Epidemiol. Infect. (2000), **124**, 91–96. Printed in the United Kingdom

Molecular epidemiology of *Helicobacter pylori*: separation of *H. pylori* from East Asian and non-Asian countries

Y. YAMAOKA[1, 2, *], M. S. OSATO[1], A. R. SEPULVEDA[1], O. GUTIERREZ[3], N. FIGURA[4], J. G. KIM[5], T. KODAMA[2], K. KASHIMA[2] AND D. Y. GRAHAM[1]

[1] *VA Medical Center and Baylor College of Medicine, Houston, TX, USA*
[2] *Third Department of Internal Medicine, Kyoto Prefectural University of Medicine, Kyoto, Japan*
[3] *Universidad Nacional de Colombia, Bogota, Colombia*
[4] *Institute of Internal Medicine, University of Siena, Siena, Italy*
[5] *Guro Hospital, Korea University College of Medicine, Seoul, Korea*

(*Accepted 5 October 1999*)

SUMMARY

The predominant *H. pylori* strain circulating among geographic locations differs with regard to the genomic structure. This study determined whether structural subtypes of the *cagA* 3′ repeat region could be used to identify the population of origin of *H. pylori* isolates. We examined 600 *cagA*-positive *H. pylori* (Colombia, 100; USA, 100; France, 100; Canada, 20; Italy, 20; Korea, 100; Japan, 100; Hong Kong, 20; Taiwan, 20; Vietnam, 20). The *cagA* 3′ region was amplified by PCR using primers specific to Japanese and Western 3′ *cagA* gene sequences. PCR using Japanese *cagA* primers resulted in PCR products in 99·6% of strains from East Asia but no non-Asian strains. Conversely, PCR using Western *cagA* primers resulted in amplicons in 100% of non-Asian strains, and only one from East Asia. *cagA* genotyping is useful for molecular epidemiological studies as strains can be completely separated by differences in the *cagA* 3′ region.

INTRODUCTION

Helicobacter pylori strains that possess the *cag* pathogenicity island are associated with enhanced mucosal inflammation and increased mucosal interleukin-8 (IL-8) levels [1–6]. Although the *cagA* gene is a marker for this pathogenicity island, its function is still unknown. The *cagA* gene product is a highly immunogenic antigen of variable size [7–10]. Variation in the size of the CagA protein has been correlated with the presence of repeat sequences located in the 3′ region of the gene [7, 10].

In a previous study, we showed that the 3′ region of the *cagA* gene of Japanese strains differed markedly from those from Western strains [10]. Analysis of sequences deposited in GenBank regarding reference strains isolated from Western countries showed that these regions consist of two types of repeat regions; 57 bp regions (first repeat region; FR region) and 102 bp regions (Western-type second repeat region; WSR region). Japanese strains had similar 57 bp regions (FR region), but the second repeat region of 162 bp (Japanese-type second repeat region; JSR region) was completely different from Western reference strains [10].

It has become evident that the predominant *H. pylori* strains in different geographic locations differ with regard to the genomic structure [11–17]. It has been suggested that East Asian strains may differ from non-Asian strains in the structure of some genes, with emphasis on *vacA* gene genotyping [14–17]. A previous study suggested the *cagA* 3′ repeat region as a candidate for a simple and reliable method to separate East Asian strains completely from non-

* Author for correspondence: Veterans Affairs Medical Center (111D), 2002 Holcombe Blvd, Houston, TX 77030 USA.

Asian strains [10]. This study attempted to confirm that hypothesis using PCR with *H. pylori* isolates obtained from different East Asian and non-Asian countries.

MATERIALS AND METHODS

H. pylori isolates

We examined 600 *cagA* gene-positive *H. pylori* isolates; 100 isolates from Colombia, 100 from the USA, 100 from France, 20 from Canada, 20 from Italy, 100 from Korea, 100 from Japan, 20 from Hong Kong, 20 from Taiwan and 20 from Vietnam. *cagA* gene-positive isolates were identified by polymerase chain reaction (PCR) using the conserved region of the *cagA* gene, as described previously [5, 6]. Each isolate was obtained from a different patient who underwent gastroscopy, usually due to symptoms of dyspepsia. We excluded the cases of gastric cancer in this study. All the individuals were living in their respective countries except the Vietnamese who had been born in Vietnam and had immigrated to the USA. Informed consent was obtained from all patients. The study protocol was approved by the local Ethics Committees.

Preparation of *H. pylori* genomic DNA

H. pylori were isolated from gastric biopsies as previously described [5, 6]. Isolates were grown at 37 °C on brain heart infusion agar plates supplemented with 7% horse blood (Cocalico Biological, Inc. Reamstown, PA) in a 12% CO_2 incubator under 100% relative humidity. Multiple studies in our laboratory have shown that the growth of *H. pylori* under microaerobic conditions (e.g. CampyPak Plus, BBL, Cockeysville, MD) and 12% CO_2 environments are identical. The organisms were identified as *H. pylori* by Gram staining, colony morphology and positive oxidase, catalase and urease reactions. Multiple colonies were collected together from the bacterial culture plates except for Italian isolates in which a single colony was tested. Genomic DNA was extracted using the QIAamp Tissue kit (QIAGEN Inc. Santa Clarita, CA) according to the manufacturer's instructions.

PCR

The oligonucleotide primers used for PCR-based typing of the *cagA* gene were designed according to the *cagA* gene sequence deposited in GenBank. We compared 4 Western strains: CCUG17824, G39, ATCC53726 and ATCC43526 (GenBank accession numbers X70038, X70039, L11714 and AB003397, respectively) with 4 Japanese strains: JK25, JK252, JK269 and JK22 (GenBank accession numbers AF043457, AF043458, AF043459 and AF043460, respectively).

Primers CAGTF (5′-ACCCTAGTCGGTAATG-GG-3′) and CAGTR (5′-GCTTTAGCTTCTGAYA-CYGC-3′, Y = C + T) were designed to include the entire 3′ repeat regions of the *cagA* gene (Fig. 1). These primer regions were conserved in both Western and Japanese strains deposited in GenBank.

The primers CAGTF/CAGWR and CAGWF/CAGTR inside the sequence specific for Western strains were used to amplify the regions that include the first repeat (FR region = 57 bp) and the second repeat (WSR region = 102 bp) regions, respectively (Fig. 1) (CAGWR: 5′-TGCCCTACAMCACCSAA-ACCAC-3′, CAGWF: 5′-AAAAATTGACCRACT-CAATC-3′, M = A + C, S = C + G, R = A + G). The expected lengths of PCR products amplified with the primers CAGTF/CAGWR and CAGWF/CAGTR were [(218 ~ 227) + 57r] bp and [(174 ~ 177) + 102m] bp respectively when an isolate contains 'r' times FR regions and 'm' times WSR regions.

Furthermore, we designed two sets of primers inside the *cagA* 3′ repeat sequence specific for Japanese strains (CAGTF/CAGJR and CAGJF/CAGTR), which resulted in the amplification of the first repeat (FR region = 57 bp) and the second repeat (JSR region = 162 bp) regions (Fig. 1) (CAGJR: 5′-GCAATTTTGTTAATCCGGTC-3′, CAGJF: 5′-GCATCAGCAGGTAAAGGAGT-3′). In Japanese strains, immediately downstream of the FR region, the R1 region (amino acid sequence EPIYA) and JSR regions were located. Therefore, we designed primers inside the JSR region and predicted multiple number of bands according to the number of JSR regions. The expected lengths of PCR products amplified with the primers CAGTF/CAGJR and CAGJF/CAGTR were [(236 ~ 242) + 57r + 162 (*n*–1)] bp (*n* = 1 to *n*) and [222 + 162 (*n*–1)] bp (*n* = 1 to *n*), respectively when an isolate contains 'r' times FR regions and '*n*' times JSR regions. According to the number of '*n*', we could detect '*n*' different sizes of PCR products. For example, when an isolate contains 1 FR region and 2 JSR regions, the PCR products using primers CAGTF/CAGJR were expected to result in 2 bands; 1 smaller band of (236 ~ 242) + 57 + 162 (1–1) =

https://doi.org/10.1017/S0950268899003209 Published online by Cambridge University Press

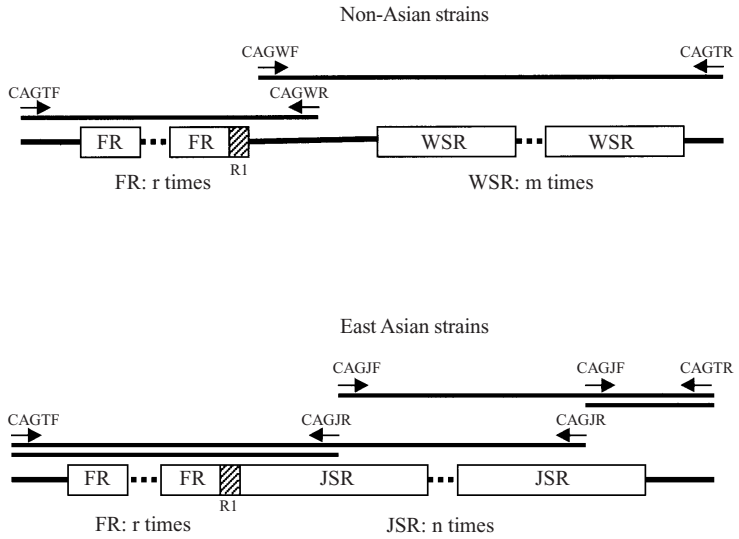

Fig. 1. Primary structure variants of the 3′ region of the *cagA* gene in non-Asian and East Asian populations. First repeat region (FR), Western second repeat region (WSR), Japanese second repeat region (JSR) and EPIYA (R1) fragments are not represented on a proportional scale. In Japanese strains, as the R1 region and JSR region immediately followed the FR region, we designed the primer inside the JSR region, predicting multiple number of bands according to the number of JSR regions. The expected lengths of PCR products amplified with the primers CAGTF/CAGWR and CAGWF/CAGTR were [(218 ~ 227) + 57r] bp and [(174 ~ 177) + 102m] bp, respectively when an isolate contains 'r' times FR regions and 'm' times WSR regions. The expected lengths of PCR products amplified with the primers CAGTF/CAGJR and CAGJF/CAGTR are [(236 ~ 242) + 57r + 162 (*n*–1)] bp (*n* = 1 to n) and [222 + 162 (*n*–1)] bp (*n* = 1 to n), respectively when an isolate contains 'r' times FR regions and '*n*' times JSR regions.

293 ~ 299 bp and another larger band of (236 ~ 242) + 57 + 162 (2–1) = 455 ~ 461 bp (when *n* = 1 and 2 are placed into the formula, respectively). Similarly, the PCR products using primers CAGTF/CAGJR were expected to result in 2 bands; 222 + 162 (1–1) = 222 bp and 222 + 162 (2–1) = 384 bp.

PCR amplification was performed as previously described [13], for 35 cycles consisting of 1 min at 95 °C, 1 min at 52 °C and 1 min at 72 °C. The final cycle included a 7 min extension step to ensure full extension of the PCR products. PCR products were analysed by gel electrophoresis in ethidium bromide containing 1·5% agarose gels, and the bands were examined under u.v. light for the presence of the amplified DNA.

RESULTS

The primers designed to amplify the entire 3′ repeat regions of the *cagA* gene (CAGTF and CAGTR) detected 99% (594/600) of *cagA* gene positive strains (Table 1). Only 2 USA (2%) strains, 1 Colombian (1%) strain and 3 French (3%) strains were not detected by these sets of primers. A possible mutation in the primer region resulted in our inability to categorize the origin of these six strains.

PCR using primers specific to the Japanese 3′ *cagA* gene sequence (CAGTF/CAGJR and CAGJF/CAGTR) resulted in PCR products in all but one (99·6%) of the East Asian strains, but none in the non-Asian strains (Fig. 2, Table 1). Conversely, PCR using the Western *cagA*-specific primers (CAGTF/CAGWR and CAGWF/CAGTR) resulted in PCR products in 100% of non-Asian strains, but in none from the Korean, Vietnamese, Taiwanese or Hong Kong strains (Fig. 2, Table 1). One Japanese strain was detected only by Western *cagA* specific primers. That person was born in India and lived in Japan only after she was 30 years old making it likely that she had acquired a non-Japanese strain in childhood.

Using our designed primers, we could also detect the number of repeats contained in non-Asian strains (FR and WSR regions) and East Asian strains (FR and JSR regions) (Table 1). In this study, 95–100% of the East Asian strains had 1 FR and 100% of these had 1 JSR. All East Asian isolates had only a single band detected on agarose gels. In contrast, non-Asian strains had some variety in WSR regions. There were more than 2 bands in 21 Colombian strains, 18 French strains and 10 USA strains. We randomly selected 25 cases with multiple bands (14 Colombian, 5 French and 6 USA strains) and picked at least 20 colonies

https://doi.org/10.1017/S0950268899003209 Published online by Cambridge University Press

Table 1. *Detection of* cagA *3′ repeat region by PCR using several specific primers*

5′ Primer	3′ Primer	Region designed	Japan ($n = 100$)	Korea ($n = 100$)	Vietnam ($n = 20$)	Hong Kong ($n = 20$)	Taiwan ($n = 20$)
CAGTF	CAGTR	Entire repeat region	100	100	20	20	20
CAGTF	CAGJR	Japanese 1st repeat	99	100	20	20	20
		1 time	97 (98%)	100 (100%)	20 (100%)	20 (100%)	19 (95%)
		Others*	2 (2%)	0	0	0	1 (5%)
CAGJF	CAGJR	Japanese 2nd repeat	99	100	20	20	20
		1 time	99 (100%)	100 (100%)	20 (100%)	20 (100%)	20 (100%)
CAGTF	CAGWR	Western 1st repeat	1†	0	0	0	0
CAGWF	CAGTR	Western 2nd repeat	1†	0	0	0	0

5′ Primer	3′ Primer	Region designed	USA ($n = 100$)	Colombia ($n = 100$)	France ($n = 100$)	Canada ($n = 20$)	Italy ($n = 20$)
CAGTF	CAGTR	Entire repeat region	98	99	97	20	20
CAGTF	CAGJR	Japanese 1st repeat	0	0	0	0	0
CAGJF	CAGJR	Japanese 2nd repeat	0	0	0	0	0
CAGTF	CAGWR	Western 1st repeat	98	98	95	20	20
		1 time	98 (100%)	96 (98%)	95 (100%)	19 (95%)	20 (100%)
		Others ‡	0	2 (2%)	0	1 (5%)	0
CAGWF	CAGTR	Western 2nd repeat	97	99	97	20	20
		0 time	5 (5%)	2 (2%)	0	0	2 (10%)
		1 time	78 (80%)	54 (55%)	66 (68%)	14 (70%)	9 (45%)
		2 times	7 (7%)	23 (23%)	13 (13%)	1 (5%)	9 (45%)
		Multiple	10 (10%)	21 (21%)	18 (19%)	5 (25%)	0§

* Two Japanese strains had 3 FR regions and 1 Taiwanese strain had no FR region.
† This strain had 1 FR and 1 WSR regions.
‡ One Colombian strain and 1 Canadian strain had 2 FR and 1 Colombian strain had 3 FR regions.
§ We picked single colonies in Italian strains, as there was no possibility of mixed infection.

per case each case confirming that these patients were infected with more than one *cagA* genotype (data not shown).

DISCUSSION

The primary gene structure of the *cagA* 3′ repeat region in *H. pylori* obtained from 5 East Asian countries differed markedly from that present in *H. pylori* obtained from 5 different non-Asian countries. We found that we could completely separate the region of origin (East Asian vs. non-Asian countries) by PCR using specific primers designed within the *cagA* 3′ repeat region suggesting that differences in *cagA* gene structure can be used as markers of the region of origin in molecular epidemiology studies.

Although our studies have focused on the 3′ repeat region, there may be other geographically conserved regions of *cagA*. For example, van der Ende and colleagues, in a study involving 24 isolates, suggested that distinct *cagA* positive *H. pylori* populations were circulating in China compared to The Netherlands [14]. Their comparison involved a 243-nucleotide part of the *cagA* gene separate from the 3′ repeat region. Geographic differences have also been noted in the *vacA* gene genotype (e.g. Portuguese strains appear to be predominantly *vacA* s1b genotype, Dutch strains are predominantly *vacA* s1a genotype, and Asian strains are predominantly *vacA* s1c genotype) [13, 15–17].

We focused on the *cagA* 3′ repeat region because it has a number of advantages for molecular epidemiology studies. For example, because the regions in East Asian isolates are completely different from those in non-Asian isolates they allow one to easily and completely separate them. In contrast, one cannot distinguish the *vacA* s1a genotype from the s1c genotype using the original primers specific for the s1a genotype [17, 19]. In addition, as the *vacA* s1a and s1b genotype, have high homology, the technical aspects of PCR methods become critical as the s1b genotype strains can be falsely identified using either s1a or s1b

https://doi.org/10.1017/S0950268899003209 Published online by Cambridge University Press

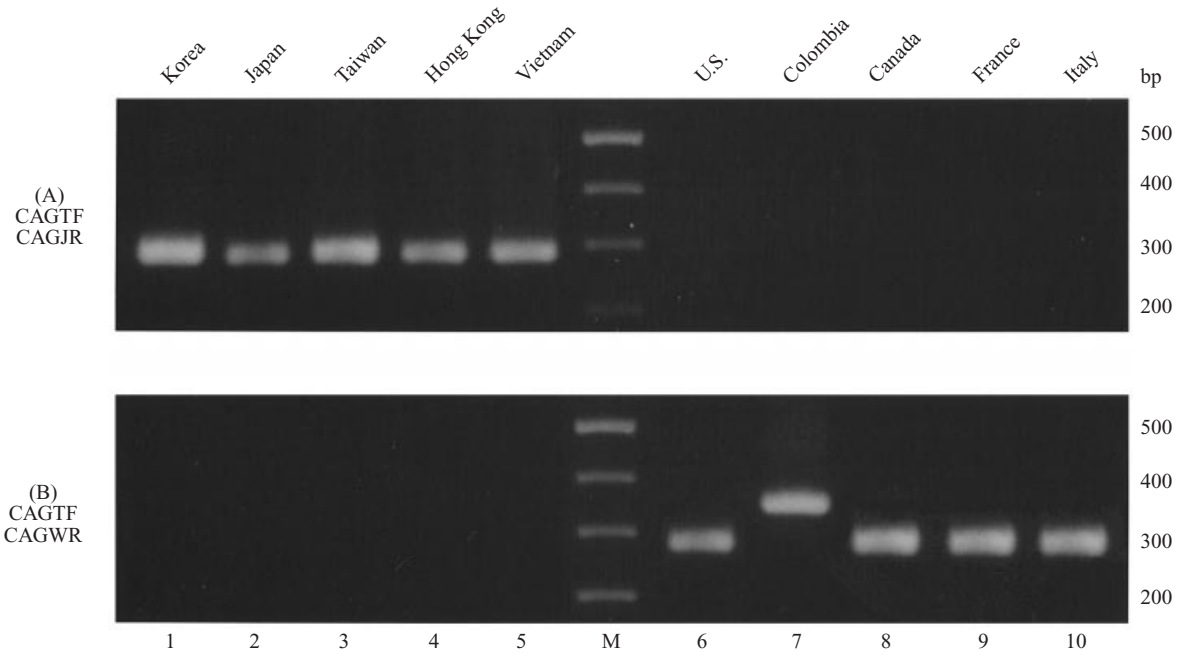

Fig. 2. Analysis of the 3′ region of the *cagA* gene by PCR. PCR products were amplified with the primers CAGTF/CAGJR (primers specific for Japanese strains for the first repeat regions) and CAGTF/CAGWR (primers specific for Western strains for the first repeat regions). Lanes 1–5: East Asian strains (lane 1, from Korea; lane 2, from Japan; lane 3, from Taiwan; lane 4, from Hong Kong; lane 5, from Vietnam). Lanes 6–10: non-Asian strains (lane 6, from USA; lane 7, from Colombia; lane 8, from Canada; lane 9, from France; lane 10, from Italy). The length of PCR product amplified with the primers CAGTF/CAGWR in lane 7 was larger than others because of the existence of two first repeat regions. M: molecular size marker.

specific primers with slightly lower annealing temperature conditions or many PCR cycles. The original reported primers specific for the *vacA* m1 genotype were found to not be useful in all countries (e.g. in Japan the original m1 primers detected only 3% (14/468) of m1 genotype positive *H. pylori* isolates [18]). A final potential advantage of examination of the *cagA* 3′ repeat region is that the region may relate to virulence [10, 20]. The designed primers in this study can clearly distinguish the number of repeat regions.

We are now using this method to determine whether differences in *cagA* genotyping can be used to examine migration patterns of different human populations. Of interest, the Vietnamese population we studied were all young people (ages 20–49, mean age 36) born in Vietnam and currently residing in the USA. All Vietnamese strains had the East Asian *cagA* genotype. A single Japanese strain detected only by Western *cagA* specific primers was from a Japanese woman who was born in India and lived in Japan after she was 30 years old. These examples are consistent with the notion that *H. pylori* infection is typically acquired in childhood and examination of the 3′ repeat region of *cagA* can be used to detect the geographic region of origin of the strain. How or when these strains mix with those of the new country is unknown. We have some preliminary data from Pakistan that showed all the isolates to be of the non-Asian type (unpublished observation). These findings are consistent with anthropological studies that Pakistani and northern Indian peoples are genetically similar to Caucasoid rather than Mongoloid peoples. Studies are underway examining *H. pylori* from different populations to answer this question.

ACKNOWLEDGEMENTS

This work was supported in part by the Department of Veterans Affairs and by NIH grants DK53659, as well as the generous support of Hilda Schwartz. Preliminary results of this study were presented in abstract form at the XIth International Workshop on Gastroduodenal Pathology and *Helicobacter pylori*, Budapest, September 1998 (Gut 1998; **43** Suppl. 2: A39; 05/126).

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 USA 1996; **93**: 14648–53.
2. Crabtree JE, Farmery SM, Lindley IJ, Figura N, Peichl P, Tompkins DS. CagA/cytotoxic strains of *Helicobacter pylori* and interleukin-8 in gastric epithelial cell lines. J Clin Pathol 1994; **47**: 945–50.
3. Crabtree JE, Covacci A, Farmery SM, et al. *Helicobacter pylori* induced interleukin-8 expression in gastric epithelial cells is associated with CagA positive phenotype. J Clin Pathol 1995; **48**: 41–5.
4. Peek RM, Miller GG, Tham KT, et al. Heightened inflammatory response and cytokine expression in vivo to *cagA*+*Helicobacter pylori* strains. Lab Invest 1995; **71**: 760–70.
5. Yamaoka Y, Kita M, Kodama T, Sawai N, Imanishi J. *Helicobacter pylori cagA* gene and expression of cytokine messenger RNA in gastric mucosa. Gastroenterology 1996; **110**: 1744–52.
6. Yamaoka Y, Kita M, Kodama T, Sawai N, Kashima K, Imanishi J. Induction of various cytokines and development of severe mucosal inflammation of *cagA* gene-positive *Helicobacter pylori* strains. Gut 1997; **41**: 442–51.
7. Covacci A, Censini S, Bugnoli M, et al. Molecular characterization of the 128-kDa immunodominant antigen of *Helicobacter pylori* associated with cytotoxicity and duodenal ulcer. Proc Natl Acad Sci USA 1993; **90**: 5791–5.
8. Tummuru MK, Cover TL, Blaser MJ. Cloning and expression of a high-molecular-mass major antigen of *Helicobacter pylori*: evidence of linkage to cytotoxin production. Infect Immun 1993; **61**: 1799–809.
9. Maeda S, Kanai F, Ogura K, et al. High seropositivity of anti-CagA antibody in *Helicobacter pylori*-infected patients irrelevant to peptic ulcers and normal mucosa in Japan. Dig Dis Sci 1997; **42**: 1841–7.
10. Yamaoka Y, Kodama T, Kashima K, Graham DY, Sepulveda AR. Variants of the 3′ region of the *cagA* gene in *Helicobacter pylori* isolates from different *H. pylori* associated diseases. J Clin Microbiol 1998; **36**: 2258–63.
11. Höök-Nikanne J, Perez-Perez GI, Blaser MJ. Antigenic characterization of *Helicobacter pylori* strains from different parts of the world. Clin Diag Lab Immun 1997; **4**: 592–7.
12. Perez-Perez GI, Bhat N, Gaensbauer J, et al. Country-specific constancy by age in *cagA*+ proportion of *Helicobacter pylori* infections. Int J Cancer 1997; **72**: 453–6.
13. van Doorn LJ, Figueiredo C, Rossau R, et al. Typing of *Helicobacter pylori vacA* gene and detection of *cagA* gene by PCR and reverse hybridization. J Clin Microbiol 1998; **36**: 1271–6.
14. van der Ende A, Pan ZJ, Bart A, et al. *cagA*-positive *Helicobacter pylori* populations in China and The Netherlands are distinct. Infect Immun 1998; **66**: 1822–6.
15. van Doorn LJ, Figueiredo C, Sanna R, et al. Expanding allelic diversity of *Helicobacter pylori vacA*. J Clin Microbiol 1998; **36**: 2597–603.
16. van Doorn LJ, Figueiredo C, Mégraud F, et al. Geographic distribution of *vacA* allelic types of *Helicobacter pylori*. Gastroenterology 1999; **116**: 823–30.
17. Yamaoka Y, Kodama T, Gutierrez O, Kim JG, Kashima K, Graham DY. Relationship between *Helicobacter pylori iceA*, *cagA*, and *vacA* status and clinical outcome: studies in four different countries. J Clin Microbiol 1999; **37**: 2274–9.
18. Yamaoka Y, Kodama T, Kita M, Imanishi J, Kashima K, Graham DY. Relationship of *vacA* genotypes of *Helicobacter pylori* to *cagA* status, cytotoxin production, and clinical outcome. Helicobacter 1998; **4**: 241–53.
19. Atherton JC, Cao P, Peek RM, Tummuru MKR, Blaser MJ, Cover TL. Mosaicism in vacuolating cytotoxin alleles of *Helicobacter pylori*. J Biol Chem 1995; **270**: 17771–7.
20. Yamaoka Y, El-Zimaity HMT, Gutierrez O, et al. Relationship between subtypes of the *cagA* 3′ repeat region, gastric histology, and susceptibility to low pH. Gastroenterology 1999; **117**: 342–9.