The organization of Helicobacter pylori cag-pathogenicity island (cagPAI) genes in multiracial population with histopathological changes of gastric mucosa

Alfizah Hanafiah (alfizah@ppukm.ukm.edu.my)
Universiti Kebangsaan Malaysia Fakulti Perubatan https://orcid.org/0000-0003-4859-1439
Shaza Azlin Razak
University Kebangsaan Malaysia
Hui-min Neoh
Universiti Kebangsaan Malaysia
Noraziah Mohamad Zin
Universiti Kebangsaan Malaysia
Bruno S. Lopes
University of Aberdeen

Research article

Keywords: Helicobacter pylori, cagPAI, histopathological scores, multi-ethnic population

Posted Date: May 14th, 2019

DOI: https://doi.org/10.21203/rs.2.9601/v1

Abstract

Background: Helicobacter pylori is a Gram-negative bacillus that colonises only the mucus layer of the human stomach and is implicated in gastric diseases. Virulent H. pylori harbouring cag-pathogenicity island (cagPAI) which encodes genes for type IV secretion system (T4SS) and CagA protein is one of the major virulence determinants involved in disease development. We examined the entire cagPAI genes in 95 H. pylori isolates from a multiracial population and examined the intactness of cagPAI region with histopathological scores of the gastric mucosa. Results: 95.8% of H. pylori isolates were cagPAI-positive with 23.2% having an intact cagPAI, whereas 72.6% had a partial/rearranged cagPAI. In our study, cag2 and cag4 were found to be significantly higher in H. pylori isolated from Malays, whereas cag4 was predominant in Chinese isolates. We also detected cag24 in significantly high proportion in isolates from the Malays and the Indians compared to the Chinese isolates. The intactness of cagPAI region showed an association with histopathological scores of the gastric mucosa. Significant association was observed between H. pylori harbouring partial cagPAI and higher density of H. pylori and neutrophil activity, whereas strains which lacked cagPAI was associated with higher inflammatory score. Conclusions: The screening of the entire cagPAI genes provides an accurate overview of the cagPAI organisation in H. pylori isolates in a multiracial population. The genotypes of H. pylori strains with various cagPAI rearrangement associated with patients' ethnicities and histopathological scores might contribute to the pathogenesis of H. pylori infection in a multi-ethnic population.

Background

Helicobacter pylori is a Gram-negative, microaerophilic, curved-shaped and flagellated bacterium frequently found in the stomach of humans [1]. It is an important pathogen that causes gastrointestinal diseases such as chronic gastritis, peptic ulcer, gastric cancer and gastric mucosa-associated lymphoid tissue (MALT) lymphoma [2,3], although most infected patients appeared asymptomatic. Hence, *H. pylori* is also classed as type I carcinogen [4]. Factors that contribute to the infected patient's disease sequelae include environmental factors such as lifestyle and diet , host genetics, host immune responses and bacterial virulence factors [4-6].

Cytotoxin-associated gene pathogenicity island (*cag*PAI) is one of the major virulence factors associated with disease outcome in infected hosts. It is approximately 40 kb in size consisting of around 28 genes [7], encoding mainly CagA protein, type IV secretion system (T4SS) and other genes for induction of host's interleukin-8 (IL-8) [7,8]. Although the mechanisms resulting in severe disease development are poorly understood, a major factor is likely to be *H. pylori*-induced gastric injury and inflammation [9]. Studies show that intactness of *cag*PAI has a significant correlation with disease severity, whereas *H. pylori* strains with partial deletions within *cag*PAI region are significantly less-pathogenic in nature [10,11]. However, the rates of severe disease development vary between human populations, and differences in *H. pylori* genotypes may partially explained these differences [12,13].

Integrity of *cag*PAI seems to have an important role in the progress of the gastroduodenal disorders, so that intact *cag*PAI could be seen in *H. pylori* strains from countries with higher rate of gastric cancer [14]. This integrity also has important effect on the induction of inflammatory response in the gastric mucosa [15]. Several studies have investigated the association of *H. pylori cag*PAI and gastroduodenal diseases [14,16], however, knowledge about the relationship between *H. pylori cag*PAI intactness and changes of the infected gastric tissue is sparse. More than 90% of *H. pylori* strains in Malaysia are *cag*PAI-positive [17] and Malaysian population consists of multi-ethnic people, therefore the interaction of *H. pylori* strains with different genotype in various host genetics may have an impact on the differences in disease development.

The organisation of *cag*PAI genes in *H. pylori* in Malaysian population which has multi-ethnic groups of people has not been well studied. There is lack of comprehensive information with regards to abundance of intact versus rearranged *cag*PAI among *H. pylori* strains in this population. Hence, in this study, we sought to characterise the genes within *cag*PAI and to determine the association of various *cag*PAI structure in *H. pylori* isolates with histopathological changes of the infected gastric mucosa. The outcome of this study may provide valuable information in order to draw association between existence of *cag*PAI genes and its association with disease sequelae in strains from multi-ethnic population and also in strains isolated in different histopathological conditions.

Results

Histopathological characteristics of the gastric mucosa in the studied populations

Histopathological scores of the gastric mucosa among different ethnic groups showed that the Malays had higher mean scores for *H. pylori* density and neutrophil activity whereas the Chinese showed higher grade of inflammation (Table S1). Higher mean score for intestinal, metaplasia was observed among the Indians, while the atrophy of higher grade was observed in the Chinese. Patients of different ethnicities were grouped into different types of disease conditions based on the histopathological changes (Table S2), i.e chronic gastritis (CG) (n=20), chronic active gastritis (CAG) (n=44) and intestinal metaplasia/atrophy (IM/Atr) (n=28). There was a significant difference in the proportion of CG and CAG between the Chinese and the non-Chinese patients. CG was diagnosed more in the Chinese patients compared to the non-Chinese (p = 0.03), whereas CAG and IM/Atr were observed more in the non-Chinese than the Chinese (p = 0.042).

Distribution of the *cag*PAI genes in *H. pylori* isolates

Detection of the *cag*PAI region in our clinical *H. pylori* isolates showed that 95.8% (n=91) of the isolates were *cag*PAI-positive. Four genes in the *cag*PAI region (*cag1*, *cag6*, *cag8* and *cag21*) were detected in all isolates whereas 35.2% isolates were *cag*2 (n=32) and 52.7% *cag14* (n=48) (Table 1). Detection of other

genes ranged from 69.2 – 98.9%. The absence of *cag2* was confirmed with 690 or 1100 bp amplicon using empty-site PCR as described by Schmidt et al., [21]. *cag14* was detected using 4 sets of primer pair as described by Ta et al., [20].

Six genes (*cag1, cag5, cag6, cag8, cag12* and *cag26*) in the *cag*PAI region were detected in all Indian isolates, whereas 12 and 19 genes were detected in all Chinese and Malay isolates, respectively (Table 1). A significant difference in detection of *cag2*, *cag4*, *cag14* and *cag24* were observed among *H. pylori* from patients with different ethnicities. Detection of *cag2* was significantly high in isolates from Malays (86.7%), followed by Indians (57.9%) and was least in Chinese isolates (9.8%) (^{2} = 36.620, df =2, p < 0.0001). The presence of *cag4* was high in isolates from Chinese (80.4%) compared to the Malays (46.7%) and Indians (63.2%) (^{2} = 7.001, df =2, p = 0.03). Significant difference was observed in the detection of *cag14* in the Malay isolates (93.3%) compared to the Chinese (39.2%) and the Indian (52.6%) isolates (^{2} = 13.603, df = 2, p = 0.001). Also, the *cag24* was significantly higher in the isolates from the Malays (93.3%) and the Indians (89.5%) compared to isolates from the Chinese patients (54.9%) (^{2} = 12.701, df = 2, p = 0.002).

We did further analyses to look for the distribution of individuals *cag*PAI genes in different disease conditions. All the *cag*PAI genes show similar distribution in CG, CAG and IM/Atr except for the *cag2* (data not shown). *cag2* was detected in 15.8% (3/19) of CG, 38.1% (16/42) of CAG and 40.7% (11/27) of IM/Atr. However, no significant difference was observed for the detection of *H. pylori* carrying *cag2* in different group of diseases (p = 0.16).

Analysis of *cag*PAI intactness in *H. pylori* isolates

The *cag*PAI was defined as intact if all the gene sets of the *cag*PAI were present including strains lacking only the *cag2* (HP0521). A previous systematic mutagenesis study showed that the HP0521 gene was not involved in the process of CagA translocation and IL-8 induction Fischer et al., [7]. In addition, NCBI database defined the HP0521 as a pseudogene (NCBI-Gene ID: 900040) (DBGET/LinkBD: an integrated database retrieval system, last accessed Oct 8, 2018). Partial *cag*PAI was defined when an isolate lacked one (other than HP0521) or more of the *cag*PAI genes, while negative/deleted *cag*PAI was defined if none of the genes were present and a product of approximately 650 bp with primers from the flanking regions was obtained. Among the 91 cagPAI-positive *H. pylori* strains, 24.2% (n=22) had intact *cag*PAI and 75.8% (n=69) exhibited partial (rearranged) *cag*PAI. Strains harbouring intact or partial *cag*PAI were not associated with patients' ethnicities (p > 0.05).

Association between *cag*PAI intactness and histopathological scores of the gastric mucosa are shown in Table 2. The presence of partial *cag*PAI was significantly related to the higher total score of *H. pylori* density (p = 0.036) and neutrophil activity (p = 0.03) compared to the intact *cag*PAI. *H. pylori* harbouring deleted *cag*PAI was significantly correlated with higher inflammatory score (mononuclear infiltration) compared to *H. pylori* with partial *cag*PAI (p = 0.002). The distribution of *H. pylori* with intact *cag*PAI was

detected more in the gastric mucosa with IM/Atr, whereas partial *cag*PAI *H. pylori* was detected more in CAG, however the difference was not significant (Table 3).

Discussion

Racial differences in the prevalence of *H. pylori* infection and disease-related severity were observed among patients from multiracial ethnicities [22,23]. Bacterial virulence factor is one of the contributing factors to the development of severe *H. pylori*-related diseases. The diversity of *cag*PAI region in the *H. pylori* genome may have a modifying effect on the pathogenic potential of the infecting strain [24].

In this study, we comprehensively determined the presence of all *cag*PAI genes in 91/95 *H. pylori* isolates from Malaysian population which were isolated from patients of different ethnic groups. The results show that more than 95% of our *H. pylori* strains were *cag*PAI-positive where 24.2% of the isolates carry all *cag*PAI genes, 75.8% exhibited partial or rearrangement in the *cag*PAI genes. In our previous study, we detected only 3.2% of the isolates carrying all the selected *cag*PAI genes [17]. The low percentage of *H. pylori* isolates harbouring intact *cag*PAI genes in our previous study is because we analysed only a subset of the *cag*PAI genes (*cag*67, *cag*10, *cag*13, *cag*T, *cag*M and *cag*E) as these genes was shown to have linkage between certain genes in the *cag*PAI region and severe disease as described by earlier studies [25,26]. In contrast, high frequency of intact *cag*PAI and low frequency of partial *cag*PAI in *H. pylori* strains isolated from similar ethnic populations was reported by Schmidt et al., [21]. In their study, few *cag*PAI genes (*cagE*, *cagL*, *cagT* and HP521) were examined to detect the intactness of *cag*PAI region. Discordant in the frequency of *cag*PAI intactness in many reports was due to the difference *cag*PAI genes that being examined [14,27,28]. Thus, results of the present study indicate that deletions can occur in all parts of the *cag*PAI and screening the entire genes in the *cag*PAI is needed to determine the accurate organization of the *cag*PAI region. For comparison with our results, we reviewed only studies that screened all the *cag*PAI genes. A previous study observed complete *cag*PAI present in 82.6% of the strains, while a partially deleted *cag*PAI in 9.6% of the strains and 7.7% lacked the entire *cag*PAI in Indian population [11]. In Swedish population, 76% of the strains carried an intact *cag*PAI, 15% had partially deleted *cag*PAI and the *cag*PAI was lacked in 9% of the strains [10]. A study by Azuma et al., [29] showed that the complete *cag*PAI was identified in all 11 Japanese isolates. Variation in the *cag*PAI positivity in different population of *H. pylori* isolates might be related to the difference in geographical origin of *H. pylori* subpopulations. Carriage of the *cag*PAI region is almost universal presence in *H. pylori* hpEastAsia and hpAfrica1 populations, intermediate presence in hpEurope and complete absence in hpAfrica2 [19]. Malaysian isolates showed a mixed subpopulation of hpEastAsia, hpAsia2 and hpEurope as indicated by multiracial communities living in the country [30,31].

Analysis of the entire *cag*PAI genes in the present study revealed that *cag1*, *cag6*, *cag8* and *cag21* were present in all isolates. These genes might represent core genes of the *cag*PAI region, however function of the *cag1*, *cag6* and *cag21* are still unknown [19]. *cag8* (HP0528, *cagX*) is a component of T4SS (VirB9) encodes a membrane protein [19]. One strain lacked *cagA* gene but had other *cag*PAI genes indicating that *cagA*-positive isolates do not necessarily have to be *cag*PAI positive. Indian isolates had more

rearrangement in the *cag*PAI region compared to the Malay and the Chinese. Studies have shown that the subpopulations of *H. pylori* Indian isolates in our country consisted of mixed populations i.e., hpEurope, hpAsia2 and hpEAsia and this might reflect the diversity of *cag*PAI genes rearrangement among the Indian isolates [30,31].

The presence of specific genes in *H. pylori* isolates associated with different ethnicities (*cag4* in the Chinese isolates and *cag2*, *cag14* and *cag24* in the non-Chinese isolates) might represent strain associated disease outcomes. The *cagA* (VirB1) is a component of T4SS, whereas the function is still unknown for *cag2*, *cag14* and *cag24* [19]. Although the difference was not statistically significant, high frequency of *cag2* was detected in gastric mucosa with CAG and IM/Atr and reflects the presence of this gene in non-Chinese isolates. These observations require further investigation to decipher the role of these genes.

We found an association of *cag*PAI intactness with histopathological scores of the gastric mucosa. *H. pylori* harbouring partial *cag*PAI were associated with higher density of *H. pylori* and neutrophil activity, whereas *H. pylori* with deleted *cag*PAI causes increased in inflammatory score. The presence of neutrophil activity in the gastric mucosa is associated with CAG and this has been shown in our study that partial *cag*PAI *H. pylori* strains was detected more in CAG groups. As strains with deleted *cag*PAI only cause inflammation of the gastric mucosa, the presence of *cag*PAI proteins encoded by *H. pylori* strains is needed to cause more severe disease such as active gastritis and intestinal metaplasia. However, no specific gene could be identified that causes severe condition. A group of genes encoded T4SS and for induction of IL-8 secretion have been shown to involve in the process of disease development [7,21].

Conclusions

Results of the present study show that *cag*PAI organisation is diverse in isolates from different ethnicities. Comprehensive screening of the entire *cag*PAI genes provides a more accurate overview of the *H. pylori cag*PAI genotype and allows better identification of the virulence traits of the organisms in our multiracial population. *H. pylori* strains harbouring partial/rearrangement of the *cag*PAI genes associated with increased colonization and recruitment of neutrophil at the site of infection and further contribute to various disease outcomes caused by different genotypes of *H. pylori* strains.

Methods

Bacterial isolates

A total of 95 non-repetitive *H. pylori* clinical isolates were obtained from patients (48 females and 47 males) recruited in the previous studies (research no. ETP-2013-042 and GUP-2011-307) between year 2011 to 2015. The patients' population comprised of different ethnicities (15 Malays, 52 Chinese, 21 Indians and 7 others), with mean age of 53.71 ± 17.24 years old and age range from 17 to 83 years old. Biopsy samples from the antrum or corpus of the stomach from the patients were cultured for *H. pylori*

isolation. These isolates were then stored at -70C in brucella broth containing 15% glycerol. *H. pylori* were subcultured from frozen stock onto Columbia blood agar (Oxoid, Basingstoke, England) supplemented with 7% sheep blood and Dent's supplement (Oxoid, Basingstoke, England) and incubated at 37°C for 5 to 7 days under microaerophilic environment. All patients had gastritis graded according to Updated Sydney Classification [18] except for two patients where the histopathological examination (HPE) results were not available.

DNA extraction

H. pylori colonies were scraped from the agar surface of Columbia blood agar plate and subjected to DNA extraction using FavorPrep^{TM} Tissue Genomic DNA Extraction Mini kit according to the manufacturer's instructions (Favorgen Biotech Corporation, Ping-Tung 908, Taiwan). DNA samples were diluted with ultrapure water to a concentration of 25 ng/µl and stored at -20°C until further processing.

Determination of *cag*PAI genes

The presence or absence of *cag*PAI in *H. pylori* strains was determined by PCR using primers for detection of the 5' and 3' flanking region of the *cag*PAI as described by Olbermann et al., [19]. The amplifications were carried out in 25 µl volume, each containing 12.5 µl mastermix (Lucigen, USA), 10 µl of each primers, 1 µl (25 ng) DNA and 10 µl DNAse and RNAse free sterile distilled water. PCR amplification for detection of *cag*PAI region consisted of initial denaturation at 95°C for 3 min, followed by 30 cycles of 95° for 30 s, 50°C for 60 s, and 72°C for 45 s, ending with final extension at 72°C for 5 min. The amplifications were performed in a PCR thermal cycler T100 Series (Bio-Rad, USA). The products were run on 1.5% agarose gel and stained with FloroSafe DNA stain (1st BASE Pte. Ltd, Singapore) and visualised with gel documentation (AlphaImager, Biosciences, CA). The *cag*PAI-positive isolates (n=91) were then subjected to subsequent PCRs for identification of all *cag*PAI genes using primers as described previously [19,20]. The deletion of HP0521 gene were confirmed using HP0521 empty site (ES) primer pair as described previously [21]. PCR amplification for *cag*PAI genes consisted of initial denaturation at 95°C for 3 min, followed by 30 cycles of 95° for 30s, annealing temperature for 60s (48C for *cag11*, 48.8C for *cag3* and 55C for *cag1*, *cag2*, *cag4*, *cag5*, *cag*, *cag6* to *cag10*, *cag12* to *cag26*), and extension at 72°C for 45 s. A final extension at 72°C for 5 min was performed for each PCR run. Representative positive PCR products (n=28) were sent for sequencing and the nucleotide sequences were blasted against NCBI databases to confirm the gene identity.

Statistical analysis

Statistical analysis was performed using SPSS software version 23 (SPSS Inc, Chicago, IL, USA). Differences between groups were evaluated using Chi-square (χ^2) test, Yate's continuity correction and Fisher's exact probability test. Independent t-test was used to compared means between different groups

of histopathological scores. Score was represented with mean standard error of mean (SE). Differences were considered significant when *p* value was <0.05.

Declarations

Acknowledgments

We would like to thank to the Universiti Kebangsaan Malaysia for providing both the permission and the facilities to conduct and publish this research and to the technical staffs of Dept. of Medical Microbiology & Immunology, Faculty of Medicine, Universiti Kebangsaan Malaysia for their technical help.

Funding

The research was funded by a grant from Ministry of Higher Education of Malaysia (grant no. FRGS/2/2014/SKK04/UKM/02/01). We also thank to Ministry of Higher Education of Malaysia for providing a studentship to SAR under the MyBrain15 program.

Availability of data and materials

Data will be shared upon request to the corresponding author alfizah@ppukm.ukm.edu.my

Authors' contribution

SAR performed all experiments and data analysis. HMN and NMZ participated in the study design and data analysis. AH involved in the design of the study, data analysis and manuscript writing. BSL participated in data analysis and manuscript writing. All authors read and approved the final manuscript.

Ethics approval and consent to participate

The research protocol was approved by the Medical Research Ethic Committee of the university (UKM1.5.3.5/244/JEP-2016-095). The present study used *H. pylori* stock cultures where the informed consent was not applicable. However, these isolates were obtained from patients in previous studies (research no. ETP-2013-042 and GUP-2011-307) where informed consent was obtained from all the individuals included in the study.

Competing interests

The authors declare that they have no conflict of interest.

References

1. Graham JR. *Helicobacter pylori*: human pathogen or simply an opportunist? Lancet. 1995;345(8957):1095-7.

2. Kusters JG, Van Vliet AH, Kuipers EJ. Pathogenesis of *Helicobacter pylori* infection. Clin Microbiol Rev. 2006;19:449-0.

3. Suerbaum S, Michetti P. *Helicobacter pylori* infection. N Engl J Med. 2002;347:1175-86.

4. Wroblewski LE, Peek RM Jr, Wilson KT. *Helicobacter pylori* and gastric cancer: factors that modulate disease risk. Clin Microbiol Rev. 2010;23:713-39.

5. Compare D, Rocco A, Nardone G. Risk factors in gastric cancer. Eur Rev Med Pharmacol Sci. 2010;14:302-8.

6. Kim SS, Ruiz VE, Carroll JD, Moss SF. *Helicobacter pylori* in the pathogenesis of gastric cancer and gastric lymphoma. Cancer Lett. 2010;305:228-38.

7. Fischer W, Puls J, Buhrdorf R, Gebert B, Odenbreit S, Hass R. Systematic mutagenesis of the *Helicobacter pylori cag* pathogenicity island: essential genes for CagA translocation in host cells and induction of interleukin-8. Mol Microbiol. 2001;42:1337-48.

8. Hatakeyama M. SagA of CagA in *Helicobacter pylori* pathogenesis. Curr Opin Microbiol. 2008;11:30-7.

9. Kao CY, Sheub BS, Wu JJ. *Helicobacter pylori* infection: An overview of bacterial virulence factors and pathogenesis. Biomed J. 2016;39(1):14-23.

10. Nilsson C, Sillén A, Eriksson L, Strand ML, Enroth H, Normark S, Falk P, Engstrand L. Correlation between *cag* pathogenicity island composition and *Helicobacter pylori*-associated gastroduodenal disease. Infect Immun. 2003;71:6573-81.

11. Patra R, Chattopadhyay S, De R, Datta S, Chowdhury A, Ramamurthy T, Balakrish Nair G, Berg ED, Mukhopadhyay AK. Intact cag pathogenicity island of *Helicobacter pylori* without disease association in Kolkata, India. Int J Med Microbiol. 2011;301:293-302.

12. Bridge DR, Merrel DS. Polymorphism in the *Helicobacter pylori* CagA and VacA toxins and disease. Gut Microbes. 2013;4(2):101-17.

13. Sahara S, Sugimoto M, Vilaichone RK, Mahachai V, Miyajima H, Furuta T, Yamaoka Y. Role of *Helicobacter pylori cagA* EPIYA motif and *vacA* genotypes for the development of gastrointestinal diseases in Southeast Asian countries: a meta-analysis. BMC Infect Dis. 2012;1:2223.

14. Lai CH, Perng CL, Lan KH, Lin HJ. Association of detection of virulence gene belonging to *cag* pathogenicity island in *Helicobacter pylori* IS605 and *cag*-PAI of *Helicobacter pylori* isolated from patients

with gastrointestinal diseases in Taiwan. Gastroenterol Res Pract. 2013:Article ID 356217.

15. Waskito LA, Miftahussurur M, Lusida MI, Syam AF, Suzuki R, Subsomwong P, Uchida T, Hamdan M, Nasronudin, Yamaoka Y. Distribution and clinical associations of integrating conjugative elements and *cag* pathogenicity islands of *Helicobacter pylori* in Indonesia. Sci Rep. 2018;8:6073.

16. Khatoon J, Prasad KN, Prakash Rai R, Ghoshal UC, Krishnani N. Association of heterogenicity of *Helicobacter pylori cag* pathogenicity island with peptic ulcer diseases and gastric cancer. Bri J Biomed Sci. 2017;74(3):121-6.

17. Alfizah H, Rukman AH, Norazah A, Hamizah R, Ramelah M. Ethnicity association of *Helicobacter pylori* virulence genotype and metronidazole susceptibility. World J Gastroenterol. 2013;19:1283-91.

18. Dixon MF, Genta RM, Yardley JH, Correa P. Classification and grading of gastritis. The updated Sydney System. International Workshop on the Histopathology of Gastritis, Houston 1994. Am J Surg Pathol. 1996;20(10):1161-81.

19. Olbermann P, Josenhans C, Moodley Y, Uhr M, Stamer C, Vauterin M, Suerbaum S, Achtman, M, Linz B. A global overview of the genetic and functional diversity in the *Helicobacter pylori cag* pathogenicity island. PLoS Genet. 2010;6:e1001069.

20. Ta LH, Hansen LM, Sause WE, Shiva O, Millstein A, Ottemann KM, Castillo AR, Solnick JV. Conserved transcriptional unit organization of the *cag* pathogenicity island among *Helicobacter pylori* strains. Frontiers Cellular Infect Microbiol. 2012;2:Article 46.

21. Schmidt HMA, Andres S, Nilsson C, Kovach Z, Kaakoush NO, Engstrand L, Goh KL, Fock KM, Forman D, Mitchell H. The *cag*PAI is intact and functional but HP521 varies significantly in *Helicobacter pylori* isolates from Malaysia and Singapore. Eur J Clin Microbiol Infect Dis. 2010;29:439-51.

22. Epplein M, Signorello LB, Zheng W, Peek RM Jr, Michel A, Williams SM, Pawlita M, Correa P, Cai Q, Blot WJ. Race, African ancestry, and *Helicobacter pylori* infection in a low-income United States population. Cancer Epidemiol Biomarkers Prev. 2011;20:826-34.

23. Latifi-Navid S, Ghorashi SA, Siavoshi F, Linz B, Massarrat S, Khegay T, Salmanian AH, Shayesteh AA, Masoodi M, Ghanadi K, Ganji A, Suerbaum S, Achtman M, Malekzadeh R, Falush D. Ethnic and geographic differentiation of *Helicobacter pylori* within Iran. PLoS One. 2010;5:e9645.

24. Yuan XY, Yan JJ, Yang YC, Wu CM, Hu Y, Geng JL. *Helicobacter pylori* with East Asian-type *cag*PAI genes is more virulent than strains with Western-type in some *cag*PAI genes. Braz J Microbiol. 2017;48:218-24.

25. Deguchi R, Igarashi M, Watanabe K, Takagi A. Analysis of the *cag* pathogenicity island and IS605 of *Helicobacter pylori* strains isolated from patients with gastric cancer in Japan. Aliment Pharmacol Ther. 2004;20(Suppl. 1):13-6.

26. Hsu PI, Hwang IR, Cittelly D, Lai KH, El-Zimaity HM, Gutierrez O, Kim JG, Osato MS, Graham DY, Yamaoka Y. Clinical presentation in relation to diversity within the *Helicobacter pylori cag* pathogenicity island. Am J Gastroenterol. 2002;97:2231-8.

27. Antonio-Rincón F, López-Vidal Y, Castillo-Rojas G, Lazcano-Ponce EC, Ponce-de-León S, Tabche-Barrera ML, Aguilar-Gutiérrez GR. Pathogenicity island *cag*, *vacA* and IS605 genotypes in Mexican strains of *Helicobacter pylori* associated with peptic ulcers. Ann Clin Microbiol Antimicrob. 2011;10:18.

28. Varda Brkić D, Katičić M, Bedenić B, Stanko AP, Plečko V. Detection of virulence gene belonging to *cag* pathogenicity island in *Helicobacter pylori* isolates after multiple unsuccessful eradication therapy in Northwest Croatia. Period Biol. 2016;118(1):45-52.

29. Azuma T, Yamakawa A, Yamazaki S, Ohtani M, Ito Y, Muramatsu A, Suto H, Yamazaki Y, Keida Y, Higashi H, Hatakeyama M. Distinct diversity of the *cag* pathogenicity island among *Helicobacter pylori* strains in Japan. J Clin Microbiol. 2004;42:2508-17.

30. Breurec S, Guillard B, Hem S, Brisse S, Dieye FB, Huerre M, Oung C, Raymond J, Tan TS, Thiberge JM, Vong S, Monchy D, Linz B. Evolutionary history of *Helicobacter pylori* sequences reflect past human migrations in Southeast Asia. PLoS One. 2011;6(7):e22058.

31. Tay CY, Mitchell H, Dong Q, Goh KL, Dawes IW, Lan R. Population structure of *Helicobacter pylori* among ethnic groups in Malaysia: recent acquisition of the bacterium by the Malay population. BMC Microbiol. 2009;9:126.

Tables

Table 1. Distribution of the *cag*PAI genes among 91 *cag*PAI-positive *H. pylori* isolates from patients with different ethnicities

Gene no. in 26695 strain	Gene name	Component of T4SS	n (%)	*Patients' ethnicity, n (%)			
				M (n=15)	C (n=51)	I (n=19)	Other (n=6)
HP0520	*cag1* (*cag*)	-	91 (100)	15 (100)	51 (100)	19 (100)	6 (100)
HP521	*cag2*	-	32 (35.2)	13 (86.7)	5 (9.8)	11 (57.9)	3 (50)
HP0522	*cag3* (*cag*)	u	90 (98.9)	15 (100)	51 (100)	18 (94.7)	6 (100)
HP0523	*cag4* (*cag*)	VirB1	66 (72.5)	7 (46.7)	41 (80.4)	12 (63.2)	6 (100)
HP0524	*cag5* (*cagβ*)	VirD4	90 (98.9)	15 (100)	51 (100)	19 (100)	5 (83.3)
HP0525	*cag*	VirB11	85 (93.4)	14 (93.3)	50 (98)	15 (78.9)	6 (100)
HP0526	*cag6* (*cagZ*)	-	91 (100)	15 (100)	51 (100)	19 (100)	6 (100)
HP0527	*cag7* (*cagY*)	VirB9	88 (96.7)	15 (100)	50 (98)	17 (89.5)	6 (100)
HP0528	*cag8* (*cagX*)	VirB6	91 (100)	15 (100)	51 (100)	19 (100)	6 (100)
HP0529	*cag9* (*cagW*)	VirB8	90 (98.9)	15 (100)	51 (100)	18 (94.7)	6 (100)
HP0530	*cag10* (*cagV*)	-	88 (96.7)	14 (93.3)	50 (98)	18 (94.7)	6 (100)
HP0531	*cag11* (*cagU*)	VirB7	77 (84.6)	15 (100)	41 (80.4)	15 (78.9)	6 (100)
HP0532	*cag12* (*cagT*)	-	90 (98.9)	15 (100)	51 (100)	18 (94.7)	6 (100)
HP0534	*cag13* (*cagS*)	-	89 (97.8)	15 (100)	51 (100)	17 (89.5)	6 (100)
HP0535	*cag14* (*cagQ*)	-	48 (52.7)	14 (93.3)	20 (39.2)	10 (52.6)	4 (66.7)
HP0536	*cag15* (*cagP*)	-	89 (97.8)	15 (100)	51 (100)	17 (89.5)	6 (100)
HP0537	*cag16* (*cagM*)	u	84 (92.3)	15 (100)	49 (96.1)	14 (73.7)	6 (100)
HP0538	*cag17*	-	86	15 (100)	48	17	6

	(*cagN*)		(94.5)		(94.1)	(89.5)	(100)
HP0539	*cag18* (*cagL*)	VirB5	87 (95.6)	15 (100)	50 (98)	16 (84.2)	6 (100)
HP0540	*cag19* (*cagI*)	-	83 (91.2)	15 (100)	48 (94.1)	16 (84.2)	4 (66.7)
HP0541	*cag20* (*cagH*)	-	89 (97.8)	14 (93.3)	51 (100)	18 (94.7)	6 (100)
HP0542	*cag21* (*cagG*)	-	91 (100)	15 (100)	51 (100)	19 (100)	6 (100)
HP0543	*cag22* (*cagF*)	-	89 (95.6)	15 (100)	51 (100)	17 (89.5)	6 (100)
HP0544	*cag23* (*cagE*)	VirB3/B4	87 (95.6)	15 (100)	50 (98)	16 (84.2)	6 (100)
HP0545	*cag24* (*cagD*)	-	63 (69.2)	14 (93.3)	28 (54.9)	17 (89.5)	4 (66.7)
HP0546	*cag25* (*cagC*)	VirB2	80 (87.9)	14 (93.3)	46 (90.2)	15 (78.9)	5 (83.3)
HP0547	*cag26 (cagA)*	effector	90 (98.9)	15 (100)	50 (98)	19 (100)	6 (100)

M; Malays, C; Chinese, I; Indians, u; unknown function

Table 2. Association of *H. pylori cag*PAI intactness with histopathological changes of gastric mucosa

Histopathological changes	Score	*cag*PAI, n (%)		
		Intact	Partial	Deleted
H. pylori density1	0	7 (31.8)	15 (22.4)	1 (25)
	1	10 (45.5)	21 (31.3)	2 (50)
	2	4 (18.2)	18 (26.9)	1 (25)
	3	1 (4.5)	13 (19.4)	0
Total score	Mean SE	0.95 0.18	1.43 0.13	1.0 0.41
MNC infiltration2	0	0	1 (1.5)	0
	1	6 (27.3)	23 (34.3)	0
	2	14 (63.6)	36 (53.7)	4 (100)
	3	2 (9.1)	7 (10.4)	0
Total score	Mean SE	1.82 0.13	1.73 0.08	2.0 0
Neutrophil activity3	0	10 (45.5)	14 (20.9)	1 (25)
	1	9 (40.9)	32 (47.8)	1 (25)
	2	2 (9.1)	15 (22.4)	2 (50)
	3	1 (4.5)	6 (9.0)	0
Total score	Mean SE	0.73 0.18	1.19 0.11	1.25 0.48
Intestinal metaplasia	0	17 (77.3)	59 (88.1)	4 (100)
	1	4 (18.2)	6 (9)	0
	2	1 (4.5)	1 (1.5)	0
	3	0	1 (1.5)	0
Total score	Mean SE	0.27 0.12	0.16 0.06	0
Atrophy	0	15 (68.2)	53 (79.1)	3 (75)
	1	5 (22.7)	10 (14.9)	1 (25)
	2	1 (4.5)	2 (3)	0
	3	1 (4.5)	2 (3)	0
Total score	Mean SE	0.45 0.17	0.30 0.08	0.25 0.25

Statistical analysis (Independent t-test):

1 Partial *vs* Intact; t = 2.166, *p* = 0.036, 95% CI (0.033-0.923)

Partial *vs* Deleted; *p* = 0.42

Deleted *vs* Intact; *p* = 0.05

2 Deleted *vs* Partial; t = 3.308, *p* = 0.002, 95% CI (0.106 – 0.431)

Deleted *vs* Intact; *p* = 0.162

Intact *vs* Partial; *p* = 0.586

3 Partial *vs* Intact; t = 2.20, *p* = 0.03, 95% CI (0.045 – 0.888)

Deleted *vs* Intact; *p* = 0.266

Deleted *vs* Partial; *p* = 0.902

Table 3. *cag*PAI intactness in *H. pylori* in patients with different disease groups

Disease group	*cag*PAI, n (%)		
	Intact (n=22)	Partial (n=66)	Deleted (n=4)
CG	7 (3.8)	12 (18.2)	1 (25)
CAG	6 (13.6)	36 (54.5)	2 (50)
IM/Atr	9 (40.9)	18 (27.3)	1 (25)

Intact vs partial: 2 = 4.992, df = 2, *p* = 0.08

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- ManuscriptcagPAITableS2M1.docx
- ManuscriptcagPAITableS1M1.docx