GUIDELINES

Chinese guidelines on the management of liver cirrhosis (abbreviated version)
Xu XY, Ding HG, Li WG, Xu JH, Han Y, Jia JD, Wei L, Duan ZP, Ling-Hu EQ, Zhuang H

MINIREVIEWS

Role of pancreatography in the endoscopic management of encapsulated pancreatic collections – review and new proposed classification
Proença IM, dos Santos MEL, de Moura DTH, Ribeiro IB, Matuguma SE, Cheng S, McCarty TR, do Monte Junior ES, Sakai P, de Moura EGH

ORIGINAL ARTICLE

Basic Study

Discovery of unique African Helicobacter pylori CagA-multimerization motif in the Dominican Republic

Nimbolide inhibits tumor growth by restoring hepatic tight junction protein expression and reduced inflammation in an experimental hepatocarcinogenesis
Ram AK, Vairappan B, Srinivas BH

Case Control Study

Altered metabolism of bile acids correlates with clinical parameters and the gut microbiota in patients with diarrhea-predominant irritable bowel syndrome
Wei W, Wang HF, Zhang Y, Zhang YL, Niu BY, Yao SK

Alteration of fecal tryptophan metabolism correlates with shifted microbiota and may be involved in pathogenesis of colorectal cancer
Sun XZ, Zhao DY, Zhou YC, Wang QQ, Qin G, Yao SK

Retrospective Cohort Study

High mortality associated with gram-negative bacterial bloodstream infection in liver transplant recipients undergoing immunosuppression reduction
Chen F, Pang XY, Shen C, Han LZ, Deng YX, Chen XS, Zhang JJ, Xia Q, Qian YB

Retrospective Study

Liver fibrosis index-based nomograms for identifying esophageal varices in patients with chronic hepatitis B related cirrhosis
Xu SH, Wu F, Guo LH, Zhang WB, Xu HX
## Clinical Trials Study

7222  Relationship between the incidence of non-hepatic hyperammonemia and the prognosis of patients in the intensive care unit  

_Yao ZP, Li Y, Liu Y, Wang HL_

## Observational Study

7232  Association between ADAMTS13 activity–VWF antigen imbalance and the therapeutic effect of HAIC in patients with hepatocellular carcinoma  


## SYSTEMATIC REVIEWS

7242  Diagnosis and treatment of iron-deficiency anemia in gastrointestinal bleeding: A systematic review  

_Cotter J, Baldaia C, Ferreira M, Macedo G, Pedroto I_

## CASE REPORT

7258  Endoscopic mucosal ablation - an alternative treatment for colonic polyps: Three case reports  

_Mendoza Ladd A, Espinosa J, Garcia C_

7263  Tuberous sclerosis patient with neuroendocrine carcinoma of the esophagogastric junction: A case report  

ABOUT COVER
Editorial Board Member of World Journal of Gastroenterology, Professor Udo Rolle is a Distinguished Professor at the University Hospital of the Goethe-University in Frankfurt, Germany. Having received his doctoral degree from the University of Leipzig in 1994, Prof. Rolle undertook his postgraduate training, first at the University of Leipzig (Germany) and then at the University Hospital Dublin (Ireland), receiving his habilitation in 2003. He rose to Full Professor and Chief Surgeon in the Department of Paediatric Surgery and Paediatric Urology at the University Hospital Frankfurt in 2008. His ongoing research interests involve neonatal GI surgery, pediatric GI motility disorders with long-term follow-up in various congenital anomalies. Currently, he serves as President of the German Association of Paediatric Surgery, General Secretary of the World Federation of Associations of Paediatric Surgery, and General Secretary of the UEMS section of Paediatric Surgery. (L-Editor: Filipodia)

AIMS AND SCOPE
The primary aim of World Journal of Gastroenterology (WJG, World J Gastroenterol) is to provide scholars and readers from various fields of gastroenterology and hepatology with a platform to publish high-quality basic and clinical research articles and communicate their research findings online. WJG mainly publishes articles reporting research results and findings obtained in the field of gastroenterology and hepatology and covering a wide range of topics including gastroenterology, hepatology, gastrointestinal endoscopy, gastrointestinal surgery, gastrointestinal oncology, and pediatric gastroenterology.

INDEXING/ABSTRACTING
The WJG is now indexed in Current Contents®/Clinical Medicine, Science Citation Index Expanded (also known as ScSearch®), Journal Citation Reports®, Index Medicus, MEDLINE, PubMed, PubMed Central, and Scopus. The 2020 edition of Journal Citation Report® cites the 2019 impact factor (IF) for WJG as 3.665; IF without journal self cites: 3.534; 5-year IF: 4.048; Ranking: 35 among 88 journals in gastroenterology and hepatology; and Quartile category: Q2.
Discovery of unique African *Helicobacter pylori* CagA-multimerization motif in the Dominican Republic

Takaaki Ono, Modesto Cruz, Hiroyuki Nagashima, Phawinee Subsomwong, Junko Akada, Takashi Matsumoto, Tomohisa Uchida, Rumiko Suzuki, Celso Hosking, José A Jiménez Abreu, Yoshio Yamaoka

**Abstract**

**BACKGROUND**

*Helicobacter pylori* (*H. pylori*) colonizes the human stomach and is a major cause of peptic ulcer disease and gastric cancer. However, although the prevalence of *H. pylori* is high in Africa, the incidence of gastric cancer is low, and this phenomenon is called to be African enigma. The CagA protein produced by *H. pylori* is the most studied virulence factor. The carcinogenic potential of CagA is...
We found the unique African CM motif in Western-type CagA and termed it Africa1-CM. The less toxicity of this motif could be one reason to explain the African enigma.

**Key Words:** *Helicobacter pylori*; CagA; CagA* 3' region; CagA-multimerization motif; Glu-Pro-Ile-Tyr-Ala motif; Dominican Republic

©The Author(s) 2020. Published by Baishideng Publishing Group Inc. All rights reserved.
INTRODUCTION

*Helicobacter pylori* (H. pylori) plays an essential role in the development of various gastroduodenal diseases including gastric cancer[8]. However, only a small proportion of people infected with *H. pylori* develop these diseases. Interestingly, the incidence of gastric cancer in Africa is much lower than that in other countries, despite the high prevalence of *H. pylori* infection in this area[9]. This contradictory phenomenon is known as the “African enigma”. Geographic differences in the incidence of gastric cancer can be explained, at least in part, by the presence of different types of *H. pylori* virulence factor, especially CagA, VacA and OipA[10].

The *H. pylori* CagA protein is one of the well-known pathogenic virulence factors involved in the pathogenesis of gastric cancer[11]. The cagA gene encodes a 120-145 kDa CagA and is located at one end of the cag pathogenicity island (PAI) that encodes the type IV secretion system (TSS)[12]. CagA can directly translocate into host gastric epithelial cells via the TSS and undergoes tyrosine phosphorylation by Src family kinases and then binds to various molecules including the Src homology 2-containing protein tyrosine phosphatase-2 (SHP-2)[13]. Gastric mucosal epithelial cells form a monolayer of polarized cells through tight junctions. When CagA binds to the kinase domain of partitioning-defective 1b (PAR1b)/microtubule affinity-regulating kinase 2 (MARK2), it destroys the tight junctions in a tyrosine phosphorylation-independent manner and causes loss of epithelial cell polarity[14]. CagA multicellularizes in cells independently of tyrosine phosphorylation, and a CagA-multimerization (CM) sequence consisting of 16 amino acids has been identified as its responsible region[15].

There is also a report that the CM sequence is called conserved repeat responsible for phosphorylation-independent activity[16]. This CM sequence is essential for binding between CagA and PAR1. PAR1 forms a homomultimer in the cell. Thus, CagA forms a multimer indirectly through binding to the PAR1 multimer and binds to SHP-2 after being tyrosine-phosphorylated[17].

The tyrosine phosphorylation site of CagA is characterized by the presence of a unique Glu-Pro-Ile-Tyr-Ala (EPIYA) motif, which is present in multiple numbers in the C-terminal region of CagA[18]. The EPIYA-repeat region is composed of various combinations of four distinct EPIYA segments, EPIYA-A, -B, -C and -D, based on the sequences surrounding each of the EPIYA motifs[19]. *H. pylori* cagA-positive strains isolated in Western countries possess the Western CagA, which contains EPIYA-A, EPIYA-B, and EPIYA-C segments, whereas that in East Asian countries possess the East Asian CagA, which contains EPIYA-A, EPIYA-B and EPIYA-D segments. We previously showed that the East Asian CagA was more virulent than the Western CagA[20]. The EPIYA-D motif of the East Asian CagA exhibits a stronger SHP-2 binding affinity than the EPIYA-C motif of the Western CagA[21]. In addition, the number of EPIYA-C sites is directly correlated to the level of tyrosine phosphorylation and SHP-2 binding activity among the Western CagA species.

Eleven of the 16 amino acids in the CM motif are well conserved between the Western and East Asian cagA species[9]. The peptide sequence of the typical Western CM motif (W-CM), which is observed in the Western cagA, is “FPLKRHDKVD DLSKGV” and that of the typical East CM motif (E-CM), which is observed in the East Asian cagA, is “FPLRSHAADLSKGV”[11,12]. In the Western cagA species, each EPIYA-C segment contains a single CM motif, and there is also another copy of the CM motif immediately distal to the last repeat of the EPIYA-C segment. In the East Asian cagA species, there is a CM motif that is located immediately distal to each repeat of the EPIYA-D segment. Thus, the Western cagA carries at least two W-CM sequences, whose number increases in parallel with the number of EPIYA-C segments. The ability of the Western CagA to bind to PAR1 is proportional to the number of W-CM sequences[11]. Meanwhile, a single E-CM sequence has almost twice the PAR1 binding activity of the single W-CM sequence.

We have previously reported that all 64 *H. pylori* strains isolated in 2011 in the Dominican Republic carried the Western cagA specific sequences[22]. However, this report did not mention the CM motif. In the subsequent study in the Dominican Republic, we have already performed a multilocus sequence typing (MLST) analysis of seven housekeeping genes for a total of 119 *H. pylori* species, which were collected in 2011 and 2016, followed by population structure analysis using STRUCTURE software[23]. Therefore, *H. pylori* strains of the Dominican Republic were divided into two populations: 68 strains with hpAfrica1 [46, 5, and 17 with hspWAfrica, hspSAfrica, and a hybrid (hspWAfrica/hpEurope), respectively], and 51 strains with hpEurope (47 and 4 with hspEurope and hspEurope, respectively). The ethics in the Dominican Republic consists of 16% of European, 11% of African, and 73% of mixed race[24]. Therefore, it is important to investigate the genetic diversity of the cagA gene in the
Dominican Republic, where human demography consists of various ancestries. In this study, we evaluated the correspondence between the CM motif and phylogeographical classification using MLST population structure in *H. pylori* in the Dominican Republic. We also examined the relationship between these types and gastric mucosal damages.

**MATERIALS AND METHODS**

**Study population and DNA extraction**

We recruited outpatients with mild dyspeptic symptoms living in the Dominican Republic. Gastric mucosal biopsy specimens were taken from 258 dyspeptic patients (158 in 2011 and 100 in 2016; 86 males and 172 females; age range, 17-91 years; mean age, 46.2 ± 15.8 years) who underwent endoscopy examination at the Digestive Disease Center (Dr Luis E. Aybar Health and Hygiene City, Santo Domingo, Dominican Republic). Patients with a history of partial gastric resection or previous treatment for *H. pylori* infection were excluded. During each endoscopy session, 4 gastric biopsy specimens were obtained (three from the antrum and one from the corpus). The three specimens from the antrum were used for *H. pylori* culture, rapid urease test and histological examination. The specimen from the corpus was used for histological examination.

Gastric biopsy specimens for histological examination were embedded in paraffin after being fixed in 40 g/L formaldehyde. Hematoxylin and eosin and May-Giemsa were selected as stain techniques. The updated Sydney system was used to analyze histological severity\(^{[16]}\). The degree of the bacterial load was classified into four grades: 0, “normal”; 1, “mild”; 2, “moderate”; and 3, “marked”. Peptic ulcer and gastric cancer were identified via endoscopy, and gastric cancer was confirmed by histopathology. Gastritis was diagnosed in the absence of peptic ulcer or gastric malignancy. For *H. pylori* culture, antral biopsy specimens were homogenized and inoculated onto antibiotic selection plates, and then subcultured on Mueller Hinton II Agar medium (Becton Dickinson, Sparks; MD, United States) supplemented with 7% horse blood without antibiotics. The plates were incubated up to 10 d at 37 °C under microaerophilic conditions (100 mL/L O\(_2\), 50 mL/L CO\(_2\), and 850 mL/L N\(_2\)). *H. pylori* isolates were identified based on colony morphology, Gram staining results, and oxidase, catalase, and urease reactions. Isolated strains were stored at −80 °C in *Brucella* broth (Becton Dickinson, Sparks; MD, United States) containing 10% dimethyl sulfoxide and 10% horse serum. Bacterial DNA was extracted using a commercially available kit (QIAGEN Inc.; Valencia, CA, United States). Eventually, 64 strains were cultured from 158 patients in 2011, and 56 strains were cultured from 100 patients in 2016, thus a total of 120 *H. pylori* strains were obtained. The ethnicity of the 120 patients, based on self-assessment at the time of medical examination, was 114 multiracial (32 males, 82 females) and 6 Africans (3 males, 3 females).

**Analysis of *H. pylori* population structure**

A detailed description of *H. pylori* population structure methods has already been published in our previous study\(^{[14]}\).

**Analysis of *H. pylori* cagA**

The *cagA* gene was determined using polymerase chain reaction PCR-based sequencing as described previously\(^{[10,17]}\). The absence of *cagA* was confirmed by the presence of *cagA* empty site, as previously described\(^{[10]}\). The *cagA* genotype (East Asian-type and Western-type) was confirmed by sequencing the PCR products as described previously\(^{[10]}\). The EPIYA-A region and CM motif from the strains were compared using the program WebLogo (version 3.7.4) (http://weblogo.threeplusone.com/).

**Statistical analysis**

Statistical significance was tested using Fisher’s exact test and Wilcoxon rank-sum test performed in the R package (version 3.4.0). A *P* value of < 0.05 was accepted as statistically significant.
RESULTS

H. pylori infection rate and endoscopic findings
We performed endoscopy for 258 dyspeptic patients. The prevalence of *H. pylori* infection determined via histological examination was 67.8% (175/258). Clinical diagnosis of them were 219 cases with chronic gastritis, 38 cases with peptic ulcer and 1 case with gastric cancer. From 258 dyspeptic patients, a total of 120 strains were isolated (120/258, 46.5%). Of these, 93 strains were isolated from subjects with chronic gastritis, 26 from peptic ulcer subjects, and 1 from a gastric cancer subject (Supplementary Table 1).

H. pylori population structure
As we have already reported[14], the Dominican Republic strains were assigned to either hpAfrica1 (57.1%, 68/119) or hpEurope (42.9%, 51/119) (Supplementary Table 1). From the result of the STRUCTURE run of linkage model at K = 4, the four major ancestral components reported in the previous studies were observed: Ancestral Africa1, ancestral Europe 1, ancestral Europe 2, and ancestral East Asia.

CagA sequences and EPIYA phosphorylation motif patterns
In total, 84 of 120 (70.0%) strains possessed the *cagA* gene. The *cagA* gene was not detected in one strain; however, this strain contained partial *cag* PAI genes, as confirmed by *cag* PAI empty site. Thus, we considered this strain as “cagA undetermined” (Supplementary Table 1). The prevalence of *cagA* was compared between the clinical outcomes of patients (Table 1). No significant difference was found between gastritis and peptic ulcer patient groups. All 84 *cagA*-positive strains were classified as Western CagA: ABC (*n* = 72), ABCC (*n* = 7), AC (*n* = 2), AABC (*n* = 1), AB (*n* = 1), and ABBC (*n* = 1). Interestingly, in 30 *cagA*-positive strains, the “GKDKGPE” motif was observed immediately upstream of the EPIYA motif in the EPIYA-A segment, which was not seen in the typical Western-type and East Asian-type CagA (Figure 1, Supplementary Table 1).

CM motifs and patterns
In total, 172 CM motifs were found in 84 *cagA*-positive strains (82 for 1st CM motif, 83 for 2nd CM motif, and 7 for 3rd CM motif). These CM motifs were classified into 27 types. Twelve types were observed in the 1st CM motif, 19 types in the 2nd CM motif, and 3 types in the 3rd CM motif (Table 2). The CM motif type that had the highest number of appearances was the typical W-CM motif (FPLKRHDKVDDLSKVG) observed in strains circulating in Western countries[20], with 45 found in the 1st CM motif, 31 found in the 2nd CM motif, and 4 found in the 3rd CM motif. Interestingly, no typical E-CM motif (FPLRRSAVNDLSKVG) was observed in strains circulating in East Asian countries. Furthermore, the Amerindian CM motif (Am-CM: xLKRxAKVDLxKxG and YTLxMHAGDxNLxKVG) of Amerindian CagA was not observed[22,23]. However, “FPLRxxAKVDLxKVG”, which was found in 21 strains similar to the typical E-CM motif, was found in the 2nd and 3rd CM motifs. Amino acid sequence comparison with the GenBank BLAST data (http://www.ncbi.nlm.nih.gov/BLAST/) indicated that the “FPLRRSKVDLxKVG” sequence mainly matched with the sequences in the strains in the African (Senegal, South Africa, and Gambia) and American (United States, Colombia, Mexico, and Cuba) countries (Supplementary Table 1). In previous studies[22,23], this motif was classified as a subtype of the typical E-CM motif. Interestingly, there were no East Asian and Amerind strains in the list of matches from the GenBank BLAST search. As reported in our previous study[14], many of the strains used in this study were African type (hpAfrica1) (Supplementary Table 1). Therefore, we hypothesized that “FPLRRSKVDLxKVG” is the typical Africa1 CM (Af1-CM) motif specific to a strain derived from Africa (hpAfrica strain). A previous study[23] noted that W-CM and E-CM varied at positions 4, 6, 7, 8 and 10 (FPLxRxVxDLxKVG). In the type classification of CM motif in this study, we focused on these five sequences, and when four or more positions of the sequences of a strain matched the typical W-CM, we designated a classification criterion that this strain belongs to W-CM. Similarly, when four or more positions of the sequences of a strain match the typical E-CM or the typical Af1-CM, we designated a classification criterion that this strain belongs to E-CM or Af1-CM, respectively. Other types of sequences not classified by these criteria were assigned as different CM (D-CM).

As shown in Table 2, in the 1st CM motif, 76 strains were W-CM and 6 were D-CM. In the 2nd CM motif, 37 strains were W-CM, 34 were Af1-CM, 1 was E-CM and 11 were D-CM. In the 3rd CM motif, 5 strains were W-CM and 2 were Af1-CM. Table 3 shows
Table 1 Association between the prevalence of cagA and clinical outcomes

<table>
<thead>
<tr>
<th></th>
<th>Total</th>
<th>Gastritis</th>
<th>Peptic ulcer</th>
<th>Gastric cancer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of strains</td>
<td>120</td>
<td>93 (76.0%)</td>
<td>26</td>
<td>1</td>
</tr>
<tr>
<td>cagA positive</td>
<td>84</td>
<td>66 (71.0%)</td>
<td>18 (69.2%)</td>
<td>0</td>
</tr>
<tr>
<td>cagA negative</td>
<td>35</td>
<td>26 (74.3%)</td>
<td>8 (30.8%)</td>
<td>1 (100%)</td>
</tr>
<tr>
<td>cagA undetermined</td>
<td>1</td>
<td>1 (1.1%)</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 2 Peptide sequences and types of CagA-multimerization motif in Helicobacter pylori strains

<table>
<thead>
<tr>
<th>First motif Peptide sequences</th>
<th>Type</th>
<th>No.</th>
<th>Second motif Peptide sequences</th>
<th>Type</th>
<th>No.</th>
<th>Third motif Peptide sequences</th>
<th>Type</th>
<th>No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>FPLKRHDKVDDLSKVG</td>
<td>W-CM</td>
<td>45</td>
<td>FPLKRHDKVDDLSKVG</td>
<td>W-CM</td>
<td>31</td>
<td>FPLKRHDKVDDLSKVG</td>
<td>W-CM</td>
<td>4</td>
</tr>
<tr>
<td>FPLKRHDKVDDLSKVG</td>
<td>W-CM</td>
<td>17</td>
<td>FPLRRSAKVDLSKVG</td>
<td>Af1-CM</td>
<td>21</td>
<td>FPLRRSAKVDLSKVG</td>
<td>Af1-CM</td>
<td>2</td>
</tr>
<tr>
<td>FPLKRHKVDDLSKVG</td>
<td>W-CM</td>
<td>4</td>
<td>FPLKHKVDDLSKVG</td>
<td>W-CM</td>
<td>7</td>
<td>FLLKRHDKVDDLSKVG</td>
<td>W-CM</td>
<td>1</td>
</tr>
<tr>
<td>FPLKHKAKVDDLSKVG</td>
<td>W-CM</td>
<td>3</td>
<td>FPLRRSAKVDLSKAG</td>
<td>Af1-CM</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FPLKHKVDDLSKVG</td>
<td>W-CM</td>
<td>3</td>
<td>FPLRRSAKVDLSKVG</td>
<td>Af1-CM</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FPLKHKVDDLSKVG</td>
<td>W-CM</td>
<td>2</td>
<td>FLLKRSAKVDLSKVG</td>
<td>D-CM</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FLLKRHDKVDDLSKVG</td>
<td>W-CM</td>
<td>1</td>
<td>FLLKRSAKVDLSKVG</td>
<td>Af1-CM</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FPLRSKVDVLDDLKVG</td>
<td>D-CM</td>
<td>1</td>
<td>FPLKRHDVLDDLKVG</td>
<td>W-CM</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FLLKRHDVLDDLKVG</td>
<td>D-CM</td>
<td>1</td>
<td>FPLKRHDVLDDLKVG</td>
<td>W-CM</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FLLKRHDVLDDLKVG</td>
<td>D-CM</td>
<td>1</td>
<td>FPLKRHDVLDDLKVG</td>
<td>W-CM</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FPLRKAKVDVLDDLKVG</td>
<td>D-CM</td>
<td>1</td>
<td>FLLKRSTKVDLSKAG</td>
<td>Af1-CM</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nothing (Deletion)</td>
<td></td>
<td>2</td>
<td>FPLRSAAVDDLSKVG</td>
<td>E-CM</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>FLLRRAVDDLSKVG</td>
<td>Af1-CM</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>FLLRSAVDDLSKVG</td>
<td>Af1-CM</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>FLLRSDVDDLSKVG</td>
<td>D-CM</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>FPDKHDVDDLSKVG</td>
<td>D-CM</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>FPDKHDVDDLSKVG</td>
<td>Af1-CM</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>FPLRSKVDVLDDLKVG</td>
<td>Af1-CM</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>84</td>
<td>Total</td>
<td>83</td>
<td>Total</td>
<td>7</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The CagA-multimerization (CM) motif types were determined based on comparison with the typical Western CM motif (FPLKRHDKVDDLSKVG), the typical East Asian CM motif (FPLRRSAVDDLSKVG), and the typical Africa1 CM motif (FPLRRSAKVDLSKVG). W-CM: Western CM motif; E-CM: East-Asian CM motif; Af1-CM: Africa1-CM motif; D-CM: Different CM motif.

Number of strains containing Af1-CM, D-CM and E-CM as observed in the GenBank BLAST data

To count the appearance frequencies of 27 kinds of CM motifs observed in the Dominican Republic in each country of the world, a GenBank BLAST search was performed. The results are shown in Supplementary Table 2. The three W-CM peptide sequences of FPLKRHDKVDDLSKVG, FPLKRHDKVDDLSKVG and FPLKHKHDVDDLKVG had more than 100 matches in the database, therefore, it was difficult to create a detailed list. However, these three peptide sequences were contained in strains around the world, with no regional bias.

The combinations of the 1st, 2nd and 3rd CM motifs. The three frequent types were W-CM:W-CM (n = 30), followed by W-CM:Af1-CM (n = 27), and W-CM:D-CM (n = 10).
Table 3 *CagA*-multimerization motif patterns in *Helicobacter pylori* strains

<table>
<thead>
<tr>
<th>CM motif pattern</th>
<th>Occurrence</th>
<th>Frequency (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>W-W</td>
<td>30</td>
<td>35.7</td>
</tr>
<tr>
<td>W-Af1</td>
<td>27</td>
<td>32.1</td>
</tr>
<tr>
<td>W-D</td>
<td>10</td>
<td>11.9</td>
</tr>
<tr>
<td>D-Af1</td>
<td>5</td>
<td>6.0</td>
</tr>
<tr>
<td>W-W-W</td>
<td>5</td>
<td>6.0</td>
</tr>
<tr>
<td>W-Af1-Af1</td>
<td>2</td>
<td>2.4</td>
</tr>
<tr>
<td>W-</td>
<td>2</td>
<td>2.4</td>
</tr>
<tr>
<td>D-D</td>
<td>1</td>
<td>1.2</td>
</tr>
<tr>
<td>W-E</td>
<td>1</td>
<td>1.2</td>
</tr>
<tr>
<td>W-</td>
<td>1</td>
<td>1.2</td>
</tr>
<tr>
<td>Total</td>
<td>84</td>
<td></td>
</tr>
</tbody>
</table>

*CagA*-multimerization (CM) motif patterns were determined based on the identification of CM sequences that were located before and after the (Glu-Pro-Ile-Tyr-Ala)-C motif. W-W: Strains with two Western CM motifs in their CagA; W-Af1: Strains with one Western and one Africa1 CM motif in their CagA; W-D: Strains with one Western and one Different CM motif in their CagA; D-Af1: Strains with one Different and one Africa1 CM motif in their CagA; W-W-W: Strains with three Western CM motifs in their CagA; W-Af1-Af1: Strains with one Western and two Africa1 CM motifs in their CagA; -W: Indicates that only one Western CM motif was observed following the Glu-Pro-Ile-Tyr-Ala (EPIYA)-C motif, in the absence of a CM motif before the EPIYA-C motif; D-D: Strains with two Different CM motifs in their CagA; W-E: Strains with one Western and one East Asian CM motif in their CagA; W-: Indicates that only one Western CM motif was observed before the EPIYA-C motif, in the absence of a CM motif following the EPIYA-C motif.

**Relationship between cagA genotypes and phylogeographical classification using MLST population structure**

Of the 68 *hpAfrica1* strains, 61 were *cagA* positive (92.8%), 6 were *cagA* negative (9.8%), and one could not determine the presence of *cagA* (Supplementary Table 1). In contrast, of the 51 *hpEurope* strains, only 23 were *cagA* positive (45.1%) and 28 were *cagA* negative (54.9%). The difference in *cagA* prevalence between the *hpAfrica1* and the *hpEurope* strains was significant (*P* < 0.001, Fisher’s exact test). Next, to verify whether the Af1-CM motif was related to the *hpAfrica1* strain, it was compared with the phylogeographical classification using MLST population structure. Among 34 strains containing Af1-CM in the 2nd CM motif, 29 were *hpAfrica1* (85.3%) and 5 were *hpEurope* (14.7%). In contrast, of the 37 strains containing the W-CM motif in the 2nd CM motif, 23 were *hpAfrica1* (62.2%) and 14 were *hpEurope* (37.8%). The difference in CM motif type between the *hpAfrica1* and the *hpEurope* strains was significant (*P* = 0.034, Fisher’s exact test). Next, among 23 *cagA*-positive strains of *hpEurope*, the ratio of the ancestral Africa1 component, which was calculated using the STRUCTURE run of linkage model, was compared between 5 strains having Af1-CM motif and 18 strains without Af1-CM motif. From the result, the 5 strains having the Af1-CM motif had a significantly higher ancestral Africa1 component than the 18 strains without the Af1-CM motif (*P* = 0.030, Wilcoxon rank-sum test; Figure 2).

**“GKDKGPE” motif unique to *hpAfrica1* strains**

Harrison *et al.* hypothesized that the “KDKGPE” motif in the EPIYA-A segment of some Western *CagA* strains could be a unique African motif. Beltrán-anaya *et al.* have also noted the appearance of “GKDKGPE” in some Mexican strains. Amino acid sequence comparison with the GenBank BLAST data ([http://www.ncbi.nlm.nih.gov/BLAST/](http://www.ncbi.nlm.nih.gov/BLAST/)) indicated that the “GKDKGPEPIY AQVNK” sequence observed in many of the strains in this study was mainly matched with the sequences of strains in the African (Senegal, Gambia, and Sudan) and American (United States, Colombia, Mexico, and Cuba) continents. To verify whether the “GKDKGPE” motif was related to the *hpAfrica1* strain, it was compared with the phylogeographical classification using MLST population structure. Among 30 strains containing the “GKDKGPE” motif, 28 were *hpAfrica1* and 2 were *hpEurope*. In contrast, after excluding the two AC-type strains, among the 52 strains without the “GKDKGPE” motif, 33 were *hpAfrica1* and 19 were *hpEurope*. The presence of the “GKDKGPE” motif was significant between the *hpAfrica1* and the *hpEurope* strains (*P* = 0.003,
Figure 1 Variation in the CagA (Glu-Pro-Ile-Tyr-Ala)-A region from the Dominican Republic strains. A: CagA Glu-Pro-Ile-Tyr-Ala (EPIYA)-A region in 52 strains without containing “GKDKGPE” motif (excluding two AC-type strains); B: CagA EPIYA-A region in 30 strains containing “GKDKGPE” motif. The reference sequences\(^{19}\) are shown at the bottom of the figure (green); C: Most frequent Western-type CagA EPIYA-A region; D: Most frequent East Asian-type CagA EPIYA-A region.

Figure 2 Box plot diagram of the ancestral Africa1 components in the hpEurope population classified by the presence or absence of the Africa1 CagA-multimerization motif. The difference in the ancestral Africa1 ratio between the bacterial populations was investigated using the Wilcoxon rank-sum test. AA1: Ancestral Africa1; AF1-CM: Africa1 CagA-multimerization.
Fisher’s exact test). In addition, another characteristic was observed in the four amino acid sequences immediately upstream of the EPIYA motif in the EPIYA-A segment among the 30 strains containing the “GKDKGPE” motif; 29 were QVNK and 1 was KVNK (Supplementary Table 1). In contrast, among 52 strains in which the “GKDKGPE” motif was not conserved, 12 strains were QVNK, and 40 strains were KVNK. The presence of the “GKDKGPE” motif was significant between the QVNK and KVNK motifs (P < 0.001, Fisher’s exact test).

**Relationship between the CM and “GKDKGPE” motifs**
Among 37 strains containing the W-CM motif in the 2nd CM motif, the “GKDKGPE” motif was conserved in 3 strains and not in 34 strains (Supplementary Table 1). In contrast, among 34 strains containing the Af1-CM motif in the 2nd CM motif, the “GKDKGPE” motif was conserved in 22 strains and not in 12 strains. The presence of the “GKDKGPE” motif was significant between the two CM motif types in the 2nd CM motif (P < 0.001, Fisher’s exact test).

**CM motif patterns and gastric disorders**
Histological scores according to the CM motif patterns are shown in Supplementary Table 3. No significant difference was observed between the two groups in all histological scores. Association of CM motif patterns with clinical outcomes is shown in Supplementary Table 4. No correlation was found between clinical results and CM motif patterns.

**Nucleotide sequencing**
Sequencing data for the cagA genes of *H. pylori* is available under DDBJ accession numbers AB860373-AB860407, AB860409-AB860411, and LC546765-LC546810.

**DISCUSSION**
All 84 cagA-positive strains in this study were classified as Western CagA. The hpAfrica1 strains had a significantly higher ratio of having the cagA gene than the hpEurope strains (P < 0.001, Fisher’s exact test). This ratio was similar to that of the previous study. Twenty-seven kinds of CM motifs were observed in 84 cagA-positive strains isolated in the Dominican Republic, which is higher in number than those of reports in many countries around the world. In contrast, this rich variety of CM motifs in the Dominican Republic was similar to that of reports in Colombia and New York.

The CM motif type with the highest number of occurrences was the typical W-CM motif, but no typical E-CM motif or Am-CM was observed. Interestingly, a large number of CM motifs (FPLRRSAKVEDLSKVG) having a sequence similar to that of the typical E-CM motif were observed in the 2nd CM motif (also in some 3rd CM motifs), and we termed it Africa1-CM (Af1-CM). The proportion of Af1-CM in the 2nd CM motif in the hpAfrica1 strains was significantly higher than that in the hpEurope strains (P = 0.034, Fisher’s exact test). This significant difference supported our hypothesis that Af1-CM is a unique CM motif in an African-derived strain (hpAfrica1). This Af1-CM does not appear in the 1st CM motif but seems to appear in the 2nd or 3rd CM motif. Interestingly, no strain with the same sequence as the five kinds of Af1-CM subtypes was found in the African continent using the GenBank BLAST search (Supplementary Table 2). Thus, although similar in sequence to the original Af1-CM motif, these strains may not be of African origin and hence further investigation is required. No typical E-CM (FPLRRSAAVNDLSKVG) was observed in the 84 cagA-positive strains isolated in the Dominican Republic, but a subtype of E-CM (FPLRRSAAVDDLSKVG), one amino acid difference from E-CM, was found. However, the fourth Arginine and the sixth Serine were also included in the sequence of the typical Af1-CM, and the tenth was Aspartic acid instead of Asparagine specific to E-CM. Therefore, if different classification criteria are set, this motif may be classified into the Af1-CM subtype. Seven kinds of CM motifs were classified into D-CM. Since these motifs contain amino acids characteristic of W-CM and Af1-CM respectively, their origin may be a recombinant strain, or a completely different ancestral strain that has not yet been surveyed. The same reason may be assigned to why many types of subtypes of W-CM, E-CM, and Af1-CM appeared.

Olbermann et al. compared the CM motifs of 38 representative cagA-positive strains of the world. Among them, strain CC42C (FPLRRSAKVEDLSKVG, hpAfrica1, South Africa) and strain D3a (FPLRRSAKVEDLSKVG, hpAfrica1, Senegal) had the same motif as Af1-CM proposed by our group. The fact that the phylogeographical
classification types of two strains are hpAfrica1 supports our Af1-CM hypothesis. Strain LSU2003-1 (FPLRKSAKVDDLSDKVG, hpAfrica1, United States) is classified as D-CM according to the classification criteria defined in our study, but this strain also contains some amino acids characteristic of Af1-CM.

Sicinschi et al.\cite{23} reported that of the 42 Western CagA (ABC) strains collected in Colombia, 13 had the E-CM motif, of which 12 were collected from the low-risk gastric cancer region (residents of Pacific coast). According to the classification criteria defined in our study, it follows that 13 strains classified as E-CM in their paper are classified as Af1-CM. Thus, these observations in Colombia support our hypothesis. Hatakeyama\cite{10,30} considered in a review paper that the 1st CM and 2nd CM of two strains of Western CagA (ABC) collected in Mexico were both E-CM, which is a rare case. However, this E-CM motif has the same sequence as the typical Af1-CM motif proposed by our group. It makes sense when we consider that these strains are descendants of African strains from the slave trade era. Ogrodnik et al.\cite{31,32} reported that 27 out of 42 Western CagA strains collected at a New York Hospital had an E-CM motif. In addition, when compared with the clinical diagnosis, the strain containing E-CM motif was less toxic than that containing W-CM motif. The main E-CM motifs of these strains are FPLRRSAKVEDLSKVVG, FPLRKSAKVDDLSDKVG, and FPLRKSAKVELSKVG, which are the same motifs as Af1-CM proposed by our group. The majority of \textit{H. pylori} infected people treated at this New York hospital are immigrants from the Caribbean, such as Jamaica, the Dominican Republic, and Haiti. Thus, it also makes sense when we consider that these strains are descendants of African strains from the slave trade era.

Ahire et al.\cite{33} created both E-CM motif-transfected and W-CM motif-transfected strains and then compared the toxicities of these strains by testing with gastric epithelial cells. The E-CM motif used in that cloning study had the same sequence as the typical Af1-CM motif proposed by our group. This study also reported that strains with the E-CM motif (Af1-CM motif in the classification criteria in our paper) induced less hummingbird elongation than strains with the W-CM motif. Nesić et al.\cite{24} showed that the third Leucine, fifth Arginine, ninth Valine, and twelfth Leucine of the 14 amino acid CM motifs were four critical amino acid residues involved in the MARK2 kinase binding domain of CagA, via crystal structure analysis using X-ray diffraction. In the Af1-CM motif that we found in our study, these four amino acids were always conserved, similar to the E-CM and W-CM motifs, and thus it does not seem to affect this interaction.

Harrison et al.\cite{34} hypothesized that the "KDKGPE" motif in the EPIYA-A segment of some Western CagA strains could be a unique African motif. Amino acid sequence comparison with the GenBank BLAST data indicated that the 'GDKGPEEPIY AQVNK' sequence observed in many strains in this study mainly matched with sequences of strains in the African and American continents. Furthermore, there was a significant association between strains with the hpAfrica1 population and those containing the "GDKGPE" motif \( (P = 0.003, \) Fisher’s exact test). These results supported the hypothesis by Harrison et al. that "GDKGPE" is a unique motif in an African-derived strain (hpAfrica1). In addition, many of the four amino acid sequences immediately downstream of the "GDKGPEEPIYA" motif were "QVNK" \( (P < 0.001, \) Fisher’s exact test). According to Xia et al.\cite{35}, "KVNK" predominates in the Western CagA (ABC), while "QVNK" predominates in the East Asian CagA (ABD). Interestingly, all CagA types in our study are Western CagA. Therefore, "QVNK" dominates the four amino acid sequences immediately downstream of the "GDKGPEEPIYA" motif of the Western CagA strain derived from Africa (hpAfrica1) having the "GDKGPE" motif.

In our study, a few hpEurope strains had the Af1-CM and "GDKGPE" motifs. \textit{H. pylori} frequently undergoes recombination among unrelated strains\cite{19,20}. Among the hpEurope strains, having the Af1-CM motif had a significantly higher ancestral African component than not having it \( (P = 0.030, \) Wilcoxon rank-sum test). This result suggests that the ancestors of the hpEurope strains that had the Af1-CM motif underwent genetic recombination with the hpAfrica1 strains in the past.

In this study, no association was found between the CM motif patterns and gastric disorders. According to previous studies\cite{12,21}, \textit{H. pylori} strains with Af1-CM are less toxic compared to strains with the W-CM motif. Thus, the less toxicity of Af1-CM motif could be one reason to explain the African enigma. Twenty-seven kinds of CM motifs were observed in 84 strains of the Western-type CagA in the Dominican Republic. In the future, it is necessary to evaluate the relationship between these various CM motifs and their toxicity by conducting tests such as hummingbird phenotype \textit{in vitro}.

To our knowledge, this report is the first study to elucidate the diverse
characteristics of the CM motif types of the Western-type CagA strains in the Dominican Republic. However, there are several limitations in this study. Firstly, the number of strains involved was small. For a better understanding, further studies should use a large number of samples, balanced for each diagnosis. Secondly, although our samples were taken at a national reference hospital in Santo Domingo city, the capital of the Dominican Republic, our results cannot be generalized across the entire region of the Dominican Republic since the physical and cultural landscape varies by the geographic region in the Dominican Republic.

CONCLUSION

In conclusion, we found the unique African (hpAfrica1-type) CM motif in the Western-type CagA in the Dominican Republic and termed it Africa1-CM. The less toxicity of this motif could be one reason to explain the African enigma. In addition, a few hpEurope strains carrying the Af1-CM motif had a significantly higher ancestral Africa1 component than those without the Af1-CM motif. This result suggests that the ancestors of these hpEurope strains, having the Af1-CM motif, underwent genetic recombination with the hpAfrica1 strains in the past.

ARTICLE HIGHLIGHTS

Research background
Helicobacter pylori (H. pylori) plays an essential role in the development of peptic ulcer disease and gastric cancer. The CagA protein produced by H. pylori is the most studied virulence factor.

Research motivation
Gastric cancer is one of the most common cancers.

Research objectives
The present study aimed to evaluate the correspondence between the CagA-multimerization (CM) motif and phylogeographical classification using multilocus sequence typing (MLST) population structure in H. pylori in the Dominican Republic, which could be helpful to elucidate the African enigma.

Research methods
The Glu-Pro-Ile-Tyr-Ala (EPIYA) pattern and CM motif genotypes were determined using a polymerase chain reaction-based sequencing. The population structure was analyzed using MLST. Peptic ulcer disease and gastric cancer were identified via endoscopy, and gastric cancer was confirmed by histopathology.

Research results
Many CM motifs, which are the amino acid sequences of "FPLRRSAKVEDLSKVG", were found. This type was significantly more frequent in strains classified as hpAfrica1 using MLST analysis ($P = 0.034$).

Research conclusions
We found the unique African CM motif in the Western-type CagA in the Dominican Republic and termed it Africa1-CM.

Research perspectives
In the future, it is necessary to evaluate the relationship between these various CM motifs and their toxicity by conducting tests such as hummingbird phenotype in vitro.

ACKNOWLEDGEMENTS

We thank members of our laboratories for the discussions and comments.
REFERENCES


Ono T et al. Discovery of unique African CagA-multimerization motif


