

Helicobacter pylori cagA gene variants in Malaysians of different ethnicity

Ramelah Mohamed · Alfizah Hanafiah · Isa M. Rose ·
Mohd Rizal A. Manaf · Shiekh Anwar Abdullah ·
Ismail Sagap · A. van Belkum · Jasmi A. Yaacob

Received: 22 January 2009 / Accepted: 25 January 2009 / Published online: 27 February 2009
© The Author(s) 2009. This article is published with open access at Springerlink.com

Abstract We have defined DNA repeat variability in the 3'-terminus of the *cagA* gene of *Helicobacter pylori* strains from Malaysian patients of different ethnicities. We identified different alleles based on the EPIYA repeats. *cagA* types A-B-D and A-B-B-D are more similar to the sequence of Japanese strains, whereas *cagA* types A-B-C, A-B-C-C, A-B and A-C displayed similarity to strain 26695 sequences. A significant association was found between *cagA* genotypes and patients' ethnicity, with *cagA* type A-B-D being predominantly isolated from Chinese patients and *cagA* type A-B-C from Malays and Indians. Our data further corroborate the possibility that variant biological activity of CagA may affect the host specificity and/or pathogenicity of *H. pylori*.

The population of Malaysia comprises three major ethnic groups: Malays, Chinese and Indians. Endoscopy-based, sero-epidemiological and molecular studies conducted in the region reported that there are differences in the prevalence of *Helicobacter pylori* infection among different ethnic groups [1–4]. Further observations also revealed differences in the distribution of non-ulcer dyspepsia (NUD), peptic ulcer disease (PUD) and (pre-)cancerous lesions in *H. pylori*-positive patients. The Chinese consistently showed a higher rate of severe disease. Although Indians have a higher prevalence of *H. pylori* infection, the frequency of peptic ulcer disease among them was relatively low, a pattern that has not changed significantly since 1994 [1, 3]. Differences in the prevalence of *H. pylori*

R. Mohamed (✉)
UKM Medical Molecular Biology Institute (UMBI),
Universiti Kebangsaan Malaysia (UKM),
7th Floor, Jalan Yaacob Latiff,
Bandar Tun Razak,
56000 Cheras Kuala Lumpur, Malaysia
e-mail: ramelah@mail.hukm.ukm.my

A. Hanafiah
Department of Medical Microbiology & Immunology,
Faculty of Medicine,
Universiti Kebangsaan Malaysia (UKM),
56000 Cheras Kuala Lumpur, Malaysia

I. M. Rose
Department of Pathology,
Faculty of Medicine,
Universiti Kebangsaan Malaysia (UKM),
56000 Cheras Kuala Lumpur, Malaysia

M. R. A. Manaf
Department of Community Health, Faculty of Medicine,
Universiti Kebangsaan Malaysia (UKM),
56000 Cheras Kuala Lumpur, Malaysia

S. A. Abdullah
Department of Medicine, Faculty of Medicine,
Universiti Kebangsaan Malaysia (UKM),
56000 Cheras Kuala Lumpur, Malaysia

I. Sagap · J. A. Yaacob
Department of Surgery, Faculty of Medicine,
Universiti Kebangsaan Malaysia (UKM),
56000 Cheras Kuala Lumpur, Malaysia

A. van Belkum
Department of Medical Microbiology and Infectious Diseases,
Erasmus MC, 's Gravendijkwal 230,
3015 GD Rotterdam, The Netherlands

infection and gastro-duodenal diseases among patients of different ethnic groups residing in the same geographic region may provide insight into the pathogenesis of the infection.

The *cagA* gene encodes the cytotoxin-associated gene A (CagA) protein and is a marker for the *cag* pathogenicity island. The *cagA* gene is diverse in its structure, especially at the 3'-terminus. Many studies reported differences in this sequence among strains from East Asian and Western countries [4–7]. This revealed that there are several distinct forms of CagA with an uneven geographical distribution, which may mark differences in virulence among *cagA*-positive *H. pylori* strains. Our previous report showed that the overall incidence of *cagA*-positive *H. pylori* isolates was 94% and that there was no significant association between *cagA* subtypes and gastroduodenal diseases [8]. Genetic variation in the *cagA* 3'-terminal region among Malaysian *H. pylori* strains has never been explored. In the present study, we characterise the *cagA* 3'-terminus in *H. pylori* strains isolated from dyspeptic patients of different ethnicity and variable gastroduodenal diseases.

A total of 110 *cagA*-positive *H. pylori* isolates were derived from gastric antrum biopsies of patients attending the Endoscopy Unit of Universiti Kebangsaan Malaysia Medical Centre (Kuala Lumpur, Malaysia). The patients (19 Malays, 68 Chinese and 23 Indians) were classified into four groups according to endoscopic findings and histopathology: normal stomach ($n=2$), gastritis ($n=52$), duodenitis ($n=10$), gastric ulcer ($n=15$), duodenal ulcer ($n=10$), gastric and duodenal ulcer ($n=3$), intestinal metaplasia ($n=23$), atrophy ($n=2$), dysplasia ($n=1$) and gastroesophageal reflux disease (GERD) ($n=1$). Patients were classified as having peptic ulcer disease when active gastric and/or duodenal ulceration was detected upon endoscopy with a mucosal break of at least 0.5 mm in one dimension. Patients with past ulcers were also included because of potential relapses and remitting conditions.

Assessment of the histological grade of severity and the presence of *H. pylori* were routinely done at the Department of Pathology. The biopsies were fixed in 10% formalin and paraffin-embedded sections were cut and stained with haematoxylin-eosin. When necessary, sections were also stained according to Warthin-Starry for better visualisation of *H. pylori*. The slides were classified by an independent pathologist. Intestinal metaplasia, atrophy, dysplasia and gastric carcinoma were determined and grouped as (pre-)cancerous lesions.

Biopsies were sub-cultured for *H. pylori* on Columbia agar base (Oxoid) containing Dent's supplement (Oxoid) and 7% ox blood. The plates were incubated at 37°C for five days under micro-aerophilic conditions. The bacterial cultures were confirmed to be *H. pylori* on the basis of colony morphology, the presence of curved or spiral-shaped

Gram-negative bacteria, and positive urease, catalase and oxidase tests. A single colony was picked from each culture plate and sub-cultured for DNA extraction.

Genomic *H. pylori* DNA was prepared using the High Pure PCR Template Preparation Kit (Roche, Mannheim, Germany) according to the manufacturer's instructions. Differences in the length of the 3'-terminal end of the *cagA* gene were identified using primers *cag1* and *cag3* as described previously [9]. The amplification products range from 500 to 800 base pairs (bp) in length. *H. pylori* ATCC 700824 (strain J99) and *H. pylori* ATCC 700392 (strain 26695) were used as positive controls. *UreC* target sequences provided the internal control. Polymerase chain reaction (PCR) products were analysed on ethidium bromide-stained agarose gels. The *cagA* PCR products were purified using the MinElute Gel Extraction Kit (QIAGEN, Hilden, Germany) and then cloned into pCR2.1 using TOPO TA cloning (Invitrogen, Carlsbad, CA). The *cagA* inserts were sequenced on both strands using M13 primers. All sequences were stored in BioEdit. Clustal W version 1.83 was used for multiple sequence alignment. Nucleotide and amino acid sequences were analysed by pair-wise sequence comparison using the Kimura distance method. The sequences used in the analyses comprised of three *H. pylori* genome sequences

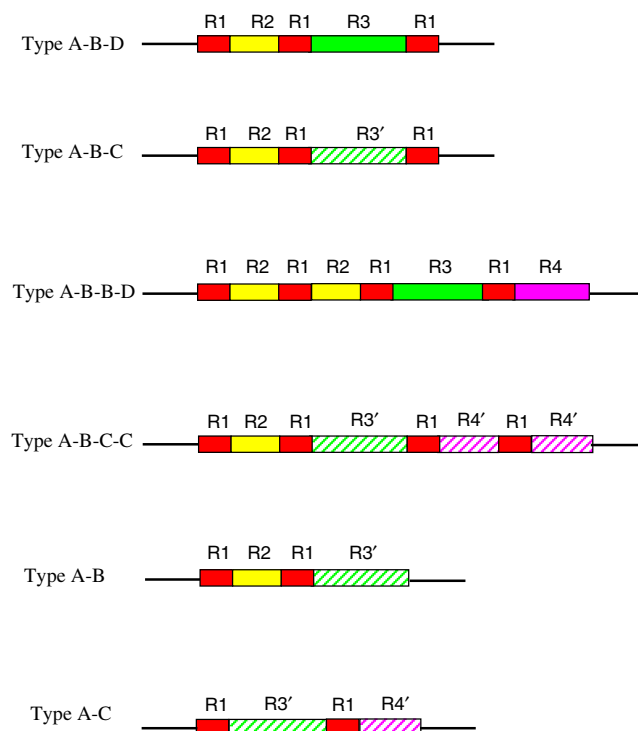


Fig. 1 Primary structures of the 3' region of the *cagA* gene in clinical *Helicobacter pylori* isolates. The fragments are not represented on a proportional scale

Table 1 Amino acid sequences in the repeat region of CagA

Repeat region	Amino acid sequence
R1	EPIYA
R2	QVNKKKTGQAASPE
R3	QVAKKVS AKIDQLNEATSAINRKIDRINKIASAGKGVGGFSGAGRSASPE
R3'	QVAKKVN AKIDRLNQIASGLGGVQAAGFPLKRHDKVDDL SKVGLSASPE
R4	TIDFDEANQAGFPLRRAAAVNDLSKVG
R4'	TIDDLGGPFPLKRHDKVDDL SKVG

obtained from GenBank, strain 26695 (ATCC 700392), strain J99 (ATCC 700824), the Japanese strain (accession no. AB 246742), and 110 sequences obtained from our *H. pylori* clinical isolates.

Statistical analysis was performed using the Statistical Package for the Social Sciences (SPSS). Differences in the distribution of genotypes of the *cagA* gene and the association between genotypes, clinical outcome and patients' ethnicity were analysed using χ^2 tests and Fisher's exact testing. A *P*-value less than 0.05 was considered to be significant.

Amino acid sequences of the C-terminus of CagA revealed six genotypes that differed in their structural organisation due to variations in the numbers and/or types of different repeats (Fig. 1) [9]. The CagA types were assigned to A-B-D, A-B-C, A-B-B-D, A-B-C-C, A-B and A-C based on the EPIYA sequence, as previously described [10]. The sequences of repeat regions R1, R2, R3, R3', R4 and R4' are shown in Table 1. R1, R2 and R3 are similar to sequences as previously described [9]. The R1 and R2 regions are conserved and are found to be similar between Eastern and Western type sequences. R3 and R4 are similar to the Japanese strain sequence, whereas R3' and R4' sequences are more similar to the Western strain sequence. The data showed that the R1 region, which is composed of the EPIYA motif, was present in all of the strains. Copies of R1 were found at various frequencies in between the R2, R3, R3', R4 and R4' repeat regions. The majority of CagA sequences had at least three EPIYA motifs, as observed in CagA types A-B-D and A-B-C. Some variants contained two repeats (types A-B and A-C) or four repeats (types A-B-B-D and A-B-C-C).

Similarity score analysis showed that CagA type A-B-D is 80% similar to the Japanese strain and 60% to the Western strain. CagA type A-B-D was found in 70 (75.3%) strains and CagA type A-B-C in 23 (24.7%) strains. One strain had CagA type A-B-B-D and looked similar to the Japanese strain, whereas 13 strains had CagA type A-B-C-C. The similarity of CagA type A-B-C-C dropped to about 70% when compared to the Japanese strain, but had a high similarity score to strain 26695 (89%). CagA type A-B (two strains) and A-C (one strain) showed a similarity of about 76% to strain 26695.

Table 2 shows the distribution of CagA type A-B-D and CagA type A-B-C among patients from different ethnic groups and with variant diseases. A large proportion of Chinese patients were infected with *H. pylori* strains carrying CagA type A-B-D, whereas infection with *H. pylori* strains carrying CagA type A-B-C were predominantly detected in Indians and Malays. Differences in the infecting sub-genotypes of *H. pylori* CagA types A-B-D and A-B-C between the Chinese and non-Chinese (Malays and Indians) were highly statistically significant ($P < 0.0005$). However, there was no association between the CagA type A-B-D and the severity of disease, even though the isolation of these sub-genotypes from predominantly infected Chinese patients was significant.

Table 2 Distribution of the subtypes of CagA types A-B-D and A-B-C among patients from different ethnic groups and patients with different disease groups

	CagA type	
	A-B-D (%) (<i>n</i> =70)	A-B-C (%) (<i>n</i> =23)
*Ethnic groups		
Chinese (<i>n</i> =67)	58 (82.9)	9 (39.1)
Malays (<i>n</i> =12)	8 (11.4)	4 (17.4)
Indians (<i>n</i> =14)	4 (1.4)	10 (43.5)
**Disease groups		
NUD	34 (48.6)	10 (45.5)
Normal stomach	2	0
Gastritis	32	10
PUD	18 (25.7)	8 (36.4)
GU	10	4
DU	7	2
GU + DU	1	2
PCL	18 (25.7)	4 (18.2)
Intestinal metaplasia	13	3
Atrophy	2	0
Dysplasia	0	1

Note: one sequence from the strain isolated from GERD patients was not included in the statistical analysis

GU = gastric ulcer; DU = duodenal ulcer

*Pearson Chi-square, 2×2 table; $\chi^2 = 16.434$, $P < 0.0005$

**Pearson Chi-square, 3×2 table; $\chi^2 = 1.103$, $P = 0.576$

We found no association between the CagA type A-B-D and the severity of disease, even though this type was predominantly isolated from Chinese patients. However, our data suggest a trend, albeit not statistically significant, indicating that a high proportion of patients from the precancerous and cancerous groups were infected with CagA type A-B-D—an observation that merits further study.

The first and second EPIYA motifs (in the repeat region R1 and R2) that are designated as EPIYA-A and EPIYA-B, respectively [7], are present in almost all CagA proteins, whereas the remaining EPIYA-C motifs were duplicates of an EPIYA containing 34 amino acids. Since the 34 amino acid sequence shows variation in copy numbers, ranging from one to three in most Western CagA proteins, it is designated as the Western CagA-specific (WSS) [10, 11]. In contrast, the amino acid sequence of East Asian CagA is quite different from that of Western CagA. Predominantly, East Asian CagA proteins do not have the WSS but, instead, possess a distinct sequence in the corresponding region, which is designated East Asian CagA-specific (ESS) [10]. ESS possesses an EPIYA motif designated as EPIYA-D.

In our study, CagA type A-B-D demonstrated a single ESS, whereas CagA type A-B-C demonstrated a single WSS. All of CagA type A-B-C-C had two WSS regions and is, thus, classified as A-B-C-C, and in a single strain (HQ326), A-B-B-D region was detected. Sequence analysis of CagA type A-B-C-C showed that the strains were similar to Western CagA type, and of which the presence of two EPIYA-C may contribute to the high phosphorylation activity and more intense cellular rearrangement. Backert et al. [12] showed that the SASEPIY sequence motif and, particularly, Y-972 is of general importance for CagA phosphorylation and proved that the phosphorylation of CagA at Y-972 has a function in actin-based cytoskeletal rearrangements. These polymorphisms within CagA might affect the biological function of the protein and might explain the lack of a consistent correlation between CagA and disease severity, as noted in the present study and others [13, 14].

We examined the CagA diversity among clinical isolates from patients of Malay, Chinese or Indian ethnicity and observed a mixed presence of *H. pylori* that had East Asian or Western type CagA, although the East Asian strain appeared to be predominately isolated from the Chinese patients. The prevalence of East Asian CagA and Western CagA strains among our local isolates was 64.5 and 36.5%, respectively, which were similar to Thailand, but quite different from Japan and China. All of the studied samples isolated from Fukui, Japan, and 94.4% from Hangzhou, China, had East Asian type CagA [15, 16]. Data from Thailand reported a prevalence of 53.7% of East Asia CagA and 26.8% of Western CagA strains in their population [17, 18]. It has been reported that large sequence differences

distinguish the CagA function of Asian strains from other strains [11, 19]. The present data showed that diversity of CagA occurs even in the same population in Malaysia, where most of the isolates from Chinese patients carried East Asian CagA type and most of the isolates from Indians and Malays carried the Western CagA type. An important unanswered question is whether *H. pylori* strains having a particular CagA structure could be linked to a defined gastroduodenal disease. Yamaoka et al. [9] have reported that *H. pylori* possessing more than three repeats (102 bp) in the 3'-terminal region of the *cagA* gene are associated with enhanced injury and reduced survival under acidic conditions. In our study, patients with peptic ulcer disease or (pre-)cancerous lesions were more likely to be infected with East Asian CagA *H. pylori* strains, although the association was shown to be not significant. Occhialini et al. [20], studying 33 *H. pylori* isolates from Costa Rica, have found no association between the number of repeated sequences at the 3' region of the *cagA* gene or the presence of tyrosine phosphorylation motifs or the clinical origin of these strains, and, furthermore, they did not find strains with three repeats. Further studies are required to assess whether these discrepancies are due to the variable region of *cagA* or if, in fact, they correspond to differences in the immune response within the infected population.

Strain-to-strain variation in the CagA sequence, particularly in the variable, highly hydrophilic and surface-exposed C-terminal region, could be a valuable tool for the bacteria to escape the immune clearance of the host. This capability could have arisen by horizontal gene transfer of the pathogenicity island, where the *cagA* gene is located and optimised by homologous recombination events among DNA from different *H. pylori* strains.

Acknowledgements This research was funded by a grant from the Ministry of Science, Technology and Innovation of Malaysia. A very special thank you goes to the staff of the Endoscopy Unit, Universiti Kebangsaan Malaysia Medical Centre, who assisted us in the collection of biopsy specimens and also to the staff at the Histopathology Unit, Department of Pathology, Faculty of Medicine, Universiti Kebangsaan Malaysia, for preparing the paraffin-embedded tissues for microscopic examination.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

1. Goh KL (1997) Prevalence of and risk factors for *Helicobacter pylori* infection in a multi-racial dyspeptic Malaysian population undergoing endoscopy. *J Gastroenterol Hepatol* 12:S29–S35. doi:10.1111/j.1440-1746.1997.tb00455.x

2. Kaur G, Naing NN (2003) Prevalence and ethnic distribution of *Helicobacter pylori* infection among endoscopic patients in North Eastern Peninsular Malaysia. *Mal J Med Sci* 10:66–70
3. Haron A, Mazlam MZ, Aminuddin A, Isa MR, Madhav VK, Che Ghani S (1994) Six year review of gastric carcinoma at the Universiti Kebangsaan Malaysia. *J Perubatan UKM* 16:13–18
4. Tao R, Fang PC, Liu HY, Jiang YS, Chen J (2004) A new subtype of 3' region of *cagA* gene in *Helicobacter pylori* strains isolated from Zhejiang Province in China. *World J Gastroenterol* 10:3284–3288
5. Dong Q, O'Sullivan M, Hall W, Herra C, Kean C, O'Morain C, Buckley M (2002) Identification of a new segment involved in *cagA* 3' region variation of *Helicobacter pylori*. *FEMS Immunol Med Microbiol* 33:51–55
6. Yamazaki S, Kato S, Matsukura N, Ohtani M, Ito Y, Suto H, Yamazaki Y, Yamakawa A, Tokudome S, Higashi H, Hatakeyama M, Azuma T (2005) Identification of *Helicobacter pylori* and the *cagA* genotype in gastric biopsies using highly sensitive real-time PCR as a new diagnostic tool. *FEMS Immunol Med Microbiol* 44:261–268. doi:10.1016/j.femsim.2004.12.011
7. Higashi H, Tsutsumi R, Fujita A, Yamazaki S, Asaka M, Azuma T, Hatakeyama M (2002) Biological activity of the *Helicobacter pylori* virulence factor CagA is determined by variation in the tyrosine phosphorylation sites. *Proc Natl Acad Sci USA* 99:14428–14433. doi:10.1073/pnas.222375399
8. Ramelah M, Aminuddin A, Alfizah H, Isa MR, Jasmi AY, Tan HJ, Rahman AJ, Rizal AM, Mazlam MZ (2005) *cagA* gene variants in Malaysian *Helicobacter pylori* strains isolated from patients of different ethnic groups. *FEMS Immunol Med Microbiol* 44:239–242. doi:10.1016/j.femsim.2005.02.001
9. Yamaoka Y, Kodama T, Kashima K, Graham DY, Sepulveda AR (1998) Variants of the 3' region of the *cagA* gene in *Helicobacter pylori* isolates from patients with different *H. pylori*-associated diseases. *J Clin Microbiol* 36:2258–2263
10. Higashi H, Tsutsumi R, Muto S, Sugiyama T, Azuma T, Asaka M, Hatakeyama M (2002) SHP-2 tyrosine phosphatase as an intracellular target of *Helicobacter pylori* CagA protein. *Science* 295:683–686
11. Yamaoka Y, Osato MS, Sepulveda AR, Gutierrez O, Figura N, Kim JG, Kodama T, Kashima K, Graham DY (2000) Molecular epidemiology of *Helicobacter pylori*: separation of *H. pylori* from East Asian and non-Asian countries. *Epidemiol Infect* 124:91–96. doi:10.1017/S0950268899003209
12. Backert S, Moese S, Selbach M, Brinkmann V, Meyer TF (2001) Phosphorylation of tyrosine 972 of the *Helicobacter pylori* CagA protein is essential for induction of a scattering phenotype in gastric epithelial cells. *Mol Microbiol* 42:631–644. doi:10.1046/j.1365-2958.2001.02649.x
13. Choi KD, Kim N, Lee DH, Kim JM, Kim JS, Jung HC, Song IS (2007) Analysis of the 3' variable region of the *cagA* gene of *Helicobacter pylori* isolated in Koreans. *Dig Dis Sci* 52:960–966. doi:10.1007/s10620-005-9030-z
14. Satomi S, Yamakawa A, Matsunaga S, Masaki R, Inagaki T, Okuda T, Suto H, Ito Y, Yamazaki Y, Kuriyama M, Keida Y, Kutsumi H, Azuma T (2006) Relationship between the diversity of the *cagA* gene of *Helicobacter pylori* and gastric cancer in Okinawa. *Jpn J Gastroenterol* 41:668–673. doi:10.1007/s00535-006-1838-6
15. Azuma T, Yamazaki S, Yamakawa A, Ohtani M, Muramatsu A, Suto H, Ito Y, Dojo M, Yamazaki Y, Kuriyama M, Keida Y, Higashi H, Hatakeyama M (2004) Association between diversity in the Src homology 2 domain-containing tyrosine phosphatase binding site of *Helicobacter pylori* CagA protein and gastric atrophy and cancer. *J Infect Dis* 189:820–827. doi:10.1086/381782
16. Zhou W, Yamazaki S, Yamakawa A, Ohtani M, Ito Y, Keida Y, Higashi H, Hatakeyama M, Si J, Azuma T (2004) The diversity of *vacA* and *cagA* genes of *Helicobacter pylori* in East Asia. *FEMS Immunol Med Microbiol* 40:81–87. doi:10.1016/S0928-8244(03)00299-2
17. Yamazaki S, Kato S, Matsukura N, Ohtani M, Ito Y, Suto H, Yamazaki Y, Yamakawa A, Tokudome S, Higashi H, Hatakeyama M, Azuma T (2005) Identification of *Helicobacter pylori* and the *cagA* genotype in gastric biopsies using highly sensitive real-time PCR as a new diagnostic tool. *FEMS Immunol Med Microbiol* 44:261–268. doi:10.1016/j.femsim.2004.12.011
18. Aras RA, Lee Y, Kim S-K, Israel D, Peek RM Jr, Blaser MJ (2003) Natural variation in populations of persistently colonizing bacteria affect human host cell phenotype. *J Infect Dis* 188:486–496. doi:10.1086/377098
19. Achtman M, Azuma T, Berg DE, Ito Y, Morelli G, Pan ZJ, Suerbaum S, Thompson SA, van der Ende A, van Doorn LJ (1999) Recombination and clonal groupings within *Helicobacter pylori* from different geographical regions. *Mol Microbiol* 32:459–470. doi:10.1046/j.1365-2958.1999.01382.x
20. Occhialini A, Marais A, Urdaci M, Sierra R, Muñoz N, Covacci A, Mégraud F (2001) Composition and gene expression of the *cag* pathogenicity island in *Helicobacter pylori* strains isolated from gastric carcinoma and gastritis patients in Costa Rica. *Infect Immun* 69:1902–1908. doi:10.1128/IAI.69.3.1902-1908.2001