa Cruz Biotechnology), respectively. The rabbit anti-OipA antibody was diluted 1:300 in 5% BSA-TBST, other 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.

Relative expression levels of phosphorylated CagA were measured as previously described[63]. Briefly, developed films were scanned to JPEG files, intensities of blots on the images were measured using ImageJ software version 1.47, ratios of p-CagA to CagA which normalized to the UreA intensity of each strain were calculated, and error bars represent standard deviations from three independent experiments.

13. DNA sequencing

Sanger dideoxy DNA sequencing was performed at Cosmo Genetech Co., Ltd. (Seoul, South Korea). The primers for DNA sequencing are listed in Table 1. The resulting DNA sequences were analyzed using Vector NTI version 9.1 (Invitrogen, Carlsbad, CA).

14. 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 results of the cell elongation assay and IL-8 induction assay were presented as the means ± standard deviation and analyzed using one-way analysis of variance. Multiple comparisons were performed using Tukey's method. *P* values were presented with F values and degrees of freedom (For between-group and within-group comparisons). 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* locus number or *oipA* status. A *p*-value less than 0.05 was considered to be statistically significant.

III. RESULTS

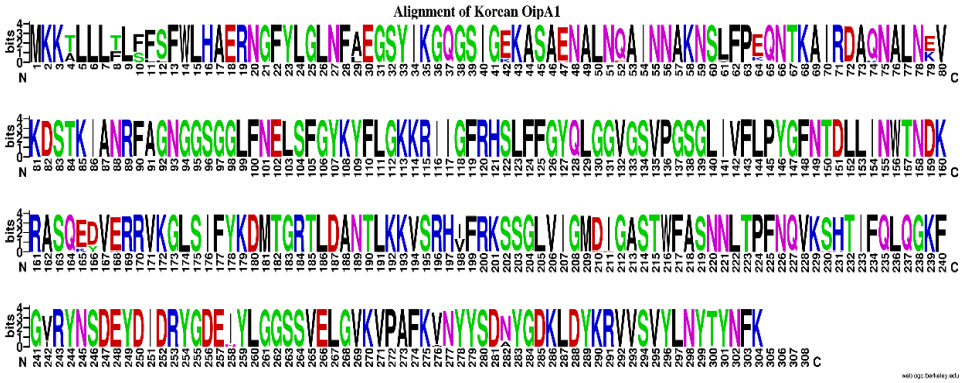
1. OipA alignment within and between Korean and US *H. pylori*

86 *oipA* genes (43 of *oipA*1 and *oipA*2, respectively) from 43 Korean *H. pylori* clinical isolates were sequenced in full length, the consensus sequences of Korean OipA1, OipA2, and US OipA were created by using the amino acid sequences deduced from these *oipA* gene nucleotide acid sequences using the CLC main workbench 20.

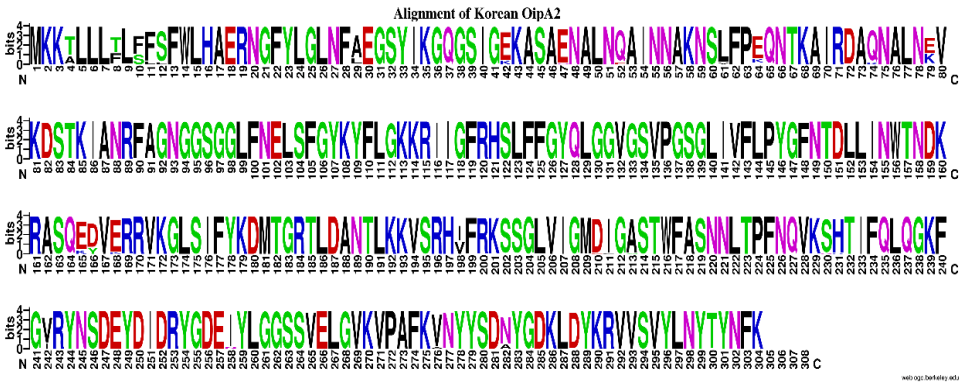
Comparisons of OipA amino acid sequences in Korean *H. pylori* (43 strains) and US *H. pylori* (19 *oipA* "on" strains) which including the comparisons within OipA1 and OipA2 of Korean *H. pylori* and the comparison within US OipA were generated using WebLogo generator as previously described[64]. The consensus sequences of Korean OipA1, OipA2, and the US OipA were used for the comparison between Korean *oipA*1 and *oipA*2, and comparison between Korean and US *oipA* using the CLC main workbench 20.

Among 43 Korean strains and 19 US strains, except for the 5' region which contains the varies of CT repeats, the OipA among different strains within the same populations are highly conserved (Figure 3). The consensus sequence of OipA1 and OipA2 are identical (Figure 4), while the OipA between Korean and the US shared 93.2% of consensus positions and 91.6% of identity positions.

A



B



C

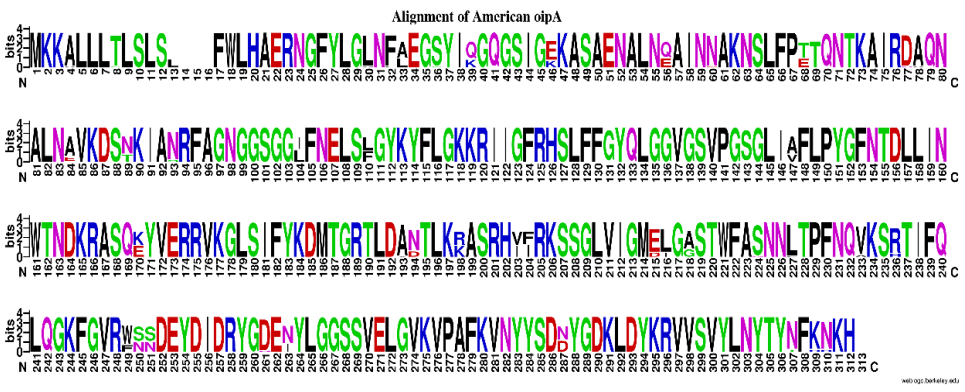


Figure 3. Alignment of Korean OipA1(A), Korean OipA2(B), and US OipA(C).

OipA from 43 Korean and 19 US clinical *H. pylori* isolates are used for alignment. These figures were created by Weblogo generator, the logo represents the alignment at each point by a stack of letters, where the height of each letter is proportional to the observed frequency of the corresponding amino acid, and the overall height of each stack is proportional to the sequence conservation, measured in bits, at that position.

A

```
OipA1     1 MKKTLLLTLFFSFWLHAERNGFYLGLNFAEGSYIKGQGSIGEKASAENALNQAINNAKNS
OipA2     1 MKKTLLLTLFFSFWLHAERNGFYLGLNFAEGSYIKGQGSIGEKASAENALNQAINNAKNS

OipA1    61 LFPEQNTKAIRDAQNALNEVKDSTKIANRFAGNGGSGGLFNELSFGYKYFLGKKRIIGFR
OipA2    61 LFPEQNTKAIRDAQNALNEVKDSTKIANRFAGNGGSGGLFNELSFGYKYFLGKKRIIGFR

OipA1   121 HSLFFGYQLGGVGSVPGSGLIVFLPYGFNTDLLINWTNDKRASQEDVERRVKGLSIFYKD
OipA2   121 HSLFFGYQLGGVGSVPGSGLIVFLPYGFNTDLLINWTNDKRASQEDVERRVKGLSIFYKD

OipA1   181 MTGRTLDANTLKKVSRHIFRKSSGLVIGMDIGASTWFASNNLTPFNQVKSHTIFQLQGKF
OipA2   181 MTGRTLDANTLKKVSRHIFRKSSGLVIGMDIGASTWFASNNLTPFNQVKSHTIFQLQGKF

OipA1   241 GVRYNSDEYDIDRYGDEIYLGGSSVELGVKVPAFKVNYYSDNYGDKLDYKRVVSVYLNYT
OipA2   241 GVRYNSDEYDIDRYGDEIYLGGSSVELGVKVPAFKVNYYSDNYGDKLDYKRVVSVYLNYT

OipA1   301 YNFK
OipA2   301 YNFK
```

B

```
Korean    1 MKKTLLLTLFFSFWLHAERNGFYLGLNFAEGSYIKGQGSIGEKASAENALNQAINNAKNS
US        1 MKKALLLTLSLSFWLHAERNGFYLGLNFAEGSYIQGQGSIGEKASAENALNQAINNAKNS

Korean   61 LFPEQNTKAIRDAQNALNEVKDSTKIANRFAGNGGSGGLFNELSFGYKYFLGKKRIIGFR
US       61 LFPTQNTKAIRDAQNALNAVKDSNKIANRFAGNGGSGGIFNELSLGYKYFLGKKRIIGFR

Korean  121 HSLFFGYQLGGVGSVPGSGLIVFLPYGFNTDLLINWTNDKRASQEDVERRVKGLSIFYKD
US      121 HSLFFGYQLGGVGSVPGSGLIAFLPYGFNTDLLINWTNDKRASQEYVERRVKGLSIFYKD

Korean  181 MTGRTLDANTLKKVSRHIFRKSSGLVIGMDIGASTWFASNNLTPFNQVKSHTIFQLQGKF
US      181 MTGRTLDANTLKRASRHIIRKSSGLVIGMELGASTWFASNNLTPFNQVKSRTIFQLQGKF

Korean  241 GVRYNSDEYDIDRYGDEIYLGGSSVELGVKVPAFKVNYYSDNYGDKLDYKRVVSVYLNYT
US      241 GVRWSSDEYDIDRYGDENYLGGSSVELGVKVPAFKVNYYSDNYGDKLDYKRVVSVYLNYT

Korean  301 YNFK---
US      301 YNFKNKH
```

Figure 4. Alignment of the consensus between Korean OipA1 and OipA2(A), Korean OipA and US OipA(B).

Consensus and multiple sequence alignments of OipA were generated from 43 Korean and 19 US clinical *H. pylori* isolates using CLC main work bench 20.0.4. Alignments were shaded using Boxshade version 3.3.1 with similar amino acids shaded in gray, identical amino acids shaded in black, while different amino acids with no shade. Consensus positions : 93.2%, identity positions : 91.6%.

2. *H. pylori* population assignment

H. pylori strains can be divided into 7 populations by employing the MLST method, the population types also showed geographical associations. They are designated as follows: hpEastAsia, hpAfrica1, hpAfrica2, hpAsia2, hpEurope, hpNEAfrica, and hpSahul. Additionally, in the present study, hpEastAsia includes four subpopulations which are EastAsia hybrid, hspAmerind, hspEAsia, and hspMaori. HpAfrica1 also includes three subpopulations, hspWAfrica, hspSAfrica, and hspCAfrica (Table 3). As shown in figure 5, except for hspCAfrica, all the population types can be found in the 613 worldwide *H. pylori* genomes. However, in the HpNEAfrica and HpSahul populations, each has only one sample.

Table 3: Multilocus sequence types of *Helicobacter pylori* according to geographical area

H. pylori population	Geographical/ethnic group
hpEastAsia	
hspEAsia	East Asians
hspMaori	Polynesians, Melanesians, Taiwan aboriginals, Indonesia
hspAmerind	Amerindians
hpAfrica1	
hspWAfrica	West Africa, Senegal, Colombia, Brazil, South America
hspSAfrica	South Africa
hspCAfrica	Cameroon
hpAfrica2	South Africa, Angola
hpAsia2	South/central Asia
hpEurope	Europe
hpNEAfrica	Northeast Africa
hpSahul	Australia, New Guinea (aboriginals)

Modified from: Table 2 of "Yamaoka Y. *Helicobacter pylori* typing as a tool for tracking human migration[J]. Clinical Microbiology and Infection, 2009, 15(9): 829-834."[65]

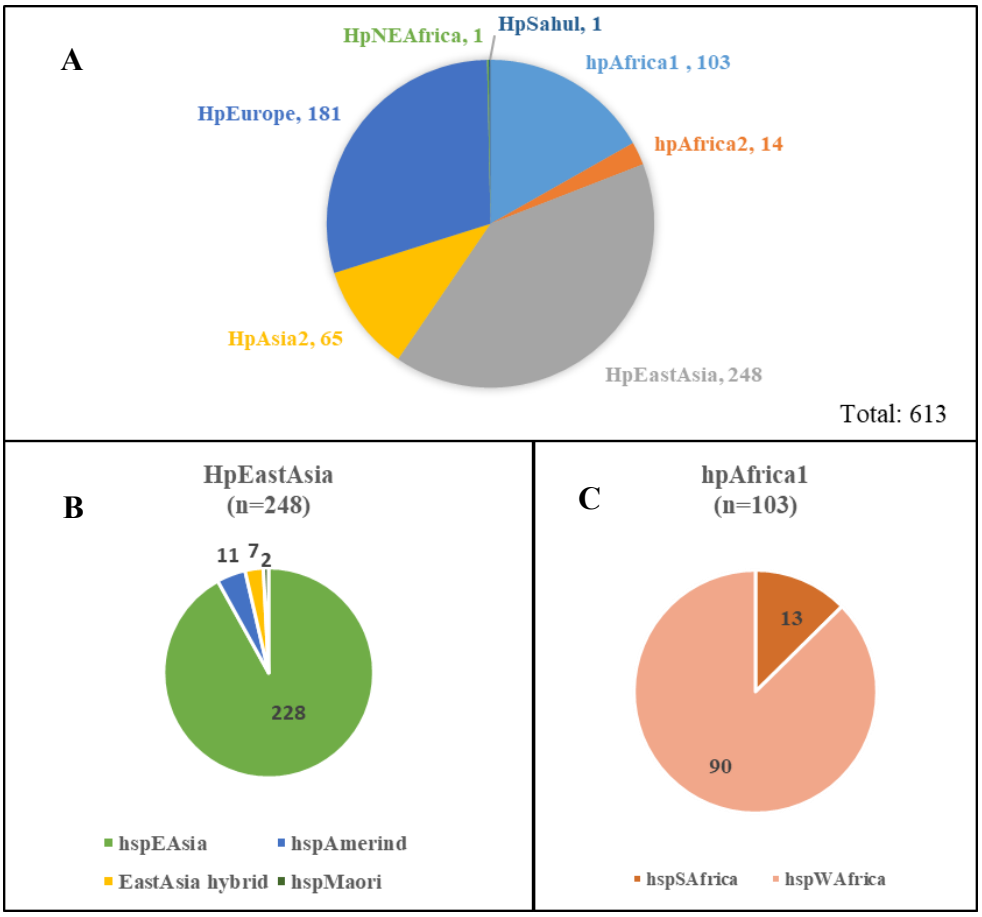


Figure 5. Worldwide *H. pylori* genome data included in this study.

A total of 613 *H. pylori* strains were included in this study, the sample number of each population are indicated.

3. Association between *oipA* copy number, *oipA* status, or CT repeat number and different *H. pylori* populations

Overall, *oipA* can be detected in all the 613 *H. pylori* genome samples, we analyzed the *oipA* genotypes of *oipA* copy number, *oipA* status, and CT-repeat number of each sample (Table 4). *oipA* copy number was categorized as 1 or 2, *oipA* status was categorized as "on" or "off", and CT-repeat number was categorized as "≥ 5" or "≤ 3", since that in theory, the slipped strand mispairing frequency is proportional to the dinucleotide-repeats number, more than five repeats may be required for the meaningful mutation rates, and less than 3 repeats may be not available for it[66].

For *oipA* copy number, among all the worldwide *H. pylori*, there are 72.4% of strains with one *oipA* gene, while 62.1% of hpEastAsia population have 2 *oipA* copies; for *oipA* status, in total, 84.1% of strains have *oipA* "on" status, while only 35.7% of hpAfrica2 and 61.8% of hpEurope population have *oipA* "on" status; for CT-repeat number, in total, the number of *oipA* which contain more than 5 or less than 3 CT-repeats are similar. However, when we take a closer look into each population, 89.7% of hpEastAsia *oipA* and 37.9% of hpAsia2 *oipA* contain less than 3 CT repeats, the ratio is much higher than other populations. And not surprisingly, *oipA* copy number, *oipA* 'on' and 'off' status, or CT-repeat number of '≤ 3' and '≥ 5' were associated significantly among *H. pylori* populations.

We further did the pairwise comparison. As shown in Table 5, more strains in hpEastAsia population (154/248, 62.1%) possess two *oipA* genes in a statistically

significant association versus 4/103 (3.9%) of hpAfrica1 (p = 0.0000), 0/14 (0%) of hpAfrica2 (p = 0.0001), 1/54 (1.5%) of hpAsia2 (p=0.0000) or 10/181(5.5%) of hpEurope (p = 0.0000) strains with two *oipA* genes. For *oipA* status, there are significant association of hpAfria2 *oipA* with "off" status (9/14, 64.3%) when compared with 14/107 (13.1%) of hpAfrica1 *oipA* (p = 0.0001), 26/402 (6.5%) of hpEastAsia *oipA* (p= 0.0000) or 2/66, (3%) of hpEastAsia2 *oipA* (p = 0.0000) with "off" status. Also, significant association of hpEurope *oipA* with "off" status (73/191, 38.2%) can be seen when compared with and compared with hpAfrica1, hpEastAsi or hpEastAsia2 *oipA* with "off" status (all p= 0.0000). In CT-repeat number, there are significant associations of hpEastAsia *oipA* with less than 3 CT repeats (358/399,89.7%) when compared with hpAfrica1(1/105, 1%), hpAfrica2 (0/12, 0%), hpAsia2 (25/66, 37.9%) or hpEurope (6/189, 3.2%) (all p = 0.0000); moreover, significant associations of hpAsia2 with less than 3 CT repeats (25/66, 37.9%) were also found when compared with hpAfrica1 (1/105, 1%) or hpEurope (6/189, 3.2%) (both p = 0.0000).

Table 4. Frequency of *oipA* genotypes with *H. pylori* populations from 613 *H. pylori* strains.

oipA genotypes	No. (%) of *H. pylori* in indicated population category							Total	*p*-value*
	Afri1	Afri2	EA	Asia2	Euro	NEA	Sah		
oipA copy number									**2.2e-16**
1	99 (96.1)	14 (100)	94 (37.9)	64 (98.5)	171 (94.5)	1 (100)	1 (100)	444 (72.4)	
2	4 (3.9)	0 (0)	154 (62.1)	1 (1.5)	10 (5.5)	0 (0)	0 (0)	169 (27.6)	
oipA status[a]									**2.2e-16**
on	93 (86.9)	5 (35.7)	376 (93.5)	64 (97)	118 (61.8)	1 (100)	1 (100)	658 (84.1)	
off	14 (13.1)	9 (64.3)	26 (6.5)	2 (3)	73 (38.2)	0 (0)	0 (0)	124 (15.9)	
CT-repeat number[b]									**2.2e-16**
≥5	104 (99)	12 (100)	41 (10.3)	41 (62.1)	183 (96.8)	1 (100)	1 (100)	390 (50.5)	
≤3	1 (1)	0 (0)	358 (89.7)	25 (37.9)	6 (3.2)	0 (0)	0 (0)	383 (49.5)	

Afri1, hpAfrica1; Ari2, hpAfrica2; EA, hpEastAsia; Asia2, hpAsia2; Euro, hpEurope; NEA, hpNEAfrica; Sah, hpSahul.

[a] *oipA* status number was counted by *oipA* gene number. If one strain has two *oipA* "on" genes, we counted the "on" status number as "2", thus, the total number of *oipA* genes is 782.

[b] The CT-repeat number was categorized as "≤ 3" or "≥ 5" which doesn't include the *oipA* genes which have 4 CT-repeats. Among the 9 *oipA* genes which have 4 CT-repeats, 2 belong to hpAfrica1, 2 belong to hpAfrica2, 3 belong to hpEastAsia, and 2 belong to hpEurope. In total, 773 *oipA* genes were included in the counting.

* *p*-value was calculated by Pearson's Chi-squared test with benjamini & hochberg correction, statistically significant values are presented in bold.

Table 5. Significant associations of population types with a variation of *oipA* genotypes

Populations compared with *oipA* genotypes	*p*-value
***oipA* copy number**[a]	
hpEastAsia vs hpAfrica1	**0.0000**
hpEastAsia vs hhpAfrica2	**0.0001**
hpEastAsia vs hpAsia2	**0.0000**
hpEastAsia vs hpEurope	**0.0000**
***oipA* status**[b]	
hpAfrica2 vs hpAfrica1	**0.0001**
hpAfrica2 vs hpEastAsia	**0.0000**
hpAfrica2 vs hpAsia2	**0.0000**
hpEurope vs hpAfrica1	**0.0000**
hpEurope vs hpEastAsia	**0.0000**
hpEurope vs hpAsia2	**0.0000**
CT repeat number[c]	
hpEastAsia vs hpAfrica1	**0.0000**
hpEastAsia vs hpAfrica2	**0.0000**
hpEastAsia vs hpEurope	**0.0000**
hpAsia2 vs hpAfrica1	**0.0000**
hpAsia2 vs hpEurope	**0.0000**

[a] The comparisons were conducted with *oipA* copy number categorized as "1" or "2".

[b] The comparisons were conducted with *oipA* status categorized as "on" or "off".

[c] The comparisons were conducted with CT repeat number categorized as "≤ 3" or "≥ 5".

* *p*-value was calculated by Pearson's Chi-squared test with benjamini & hochberg correction, statistically significant values are presented in bold.

4. Nature *oipA* phase variation in AH868 didn't affect *H. pylori* virulent phenotypes

To address how the nature phase variation of *oipA* status will affect the virulence genotypes of *H. pylori*, we chose one of the US Clinical *H. pylori* AH868 as the parent strain for screening the single colonies since it contains the highest 12 CT repeats in *oipA* among all the *H. pylori* strains in our lab. AH868 was isolated from a black 52-year old male gastric cancer patient, hundred single colonies were selected for CT repeat number screening by sequencing, and only one of them was found to change from 12 CT-repeats to 11 CT-repeats in *oipA* (Table 6), resulting in the *oipA* status changed from 'on' to 'off', We named this colony as AH868/CT^{11} (Figure 6).

We used the AH868/CT^{11} to infect AGS cells, and the phenotypes of cell elongation and IL-8 induction were used as determinants of *H. pylori* T4SS dependent virulence. Surprisingly, AH868/CT^{11}, which is "*oipA*" off, still induced a wild-type level of cell elongation and IL-8 (Figure 7).

Table 6. CT-repeat number of *oipA* in single colonies of AH868

Original strain	No. of isolates in indicated CT-repeat number category		Total
	12 CT-repeats	11 CT-repeats	
AH868[a]	99	1	100

[a] AH868, a United States Clinical *H. pylori*, with 12 CT repeats and "on" status in *oipA*.

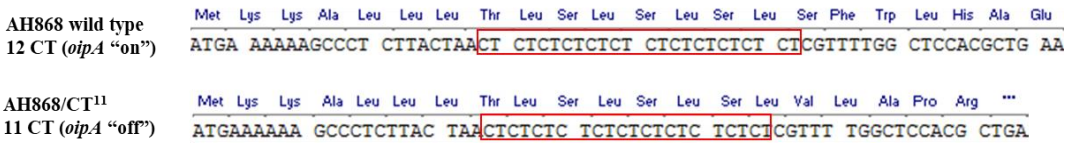


Figure 6. *oipA* Sequence of AH868 wild type and AH868/CT^{11} at 5' region.

Red boxes indicate the CT-repeat part of each *oipA*, the AH868 wild type has 12 CT-repeats, while AH868/CT^{11} has 11 CT-repeats in *oipA*, when CT repeat number changed from 12 to 11, the *oipA* status changed from "on" to "off".

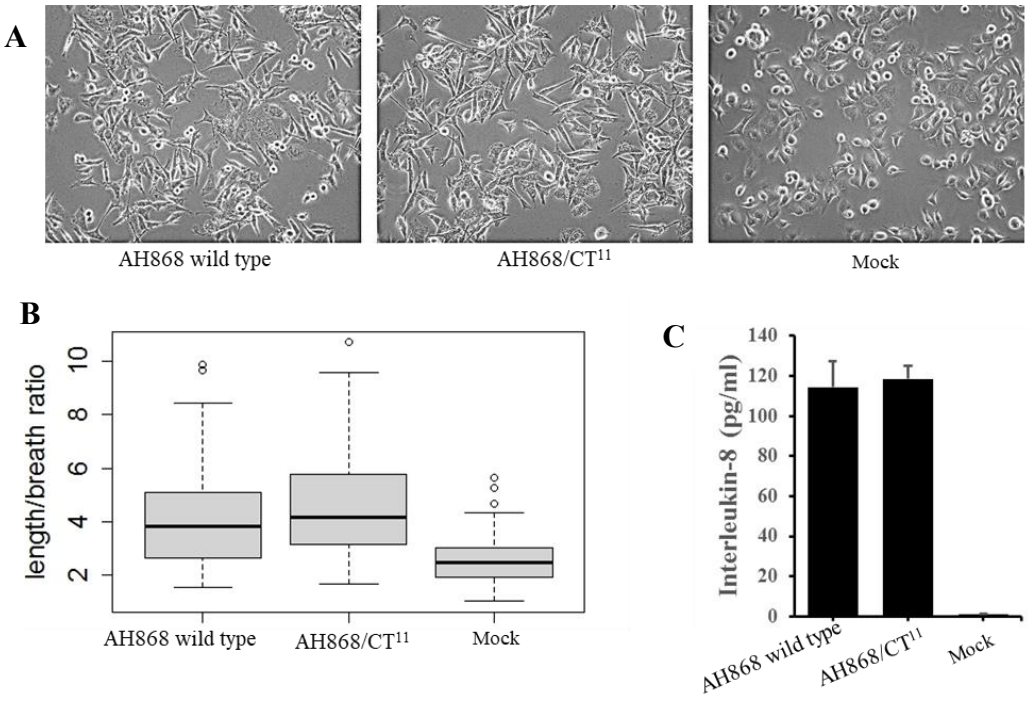


Figure 7. *oipA* nature phase variation in AH868 didn't affect the *H. pylori* virulent phenotypes.

(A) AGS cells were infected by AH868 wild-type or AH868/CT^{11} at an MOI of 100 for 6 h. Micrographs were obtained under 200× magnification. (B) Cell elongation was calculated as the ratio of length to breadth of a cell. The cell elongation induced by each *H. pylori* strain was graphed using box plots. Thick center lines represent medians; box limits indicate the 25th and 75th percentiles; whiskers extend 1.5 times the interquartile range from the 25th and 75th percentiles, and outliers are represented by open-circle dots (n = 100 for each group). (C) Secretion of IL-8 was measured at 6 h post-infection. The

bar graphs indicate average levels of IL-8 secretion of AGS cells infected with each strain, and error bars represent standard deviations, derived from the results of three independent experiments.

5. K74 *oipA* isogenic mutants induced wild-type level of IL-8 and cell elongation in AGS cells

Since the rate of nature phase variation in *oipA* was quite low, and some reports speculated that the *oipA* function is strain-dependent which may be one of the reasons for the *oipA* function controversy[37, 55]. Thus, we chose a hpEastAsia strain K74 to make Δ*oipA* isogenic mutants. Since it has two *oipA* gene copies, after transformed with Δ*oipA* construct with *cat* cassette, we got three kinds of mutants, which are K74 Δ*oipA*1, K74 Δ*oipA*2, and K74 Δ*oipA*1/2. (Figure 8).

In total, we picked 18 clones of K74 *oipA* isogenic mutants, only one of them got *oipA* double deletion which is K74 Δ*oipA1/2* #8. After infected with AGS cells, we found that all of K74 *oipA* isogenic mutants still induce wild-type levels of IL-8 expression and cell elongation in AGS cells. (Figure 9).

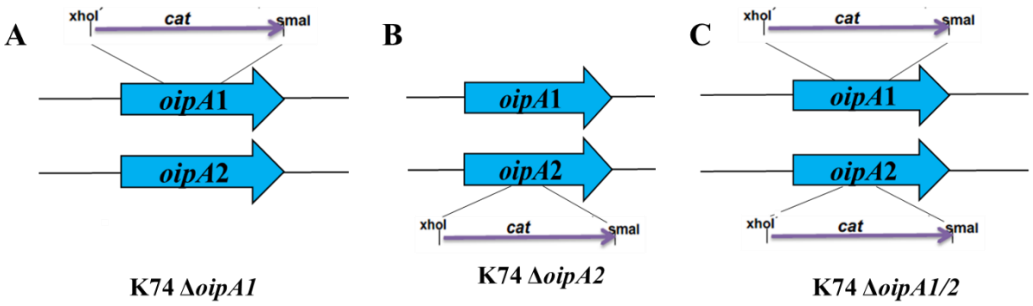
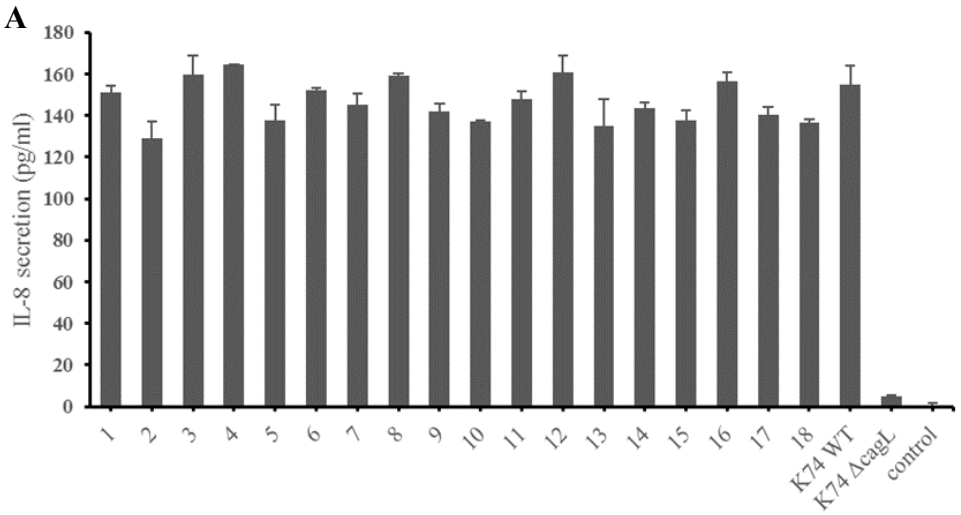


Figure 8. Schematic of K74 Δ*oipA* isogenic mutants.

The *oipA* gene in K74 was deleted with a *cat* cassette, resulting in 3 kinds of mutants which are K74 Δ*oipA*1 (A), K74 Δ*oipA*2 (B), and K74 Δ*oipA*1/2 (C), respectively.



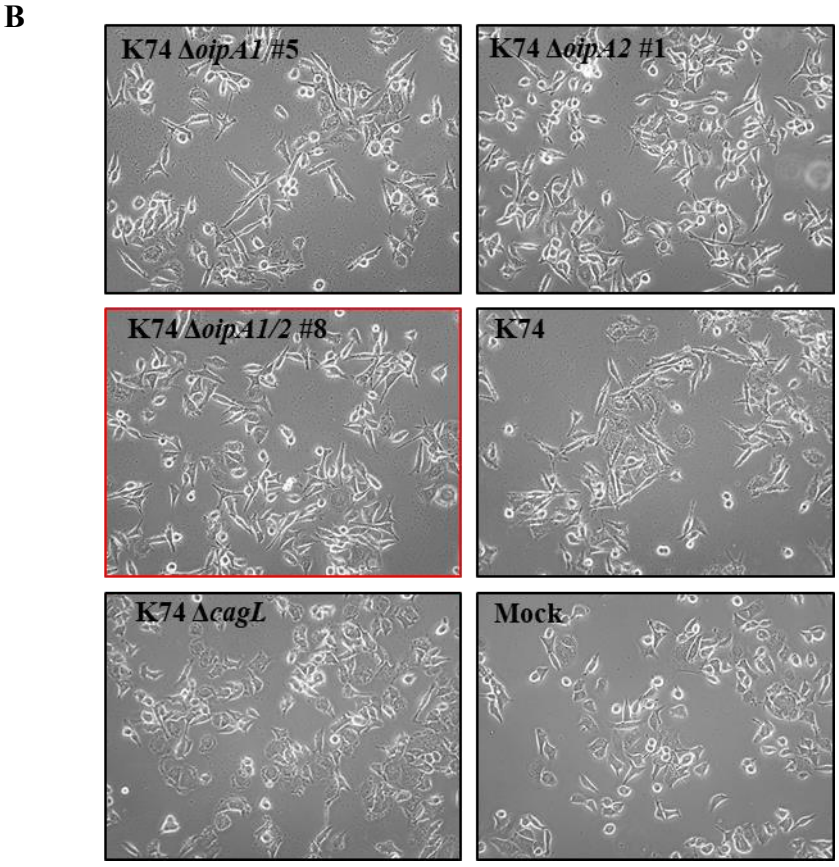


Figure 9. Cell elongation and IL-8 secretion in AGS cells infected by K74 *oipA* isogenic mutants.

(A) AGS cells were infected with indicated K74 *oipA* isogenic mutants (n = 18), secretion of IL-8 was measured at 6 h post-infection. The bar graphs indicate average levels of IL-8 secretion of AGS cells infected with each strain, and error bars represent standard deviations, derived from the results of three independent experiments. (B) The morphology of AGS cells was investigated (magnification 200×), and one representative picture from each kind of K74 *oipA* isogenic mutants was showed, K74 and K74Δ*cagL* were used as the positive and negative control, respectively. K74 Δ*oipA1/2* #8: The only mutant with double *oipA* deletion.

6. IL-8 and cell elongation phenotypes induced by G27 Δ*oipA* mutants

Next, we chose hpEurope strain G27, one of the most widely used *H. pylori* strains for *in vitro* studies, *kan-sac*B cassette was used to insert into the *oipA* of G27 to make Δ*oipA* mutants. Three Δ*oipA* transformants were selected and characterized for cell elongation and IL-8 secretion. One G27 Δ*oipA* transformant showed defected virulence phenotypes of cell elongation and IL-8 secretion while two G27 Δ*oipA* transformants showed similar induction levels of cell elongation and IL-8 secretion to G27 wild type (Figure 10). However, the complement of G27 Δ*oipA* transformant with G27 *oipA* didn't recover the ability to induce cell elongation and IL-8 secretion (Figure 11).

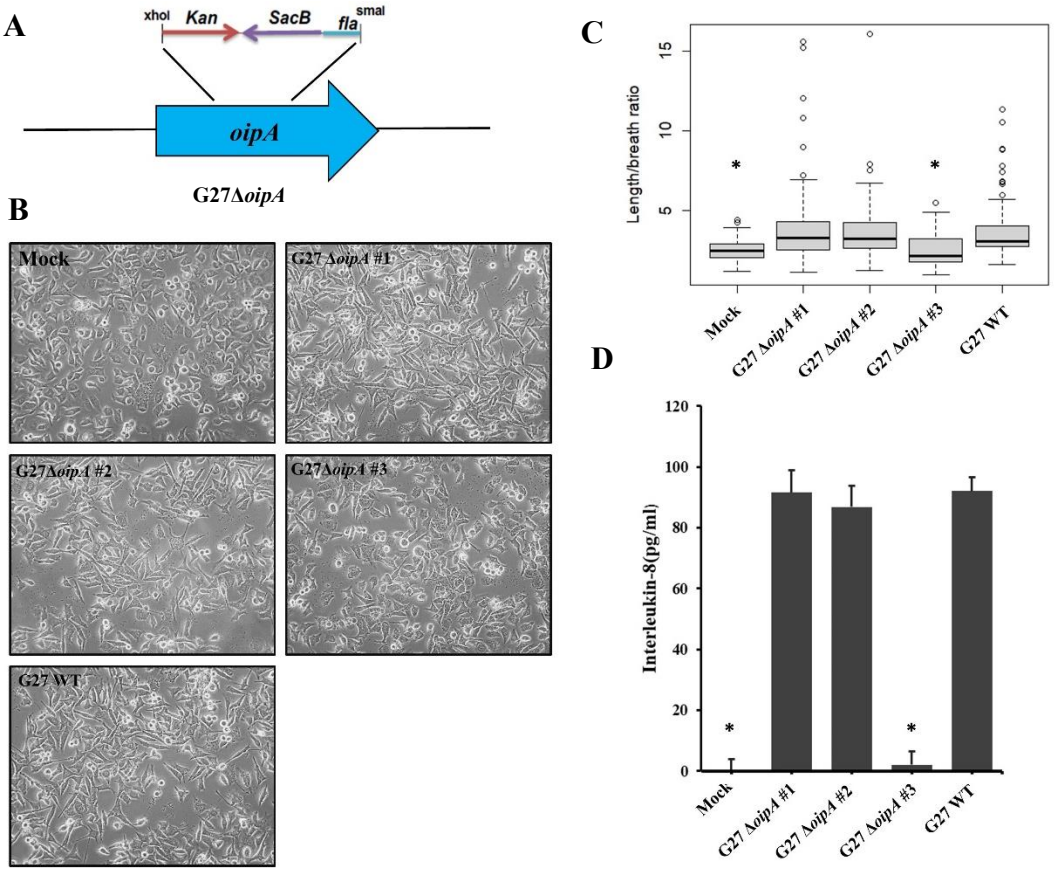


Figure 10. Cell elongation and IL-8 secretion in AGS cells infected by G27 Δ*oipA* isogenic mutants.

(A) Schematic of G27 Δ*oipA*. The *oipA* in G27 was interrupted with *kan-sac*B cassette, resulted in G27 Δ*oipA*. (B) AGS cells were infected with indicated G27 Δ*oipA* mutants, the morphology of AGS cells was investigated at 6 h post-infection (magnification 200×), with one representative picture from each G27 Δ*oipA* mutants were showed, G27 wild-type and Mock were used as the positive and negative control, respectively. (C) Cell

elongation was calculated as the ratio of length to breadth of a cell. The cell elongation induced by each *H. pylori* strain was graphed using box plots. Thick center lines represent medians; box limits indicate the 25^{th} and 75^{th} percentiles; whiskers extend 1.5 times the interquartile range from the 25^{th} and 75^{th} percentiles, and outliers are represented by open-circle dots. n= 100 for each group. (D) Secretion of IL-8 was measured at 6 h post-infection. The bar graphs indicate average levels of IL-8 secretion of AGS cells infected with each strain, and error bars represent standard deviations, derived from the results of 3 independent experiments. *, $p < 0.05$ (compared to the results of G27).

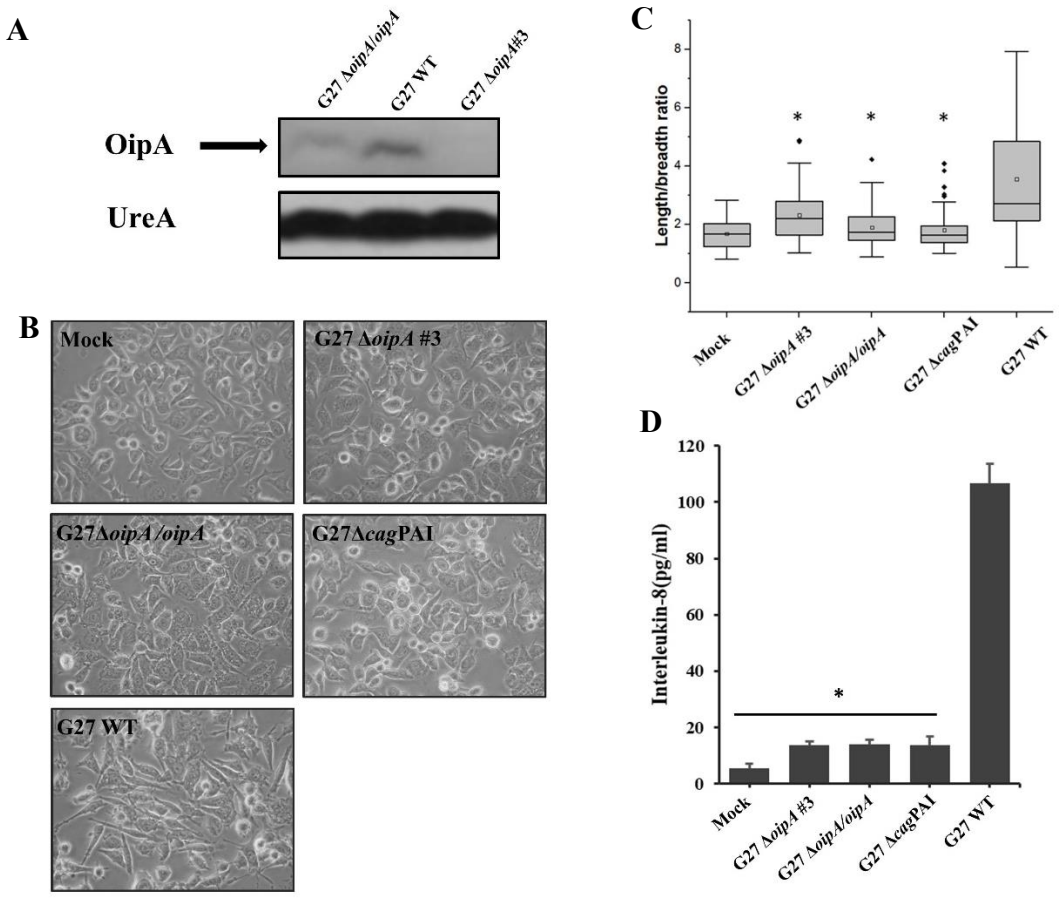


Figure 11. Cell elongation and IL-8 secretion in AGS cells infected by G27 Δ*oipA* complementation mutants.

(A) Western blotting analysis of OipA in G27 Δ*oipA* #3 and its complementation mutant, the proteins were isolated from indicated *H. pylori* strains. The blot was probed with OipA-antibody[67] and UreA-antibody. (B) AGS cells were infected with indicated G27 Δ*oipA* and Δ*oipA*/*oipA* mutant, the morphology of AGS cells was

investigated (magnification 400×) at 6 h post-infection, one representative picture from each sample were showed, G27 WT and G27Δ*cag*PAI were used as the positive and negative control, respectively. (C) Cell elongation was calculated as the ratio of length to breadth of a cell. The cell elongation induced by each *H. pylori* strain was graphed using box plots. Thick center lines represent medians; box limits indicate the 25^{th} and 75^{th} percentiles; whiskers extend 1.5 times the interquartile range from the 25^{th} and 75^{th} percentiles, and outliers are represented by open-circle dots. n= 100 for each group. (D) Secretion of IL-8 was measured at 6 h post-infection. The bar graphs indicate average levels of IL-8 secretion of AGS cells infected with each strain, and error bars represent standard deviations, derived from the results of 3 independent experiments. *, $p < 0.05$ (compared to the results of G27).

7. 9.6% of G27 *oipA* off mutants showed the defective ability of cell elongation induction in AGS cells.

Next, we chose hpEurope strain G27, one of the most widely used *H. pylori* strains for *in vitro* studies, to make *oipA* off mutagenesis mutants (Figure 12). We picked up to 31 clones for the G27 *oipA* off mutants and found all of them induced similar levels of IL-8 to G27 wild type (Figure 13). However, when it comes to the cell elongation phenotype in

AGS cells, three transformants, G27 *oipA* off #3, #4, and #22, which only accounts for 9.6% of the total 31 G27 *oipA* off mutants, induced significantly less cell elongation in AGS cells *in vitro* when compared with G27 wild-type after 6 h of infection (Figure 14).

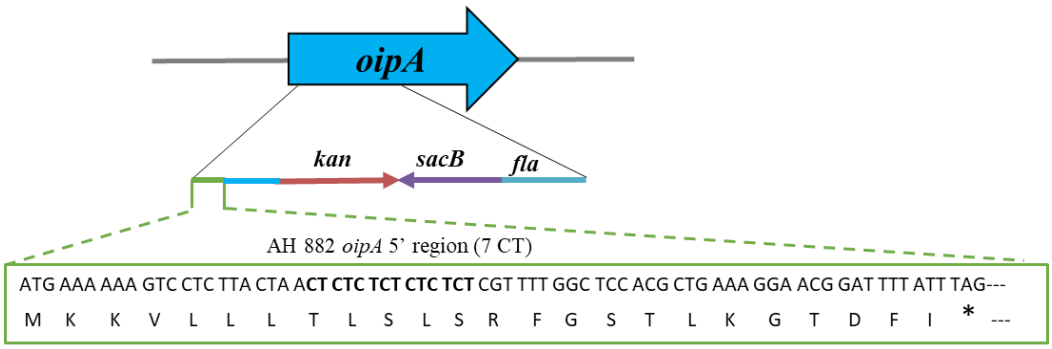


Figure 12. Schematic of G27 *oipA* off isogenic mutant.

The *oipA* of The 5' region of *oipA* in G27 was replaced with the 5' region from AH882, and a *kan*-*sac*B cassette was used to interrupt the *oipA*, resulting in G27 *oipA* off mutant, AH882 is a United states clinical *H. pylori* which has *oipA* "off" status with 7 CT-repeats, a premature stop codon appears after 24 amino acids.

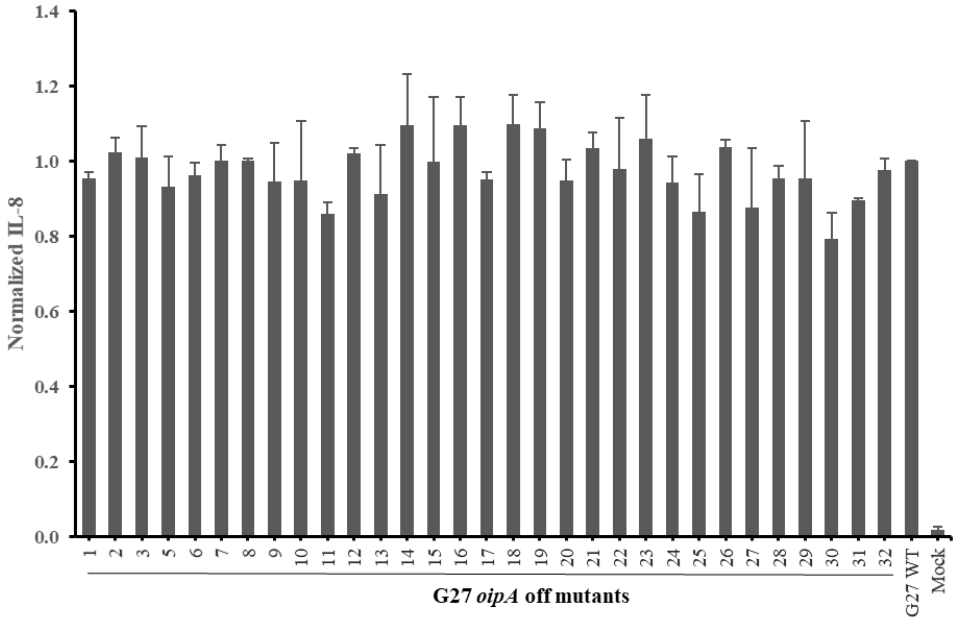
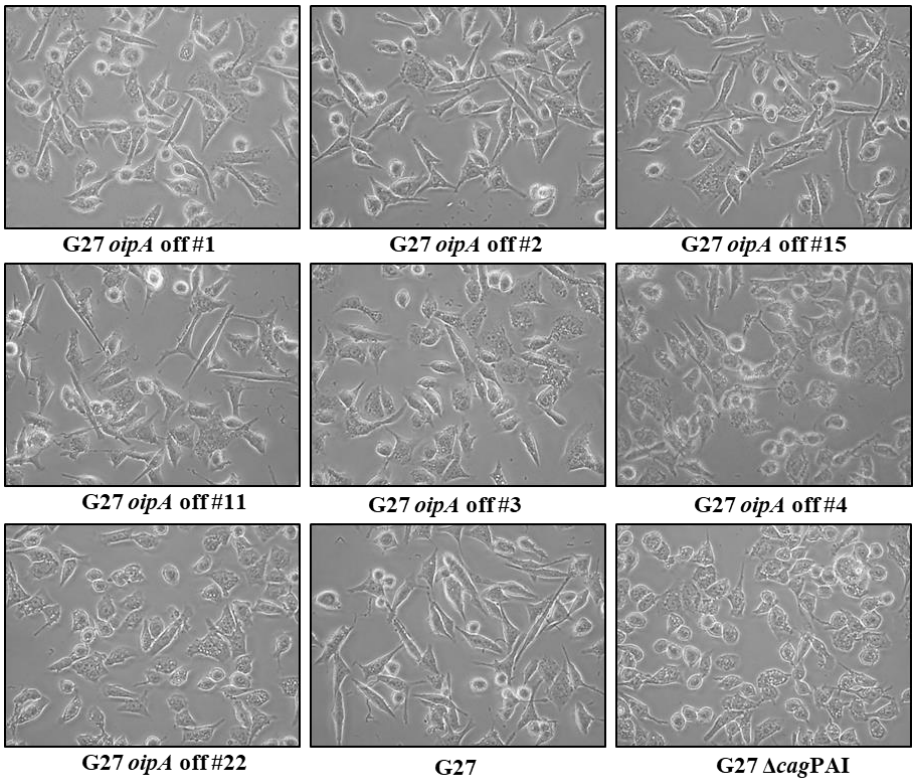


Figure 13. *oipA* off in G27 didn't affect the phenotype of IL-8 induction.

AGS cells were infected with indicated G27 *oipA* off mutants (n = 31) at an MOI of 100, secretion of IL-8 was measured at 6 h post-infection. The bar graphs indicate average normalized levels (by the level induced by G27 WT) of IL-8 secretion of AGS cells infected with each strain, and error bars represent standard deviations, derived from the results of 3 independent experiments.

A



B

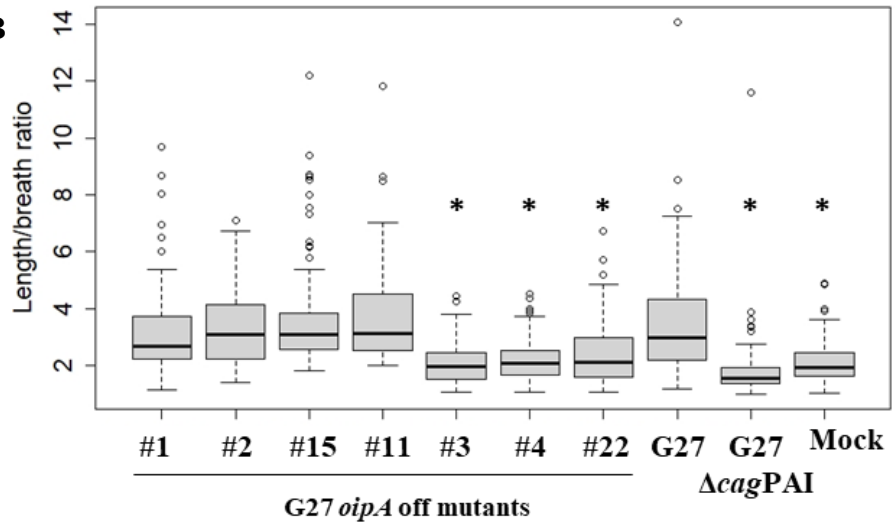


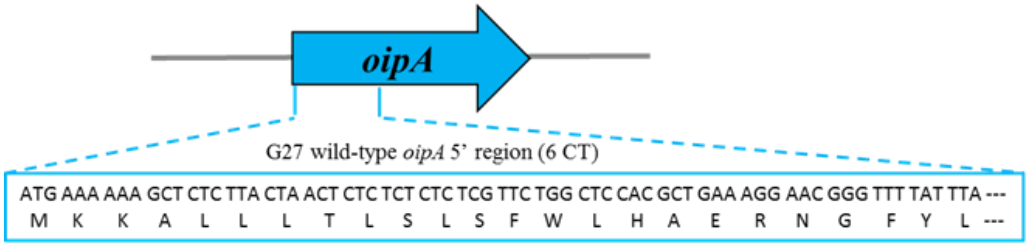
Figure 14. Morphology of AGS cells infected with G27 *H. pylori oipA* off isogenic mutants.

(A)The morphology of AGS cells was investigated (magnification 400×), the red boxes indicate the G27 *oipA* off mutants which have a significant decrease of cell elongation induction in AGS cells. (B) Cell elongation was calculated as the ratio of length to breadth of a cell. The cell elongation induced by each *H. pylori* strain was graphed using box plots. Thick center lines represent medians; box limits indicate the 25^{th} and 75^{th} percentiles; whiskers extend 1.5 times the interquartile range from the 25^{th} and 75^{th} percentiles, and outliers are represented by open-circle dots. n =100 for each group. *, p <0.05 (compared to the results of G27).

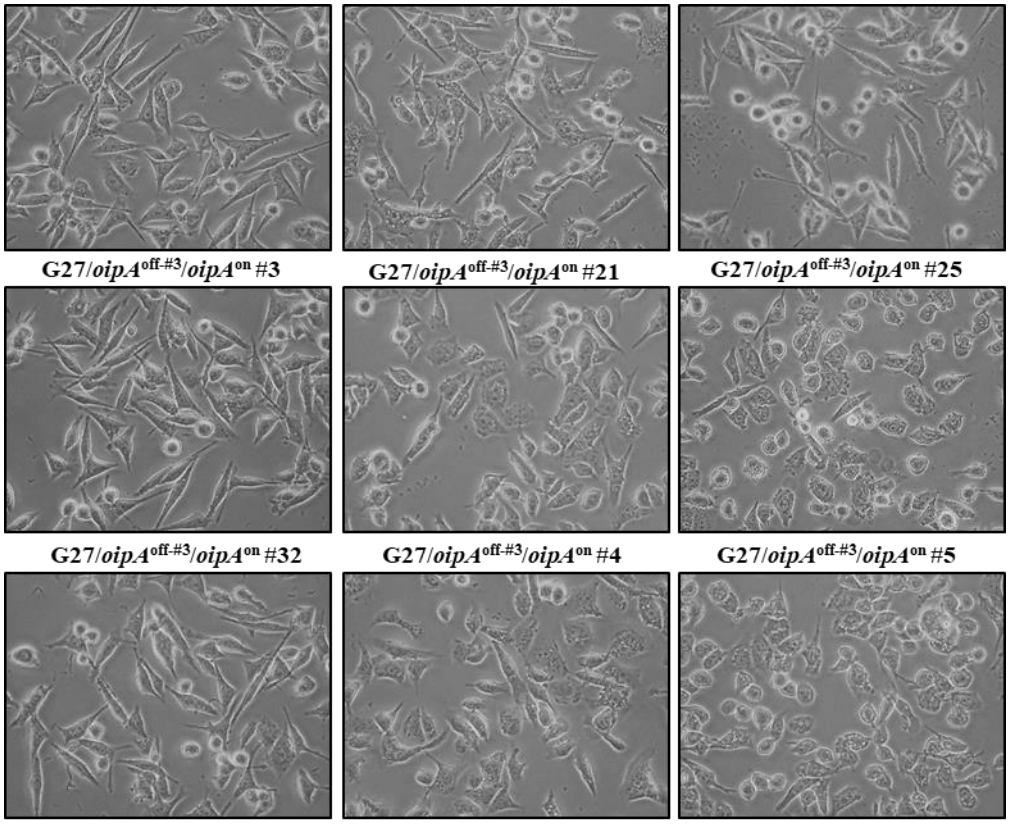
8. Complementation of G27 *oipA* off #3 with *oipA* "on" gene restored the capacity of cell elongation induction at a rate of 12.5%

To exclude the possibility that secondary mutagenesis or random mutagenesis other than *oipA* caused the cell elongation defect in G27 *oipA* off mutants, we did the complementation study. Firstly, the *oipA* in G27 *oipA* off #3 was replaced with *oipA* from G27 wild-type, which is *oipA* "on", resulting in G27 *oipA* off-#3 /*oipA*^{on} mutants. Four among 32 mutants, which accounts for 12.5% of the population, recovered the cell elongation phenotype in AGS cells (Figure 15).

A



B



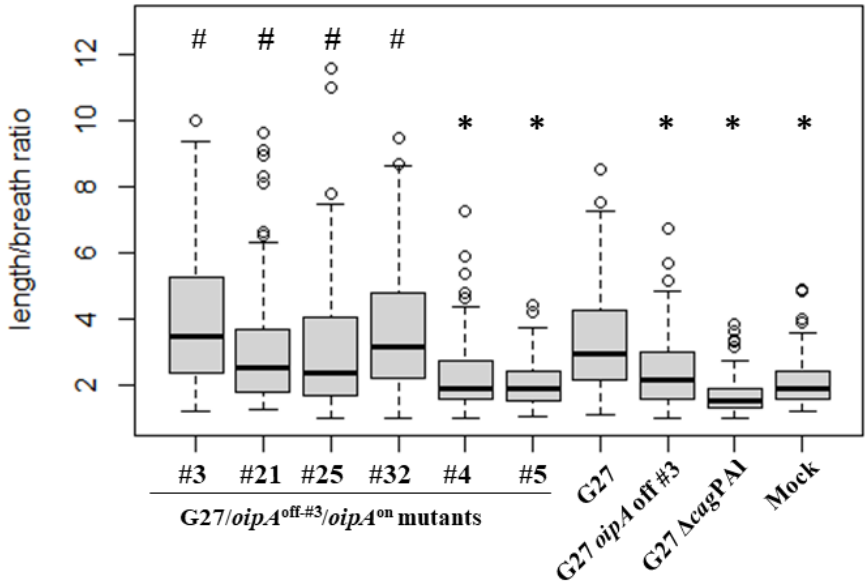


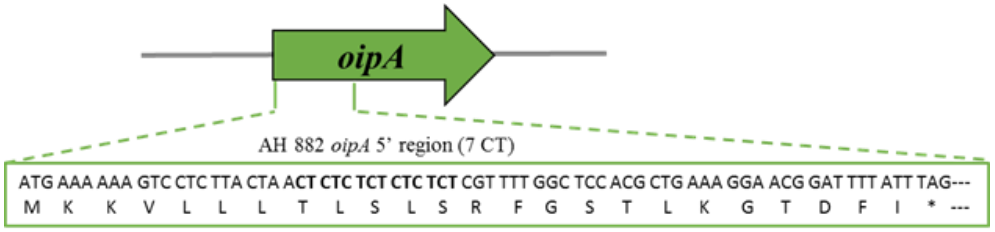
Figure 15. Morphology of AGS cells infected with G27 *oipA*$^{\text{off-\#3}}$/*oipA*$^{\text{on}}$ isogenic mutants.

(A) Schematic of G27 *oipA* off-#3/*oipA*$^{\text{on}}$. The *oipA* in G27 *oipA* off #3 was replaced with *oipA* from G27 wild-type, resulting in G27 *oipA* off-#3/*oipA*$^{\text{on}}$. (B) The morphology of AGS cells was investigated (magnification 400X). Cell elongation was calculated as the ratio of length to breadth of a cell. (C) The cell elongation induced by each *H. pylori* strain was graphed using box plots. n =100 for each group. *, $p < 0.05$ (compared to the results of G27 WT); #, $p < 0.05$ (compared to the result of G27 *oipA* off #3).

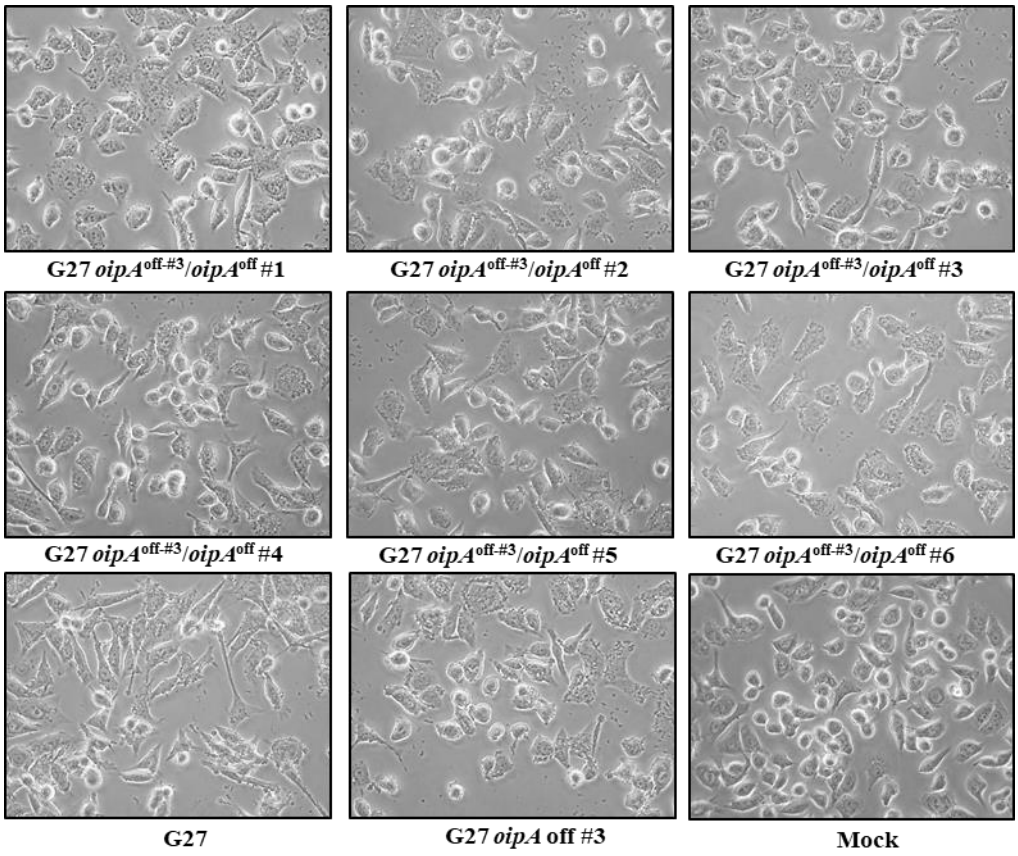
9. Complementation of G27 *oipA* off #3 with *oipA* "off" gene failed to restore the defective capacity of cell elongation induction

As the control, we replaced the *oipA* in G27 *oipA* off #3 with the *oipA* from AH882 which is *oipA* "off" status, resulting in G27 *oipA* off-#3/*oipA*^{off}. We picked 32 clones and none of them recovered the cell elongation phenotype in AGS cells by infection (Figure 16).

A



B



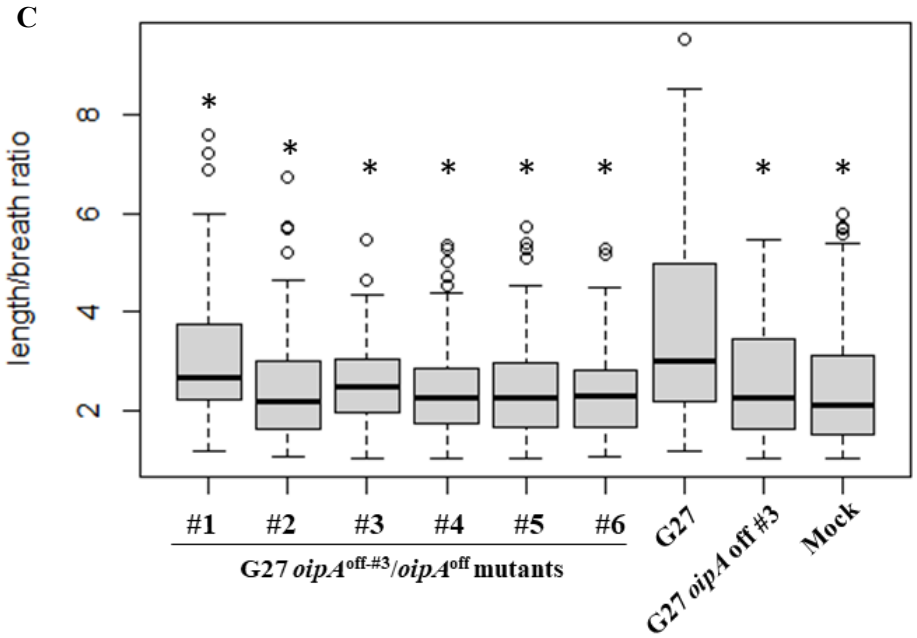


Figure 16. Morphology of AGS cells infected with G27 *oipA*$^{off\text{-}\#3}$/*oipA*^{off} complement isogenic mutants.

(A) Schematic of G27 *oipA*$^{off\text{-}\#3}$/*oipA*^{off}. The *oipA* in G27 *oipA* off #3 was replaced with *oipA* from AH882 wild-type which has an *oipA* "off" status, resulting G27 *oipA*$^{off\text{-}\#3}$/*oipA*^{off} (n = 32). (B) The morphology of AGS cells was investigated (magnification 400X). Cell elongation was calculated as the ratio of length to breadth of a cell. (C) The cell elongation induced by each *H. pylori* strain was graphed using box plots. n =100 for each group. *, $p < 0.05$ (compared to the results of G27 WT).

10. Variation of cell elongation induction abilities in G27 *oipA* off isogenic mutants were associated with different level of CagA phosphorylation

The cell elongation phenotype is strictly dependent on the delivery of CagA into host cells and its subsequent tyrosine phosphorylation by host kinases. Thus, for G27 *oipA* off mutants, lysates of AGS cells that were infected by G27 *oipA* off mutants with wild-type cell elongation induction (#1, #2, #5, #11) and G27 *oipA* off mutants with defective cell elongation induction capability (#3, #4, #22), for G27 *oipA* off #3/*oipA*^{on} mutants (Figure 17), lysates of AGS cells that were infected by G27 *oipA* off #3/*oipA*^{on} mutants with restored wild-type cell elongation induction (#3, #21, #25, #32) and G27 *oipA* off #3/*oipA*^{on} mutants with defective cell elongation induction capability (#4, #5) (Figure 18) were immunoblotted for phosphorylated CagA, total CagA, GAPDH and UreA. We found that the G27 *oipA* off isogenic mutants showed lower elongation with a relatively lower level of phosphorylated CagA.

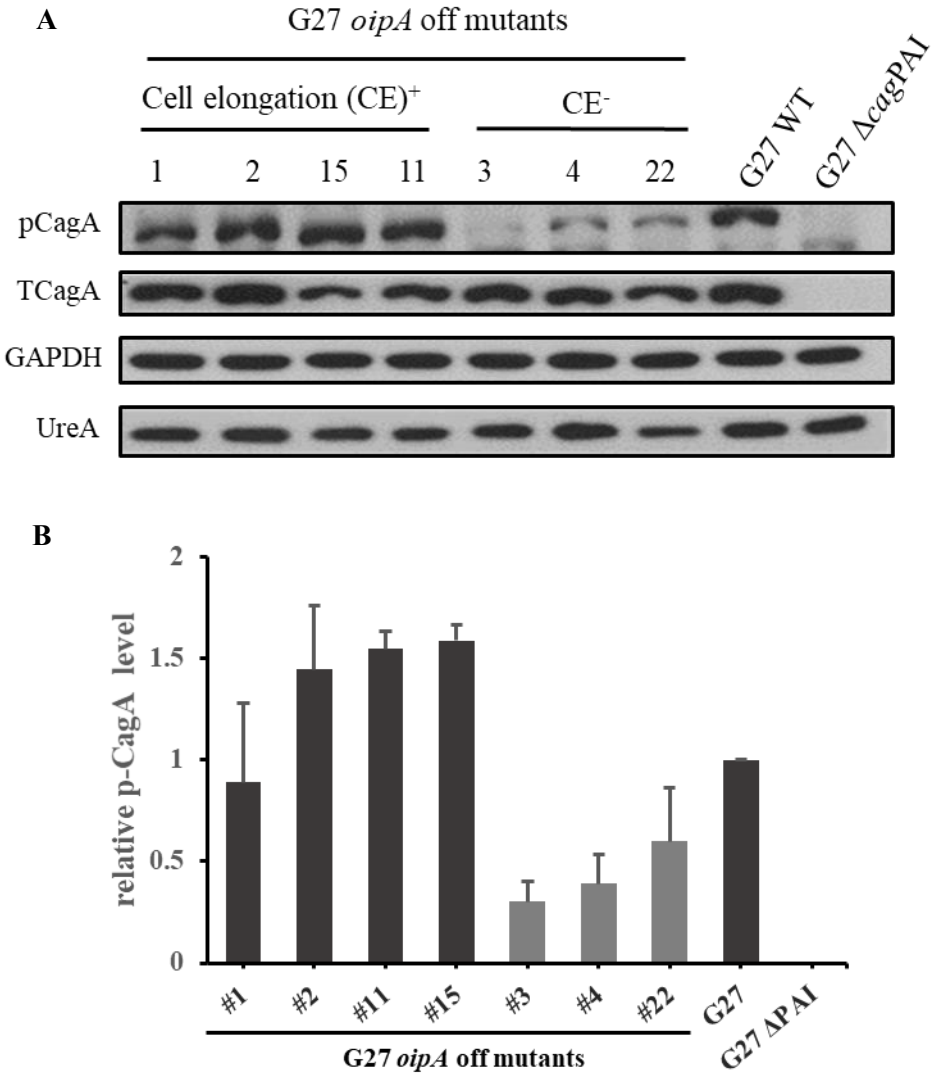


Figure 17. Relative levels of p-CagA protein in G27 *oipA* off mutants.

(A) Lysates of AGS cells that were infected by G27 *oipA* off mutants with or without cell elongation induction ability, G27 wild-type, and G27 Δ*cag*PAI were immunoblotted for phosphorylated CagA (p-CagA), Total CagA, glyceraldehyde-3-phosphate

dehydrogenase (GAPDH), and UreA. GAPDH and UreA were used as controls. (B) Ratios of p-CagA to CagA which normalized to the UreA intensity of each strain were calculated, and each value was normalized to the value calculated for G27 to determine relative CagA protein levels. The bar graphs indicate average levels of relative p-CagA expression of each strain, and error bars represent standard deviations, derived from the results of three independent experiments.

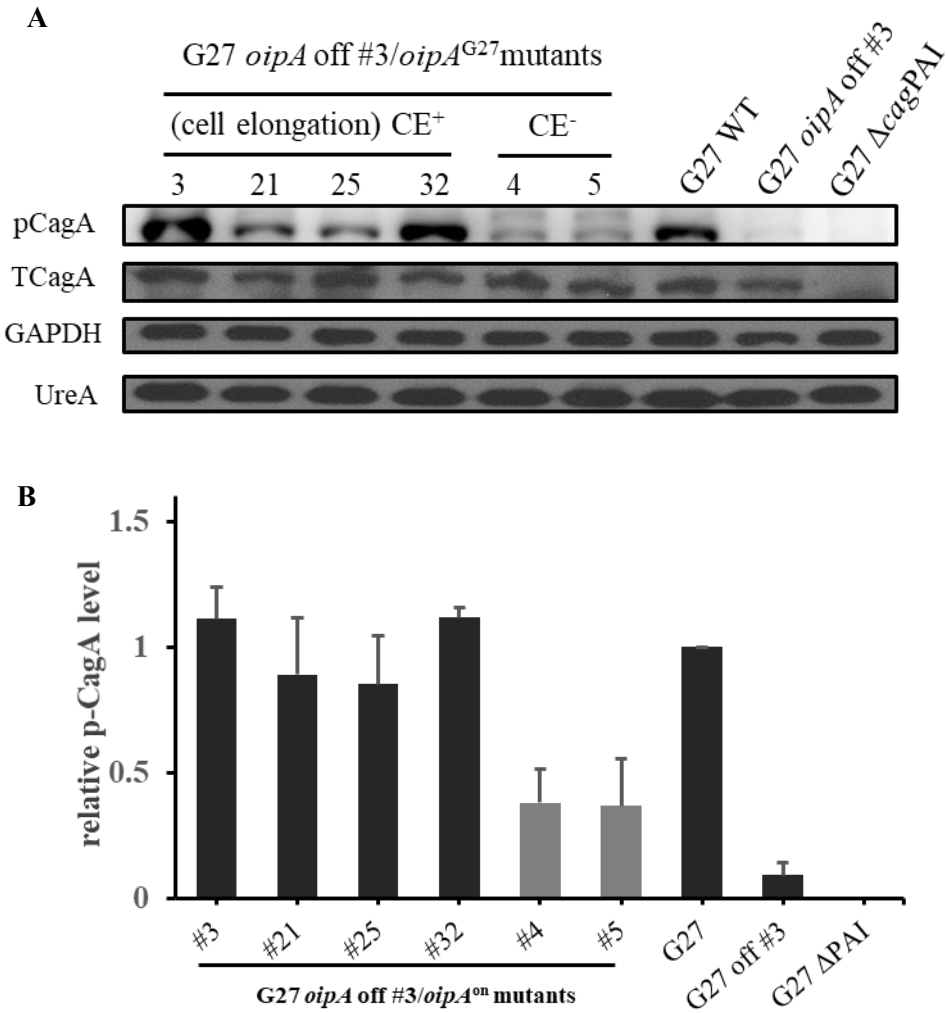


Figure 18. Relative levels of p-CagA protein in G27 *oipA*$^{\text{off-\#3}}$/*oipA*$^{\text{on}}$ mutants.

(A) Lysates of AGS cells that were infected by G27 *oipA* off-#3/*oipA*^{G27} mutants with or without cell elongation induction ability, G27 wild-type, and G27 Δ*cag*PAI were immunoblotted for phosphorylated CagA (p-CagA), Total CagA, GAPDH, and UreA.

GAPDH and UreA were used as controls. (B) Ratios of p-CagA to CagA which normalized to the UreA intensity of each strain were calculated, and each value was normalized to the value calculated for G27 to determine relative CagA protein levels. The bar graphs indicate average levels of relative p-CagA expression of each strain, and error bars represent standard deviations, derived from the results of three independent experiments.

11. Variation of cell elongation induction abilities in G27 *oipA* off isogenic mutants were not caused by the adherence ability defect

Since one of the reported *oipA* function is *H. pylori* adherence, next, we checked whether the variation of cell elongation in G27 *oipA* off isogenic mutants were caused by the different adherence ability. The result showed that all the G27 *oipA* off isogenic mutants with different cell elongation induction abilities showed a similar capacity of adherence (Figure 19), which indicates that the variation of cell elongation induction ability is not caused by adherence ability variance.

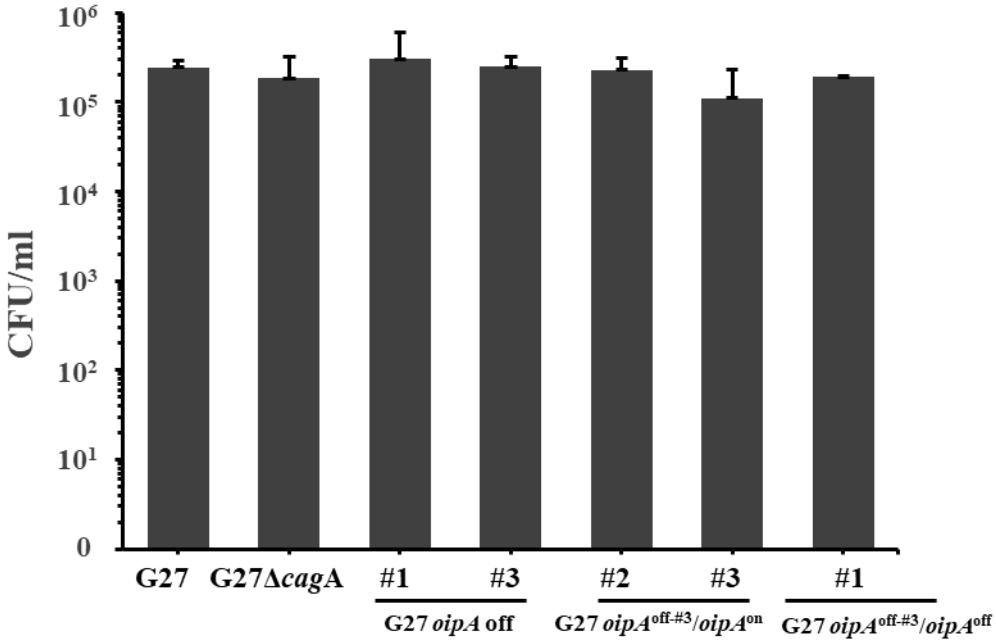


Figure 19. Bacterial adherence assay.

AGS cells were infected with indicated G27, G27 Δ*cagA*, and *oipA* isogenic mutants, and the number of attached bacteria was enumerated three minutes after initial attachment. Data represent biological duplicates for each strain. CFU: Colony-forming unit.

IV. DISCUSSION

As compared with other bacterial pathogens, *H. pylori* have a significantly higher genetic diversity which enables them to adapt to the tough environment of the human stomach and colonize on gastric epithelium even for the lifetime[68-70]. However, clinical demonstrations for chronic infection vary among populations[71]. For example, the gastric cancer incidence in East Asian countries such as Korea, Japan, and China is much higher than in Western countries[2, 72, 73]. The remarkably different rate of gastric cancer can, at least in part, be explainable by the genetic diversity of some virulent factors in *H. pylori*, as well as the variety in host factors and environmental influences such as diet and stress.

The initial colonization of *H. pylori* to the gastric epithelium is vital for the pathogenicity of *H. pylori*[74], and since OMPs play a critical role in the colonization, the genetic polymorphism of OMPs is proposed to be one of the virulence factors for developing gastric diseases. OipA, one of OMPs, has been drawn intensive attention due to its potential role in inflammation and phase variation. The phase variation was regulated by SSM mechanism based on the CT nucleotide repeat number in its 5' region. Additionally, the second *oipA* gene was found in some Eastern *H. pylori* strains, in our previous study, 231/233 (99.1%) of the Korean strains have two *oipA* genes. Thus, in the present study, we aimed to characterize the genetic variability of *oipA* gene in terms of status, gene number, and CT-repeat number in worldwide *H. pylori* populations, and its role in virulence phenotypes such as cell elongation and IL-8 induction in *H. pylori* strains.

The OipA consensus amino acid sequences of 19 US or 43 Korean *H. pylori* populations were highly conserved (Figure 3), and in the case of Korean strains, two *oipA* genes at different loci were also highly conserved (Figure 4), suggesting that *H. pylori* have been subjected to a strong selective force to conserve the sequence during evolution.

For the comparative genome analysis, the 613 world-wide *H. pylori* strains were grouped into seven population types by using MLST. It showed that the *oipA* copy number, *oipA* "on" and "off" status, or CT-repeat number were associated significantly among *H. pylori* populations.

For CT-repeat number, there is a significant difference among *H. pylori* populations. More strains in hpEastAsia population (154/248, 62.1%) possess two *oipA* genes in a statistically significant association versus 4/103 (3.9%) of hpAfrica1 (p=0.0000), 0/14 (0%) of hpAfrica2 (p=0.0001), 1/54 (1.5%) of hpAsia2 (p=0.0000) or 10/181(5.5%) of hpEurope (p=0.0000) strains with two *oipA* genes. The variation of gene copy number is reported to be a major driving force in evolution, changes in copy number may change the protein expression level encoded by the gene, and offers a selective advantage[75]. This mechanism also can be seen in some genes of *H. pylori*, for example, the dynamic expansion and contraction of *cagA* gene number was discovered in some *H. pylori* strains by our lab in 2017, the higher the *cagA* copy number resulted in the higher level of CagA expression, thus, the more virulence of *H. pylori*[60]. From the current study, the presence of two *oipA* gene copies are significantly associated with hpEastAsia population and based on our unpublished data, by using *oipA* isogenic mutants of Korean strain K74, Western blot

showed both of the *oipA* genes can be expressed. Additionally, wild-type K74 could express a higher level of OipA than its *oipA* isogenic mutant K74 Δ*oipA*1 or K74 Δ*oipA*2. Therefore, the duplication of *oipA* gene may be a mechanism *H. pylori* used to increase the expression of OipA in East-Asian *H. pylori* populations. It may be suggesting that OipA is indispensable and probably higher OipA expression is more beneficial to *H. pylori*.

For *oipA* status, generally, more than half of the *oipA* in the worldwide *H. pylori* genome showed to possess *oipA* "on" status. Interestingly, more *oipA* in hpEastAsia (376/402, 93.5%) possess "on" status with a significant association versus 118/181 (61.8%) of hpEurope *oipA* have "on" status (p = 0.0000). From our previous result, a significant association of Korean strains with *oipA* "on" status (457/462, 98.9%) was also found when compared with that in US (50/89, 56.2%). Therefore, East-Asia *H. pylori* have a higher possibility of possessing functional *oipA* genotype than Western *H. pylori* do.

The *oipA* CT-repeat number also revealed geographic characteristics. The *oipA* in hpEastAsia population was significantly more likely to have less than 3 CT-repeats (358/399,89.7%) when compared with hpAfrica1(1/105, 1%), hpAfrica2 (0/12, 0%), hpAsia2 (25/66, 37.9%) or hpEurope (6/189, 3.2%) (all p = 0.0000); moreover, significant associations of hpAsia2 with less than 3 CT repeats (25/66, 37.9%) were also found when compared with hpAfrica1 (1/105, 1%) or hpEurope (6/189, 3.2%) (both p = 0.0000). CT-repeat number-based SSM was reported to regulate the phase variation of *oipA*, the frequency of SSM is proved to be proportional to the dinucleotide-repeats number[76]. In the *H. pylori* genome, there are more than 260 locations with 3 CT repeats, thus, we can

reasonably infer that the less than 3 CT repeats in the East-Asian *oipA* might reduce the ability of phase variation, otherwise it will significantly impair the genome stability of *H. pylori*. However, as the previous data showed, most of Western *oipA* still possess more than 5 CT repeats in *oipA* which enables them to do the *oipA* phase variation more frequently than *H. pylori* in Eastern populations, suggesting that Western *H. pylori* need to adapt to a more diverse host.

To clear the controversy of the *oipA*'s role in *H. pylori*-related inflammation, we did several trials. The phenotypes of cell elongation and IL-8 induction level in host cells were used as determinants of *H. pylori* T4SS dependent pathogenesis ability.

Surprisingly, our result showed the association between *oipA* status with the cell elongation phenotype. To mimic the nature phase variation of *oipA* status by changing the CT-repeats number through the SSM mechanism, we used hpEurope *H. pylori* strain G27[77], which has been used extensively in *H. pylori* research, to made *oipA* off mutagenesis. Among 31 G27 *oipA* off mutants, three (9.6%) of them showed the defect in cell elongation induction ability, which was reversible by complementation of *oipA* to "on" also at a low rate (4/32, 12.5%). As a control, we complemented the G27 *oipA* off #3 mutant with the *oipA* "off" gene, 0/32 (0%) of G27 $oipA^{off\text{-}\#3}/oipA^{off}$ mutants restored the cell elongation ability. Thus, from the *oipA* complementation study, we could confirm that the phenotype changes of cell elongation among G27 *oipA* off as well as G27 $oipA^{off\text{-}\#3}/oipA^{on}$ mutants were caused by the variation of *oipA* status. Even though the phenotype change frequency by *oipA* off mutagenesis is low, it is worth noting that the phenotype was measured under

in vitro rich media culture conditions, and thus the frequency may be different under *in vivo* conditions. The exact mechanism of the low rate of cell elongation phenotype change among G27 *oipA* off mutants remains to be elucidated.

However, we could not observe the same phenotype change in AH868/CT[11] or K74 *oipA* isogenic mutants. Even though the discordances may be explained by the strain-specific of *oipA* function in *H. pylori*-induced cell elongation, we believe the lack of enough *oipA* off strains or mutants was probably the reason for the inconsistency. For AH868, even though it has 12 CT repeats in *oipA*, only 1/100 (1%) of the single colonies got the phase variation "on" to "off", which indicates that under *in vitro* conditions, the frequency of phase variation is low. Therefore, we could not have many *oipA* "off" colonies. The same problem was encountered for K74 *oipA* isogenic mutants, K74 is a Korean clinical isolate which found to have two *oipA* genes, we deleted the *oipA* genes in Chromosome with the method of nature transformation, only 1/18 (5.6%) of K74 *oipA* isogenic mutants was found to get *oipA* double deletion.

The Western blot analysis suggested that the defective cell elongation phenotype in G27 *oipA* off mutants as well as G27 $oipA^{off\text{-}\#3}$/$oipA^{on}$ mutants might be correlated with the decrease of CagA phosphorylation, indicating that the *oipA* phase variation is associated with cell elongation phenotype, probably in an indirect way. The association seems to be correlated with the capability of CagA translocation. Significant associations between functional *oipA* and *cagA* were found by some studies[38, 48, 54], besides, some groups reported the role of *oipA* in CagA translocation[77], however, in some other studies, this interaction of

oipA status and CagA translocation has not been found[78]. Since we did the *oipA* complementation study, our result may be helpful to explain the controversy.

Although some studies reported the linkage of *oipA* in *H. pylori* adherence[37, 79], in our study, the variation of cell elongation induction capability among G27 *oipA* off isogenic mutants is not caused by the differences in adherence ability, suggesting that *oipA*'s role in *H. pylori*-induced cell elongation is independent of adherence. Also, in some other studies, the mutagenesis of *oipA* in *H. pylori* didn't affect the adherence ability[36]. *H. pylori* are reported to have over 60 OMPs[35], several of them are served as adhesins which mediate the attachment of *H. pylori* to host cells, most of the well-characterized adhesins belongs to the Hop subfamily[35, 80]. Our results suggest that knockout of *oipA* may not substantially reduce the adherence ability of *H. pylori*, and multiple OMPs may cooperate in adhering to the epithelial cells.

The role of *oipA* in stimulating IL-8 expression of host cells when co-cultured with *H. pylori in vitro* has been studied by several groups. In agreement with some other groups, all the strains we used in this study didn't show the effect of *oipA* on IL-8 expression from gastric cells *in vitro*[37, 42, 55]. However, in one of the first studies of *oipA*, the ability of *oipA* to induce IL-8 expression from gastric epithelium cells was reported[36]. The discordances may be explained by the strain-specific of *oipA* function in *H. pylori*-induced IL-8 induction. To solve this controversy, it will be of interest to see if the *oipA* complementation in the *oipA* isogenic mutants will be able to restore the IL-8 induction ability.

Taken together, our data showed a large degree of diversity in *oipA* genotypes in regards to *oipA* gene number, CT-repeat number, and *oipA* status among worldwide *H. pylori* populations. Such diversity is suggested to be beneficial in the pathogenesis of *H. pylori* since it enables the rapid adaptation of bacterium in the diverse gastric environment in hosts. The hpEastAsia population tends to express more OipA by increasing the *oipA* gene number, and the *oipA* status was more likely to be fixed at "on" status by restricting the CT-repeat number when compared with hpEurope population. While the higher frequency of *oipA* phase variation in hpEurope population allows them to adapt to a more diverse host. Moreover, our results showed that the *oipA* phase variation is associated with cell elongation phenotype, probably in an indirect way and the association might be correlated with the CagA phosphorylation. Since the important role of CagA in the development of gastric cancer, it is tempting to speculate that these *oipA* genotype differences between hpEastAsia and hoEurope may contribute to the distinct gastric cancer rates between these two populations.

V. REFERENCES

1. Ferlay, J., et al., *Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012.* International journal of cancer, 2015. **136**(5): p. E359-E386.
2. Sung, H., 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): p. 209-249.
3. Forman, D., et al., *EPIDEMIOLOGY OF, AND RISK-FACTORS FOR, HELICOBACTER-PYLORI INFECTION AMONG 3194 ASYMPTOMATIC SUBJECTS IN 17 POPULATIONS.* Gut, 1993. **34**(12): p. 1672-1676.
4. Yamaoka, Y., M. Kato, and M.J.I.m. Asaka, *Geographic differences in gastric cancer incidence can be explained by differences between Helicobacter pylori strains.* 2008. **47**(12): p. 1077-1083.
5. Marshall, B.J. and J.R. Warren, *UNIDENTIFIED CURVED BACILLI IN THE STOMACH OF PATIENTS WITH GASTRITIS AND PEPTIC-ULCERATION.* Lancet, 1984. **1**(8390): p. 1311-1315.
6. Moss, S.F., *The clinical evidence linking Helicobacter pylori to gastric cancer.* Cellular and molecular gastroenterology and hepatology, 2017. **3**(2): p. 183-191.
7. Group, T.E.S., *Epidemiology of, and risk factors for, Helicobacter pylori infection among 3194 asymptomatic subjects in 17*

populations. Gut, 1993. **34**: p. 1672–1676.

8. Lancet, E.S.G.J.T., *An international association between Helicobacter pylori infection and gastric cancer.* 1993. **341**(8857): p. 1359-1363.
9. Blaser, M.J., *Helicobacter pylori and gastric diseases.* Bmj, 1998. **316**(7143): p. 1507-1510.
10. IARC, *Schistosomes, liver flukes and Helicobacter pylori. IARC Working Group on the Evaluation of Carcinogenic Risks to Humans. Lyon, 7-14 June 1994.* IARC Monogr Eval Carcinog Risks Hum, 1994. **61**: p. 1-241.
11. Beswick, E.J., G. Suarez, and V.E. Reyes, *H pylori and host interactions that influence pathogenesis.* World journal of gastroenterology: WJG, 2006. **12**(35): p. 5599.
12. Censini, S., 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, 1996. **93**(25): p. 14648-14653.
13. Segal, E.D., et al., *Induction of host signal transduction pathways by Helicobacter pylori.* Proceedings of the National Academy of Sciences, 1997. **94**(14): p. 7595-7599.
14. Backert, S., et al., *Functional analysis of the cag pathogenicity island in Helicobacter pylori isolates from patients with gastritis, peptic ulcer, and gastric cancer.* Infection and immunity, 2004. **72**(2): p. 1043.

15. Johnson, E.M., et al., *Genes required for assembly of pili associated with the Helicobacter pylori cag type IV secretion system.* Infection and immunity, 2014. **82**(8): p. 3457.

16. Ta, L.H., et al., *Conserved transcriptional unit organization of the cag pathogenicity island among Helicobacter pylori strains.* Frontiers in cellular and infection microbiology, 2012. **2**: p. 46.

17. Kutter, S., et al., *Protein subassemblies of the Helicobacter pylori Cag type IV secretion system revealed by localization and interaction studies.* Journal of bacteriology, 2008. **190**(6): p. 2161.

18. Viala, J., et al., *Nod1 responds to peptidoglycan delivered by the Helicobacter pylori cag pathogenicity island.* Nature immunology, 2004. **5**(11): p. 1166-1174.

19. Odenbreit, S., et al., *Translocation of Helicobacter pylori CagA into gastric epithelial cells by type IV secretion.* Science, 2000. **287**(5457): p. 1497-1500.

20. Backert, S., N. Tegtmeyer, and W. Fischer, *Composition, structure and function of the Helicobacter pylori cag pathogenicity island encoded type IV secretion system.* Future microbiology, 2015. **10**(6): p. 955-965.

21. Frick-Cheng, A.E., et al., *Molecular and structural analysis of the Helicobacter pylori cag type IV secretion system core complex.* MBio, 2016. **7**(1).

22. Chang, Y.-W., et al., *In vivo structures of the Helicobacter pylori cag type IV secretion system.* Cell reports, 2018. **23**(3): p. 673-681.

23. Backert, S., et al., *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): p. 631-644.

24. Ren, S., et al., *Structural basis and functional consequence of Helicobacter pylori CagA multimerization in cells.* Journal of Biological Chemistry, 2006. **281**(43): p. 32344-32352.

25. Higashi, H., 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, 2002. **99**(22): p. 14428-14433.

26. Stein, M., R. Rappuoli, and A. Covacci, *Tyrosine phosphorylation of the Helicobacter pylori CagA antigen after cag-driven host cell translocation.* Proceedings of the National Academy of Sciences, 2000. **97**(3): p. 1263-1268.

27. Jones, K.R., et al., *Polymorphism in the CagA EPIYA motif impacts development of gastric cancer.* Journal of clinical microbiology, 2009. **47**(4): p. 959.

28. Saadat, I., et al., *Helicobacter pylori CagA targets PAR1/MARK kinase to disrupt epithelial cell polarity.* Nature, 2007. **447**(7142): p. 330-333.

29. Selbach, M., et al., *Src is the kinase of the Helicobacter pylori CagA protein in vitro and in vivo.* Journal of Biological Chemistry, 2002. **277**(9): p. 6775-6778.

30. Segal, E., et al., *Altered states: involvement of phosphorylated CagA in the induction of host cellular growth changes by Helicobacter pylori.* Proceedings of the National Academy of Sciences, 1999. **96**(25): p. 14559-14564.

31. Higashi, H., et al., *SHP-2 tyrosine phosphatase as an intracellular target of Helicobacter pylori CagA protein.* Science, 2002. **295**(5555): p. 683-686.

32. Yamaoka, Y., et al., *Helicobacter pylori outer membrane proteins and gastroduodenal disease.* Gut, 2006. **55**(6): p. 775-781.

33. Matsuo, Y., Y. Kido, and Y. Yamaoka, *Helicobacter pylori Outer Membrane Protein-Related Pathogenesis.* Toxins (Basel), 2017. **9**(3).

34. Cover, T.L., *Role of Helicobacter pylori outer membrane proteins in gastroduodenal disease*. 2006, The University of Chicago Press.

35. Alm, R.A., et al., *Comparative genomics of Helicobacter pylori: analysis of the outer membrane protein families.* Infection and immunity, 2000. **68**(7): p. 4155.

36. Yamaoka, Y., D.H. Kwon, and D.Y. Graham, *A Mr 34,000 proinflammatory outer membrane protein (oipA) of Helicobacter pylori.* Proceedings of the National Academy of Sciences, 2000. **97**(13): p. 7533-7538.

37. Dossumbekova, A., et al., *Helicobacter pylori HopH (OipA) and bacterial pathogenicity: genetic and functional genomic analysis of hopH gene polymorphisms.* The Journal of infectious diseases, 2006. **194**(10): p. 1346-1355.

38. Farzi, N., et al., *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.* Infect Genet Evol, 2018. **60**: p. 26-34.

39. Ando, T., et al., *Polymorphisms of Helicobacter pylori HP0638 reflect geographic origin and correlate with cagA status.* J Clin Microbiol, 2002. **40**(1): p. 239-46.

40. Furuta, Y., et al., *Birth and death of genes linked to chromosomal inversion.* Proc Natl Acad Sci U S A, 2011. **108**(4): p. 1501-6.

41. Kawai, M., et al., *Evolution in an oncogenic bacterial species with extreme genome plasticity: Helicobacter pylori East Asian genomes.* BMC Microbiology, 2011. **11**(1).

42. Akanuma, M., et al., *The evaluation of putative virulence factors of helicobacter pylori for gastroduodenal disease b Use of a short-term Mongolian gerbil infection model.* 2002. **185**(3): p. 341-347.

43. Yamaoka, Y., et al., *Helicobacter pylori infection in mice: role of outer membrane proteins in colonization and inflammation.* Gastroenterology, 2002. **123**(6): p. 1992-2004.

44. Keates, S., et al., *Helicobacter pylori infection activates NF-kappa B in gastric epithelial cells.* Gastroenterology, 1997. **113**(4): p. 1099-1109.

45. Crabtree, J., 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): p. 41-45.

46. Lee, K.E., et al., *Helicobacter pylori and interleukin-8 in gastric cancer.* World Journal of Gastroenterology: WJG, 2013. **19**(45): p. 8192.

47. Sokolova, O., et al., *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): p. 548-552.

48. Yamaoka, Y., et al., *Importance of Helicobacter pylori oipA in clinical presentation, gastric inflammation, and mucosal interleukin 8 production.* Gastroenterology, 2002. **123**(2): p. 414-24.

49. Franco, A.T., et al., *Regulation of gastric carcinogenesis by Helicobacter pylori virulence factors.* Cancer research, 2008. **68**(2): p. 379-387.

50. Graham, D., et al., *Challenge model for Helicobacter pylori infection in human volunteers.* Gut, 2004. **53**(9): p. 1235-1243.

51. Liu, J., et al., *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): p. 1-10.

52. Sallas, M.L., et al., *Status (on/off) of oipA gene: their associations with gastritis and gastric cancer and geographic origins.* Arch Microbiol, 2019. **201**(1): p. 93-97.

53. Souod, N., 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**(Suppl1): p. S47.

54. Yamaoka, Y., et al., *Role of interferon-stimulated responsive element-like element in interleukin-8 promoter in Helicobacter pylori infection.* Gastroenterology, 2004. **126**(4): p. 1030-43.

55. Ando, T., et al., *Host cell responses to genotypically similar Helicobacter pylori isolates from United States and Japan.* Clin Diagn Lab Immunol, 2002. **9**(1): p. 167-75.

56. Achtman, M., et al., *Recombination and clonal groupings within Helicobacter pylori from different geographical regions.* Molecular Microbiology, 1999. **32**(3): p. 459-470.

57. Falush, D., M. Stephens, and J.K. Pritchard, *Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies.* Genetics, 2003. **164**(4): p. 1567-1587.

58. Maiden, M.C., 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, 1998. **95**(6): p. 3140-3145.

59. Oleastro, M., R. Rocha, and F.F. Vale, *Population genetic structure of Helicobacter pylori strains from Portuguese-speaking countries.* Helicobacter, 2017. **22**(4): p. 10.

60. Jang, S., et al., *Dynamic expansion and contraction of cagA copy number in Helicobacter pylori impact development of gastric disease.* MBio, 2017. **8**(1).

61. Bourzac, K.M., L.A. Satkamp, and K. Guillemin, *The Helicobacter pylori cag pathogenicity island protein CagN is a bacterial membrane-associated protein that is processed at its C terminus.* Infection and immunity, 2006. **74**(5): p. 2537.

62. Fung, C., et al., *High-resolution mapping reveals that microniches in the gastric glands control Helicobacter pylori colonization of the stomach.* PLoS biology, 2019. **17**(5): p. e3000231.

63. Taylor, S.C., et al., *A defined methodology for reliable quantification of Western blot data.* Molecular biotechnology, 2013. **55**(3): p. 217-226.

64. Crooks, G.E., et al., *WebLogo: a sequence logo generator.* Genome research, 2004. **14**(6): p. 1188-1190.

65. Yamaoka, Y., *Helicobacter pylori typing as a tool for tracking human migration.* Clinical Microbiology and Infection, 2009. **15**(9): p. 829-834.

66. Eckert, K.A. and G. Yan, *Mutational analyses of dinucleotide and tetranucleotide microsatellites in Escherichia coli: influence of sequence on expansion mutagenesis.* Nucleic acids research, 2000. **28**(14): p. 2831-2838.

67. Kudo, T., et al., *Correlation between Helicobacter pylori OipA protein expression and oipA gene switch status.* Journal of clinical microbiology, 2004. **42**(5): p. 2279-2281.

68. Kusters, J.G., A.H. Van Vliet, and E.J. Kuipers, *Pathogenesis of Helicobacter pylori infection.* Clinical microbiology reviews, 2006. **19**(3): p. 449.

69. Blaser, M.J. and J.C. Atherton, *Helicobacter pylori persistence: biology and disease.* The Journal of clinical investigation, 2004. **113**(3): p. 321-333.

70. Atherton, J.C. and M.J. Blaser, *Coadaptation of Helicobacter pylori and humans: ancient history, modern implications.* The Journal of clinical investigation, 2009. **119**(9): p. 2475-2487.

71. Amieva, M.R. and E.M. El–Omar, *Host-bacterial interactions in Helicobacter pylori infection.* Gastroenterology, 2008. **134**(1): p. 306-323.

72. Fock, K.M. and T.L. Ang, *Epidemiology of Helicobacter pylori infection and gastric cancer in Asia.* Journal of gastroenterology and hepatology, 2010. **25**(3): p. 479-486.

73. Miwa, H., M.F. Go, and N. Sato, *H. pylori and gastric cancer: the Asian enigma.* The American journal of gastroenterology, 2002. **97**(5): p. 1106-1112.

74. Guruge, J.L., et al., *Epithelial attachment alters the outcome of Helicobacter pylori infection.* Proceedings of the National Academy of Sciences, 1998. **95**(7): p. 3925-3930.

75. Hastings, P.J., et al., *Mechanisms of change in gene copy number.* Nature Reviews Genetics, 2009. **10**(8): p. 551-564.

76. Vogler, A.J., et al., *Effect of repeat copy number on variable-number tandem repeat mutations in Escherichia coli O157: H7.* Journal of bacteriology, 2006. **188**(12): p. 4253.

77. Xiang, Z., 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): p. 94-98.

78. Odenbreit, S., et al., *CagA tyrosine phosphorylation and interleukin-8 induction by Helicobacter pylori are independent from alpAB, HopZ and bab group outer membrane proteins.* Int J Med Microbiol, 2002. **292**(3-4): p. 257-66.

79. Zhao, Q., et al., *Outer inflammatory protein of Helicobacter pylori impacts IL-8 expression, adherence, cell apoptosis and cell cycle of gastric cells independent of its copy number.* Medical Microbiology and Immunology, 2020. **209**(5): p. 621-630.

80. Zanotti, G. and L. Cendron, *Structural and functional aspects of the Helicobacter pylori secretome.* World Journal of Gastroenterology: WJG, 2014. **20**(6): p. 1402.

국문요약

*Helicobacter pylori*의 IV 형 분비 시스템에서 *oipA*의 역할

지도교수 차정헌

연세대학교 대학원 응용생명과학과

Jing Lai

헬리코박터 파일로리(*H. pylori*)는 그람 음성 병원성 세균으로, 위암을 비롯한 위 질환을 유발하는 주요 원인으로 알려져 있다. 외막 염증 단백질 A (OipA)는 *H. pylori* 의 독성 인자로써 *oipA* 유전자에 의해 발현되고, *H. pylori*가 질병을 일으키는 기전 중 하나로 보고되었다.

OipA 의 발현은 *oipA* 유전자의 5’영역에서 사이토신-티민(CT) 디뉴클레오타이드 반복 수를 기반으로 한 SSM (slipped strand mispairing) 매개의 위상 변이에 의해 조절; “on” 위상 변이는 기능적인 OipA의 전체 길이(통상적으로 305 아미노산)를 보여주며, “off” 위상 변이는 일부가 생략된 비기능적인 OipA (약 24개의 아미노산)를 발현하도록 OipA 유전자 상의 오픈 리딩 프레임을 이동 시킨다.

본 연구실의 이전에 연구한 미공개 자료에 의하면, 233개의 한국인 임상 분리균주와 89개의 미국인 임상 분리균주에서 *H. pylori* 집단간의 유전자 복제 수, 위상 변이 및 CT 반복 수와 같은 다양한 *oipA* 유전자형이 있음을 발견했다. 이와 더불어, 발병 기전에서 *oipA* 유전자형의 역할은 아직 연구되어 있지 않다.

본 연구의 목적은 613개의 *H. pylori* 게놈 정보를 바탕으로 비교 게놈 분석을 통해 전 세계 인구에서 *oipA* 유전자형을 확인하고 또한, 세포 신장 및 인터루킨-8 유도의 독성 분석을 통해 *oipA* 기능을 분석하고자 한다. 한국과

미국의 *H. pylori* 집단의 OipA 공통 아미노산 서열은 43 개의 한국 균주와 19 개의 미국 균주의 시퀀싱을 통해 결정하였으며, 두 집단에서 고도로 보존된 *oipA* 유전자를 확인하였다. 비교 게놈 분석을 위해 전 세계 *H. pylori* 균주를 7 개의 집단 유형으로 분류하였고, *oipA* 복제 수, *oipA* "on" "off" 그리고 CT 반복 수가 *H. pylori* 집단 간에 유의한 상관관계가 있음을 보여주었다. 놀랍게도 hpEastAsia 집단에서 상대적으로 더 많은 균주 (154/248, 62.1%)가 통계적으로 유의한 연관성으로 두개의 *oipA* 유전자를 보유하고 있었으며 반면, hpAfrica1 집단 (p = 0.000)에서는 4/103 (3.9%), hpAfrica2 집단 (p = 0.0001)에서는 0/14 (0%), hpAsia2 집단 (p = 0.0000)에서는 1/54 (1.5%), hpEurope 집단 (p = 0.0000)에서는 10/181(5.5%)만이 두개의 *oipA* 유전자를 보유하고 있었다. CT 반복 수의 3 회 미만 반복에서, hpEastAsia(358/399, 89.7%) *oipA*와 다른 집단 (hpAfrica1, 1/105, 1%), (hpAfrica2, 0/12, 0%), (hpAsia2, 25/66, 37.9%), (hpEurope, 6/189, 3.2%) (모두 p=0.0000)을 비교하였을 때 유의한 상관관계를 보였다; 또한, hpAsia2 의 CT 반복 수의 3 회 미만 반복 (25/66, 37.9%)과 hpAfrica1(1/105, 1%), hpEurope(6/189, 3.2%) (둘다 p = 0.0000)를 비교하였을 때, 유의한 상관관계가 있음을 발견하였다. 우리는 *H. pylori* 집단 사이의 *oipA* 유전자 수, CT 반복 수 그리고 *oipA* 위상 변이 상태에서 큰 다양성을 발견했다. 이 다양성은 아마도 *H. pylori*의 다양한 숙주에 대한

적응력에 도움을 주는데, 왜냐하면 많은 유전자들이 외막단백질에 속하고 유전자 발현의 매우 역동적인 변화에 의한 적응에 관여할 것이다.

다음으로 자연적으로 발생한 *oipA* “on” “off” 위상 변이와 유전적으로 조작된 *oipA*를 가진 *H. pylori*를 독성 관점에서 *oipA*의 특징적인 역할을 확인했다. 우리는 US AH868 *H. pylori*에서 자연적으로 발생한 11 개의 CT 반복 *oipA* "off" 위상 변이 집락과 12 개의 CT 반복 *oipA* “on” 위상 변이 집락을 확인하였고, 세포 신장 및 인터루킨-8 (IL-8) 분석에서 변화가 없음을 확인하였다. 한국 균주 K74 의 경우, Δ*oipA* 동질유전자 돌연변이 18 개의 형질 전환체를 선별하여, 다른 유전자 좌위에서 Δ*oipA*1/Δ*oipA*2 의 이중 결실을 확인하였다. 이 중에서 단 하나만이 이중 Δ*oipA* 형질 전환체 K74Δ*oipA*1/Δ*oipA*2 였고 나머지 17 개는 K74Δ*oipA*1/*oipA*2 또는 *oipA*1 / Δ*oipA*2 가 생성되었으며, 18 개의 형질 전환체 모두 K74 야생형과 유사한 세포 신장과 IL-8 분비 정도를 나타냈다. hpEurope 균주 G27 에서 유전자 조작에 의한 3 개의 Δ*oipA* 형질 전환체를 선별하고, 세포 신장과 IL-8 분비 정도를 확인하였다. 예기치 않게 하나의 G27Δ*oipA* 형질 전환체에서 세포 신장과 IL-8 분비가 결손된 독성 표현형을 보인 반면, 2 개의 G27Δ*oipA* 형질 전환체에서는 야생형과 유사한 수준의 세포 신장과 IL-8 분비를 보였다. 그러나 G27 *oipA* 와 G27Δ*oipA* 상보성 균주는 세포 신장과 IL-8 분비 유도 능력을 회복하지 못하였고, 이는 결손된 독성 표현형이 Δ*oipA* 이외의 2 차

돌연변이에 의해 발생하였음을 시사한다. G27 의 결과는 세포 신장 및 IL-8 분비의 일관되지 않은 Δ*oipA* 표현형을 보여 주었고, 자연적인 *oipA* “off” 위상 변이를 모방하였기 때문에 독성 표현형에서 *oipA* 위상 변이의 역할을 조사하기 위해 다른 접근법을 사용하였다. 우리는 상동재조합을 통해 $oipA^{off}$ 구조를 변형시킴으로써 G27 *H. pylori* 에서 영구적인 *oipA* “off” 위상 변이 형질 전환체를 제작하였다. 이전의 *oipA* 분석에 대한 성향을 회피하기 위해 많은 수의 $G27^{off}$ 형질 전환체를 선별하였고, 이에 대한 독성 표현형을 확인하였다.

31 개의 G27 $oipA^{off}$ 형질 전환체 중 3 개 (9.6%)에서 야생형에 비해 세포 신장이 유의한 감소를 보였지만, IL-8 분비는 유사한 수준이었다. 나머지 G27off 형질 전환체 (28/31, 90.4%)는 야생형과 유사한 수준의 세포 신장과 IL-8 분비 수준을 나타냈다. 형질 전환체의 9.6%가 세포 신장 유도에 결함을 보였고, 이에 예상치 못한 2 차 돌연변이가 결함의 원인이라 판단하였다. 따라서, 결함이 *oipA* “off” 위상 변이로 인한 것인지 확인하기 위해 G27 $oipA^{off\#3}$ 를 G27 야생형 *oipA* ”on”으로 보완하여 G27 $oipA^{off\#3}/oipA^{on}$ 을 제작하였다. 다수의 형질 전환체를 분석한 결과, 32 개의 형질 전환체 중 4 개 (12.5%)는 세포 신장 유도 능력이 회복되었고, 나머지 87.5%의 형질 전환체는 세포 신장 감소를 유지하는 양상을 나타내었다. 대조군으로써 G27 $oipA^{off}$#3 의 *oipA* “off” 구조체는 *kan-sac*B 카세트 없이 *oipA* “off”로 대체되었으며,

32 개의 형질 전환체 중 회복된 균주는 없었다. 추가로 시행된 Western blot 분석을 통해 *oipA* 위상 변이와 세포 신장 표현형 사이의 연관성이 CagA 인산화와 관련이 있고 따라서, *oipA* 위상 변이가 아마도 간접적인 방식으로 세포 신장의 독성 표현형과 관련이 있음을 강력히 시사한다.

*oipA*는 위암 발병에 중요한 역할을 하는 차등적인 CagA 인산화에 의한 세포 신장의 독성 표현형과 관련이 있기 때문에, 집단 사이에서 상이한 *oipA* 유전자형이 위암 비율에 기여할 수 있다고 추측한다.

핵심되는 말 : 헬리코박터 파일로리, *oipA*, 위상 변이, 병인