

# 不同地区幽门螺杆菌cagA基因羧基端可变区及其蛋白序列差异分析

吴莺, 张尤历, 王文兵, 陈劲频, 沈琰

## ■背景资料

目前的研究表明, *H. pylori*与慢性胃炎、消化性溃疡和黏膜相关性淋巴瘤组织淋巴瘤发病关系密切, 并且涉及胃癌流行病学研究, *H. pylori*被WHO认定为I类致癌物质。

吴莺, 张尤历, 陈劲频, 沈琰, 江苏大学附属医院消化科 江苏省镇江市 212001  
王文兵, 江苏大学生命科学院 江苏省镇江市 212001  
江苏省科学技术厅社会发展项目, No. BS2002023  
通讯作者: 张尤历, 212001, 江苏省镇江市解放路438号, 江苏大学附属医院消化科. zjyouli@yahoo.com.cn  
电话: 0511-5011787  
收稿日期: 2006-09-28 接受日期: 2007-01-04

## Diversity of cagA gene 3' terminal variable region and its corresponding protein in *Helicobacter pylori*

Ying Wu, You-Li Zhang, Wen-Bing Wang, Jin-Pin Chen, Yan Shen

Ying Wu, You-Li Zhang, Jin-Pin Chen, Yan Shen, Department of Gastroenterology, the Affiliated Hospital of Jiangsu University, Zhenjiang 212001, Jiangsu Province, China

Wen-Bing Wang, Institute of Life Science, Jiangsu University, Zhenjiang 212001, Jiangsu Province, China

Supported by the Social Development Program of Jiangsu Science and Technology Department, No. BS2002023

Correspondence to: You-Li Zhang, Department of Gastroenterology, the Affiliated Hospital of Jiangsu University, 438 Jiefang Road, Zhenjiang 212001, Jiangsu Province, China. zjyouli@yahoo.com.cn

Received: 2006-09-28 Accepted: 2007-01-04

## Abstract

**AIM:** To study the diversity and structural characteristics of cagA (cytotoxin-associated gene A) gene 3' terminal variable region in *Helicobacter pylori* as well as their correlations with the outcome of *H. pylori*-related diseases.

**METHODS:** PCR products with positive *H. pylori* cagA were obtained from 20 cases and then sequenced by ExPASy-Translation software. Meanwhile, 3 to 6 cases of *H. pylori* 3' terminal variable region and amino acid sequence, collected from NCBI database in different countries, were comparatively analyzed.

**RESULTS:** The amino acid sequences were found with obvious regional characteristics,

and could be classified into eastern-asian type and western type. Western-type aberrations occurred in some eastern-asian strains. Totally, 87.4% *H. pylori* strains contained 3 EPIYA motifs in series, and 8.2% strains had 4 or more EPIYA motifs. A new sequence containing 3 amino acids was found. There was no correlation between the numbers of EPIYA motifs in the amino acid sequences of cagA protein and the clinical outcomes.

**CONCLUSION:** There exists marked diversity in cagA gene and protein of *H. pylori*, which can be distinguished by the main sequences of *H. pylori* cagA 3' terminal variable region. The number of EPIYA motifs in cagA 3' terminal variable region is not correlated with the clinical outcomes of diseases.

**Key Words:** *Helicobacter pylori*; Cytotoxin-associated gene A; EPIYA motif; Diversity

Wu Y, Zhang YL, Wang WB, Chen JP, Shen Y. Diversity of cagA gene 3' terminal variable region and its corresponding protein in *Helicobacter pylori*. *Shijie Huaren Xiaohua Zazhi* 2007;15(7):746-749

## 摘要

**目的:** 比较不同地区*H. pylori* cagA 3'端可变区序列差异及序列的地域特征, 分析差异与疾病的相关性。

**方法:** 选择本实验室所收集的20例cagA<sup>+</sup>的PCR产物进行测序, 并通过ExPASy-Translation软件推测其氨基酸序列, 搜索并收集GenBank中已公布的不同地区3-6个*H. pylori* cagA 3'端可变区序列及其氨基酸序列进行比较分析。

**结果:** cagA蛋白的氨基酸序列可分为具有明显地域特征的东亚型和西方型两类; 部分东亚型菌株表现出西方型变异。87.4%的菌株具有3个串联排列的EPIYA基序, 8.2%的菌株具有4个或以上EPIYA基序。新发现1条具有区

域特征的含有3个氨基酸的序列. cagA蛋白氨基酸序列中的EPIYA数目与临床结果没有相关性.

**结论:** *H pylori* cagA基因及其cagA蛋白序列具有明显多态性, 可以根据*H pylori* cagA 3'端可变区的主要序列差异进行地域分型, cagA 3'端可变区的EPIYA基序的数目与临床结果没有相关性.

**关键词:** 幽门螺杆菌; 细胞毒性相关基因A; EPIYA基序; 差异

吴莺, 张允历, 王文兵, 陈劲频, 沈琰. 不同地区幽门螺杆菌cagA基因羧基端可变区及其蛋白序列差异分析. 世界华人消化杂志 2007;15(7):746-749  
<http://www.wjgnet.com/1009-3079/15/746.asp>

## 0 引言

幽门螺杆菌(*H pylori*)的细胞毒性相关基因A(cytotoxin-associated gene A, cagA), 是研究最多的*H pylori*的基因之一. 由cagA基因编码的cagA蛋白, 经由IV分泌系统进入细菌黏附的胃上皮细胞, 在其羧基端进行酪氨酸磷酸化, 从而引起细胞骨架的重排、诱导细胞形态学的变化和细胞动力的增强, 造成细胞异常的增殖和运动, 最终导致癌变. 因此cagA阳性(cagA<sup>+</sup>)的*H pylori*菌株较cagA<sup>-</sup>的*H pylori*菌株毒力强<sup>[1-5]</sup>. 但是*H pylori*较其他细菌有更大的遗传变异, 从而导致感染后的不同临床后果: 从无症状携带到慢性胃炎、消化性溃疡、胃窦腺癌. 造成*H pylori*菌株cagA蛋白的功能差异的原因是cagA基因3'端可变区上重复序列的差异, 并由此造成了cagA基因结构的多态性和cagA蛋白的生物学差异, 这个差异还与萎缩性胃炎和胃癌的发病相关<sup>[1,6]</sup>. cagA蛋白羧基端的EPIYA基序(EPIYA基序为谷氨酸-脯氨酸-异亮氨酸-酪氨酸-丙氨酸序列)的酪氨酸残基是cagA蛋白进行酪氨酸磷酸化的位点, 有研究认为, 具有更多EPIYA基序的cagA蛋白能诱导更多的酪氨酸磷酸化和白介素-8的分泌, 因此具有更强的细胞毒性<sup>[6-7]</sup>. 我们试图通过比较不同地区*H pylori*菌株cagA 3'端可变区及其cagA蛋白序列多态性的分析以及与临床结果相关性的研究, 试图找出*H pylori*的区域特征以及可能存在的结构差异、以及这种差异与感染后不同的临床转归之间是否存在关联.

## 1 材料和方法

**1.1 材料** 以“*Helicobacter pylori*”, “cagA”主题词检索Pubmed的Nucleotide数据库. 中国的20例为我们自己检测: 我们所搜集的60株*H pylori*, 分离自中国镇江地区胃肠病患者胃黏膜, 通过PCR检测cagA, 在56例cagA<sup>+</sup>的菌株中, 分别选择胃癌菌株4例、消化性溃疡9例、慢性胃炎7例, 对其PCR产物用PCR产物纯化试剂盒回收, 再与pMD18-T载体连接, 转化感受态细胞进行蓝白斑筛选. 挑出白色单菌落扩增, 用SDS-碱裂解法抽提质粒, 分别用BamH I, EcoR I酶切, 并电泳鉴定. 对阳性克隆进行测序. 测序由上海生物工程公司完成.

**1.2 方法** 以“*Helicobacter pylori*”, “cagA”主题词检索Pubmed的Nucleotide数据库, 共有1265个结果. 按检索出来的顺序, 选择日本菌株cagA 3'端可变区序列6个, 韩国3个, 越南4个, 美国3个, 哈萨克斯坦3个, 印度3个, 爱尔兰3个, 哥斯达黎加4个. 各菌株cagA基因核酸序列接受号分别为: 1-6为日本菌株: 1-AB057039(JapanO873), 2-AB057038(JapanO552), 3-AB057011(JapanK30), 4-AB057036(JapanK260), 5-AB057020(JapanK46), 6-AB057012(JapanK31); 7-10为越南菌株: 7-AB057057(Vietnam2705), 8-AB057058(Vietnam2706), 9-AB057056(Vietnam2704), 10-AB057055(Vietnam2692); 11-13为美国菌株: 11-AB057100(USA2754), 12-AB057099(USA2791), 13-AB057095(USA35); 14-16为韩国菌株: 14-AB057044(Korea23), 15-AB057043(Korea12), 16-AB057042(Korea9); 17-19为哈萨克菌株: 17-AB057096(Kazak1), 18-AB057097(Kazak2), 19-AB057098(Kazak3); 20-22为爱尔兰菌株: 20-AF427098、21-AF427099、22-AF427100; 23-26为哥斯达黎加菌株: 23-AF289463、24-AF289464、25-AF289462、26-AF289461; 27-29为印度菌株: 27-AF222808, 28-AF222807, 29-AF222809.

相对应的蛋白序列GenBank接受号为: 1-6为日本菌株: 1-BAB87406.1, 2-BAB87405.1, 3-BAB87378.1, 4-BAB87403.1, 5-BAB87387.1, 6-BAB87379.1; 7-10为越南菌株: 7-BAB87424.1, 8-BAB87425.1, 9-BAB87423.1, 10-BAB87422.1; 11-13为美国菌株: 11-BAB87467.1, 12-BAB87466.1, 13-BAB87462.1; 14-16为韩国菌株: 14-BAB87411.1, 15-BAB87410.1, 16-BAB87409.1;

## ■ 相关报道

目前的报道主要集中在cagA蛋白的遗传差异上. 主要涉及cagA蛋白的羧基端可变区的差异和分类: 个体菌株之间的差异、地区间菌株的差异、以及这种差异与临床结果之间的关系. 这种差异与EPIYA基序的数目和其前后氨基酸的序列有关. 研究者试图找到其中的规律以及临床结果的关联.

## ■应用要点

本文通过cagA蛋白的氨基酸序列分析,研究了cagA蛋白的个体差异,研究者以后可以应用类似的方法,继续研究*H pylori*的其他毒力因子之间的差异,以期找到造成*H pylori*感染不同结局的原因。

表 1 不同地区*H pylori*的菌株数与其cagA 3'端可变区的序列特征

国家	菌株数	cagA 3'端可变区主要特征序列	EPIYA基序的数目(n)		
			2(%)	3(%)	4(%)
中国	20	KIASAGKGVGGFSGA DFD		20(100%)	
日本	6	KIASAGKGVGGFSGA DFD		6(100%)	
韩国	3	KIASAGKGVGGFSGA DFD		3(100%)	
越南	4	KIASAGKGVGGFSGA DFD		4(100%)	
美国	3	FPLKRHDKVDDL SKV DDL	1(33.3%)	2(66.7%)	
哈萨克斯坦	3	FPLKRHDKVDDL SKV DDL		2(66.7%)	1(33.3%)
印度	3	FPLKRHDKVDDL SKV DDL		1(33.3%)	2(66.7%)
爱尔兰	3	FPLKRHDKVDDL SKV EDL	1(33.3%)	1(33.3%)	1(33.3%)
哥斯达尼加	4	FPLKRHDKVDDL SKV DDL		4(100%)	
总计	49		2(4.1%)	43(87.8%)	4(8.2%)

\*表示为5个EPIYA基序。

17-19为哈萨克菌株: 17-BAB87463.1, 18-BAB87464.1, 19-BAB87465.1; 20-22为爱尔兰菌株: 20-AAL28091.1, 21-AAL28092.1, 22-AAL28093.1; 23-26为哥斯达黎加菌株: 23-AAG09912.1, 24-AAG09913.1, 25-AAG09911.1, 26-AAG09910.1; 27-29为印度菌株: 27-AAF35341.1, 28-AAF35340.1, 29-AAF35342.1。

## 2 结果

2.1 cagA 3'端可变区的区域性结构特点 中国和日本、韩国、越南共有33个菌株, 菌株的cagA 3'端可变区的结构相似, 每个菌株具有1个特征性的东亚主要氨基酸序列KIASAGKGVGGFSGA, 100%的菌株为3个串联排列的EPIYA序列(33/33, 100.0%), 紧随第3个EPIYA序列后的1个DFD序列; 美国、哈萨克斯坦、印度、爱尔兰、哥斯达黎加共有16个菌株, 菌株的cagA 3'端可变区的结构相似, 每个菌株具有2个以上特征性的西方主要氨基酸序列FPLKRHDKVDDL SKV, 68.8%(10/16)的菌株具有3个EPIYA基序, 25.0%(4/16)菌株具有4个和4个以上EPIYA基序, 以及紧随第3个EPIYA序列后的1个DDL序列或EDL序列(表1)。

2.2 *H pylori*菌株的cagA 3'端可变区长短的变化 从513-1335 bp不等, 印度菌株最长, 3个菌株cagA 3'端可变区为1224-1335 bp。

2.3 EPIYA基序的个数与临床结果的关系 中国镇江地区20株用于测序的*H pylori*菌株, 胃炎7例, 消化性溃疡9例, 胃腺癌4例, 测序结果cagA蛋白羧基端的EPIYA的数目均为3个, 没有显示cagA蛋白羧基端的EPIYA数目与临床结果相关。

## 3 讨论

Dangeruta通过PCR方法, 研究了来自5个地区500多株*H pylori*的DNA基序, 发现*H pylori*的cagPAI由于存在基序的缺失、插入、替换, 而使cagPAI表现为5个基因类型, 其中3个基因型最常见: I型主要分布在西班牙人、秘鲁人、危地马拉人、非洲人、美国人中; II型主要分布在日本人和中国人中; III型主要分布在印度人中<sup>[8]</sup>。显示了*H pylori*基因结构的多态性具有明显的地域分布特征及中国菌株与日本菌株的相似性。我们对来自不同国家和地区的*H pylori* cagA蛋白的序列研究发现, 可以根据cagA蛋白羧端可变区主要重复序列是KIASAGKGVGGFSGA序列抑或是 FPLKRHDKVDDL SKV序列, 分为2大类: 东亚型(拥有KIASAGKGVGGFSGA序列), 见于中国、日本、韩国、越南; 西方型(拥有FPLKRHDKVDDL SKV序列), 见于美国、哈萨克斯坦、印度、爱尔兰、哥斯达黎。这与Higashi *et al*<sup>[9]</sup>的研究结果相符。我们还发现了1个新的具有区域特征的序列, 即紧随第3个EPIYA后的1个含有3个氨基酸的序列: 在东亚型菌株中是DFD序列、在西方型菌株中是DDL序列、在爱尔兰菌株中是EDL序列。我们还发现, 亚洲菌株可以表现为西方型的变异。比如, 印度, 他的3个菌株均具有西方型菌株的特征, 所以, 其cagA蛋白属于西方型。我们也曾经在测序的菌株中, 发现1株来源于十二指肠球部溃疡患者的菌株, 表现为西方菌株的特征, 属于西方型。但并没有发现西方型菌株向东亚型菌株的变异。这种情况是否与人群的迁徙而发生的*H pylori*的进化有关呢? 尚需进一步的研究证明。

近来的研究发现, 不同*H pylori*菌株的cagA

蛋白细菌毒力并不相同, 这主要涉及菌株的cagA 3'端可变区重复序列的数目的变化, 而重复序列数目的变化主要是指cagA蛋白羧端串联排列的EPIYA基序数目的变化, 因为EPIYA序列上的酪氨酸残基是cagA蛋白酪氨酸磷酸化的位点。在cagA蛋白串联排列的EPIYA基序中, 第1、第2个EPIYA基序分别称为EPIYA-A、EPIYA-B, 第3个EPIYA基序, 在西方型菌株中称为EPIYA-C, 在东亚型菌株中则被称为EPIYA-D。cagA蛋白进入宿主上皮细胞后, cagA在其EPIYA-C或EPIYA-D位点被磷酸化, 然后在此位点特异性的连接并激活SHP-2; cagA在其EPIYA-A或EPIYA-B位点被磷酸化后, 在此位点连接并激活Csk<sup>[9-11]</sup>。因此具有较多EPIYA基序的cagA蛋白具有更强的磷酸化能力, 因而, 具有更强的生物活性。虽然中国和日本均有有2个和4个EPIYA基序的报道<sup>[12-13]</sup>, 可是我们的研究结果显示, 我们的20株中国菌株和GenBank中收集的12株东亚菌株, 均为3个EPIYA基序。我们的20个测序的菌株中, 有4个来自胃癌、4个来自胃溃疡, 也均表现为3个EPIYA基序。因此我们认为cagA蛋白的毒力强弱除了与EPIYA基序的数目有关外, 应该还有其他我们未知的因素参与<sup>[14-15]</sup>。Yamaoka *et al*<sup>[11]</sup>根据cagA基因PCR结果及PCR产物测序结果, 将东亚型菌株分为A型、B型、C型、D型4型。我们的序列研究发现: 我们的20株中国菌株和GenBank中收集的12株东亚菌株均表现为A型, 没有B型、C型、D型的发现, 显示东亚型菌株主要表现为A型, 与Yamaoka的报道一致。同时发现, 西方型菌株不能按此标准进行分型。

总之, 我们的结果认为不同地区的*H. pylori*可以根据cagA蛋白羧端的重复序列是KIASAGKGVGFGSGA序列抑或是FPLKRHDKVDDLSKV序列、DFD序列抑或是DDL序列, 分为东亚型和西方型, 西方型菌株根据是DDL序列抑或是EDL序列, 分为非爱尔兰型和爱尔兰型。87.8%的*H. pylori*菌株具有3个EPIYA基序, 8.2%的菌株具有4个或以上EPIYA基序, 没有发现EPIYA基序的数目与临床结果相关。

#### 4 参考文献

- 1 Hatakeyama M, Higashi H. *Helicobacter pylori* CagA: a new paradigm for bacterial carcinogenesis. *Cancer Sci* 2005; 96: 835-843
- 2 Asahi M, Azuma T, Ito S, Ito Y, Suto H, Nagai Y, Tsubokawa M, Tohyama Y, Maeda S, Omata M,

- Suzuki T, Sasakawa C. *Helicobacter pylori* CagA protein can be tyrosine phosphorylated in gastric epithelial cells. *J Exp Med* 2000; 191: 593-602
- 3 Ando T, Goto Y, Maeda O, Watanabe O, Ishiguro K, Goto H. Causal role of *Helicobacter pylori* infection in gastric cancer. *World J Gastroenterol* 2006; 12: 181-186
- 4 佟书娟, 陈军, 詹臻, 杨丹丹, 刘亚平. 胃癌组织中幽门螺杆菌cagA和vacA的表达及与其感染的相关性. 世界华人消化杂志 2006; 14: 2632-2635
- 5 杜雅菊, 赵磊, 赵瑞波, 李宝杰. 幽门螺杆菌CagA蛋白与胃癌组织中Bcl-2、p53蛋白表达的关系. 世界华人消化杂志 2003; 11: 554-557
- 6 Naito M, Yamazaki T, Tsutsumi R, Higashi H, Onoe K, Yamazaki S, Azuma T, Hatakeyama M. Influence of EPIYA-repeat polymorphism on the phosphorylation-dependent biological activity of *Helicobacter pylori* CagA. *Gastroenterology* 2006; 130: 1181-1190
- 7 Zhang Y, Argent RH, Letley DP, Thomas RJ, Atherton JC. Tyrosine phosphorylation of CagA from Chinese *Helicobacter pylori* isolates in AGS gastric epithelial cells. *J Clin Microbiol* 2005; 43: 786-790
- 8 Kersulyte D, Mukhopadhyay AK, Velapatino B, Su W, Pan Z, Garcia C, Hernandez V, Valdez Y, Mistry RS, Gilman RH, Yuan Y, Gao H, Alarcon T, Lopez-Brea M, Balakrish Nair G, Chowdhury A, Datta S, Shirai M, Nakazawa T, Ally R, Segal I, Wong BC, Lam SK, Olfat FO, Boren T, Engstrand L, Torres O, Schneider R, Thomas JE, Czinn S, Berg DE. Differences in genotypes of *Helicobacter pylori* from different human populations. *J Bacteriol* 2000; 182: 3210-3218
- 9 Higashi H, Tsutsumi R, Fujita A, Yamazaki S, Asaka M, Azuma T, Hatakeyama M. Biological activity of the *Helicobacter pylori* virulence factor CagA is determined by variation in the tyrosine phosphorylation sites. *Proc Natl Acad Sci USA* 2002; 99: 14428-14433
- 10 Tsutsumi R, Takahashi A, Azuma T, Higashi H, Hatakeyama M. Focal adhesion kinase is a substrate and downstream effector of SHP-2 complexed with *Helicobacter pylori* CagA. *Mol Cell Biol* 2006; 26: 261-276
- 11 Higashi H, Yokoyama K, Fujii Y, Ren S, Yuasa H, Saadat I, Murata-Kamiya N, Azuma T, Hatakeyama M. EPIYA motif is a membrane-targeting signal of *Helicobacter pylori* virulence factor CagA in mammalian cells. *J Biol Chem* 2005; 280: 23130-23137
- 12 Yamaoka Y, Kodama T, Kashima K, Graham DY, Sepulveda AR. Variants of the 3' region of the cagA gene in *Helicobacter pylori* isolates from patients with different *H. pylori*-associated diseases. *J Clin Microbiol* 1998; 36: 2258-2263
- 13 朱永良, 杜勤, 钱可大, 方平楚, 郑树. 幽门螺杆菌CagA C-端功能域特征及其生物学功能研究. 中华消化杂志 2004; 24: 278-281
- 14 Yan J, Wang Y, Shao SH, Mao YF, Li HW, Luo YH. Construction of prokaryotic expression system of 2 148-bp fragment from cagA gene and detection of cagA gene, CagA protein in *Helicobacter pylori* isolates and its antibody in sera of patients. *World J Gastroenterol* 2004; 10: 1183-1190
- 15 Yang GF, Deng CS, Xiong YY, Gong LL, Wang BC, Luo J. Expression of nuclear factor-kappa B and target genes in gastric precancerous lesions and adenocarcinoma: association with *Helicobacter pylori* cagA (+) infection. *World J Gastroenterol* 2004; 10: 491-496

#### ■名词解释

EPIYA基序: 位于cagA蛋白羧基端可变区, 是一个含有5个氨基酸的序列, 即谷氨酸-脯氨酸-异亮氨酸-酪氨酸-丙氨酸序列, 为cagA蛋白的磷酸化位点。