KF To 杜家輝
WK Leung 梁偉強
KW Ng 吳國偉
JHM Tong 唐紅文
RWM Lung 龍偉明

Sequencing analysis of the 3′ region of the cagA gene in *Helicobacter pylori* isolated from Hong Kong Chinese patients

Key Messages

1. The cagA gene is suggested as a marker for *Helicobacter pylori* strains with enhanced virulence, and diversity in the 3' region of the gene may be associated with different biological activities of various strains.
2. We conducted the largest and most comprehensive series of sequence analyses of the *H pylori* cagA 3' region and documented sequence variants prevailing in Hong Kong.
3. Three subtypes (1, 2, and 3) of the cagA gene were classified based on the type and number of amino acid repeats in the 3' region. Although subtypes 2 and 3 (containing 4 copies of R1 EPIYA repeats) significantly correlated with aggressive disease, subtype 1 remained the most common strain in Hong Kong (96% of the samples).
4. 964-T/S/V variants carrying mutations in a residue adjacent to the tyrosine phosphorylation site were detected in benign lesions only. This observation should be confirmed in a larger sample, together with a functional analysis to elucidate the biological significance of these variants.

Hong Kong Med J 2010;16(Suppl 4):S8-12

The Chinese University of Hong Kong:
Department of Anatomical and Cellular Pathology
KF To, JHM Tong, RWM Lung
Department of Medicine and Therapeutics
WK Leung
Department of Surgery
KW Ng

RFCID project number: 03040552

Principal applicant and corresponding author:
Prof Ka Fai To
Department of Anatomical and Cellular Pathology, The Chinese University of Hong Kong, Prince of Wales Hospital, Shatin, NT, Hong Kong SAR, China
Tel: (852) 2632 2361
Fax: (852) 2637 6274
Email: kfto@cuhk.edu.hk

Introduction

Helicobacter pylori (HP) is the aetiological agent causing chronic gastritis, peptic ulcer and gastric cancer. Strain-specific genetic diversity of HP is postulated to influence the clinical outcome of HP infection, but with considerable geographic differences. The cagA gene (cytotoxin-associated gene A) is considered a marker for a pathogenicity island of about 40 kbs.[1] The presence of cagA is associated with duodenal ulcer, gastric ulcer, gastric mucosal atrophy and gastric cancer. Variations in the number of repeats of the amino acid sequence R1 (EPIYA) have been identified in Japanese HP isolates.[2] The cagA genotype containing four or more R1 sequences is associated with atrophic gastritis and gastric cancer. Hong Kong has a high prevalence of HP infection and a high incidence of gastroduodenal diseases. About 90% of the HP strains are cagA positive. Since allelic variation in cagA exists and distinct HP subgenotypes may circulate in different regions, such differences may provide markers for differences in virulence among cagA-positive HP strains. Data on the sequence variations of HP cagA gene in Hong Kong inhabitants are scanty. A large-scale, comprehensive study is required to determine the genotypes of HP cagA gene in Hong Kong.

Methods

This study was conducted from January 2006 to January 2007. DNA of HP was extracted from both HP isolates and directly from gastric biopsy samples of subjects presenting with gastrointestinal symptoms (including dyspepsia and gastro-intestinal bleeding) and asymptomatic subjects for screening. Endoscopic features and biopsy findings were reviewed and samples were classified according to different HP-associated diseases. The presence of the HP genome and cagA were further confirmed by polymerase chain reaction (PCR). Direct DNA sequencing was performed to investigate sequencing variations of the cagA gene. The results were correlated with biopsy assessments and with different disease groups.

A total of 278 consecutive samples from subjects with HP-associated diseases were included. After reviewing relevant clinical data and biopsy features, the samples were classified into six disease categories: minimal-to-mild gastritis (n=17), moderate-to-severe gastritis (n=51), gastric ulcer (n=48), duodenal ulcer (n=88), severe atrophic gastritis with marked intestinal metaplasia (n=42), and gastric cancer (n=32).

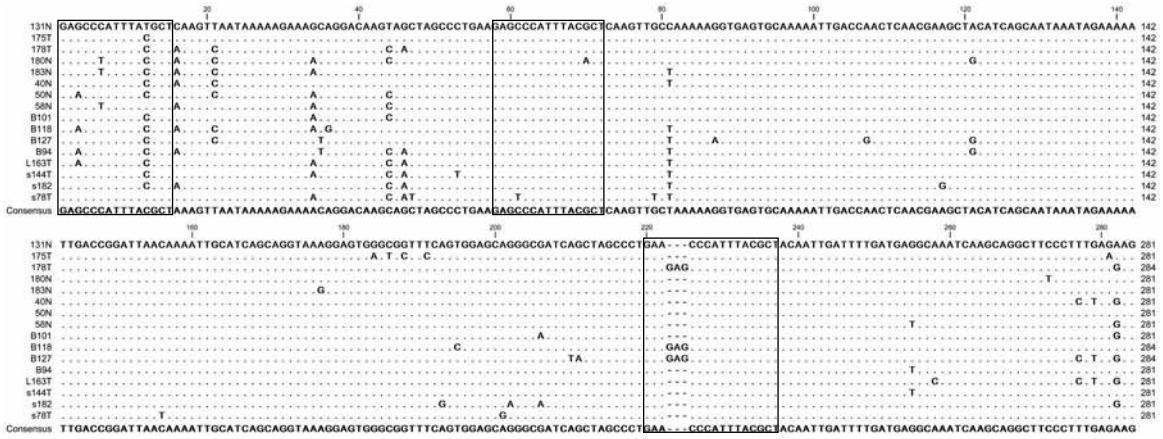
DNA was extracted from clinical samples using a High Pure PCR Template Preparation Kit (Roche) according to the manufacturer's instructions. Primer sets specific for the HP 16S rRNA gene were used for the detection of HP. The diversity of the cagA gene was examined by sequencing its 3' region.

Results

Of the 278 HP positive samples, 225 (81%) containing the cagA gene were classified as minimal-to-mild gastritis (n=17), moderate-to-severe gastritis

(a) Subtype 1 of HP cagA gene

Alignments of the nucleotide sequences



Alignments of the amino acid sequences

(b) Subtype 2 of HP cagA gene

Alignments of the nucleotide sequence

Alignments of the amino acid sequences

(c) Subtype 3 of HP cagA gene

Alignments of the nucleotide sequences

Alignments of the amino acid sequences

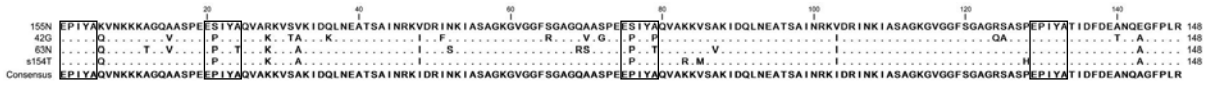


Fig 1. Alignments of the nucleotide and amino acid sequences of subtypes of the HP cagA gene

(n=33), gastric ulcer (n=41), duodenal ulcer (n=67), severe atrophic gastritis with marked intestinal metaplasia (n=37), and gastric cancer (n=30).

Sequencing analysis of the 225 HP cagA positive samples revealed that all the variants demonstrated characteristics of East Asian strains, which differed markedly from western strains.[3] The primary structure of the 3' region of the cagA gene was composed of a variable number of repeat sequences, including R1 (15-bp, 5 amino acids EPIYA), R2 (42-bp, 14 amino acids KVNKKKTGQVASPE) and R3 (147-bp, 49 amino acids, QVAKKVKAKIDQLNE ATSAINRKIDRINKIASAGKGVGGFSGAGQSASP). The structural organisation could be divided into three subtypes: subtype 1 (R1-R2-R1-R3-R1), subtype 2 (R1-R2-R1-R2-R1-R3-R1) and subtype 3 (R1-R2-R1-R3-R1-R3-R1). Alignments of the representative nucleotide and deduced amino acid sequences for each subtype are shown in Fig 1.

We investigated the diversity of the cagA gene among Hong Kong strains and the association between sequence diversity and clinical status. The distribution of cagA 3' region subtypes in different gastrointestinal diseases is shown in the Table. Subtype 1 strain with three copies of R1 (EPIYA) repeats was the most common (217/225, 96%). Both subtypes 2 and 3 contained four copies of R1 (EPIYA) repeats. Subtype 2 strain was present in four (2%) cases (one gastric cancer, two severe atrophic gastritis and one duodenal ulcer). Subtype 3 strain was present in another four (2%) cases (three gastric cancer and one duodenal ulcer).

The frequencies of the cagA genotype with four copies of R1 (EPIYA) repeats (subtypes 2 and 3) were significantly higher in isolates from patients with gastric cancers than with ulcers and gastritis (p=0.006, Fisher's exact test). As severe atrophic gastritis with marked intestinal metaplasia is regarded as pre-cancerous, we combined gastric cancer and severe atrophic gastritis as one disease group and compared corresponding cagA genotypes from persons with benign lesions (ulcer and gastritis). A significantly higher frequency of 4xR4 was also noted in the aggressive as opposed to the benign lesion group (odds ratio=7.7, 95% confidence interval=1.35-56.65, p=0.0096).

Biological diversities among different cagA proteins are caused by variations in the number and sequences of tyrosine phosphorylation sites in the molecule.[3] Thus, we tried to look for cagA strains with sequence variations adjacent to the tyrosine phosphorylation sites. CagA strains with amino acid changes from alanine to threonine/serine/valine at residue 964 (A964T/S/V) corresponding to East Asian strain F32-cagA (GenBank accession number AF202972) were identified. This was located at the eighth residue, immediately following phosphotyrosine (pY) within the last R1 (EPIYA) repeat. These strains were designated 964-T, 964-S and 964-V. The prototype was designated 964-A (Fig 2). The 964 variants were disproportionately found in patients with aggressive versus benign lesions. The 964-T variants were present in 12 (5%) cases (two gastritis, eight duodenal ulcers, and two gastric ulcers), whereas the 964-S variant was noted in one duodenal ulcer subject and 964-V variants were found in one patient with a duodenal ulcer and one with mild gastritis. No variants were detected in subjects with atrophic gastritis or gastric cancer. There was a trend towards a higher frequency of 964-T/S/V variants in persons with benign lesions (15/158 with ulcer and gastritis, 10%) rather than aggressive lesions (0/67 with atrophic gastritis and gastric caner) [P=0.007, Fisher's exact test].

Discussion

This study is the largest and most comprehensive series analysed for sequence variations at the 3' region of the HP cagA gene. A wide range of HP-associated gastrointestinal diseases from different subjects with various pathologies were studied. The cagA gene was detected in 225 (81%) out of the 278 HP positive strains. The frequency was consistent with reported frequencies from Korea and China, whereas in western countries, the percentage was much lower.

Sequencing analysis revealed three types of cagA gene (subtypes 1 to 3), which were identical to the variants from Zhejiang province, China.[4] Four subtypes of cagA isolated (types A to D) were reported from Japanese patients.[3] The Japanese types A and C were identical to our subtypes 1 and 3, respectively. However, we were not able to identify Japanese types B and D in our series. Similar findings were also reported elsewhere,[4] indicating distinct HP strains circulated in different geographic regions.

The HP cagA gene was heterogeneous with respect to its potential for undergoing tyrosine phosphorylation, SHP-2

Table. Distribution of the subtypes of the 3' region of the cagA gene in different gastroduodenal diseases

Subtype of 3' region of cagA gene	No. (%) of cagA positive cases						
	Gastric cancer (n=30)	Severe atrophic gastritis (n=37)	Gastric ulcer (n=41)	Duodenal ulcer (n=67)	Moderate-to-severe gastritis (n=33)	Minimal-to-mild gastritis (n=17)	Total (n=225)
Subtype 1	26 (87)	35 (95)	41 (100)	65 (97)	33 (100)	17 (100)	217 (96)
Subtype 2	1 (3)	2 (5)	0 (0)	1 (2)	0 (0)	0 (0)	4 (2)
Subtype 3	3 (10)	0 (0)	0 (0)	1 (2)	0 (0)	0 (0)	4 (2)
Subtypes 2 and 3	4 (13)	2 (5)	0 (0)	2 (3)	0 (0)	0 (0)	8 (4)

binding, and induction of cellular morphological changes. Such biological diversities among different cagA proteins were caused by variations in the number and sequences of tyrosine phosphorylation sites of the molecule. The R1 EPIYA repeat motifs in the 3' cagA region were potential targets of tyrosine phosphorylation. The cagA protein with more EPIYA repeats were expected to be more active biologically than those with a small number of repeats because they interacted more effectively with SHP-2 phosphatase and more severely perturb SHP-2-dependent signalling pathways. In keeping with this theory, our results demonstrated HP cagA strains carrying cagA protein with more repeat sequences (four versus three copies of R1 EPIYA) were associated with severe atrophic gastritis and gastric cancer.

Our results indicated that cagA strain subtypes 2 and 3 with four copies of R1 EPIYA repeat sequences correlated with aggressive lesions. Subtype 1 with three copies of R1 EPIYA repeats remained the most common pattern in Hong Kong, even in patients with severe atrophic gastritis and gastric cancer. Other virulence factors may contribute to different disease outcomes.

SH2 domains recognise phosphopeptide motifs composed of phosphotyrosine (pY) followed by several C-terminal residues. Alterations in the ligand-binding motifs might reduce the binding affinity to SHP-2. We identified two strain variants with amino acid changed from alanine to threonine/serine/valine at the residue 964 (A964T/S/V), the eighth residue immediately following phosphotyrosine (pY) within the last R1 (EPIYA) repeat. 964-T/S/V variants were detected in benign lesions only. Based on this observation, we proposed a previously uncharacterised model for the cagA-SHP-2 interaction. Change of the pY+8 residue in these variants might alter the SHP-2-binding affinity, thus reducing the ability to form a complex with SHP-2 compared with the complex-forming activity of wild-type cagA protein. Therefore, 964-T/S/V variants were less active biologically than wild-type counterparts. Interestingly, one

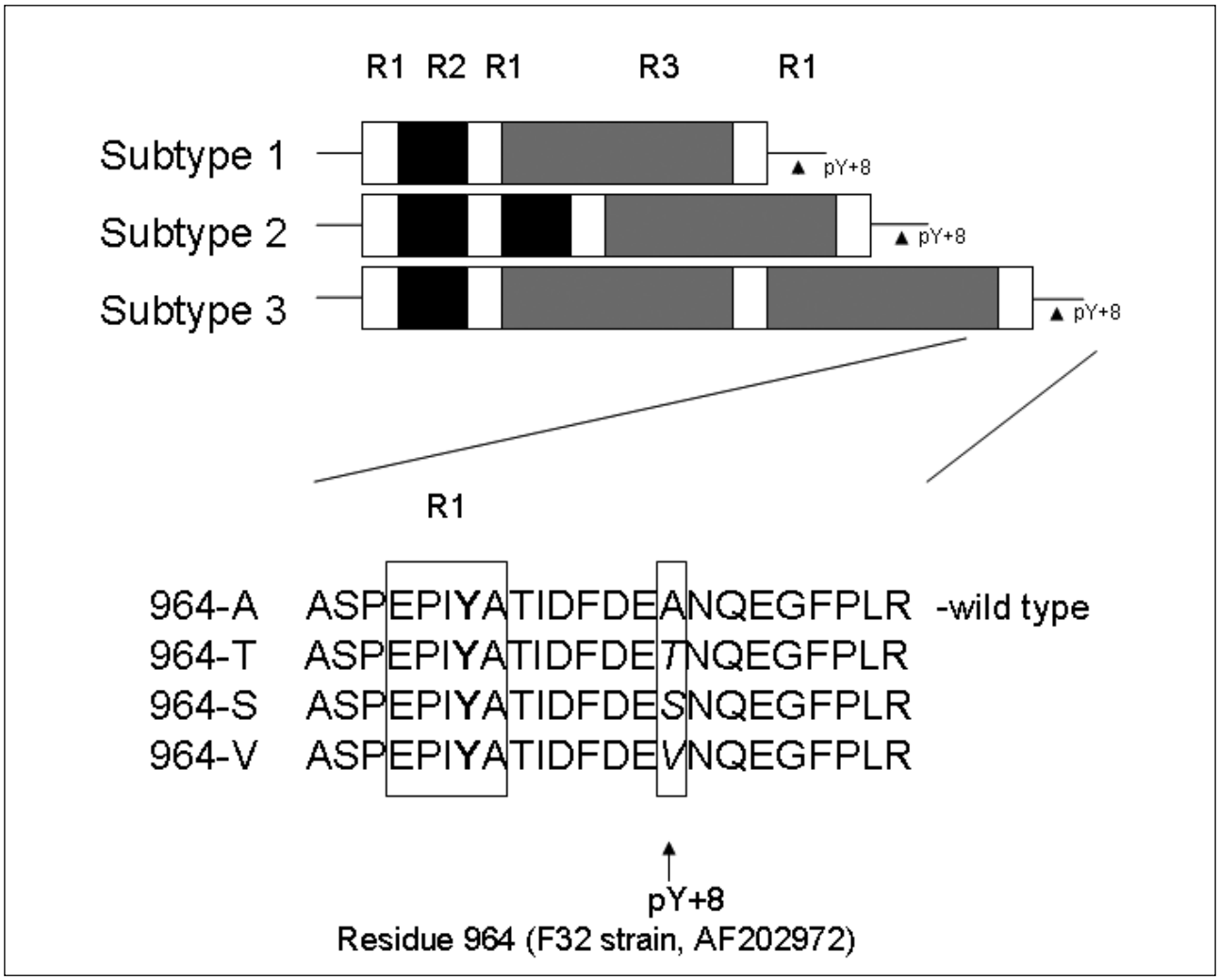


Fig 2. Schematic demonstration of pY+8 variant strains in three subtypes of the cagA protein
Arrowheads indicate the location of pY+8 residue immediately following the last R1 EPIYA repeats in each of the cagA subtypes. The phosphotyrosines are bolded in the sequence alignment of the three variants. The EPIYA repeats and pY+8 residues are boxed

of the 964-T variants belonged to the cagA subtype 3 with four copies of R1 EPIYA repeats. This strain was isolated from a patient with duodenal ulcer. cagA proteins with four copies of R1 EPIYA repeats were presumably more active biologically. The simultaneous possession of 964-T might reduce the SHP-2 binding affinity of this particular variant, rendering it less active. However, these observations need to be confirmed in studies with larger samples together with investigation of functional aspects, so as to further elucidate our hypothesis.

Acknowledgement

This study was supported by Research Fund for the Control of Infectious Diseases, Food and Health Bureau, Hong Kong SAR Government (#03040552).

References

1 Censini S, Lange C, Xiang Z, et al. cag, a pathogenicity island of *Helicobacter pylori*, encodes type I-specific and disease-associated virulence factors. Proc Natl Acad Sci U S A 1996;93:14648-53.

2 Azuma T, Yamakawa A, Yamazaki S, et al. Correlation between variation of the 3' region of the cagA gene in *Helicobacter pylori* and disease outcome in Japan. J Infect Dis 2002;186:1621-30.

3 Higashi H, Tsutsumi R, Fujita A, et al. Biological activity of the *Helicobacter pylori* virulence factor CagA is determined by variation in the tyrosine phosphorylation sites. Proc Natl Acad Sci U S A 2002;99:14428-33.

4 Tao R, Fang PC, Liu HY, Jiang YS, Chen J. A new subtype of 3' region of cagA gene in *Helicobacter pylori* strains isolated from Zhejiang Province in China. World J Gastroenterol 2004;10:3284-8.