

## Plasticity of Repetitive DNA Sequences within a Bacterial (Type IV) Secretion System Component

Rahul A. Aras,<sup>1</sup> Wolfgang Fischer,<sup>2</sup> Guillermo I. Perez-Perez,<sup>1</sup> MariaLuisa Crosatti,<sup>1</sup> Takafumi Ando,<sup>3</sup> Rainer Haas,<sup>2</sup> and Martin J. Blaser<sup>1</sup>

<sup>1</sup>Department of Medicine and Department of Microbiology, New York University School of Medicine, and VA Medical Center, New York, NY 10016

<sup>2</sup>Max von Pettenkofer-Institut für Hygiene und Medizinische Mikrobiologie, Ludwig-Maximilians-Universität, 80336 Munich, Germany

<sup>3</sup>First Department of Internal Medicine, Nagoya University School of Medicine, 466-8550 Nagoya, Japan

### Abstract

DNA rearrangement permits bacteria to regulate gene content and expression. In *Helicobacter pylori*, *cagY*, which contains an extraordinary number of direct DNA repeats, encodes a surface-exposed subunit of a (type IV) bacterial secretory system. Examining potential DNA rearrangements involving the *cagY* repeats indicated that recombination events invariably yield in-frame open reading frames, producing alternatively expressed genes. In individual hosts, *H. pylori* cell populations include strains that produce CagY proteins that differ in size, due to the predicted in-frame deletions or duplications, and elicit minimal or no host antibody recognition. Using repetitive DNA, *H. pylori* rearrangements in a host-exposed subunit of a conserved bacterial secretion system may permit a novel form of antigenic evasion.

Key words: *Helicobacter pylori* • antigenic variation • genome plasticity • pathogenicity island • type IV secretion system

### Introduction

The ability of microbial parasites to persist in vertebrate hosts protected by adaptive immunity is a central biological problem. Antigenic variation of surface components is one strategy by which microbial parasites facilitate evasion of host immune responses during long-term colonization (1). Although microbial populations are constantly changing in all colonized niches (2), generation of antigenic diversity at rates faster than the immune system can respond is a challenge that many successful long-term colonizers overcome (3). DNA recombination between paralogous loci containing variable regions is an efficient mechanism for generating diversity, and is used by a wide range of microbes. Such recombination typically involves separate loci and multiple genes (3–6), or tandem repeats within an individual open reading frame (ORF: 7, 8).

*Helicobacter pylori*, Gram-negative, curved bacteria that colonize the human stomach for decades, if not for life, increase risk for peptic ulcer disease and gastric adenocarcinoma (9). Although *H. pylori* elicits vigorous host immune responses, ongoing over the full course of colonization (10,

11), the mechanisms for avoiding immune-based clearance remain unknown. *H. pylori* strains may contain a 40-kb pathogenicity island (termed the *cag*-island; 9, 12, 13). Strains that are *cag*<sup>+</sup> induce increased IL-8 production by gastric epithelial cells in vitro (13, 14) and are associated with increased gastric mucosal cytokine production in vivo (15–17) and increased risk of developing peptic ulcer disease and gastric adenocarcinoma (18). The *cag*-island includes genes encoding subunits of bacterial type IV secretory systems (12–14, 19) that inject their substrate, the CagA protein, into host epithelial cells (20–22) altering specific signal transduction pathways (23, 24).

Type III and IV bacterial secretion systems typically involve the direct injection into host cells of macromolecules that are crucial to the lifestyle of the bacteria (25, 26). This interface represents a frontier important to understanding the evolution of virulence. One component of the *H. pylori* secretion system, *cagY* (HP0527), whose product has homol-

Address correspondence to Rahul A. Aras, Cold Spring Harbor Laboratory, P.O. Box 100, Cold Spring Harbor, NY 11724. Phone: (516) 367-6885; Fax: (516) 367-8435; email: aras@cshl.edu

Abbreviations used in this paper: FCR, 5' conserved region; FRR, 5' repeat region; MRR, middle repeat region; ORF, open reading frame; RFLP, restriction fragment length polymorphism; RAPD, random amplification of polymorphic DNA; TCR, 3' conserved region; VHR, VirB10 homology region.

ogy to *Agrobacterium tumefaciens* VirB10 (27), is required for gastric epithelial cell IL-8 induction (13) and CagA translocation into epithelial (28), indicating its important role in the *H. pylori*-host interaction.

Why is the host unable to mount an effective response against a bacterial constituent that so profoundly affects its physiology? An organism may reside in an immunologically privileged locale, preventing significant host immune recognition, but such is not the case with *H. pylori* (10). Alternatively, *H. pylori* may evade immune clearance through altering its superficial structures. The deduced *cagY* product from *H. pylori* strains 26695, J99, and NCTC11638 has striking NH<sub>2</sub>-terminal and middle region amino acid repeat patterns, and a high multiplet count dominated by lysine- and glutamate-containing doublets (29). Therefore, we hypothesized that the DNA sequences of these repeat structures (29) represent opportunities for *H. pylori* cells to alter genotype and thus phenotype by deletion or duplication of intervening regions during replication (30). Such DNA sequence changes could affect virulence or immune evasion through its effects on gene content (1).

To test the hypothesis that *H. pylori* may use the repetitive *cagY* sequences to facilitate immune evasion, we examined the effect of potential DNA recombination events on DNA sequence and protein expression in relevant *H. pylori* isolates. We show that *H. pylori* use of repetitive *cagY* DNA for recombination allows alteration of amino acid composition of a surface-exposed region of the type IV secretion system pilus (CagY), which would facilitate immune evasion of a potentially vulnerable and critical element of the persistent bacterial interaction with its host.

## Materials and Methods

**Bacterial Strains.** In this study, 90 *H. pylori* isolates from persons born in different regions of the world (United States, 31; Netherlands, 26; Mexico, 5; Argentina, 3; Peru, 4; Columbia, 4; China, 8; Thailand, 3; Korea, 5; United Kingdom, 1) were obtained from a stock collection of *H. pylori* isolates from the NYU *Helicobacter/Campylobacter* strain reference center stored at -70°C in our laboratory. From *H. pylori* strain 84-183, 249 (107 low in vitro passage and 142 high in vitro passage) single colony isolates were examined. 14 strains belonging to a single clonal group had been obtained from 7 members of an extended family from The Netherlands (31). 12 other strains from The Netherlands were obtained from 6 patients from whom *H. pylori* was isolated from antral biopsy specimens at both an initial and follow up endoscopy 7-10 yr apart (32). We also studied 15 isolates of USA strain J166 that had been recovered from the stomach of a male rhesus monkey 10 mo after experimental challenge (33). Chromosomal DNA from each of the strains studied was prepared using a phenol extraction method (34). In addition, isolates of strain B128 (35) used in experimental mouse challenge were studied. Strains included the prechallenge isolate as well as gastric isolates from two FVB/N mice 8 mo after experimental challenge (36).

**PCR Amplification and Restriction Fragment Length Polymorphism (RFLP) Analysis.** Oligonucleotide primers and PCR conditions were used to amplify the 1.5-*flaA* and 1.7-kb *ispA-glmM* gene segments, as previously described (37). The FRRF (5'-ATGAATGAAGAAAACGATAAATTG), FRRR (5'-CACT-

TGAACTTTTTGTTGGTTTCAG), MRRF (5'-GCTTACTG-AACCAACAAAAAGTTCA), and MRRR (5'-CGCTCAAA-CCATCCAAACTTC) primers were designed based on the *cagY* sequence of strain 26695 (27). PCR amplification was performed with a reaction mixture containing 100 ng template DNA, 200 ng of each primer, and 0.5 U *Taq* polymerase (QIAGEN) in a 100- $\mu$ l volume. PCR included 45 cycles at 94°C for 1 min, 54°C for 1 min, and 72°C for 3 min. RFLP analysis was performed as previously described (37).

**Analysis of Genomic Repeat Sequences.** The Genetics Computer Group program, Repeat (Wisconsin Package Version 9.1), was used to assess the presence of DNA direct repeats  $\geq 10$  bp in five arbitrarily selected 5,000 bp regions in *H. pylori* strain 26695. Repeats of 10 (mean = 44.8  $\pm$  4.7), 11 (mean = 14.2  $\pm$  4.0), 12 (mean = 4.6  $\pm$  1.1), 13 (mean = 1.2  $\pm$  0.8), 14 (mean = 0.2  $\pm$  0.4), and 15 bp (mean = 0.2  $\pm$  0.4) were found. However, no repeat sequences  $\geq 16$  bp were found within the 25,000 bp examined, indicating that use of a threshold  $\geq 16$  bp minimizes or eliminates background repeats. Sequences of *cagY* in three independently isolated strains, 26695, J99, and NCTC11638 obtained from GenBank were examined using Repeat to identify identical direct repeats  $\geq 16$  bp flanking or within the gene. Locations of the repeat sequences as well as areas subject to deletion were mapped using Microsoft Excel.

***cagY* Sequence.** For analysis of specific *cagY* sequences, relevant PCR products generated with primers FRRF and FRRR or MRRF and MRRR were purified using the QIAquick gel extraction kit (QIAGEN). The purified PCR product was then sequenced on both strands using an automated Applied Biosystems sequencer in the New York University Cancer Center Core Laboratory, and analyzed using Sequencer 3.1.1 (Gene Code Corp., Inc.).

**Immunoblot Assay.** *H. pylori* cells were harvested from 48-h plate cultures, resuspended in 500  $\mu$ l loading buffer (5% SDS, 0.003% bromophenol blue, 20% glycerol, 0.5% DTT, 1.57% Tris base), and 10  $\mu$ l protein samples were assayed by electrophoresis on a 7% sodium-dodecyl sulfate-polyacrylamide gel. CagY proteins were detected by chemiluminescence with polyclonal rabbit antiserum (1:3,000 dilution) against the *H. pylori* strain J99 CagY subunit (38). The secondary antibody (1:2,000 dilution) was goat  $\alpha$ -rabbit immunoglobulin G-alkaline phosphatase (Boehringer).

**ELISA.** All assays were performed as previously described (32). In brief, microtiter plates were coated with 500 ng of a recombinant 1,000 amino acid CagY fragment isolated from *Escherichia coli* strain 2136 (2136/CagY-J99; reference 38).  $\alpha$ -CagY hyperimmune rabbit serum or sera from *H. pylori*<sup>+</sup> or *H. pylori*<sup>-</sup> patients was used as the primary antibody. The secondary antibody was goat  $\alpha$ -human IgG coupled with horseradish peroxidase (Sigma-Aldrich). In addition to controls without primary or secondary antibody, cells of *E. coli* strain 2136 were included in each assay. To determine specific binding, the optical density values from wells coated with *E. coli* 2136 were subtracted from wells containing *E. coli* 2136/CagY-J99 at the same protein concentration (500 ng per well), and all specimens were examined at least in duplicate in at least three separate assays.

## Results

***H. pylori cagY* Possesses an Extraordinary Number of DNA Repeats.** Addressing whether the extensively repeated CagY amino acid sequence (29) was reflected at the nucleotide level in three strains studied, a mean of 264  $\pm$  34

**Table I.** Characteristics of *H. pylori* *cagY* MRR Direct DNA Repeat Sequences and Their Recombination Potential<sup>a</sup>

<i>H. pylori</i> strain	Repeat sequences <sup>b</sup>			Recombination potential <sup>c</sup>	
	Total number (independent)	Mean $\pm$ SD copies	Mean $\pm$ SD size (bp)	Number of regions <sup>d</sup>	Range in fragment size (bp)
26695	302 (99)	3.1 $\pm$ 1.6	36.1 $\pm$ 22.8	443	111–2,187
J99	235 (78)	3.0 $\pm$ 1.7	48.2 $\pm$ 42.7	347	114–2,205
NCTC11638	255 (82)	3.1 $\pm$ 2.0	34.9 $\pm$ 23.5	425	114–2,544

<sup>a</sup>The MRR of *cagY* in the three strains are shown in Fig. 1.

<sup>b</sup> $\geq 16$  nucleotides (refer to Materials and Methods).

<sup>c</sup>Recombination potential defined as number of regions flanked by direct DNA repeats, and fragment size as intervening segment plus one repeat.

<sup>d</sup>All potential recombination events involving the 1,215 regions identified (leading to either DNA deletion or duplication) result in in-frame ORF ( $P < 0.001$ ).

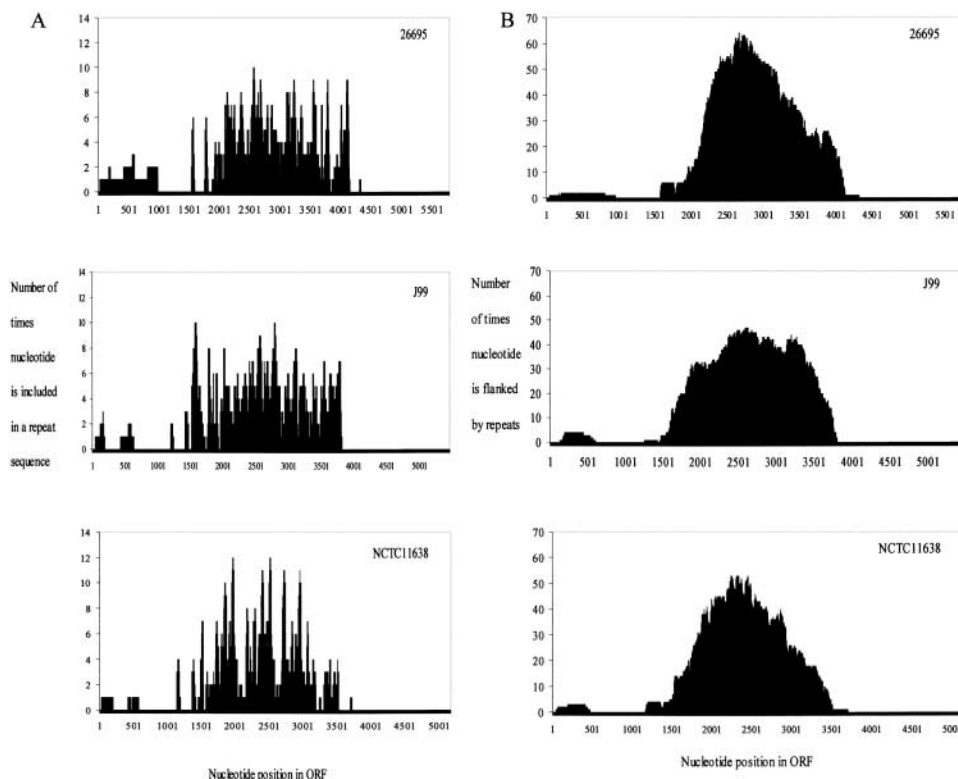
(86.3  $\pm$  11.2 independent) direct repeat sequences  $\geq 16$  nucleotides was found in *cagY* (Table I). Many of the repeat sequences appeared more than twice (mean = 3.07  $\pm$  1.76), with the most common occurring up to 13 times. Although 29 (13%) of the repeat sequences are present in more than one strain, 192 (74, 64, and 54 in strains 26695, J99, and NCTC11638, respectively) others occur in only one of the strains studied. That *cagY* contains extraordinary numbers of both strain-specific and shared direct DNA repeats and that several repeats overlap in multiple regions of the gene, produces enormous numbers of potential variants. The number of potential permutations makes it impossible to define allelic patterns, using a nomenclature similar to that used to describe the CagY amino acid structure (29).

To determine whether the *cagY* repeat structures are characteristic of other pilus or outer membrane protein genes, we examined for the presence of repeat sequences  $\geq 16$  nucleotides in representative genes: *E. coli* *fimA* (39), *papA* (40), and *sfa* (41) and *S. pyogenes* *emm1* (42). *H. pylori* *cagY* from strains 26695, J99, and NCTC11638 contain 52, 43, and 49 repeat sequences  $\geq 16$  bp/KB, respectively. That *fimA* (sequence data are available from GenBank/EMBL/DBJ under accession no. AF490890), *papA* (accession no. NC\_004431), *sfa* (accession no. U38541), and *emm1* (accession no. NC\_002737) contain 0, 0, 0, and 2.6 repeat sequences  $\geq 16$  bp/KB, respectively, provide evidence that the number of repeat sequences in *cagY* is extraordinary and could play a unique functional role in the surface-exposed gene product (38).

*Repeat Sequences in cagY Map to Two Distinct Regions.* The DNA repeats within *cagY* are clustered within two highly conserved areas, that we have termed the 5' repeat region (FRR) and the middle repeat region (MRR; Fig. 1). Within each, the repeats frequently overlap so that a single *cagY* nucleotide in strain 26695 might be part of up to 10 different repeat sequences (Fig. 1 A). However, the downstream copies of the repeat sequences have entirely different overlap patterns. Based on our analyses, *cagY* can be divided into five regions conserved within the three strains: FRR, 5' conserved region (FCR), MRR, 3' con-

served region (TCR), and VirB10 homology region (VHR; Fig. 1 C). Although VirB10 from *A. tumefaciens* and nine other organisms have high homology (BLAST E values  $< 10^{-20}$ ) to the region encoded by the *cagY* VHR, multiple alignments indicate no significant homologies in GenBank to the regions upstream of the VHR. Similarly, separate BLAST searches of the four upstream regions revealed no significant homologies. G+C content of *cagY* upstream of the MRR is lower (32–35%) than the remainder of the gene (38–40%; Fig. 1 C) and the entire *H. pylori* chromosome (39%). The *cagY* ORFs in the three strains range from 5,172 to 5,784 bp. The major differences are a 329-bp deletion in the FRR of strain J99 and NCTC11638, as well as smaller MRR deletions that vary from strain to strain (Fig. 1 C). The FCR, TCR, and VHR each are highly conserved (DNA sequence identity of 96.3–97.3%) in the three strains, similar to the level observed amongst *H. pylori* housekeeping genes (43). Thus, although annotated as a VirB10 homologue (19, 27), further sequence analysis suggests a complex evolutionary history and the presence of other functional domains.

*cagY* Intragenic Recombination Events Always Yield In-Frame Gene Products. Accordingly, we next examined the regions with multiple DNA repeats. The presence of direct DNA repeats allows for deletion (or duplication) of the intervening region plus one copy of the repeat through DNA recombination, involving either slipped strand misalignment or intrachromosomal exchange (44). To understand the potential of the repeats to permit *cagY* gene structure changes, sequences subject to change in copy number by their location between repeats were determined. Two regions mapping to the FRR and MRR were observed (Fig. 1 B). Importantly, each is independent because no sequences  $\geq 16$  bp are present in both the FRR and MRR. Each strain has  $> 300$  different potential recombination events involving the MRR (Table I), predominantly mapping to the middle of the MRR (Fig. 1). Because bacteria often use recombination between repeat sequences for phase-variable gene regulation (1), random distances between paired repeats would yield in-frame deletions one



**Figure 1.** Direct nucleotide repeats in the structure of *H. pylori cagY*. (A) The abscissa represents the nucleotide position in each ORF and the ordinate for each of three strains (26695, J99, NCTC11638) represents the number of times each nucleotide is part of a direct repeat sequence  $\geq 16$  nucleotides. For each strain, the repeats are clustered in two regions. (B) The abscissa represents the nucleotide position in each ORF and the ordinate represents the number of times that identical repeats  $\geq 16$  bp flank each nucleotide. The presence of a nucleotide between identical repeats indicates that it could be deleted or duplicated if recombination occurred between the repeats. (C) For the three strains analyzed, *cagY* might be separated into five defined regions present in each sequence, which we termed the FRR, the FCR, the MRR, the TCR, and the VHR. None of the repeats are shared in both the FRR and MRR. Downstream of both the FRR and MRR are highly conserved regions of  $\sim 550$  bp (FCR and TCR), followed by the VHR portion of *cagY*, encoding a polypeptide highly homologous to the *A. tumefaciens virB10* product. The FCR, TCR, and VHR have high levels of DNA identity between strains, with FCR size differences due to small variations in the boundaries of the adjacent repeat regions. FRR size variation is largely due to a 329-bp deletion (hatched region). The G+C content is lowest and deviates substantially from the remainder of the genome in the portion of *cagY* upstream of the MRR.

C

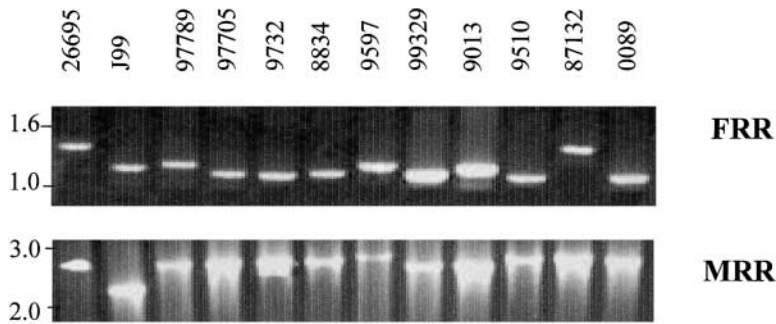
	5' repeat region (FRR)	5' conserved region (FCR)	Middle repeat region (MRR)	3' conserved region (TCR)	VirB10 homology region (VHR)
Strain 26695	1-976	977-1533	1534-4326	4327-4664	4665-5784
J99	1-625	626-1194	1195-4002	4003-4340	4341-5460
NCTC11638	1-574	575-1131	1132-3714	3715-4052	4053-5172
Size range of region (bp):	574-976	556-568	2582-2807	337	1119
Average G+C content (%):	35.5	32.6	38	40	39

third of the time. However, in the three strains studied, every one of the 1,215 potential recombination events produced an in-frame *cagY*, a highly ( $P = 4.0 \times 10^{-243}$ ) non-random phenomenon. Therefore, with two independent regions, the number of potential recombination events within each gene, in each case producing a full-length protein product, is enormous.

**Variation in *cagY* Repeat Region Sizes.** To determine the extent of intergenic *cagY* repeat region variation, PCRs were performed using primers that flank either the FRR or MRR using 62 strains from locations around the world. FRR products ranging from 1.1 to 2.0 kb were amplified (Fig. 2). For most strains, MRR products were  $\sim 2.8$  kb, but all varied between 2 and 3 kb. The geographic origins of individual strains and FRR or MRR size showed no correlation (unpublished data). Thus, whereas *cagY* size varies amongst different strains, there appear to be length boundaries on both FRR and MRR. To examine *cagY* stability in

vitro, MRRs of laboratory-passaged (since 1984) strain 84-183 (84-183HP) and the original 84-183 (84-183LP;  $< 10$  passages since 1984) were examined by PCR. PCR of the *cagY* MRR from all 107 single colony 84-183LP isolates yielded a 2.5-kb product, whereas 142 single colony 84-183HP isolates yielded 74 (52%) products of 2.5 kb, 64 (45%) products of 2.0 kb, and 4 (3%) colonies yielded 2.0, 2.2, and 2.5 kb products. Immunoblots, using  $\alpha$ -CagY serum, performed on cell lysates from the three different 84-183HP colony types yielded CagY products whose sizes were consistent with the corresponding *cagY* PCR products from that colony (Fig. 3 A). These results indicate that during in vitro passage, an *H. pylori* population develops that is heterogeneous in its expressed CagY products, potentially through deletions and/or duplications in *cagY*. To examine in vivo size variability, the MRR from 14 highly related isolates obtained from 7 members of an extended family in The Netherlands reflecting a total of  $> 150$  person-years



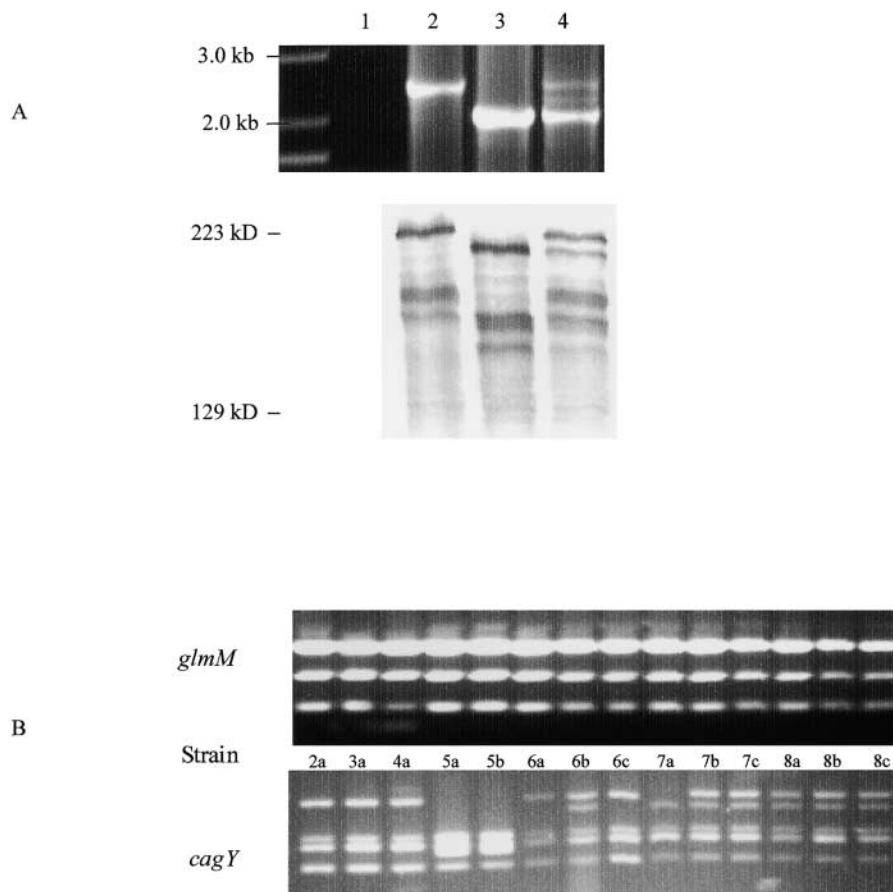


**Figure 2.** Size and sequence diversity in the *cagY* MRR of unrelated *H. pylori* isolates. PCR was performed using 26695 primers FRRF and FRRR that flank the FRR, or using primers MRRF and MRRR that flank the MRR, and products resolved on a 1% agarose gel. The variation in these 12 strains is representative of studies of a global collection of 62 unrelated *H. pylori* strains.

isolation from one another (33) were analyzed. Using 3 different restriction digestions, RFLP patterns for segments of *flaA* (unpublished data) and *glmM* (Fig. 3 B) were nearly identical for the 14 strains (Fig. 3 B), reconfirming the clonal origin of the familial isolates (31, 45). Although PCR of the *cagY* MRR yielded 2.8 kb amplicons for all 14 isolates, RFLP analysis of the products indicated at least 4 different patterns (Fig. 3 B). In two patients (6 and 7), isolates show differing MRR-RFLP patterns. These data provide evidence that the *cagY* MRR undergoes diversification in vivo, during the decades-long human colonization.

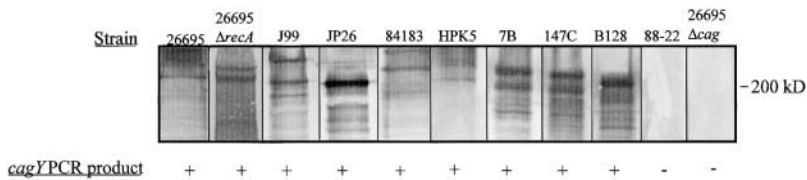
*Intrahost MRR Diversity Resulting from Deletion between Direct DNA Repeats.* Next, we sought to determine whether the direct DNA repeats detected by in silico anal-

ysis were in fact involved in size variation. Parallel studies were performed on pairs of isolates from antral biopsies from six patients over 7–10-yr intervals. Sequence, random amplification of polymorphic DNA (RAPD)-PCR, and AFLP analyses indicate that each pair of strains are clonal variants (32). From patient 13, PCR of the *cagY* MRR yielded a 2.8-kb product from the initial isolate (13aqs) and a 2.5-kb product from the isolate (13bqs) obtained 7.4 yr later (Fig. 4 A). Sequence analysis showed that in 13bqs, a 168-bp segment plus one of the flanking 156-bp direct repeats was deleted (Fig. 4 B). As predicted from analyses of *cagY* in the sequenced strains (Table I), deletion of this 324-bp segment allows for expression of an in-frame protein (Fig. 4 C). Analysis using PCR and immunoblots of *H.*



**Figure 3.** In vitro (A) and in vivo (B) *cagY* allelic diversity among *H. pylori* isolates having the same clonal origin. (A) PCR using primers MRRF and MRRR performed on 142 84–183HP (high passage) single colony isolates identified three *cagY* allele populations with products of 2.5 kb (lane 2), 2.0 kb (lane 3), and 2.0, 2.2, and 2.5 kb (lane 4) compared with identification of a single *cagY* allele population (unpublished data) within strain 84–183LP. Immunoblots of the three representative alleles using  $\alpha$ -CagY serum show the expected protein size variation (lanes 2 and 4) and the mixed alleles (lane 4). PCR and immunoblots performed on a control *cagY*<sup>-</sup> strain, 88–221, yielded no product (lane 1). (B) RFLP patterns in the *cagY* MRR in 14 *H. pylori* isolates from 7 members (patients 2–8) of an extended family from The Netherlands (reference 31). *H. pylori glmM* (control) and the *cagY* MRR were PCR amplified from the isolates and the products were digested with *MboI* for 3 h and electrophoresed on a 1% agarose gel.





**Figure 6.** CagY protein size varies in *H. pylori* strains. Whole cell lysates from nine wild-type *H. pylori* strains were subject to immunoblot using  $\alpha$ -CagY serum. For each *cag*<sup>+</sup> strain, multiple CagY products ranging from  $\sim$ 180–220 kD were identified. In these strains, the CagY bands differ in size, number, and intensity. Strain 26695 $\Delta$ recA yielded multiple CagY bands, indicating that variation does not require RecA-dependent *cagY* deletion or duplication. The complete absence of bands in *cag*<sup>-</sup> strains 88–22 and in 26695 $\Delta$ cag confirms the specificity of the antibodies.

product from mouse 6.16 compared with the other two strains (Fig. 5 A), indicating that deletion of a fragment of the *cagY* MRR had occurred during colonization. Sequence analysis of strain 6.16 identified a 441-bp deletion (including a 33-bp direct repeat of the immediately upstream sequence) in relation to strain 6.11 (Fig. 5 B). Immunoblots using  $\alpha$ -CagY serum indicated that strain 6.16 expressed smaller CagY protein products than those produced by strain B128 or 6.11 (Fig. 5 C), consistent with the sequence data. That for each of the three strains two major CagY protein products were identified, indicates that within a population of cells, the potential exists to produce CagY protein products that differ in size. These comparisons confirm that in vivo recombination between *cagY* direct repeats that result in deletions alters the composition of the in-frame-expressed protein.

**Intrahost FRR Diversity Resulting from DNA Duplication.** The presence of repetitive sequences allows for duplication as well as deletion (44). Examination of a rhesus monkey 10 mo after experimental challenge with seven *H. pylori* strains including strain J166 (33) identified an isolate, 98–147, from which a 1.8-kb FRR PCR product was amplified that was  $\sim$ 0.4 kb larger than the FRR product amplified from the prechallenge J166 strain (unpublished data). RAPD identified the two strains as being of the same clonal origin (33). Sequence analysis of the FRR from these two isolates showed that J166 contained three copies of a 393-bp tandem repeat that begins 23 bp downstream of the start codon, whereas strain 98–147 contained four copies of the repeat. The sequences were otherwise identical, indicating that the size variation resulted from duplication rather than acquisition of another allele through horizontal gene transfer. Immunoblots using the  $\alpha$ -CagY serum indicated that strain 98–147 expressed CagY protein products larger in size than those expressed by strain J166 (not depicted). These in vivo results confirm that recombination between repetitive DNA allows *H. pylori* to undergo chromosomal deletion or duplication, which alters CagY amino acid composition.

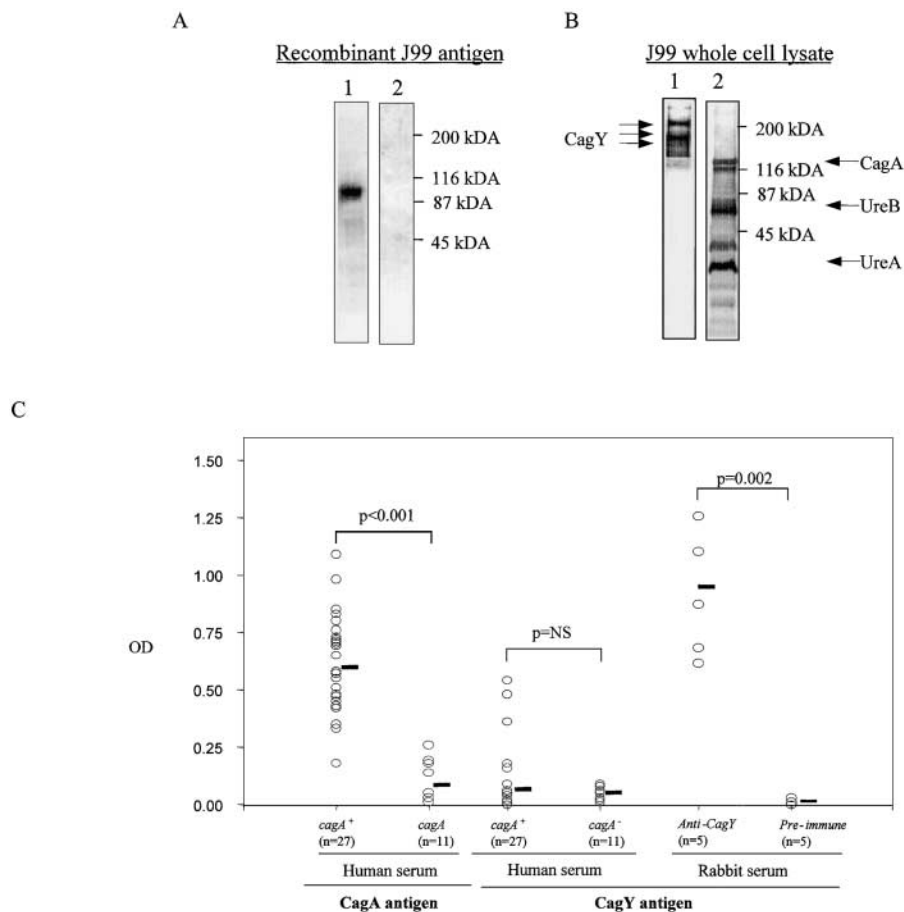
***H. pylori* Strains Produce Varied CagY Proteins.** To determine whether the observed size differences at the DNA level affect CagY protein expression and to assess interstrain and intrastrain CagY size diversity, wild-type *H. pylori* strains were examined by immunoblot using  $\alpha$ -CagY serum. Consistent with the deduced CagY size (19, 27), the  $\alpha$ -CagY serum recognized proteins of  $\sim$ 180–220-kD in all *cagY*<sup>+</sup> strains. Each strain had one to two major bands as

well as multiple minor bands, varying in both size and band intensity (Fig. 6). That the antiserum did not recognize any proteins in *cagY*<sup>-</sup> control strain 88–22 (Fig. 6) or in a *cagY* deletion mutant (not depicted) indicates that the minor bands in the *cagY*<sup>+</sup> strains are CagY specific and is consistent with *H. pylori* cultures containing heterogeneous *cagY* populations (Fig. 3, A and B). Differences in the intensity of CagY products identified in immunoblots may result from differential antibody recognition of particular products, variation in *cagY* transcription levels, or stability of the *cagY* RNA and protein products. To determine whether CagY protein size variation resulted from RecA-dependent *cagY* recombination, immunoblots were performed on a *recA* mutant (26695 $\Delta$ recA; Fig. 6). The presence of multiple CagY products in strain 26695 $\Delta$ recA indicates that CagY size differences do not require RecA-dependent events.

**CagA Translocation and IL-8 Induction by Paired *cagY* Variant Strains.** To examine the effect of *cagY* diversity on type IV secretion system injection of the CagA substrate into host cells (20–22) and cytokine induction (10, 14), CagA translocation and IL-8 induction were examined in AGS cells cocultured with *H. pylori* strains 13abqs, 13bqs, J166, 98–147, and 88–22. For the AGS cells cocultured with each of the four *H. pylori* *cag*<sup>+</sup> strains, immunoblots using  $\alpha$ -CagA antibodies identified  $\sim$ 130 kD cell-associated products (not depicted) indicating CagA injection into the cells and thus, as expected, function of the type IV secretion system in each, and no products were identified in immunoblots of cells cocultured with *cag*<sup>-</sup> strain 88–22 or without coculture. These findings indicate that in-frame *cagY* deletions and duplications do not affect CagA translocation. For *cagY* MRR variant paired strains 13aqs and 13bqs (Fig. 4), coculture with AGS cells yielded significantly different IL-8 induction by 13aqs ( $4.7 \pm 2.9$  ng/AGS cell) and 13bqs ( $0.2 \pm 0.2$  ng/AGS cell). In contrast, there was no significant difference in IL-8 induction between *cagY* FRR variant paired strains J166 ( $3.7 \pm 2.5$  ng/AGS cell) and 98–147 ( $3.8 \pm 3.1$  ng/AGS cell). These results suggest that intragenomic deletions and duplications in the CagY MRR, but not the FRR, could affect IL-8 production by host cells, or that other, unidentified, genetic differences between paired strains 13aqs and 13bqs may influence IL-8 induction.

**CagY Evades Humoral Immunity.** To assess whether surface-exposed CagY domains (38) are recognized by the immune system, IgG responses to a recombinant CagY MRR fragment were examined in sera from persons carrying





**Figure 7.** Serum IgG responses to *H. pylori* CagA and CagY antigens. (A) By immunoblot, the recombinant J99 CagY antigen was recognized by rabbit  $\alpha$ -CagY serum (lane 1), but not by serum from the patient from whom strain J99 was isolated (lane 2). (B) Similarly, using whole cell lysates from strain J99, the rabbit  $\alpha$ -CagY serum (lane 1) recognized multiple high molecular weight CagY bands, whereas patient J99 (lane 2) recognized multiple bands (including CagA, UreB, and UreA), but not CagY. (C) That  $\alpha$ -CagY immune rabbit sera reacted with the CagY antigen whereas control rabbit sera did not, indicates the antigenicity of CagY in the ELISA format. To determine whether the CagY protein is recognized by immune responses in persons with persistent *H. pylori* colonization, sera were examined from 38 *H. pylori*<sup>+</sup> persons for IgG antibodies to CagY and to CagA (as a control). As expected, sera from 27 patients carrying *cagA*<sup>+</sup> strains reacted more strongly ( $P < 0.001$ ) to a recombinant CagA antigen than did sera from 11 patients carrying *cagA*<sup>-</sup> strains. Comparison of IgG responses to the CagY antigen between persons carrying *cagA*<sup>+</sup> and *cagA*<sup>-</sup> strains showed heterogeneous, but generally low level, responses. Although several patients had limited recognition of the recombinant CagY protein, the two groups of patients were not significantly different ( $P = 0.11$ ).

*cagY*<sup>+</sup> strains and from controls. First, because CagY proteins show strain-specific variation (Fig. 6), serum from the person from whom genome strain J99 was isolated (46) was examined. This index patient showed no IgG recognition of either the recombinant CagY fragment (Fig. 7 A), nor of the native CagY in lysates of J99 cells (Fig. 7 B), despite CagY expression by that strain and despite IgG recognition of multiple other *H. pylori* antigens (Fig. 7 B). Thus, despite documented years of carriage of CagY<sup>+</sup> *H. pylori* strains by this patient (46), and his robust responses to other antigens (Fig. 7 B), the CagY protein evaded immune (IgG) recognition.

Next, we sought to determine whether this phenomenon was generalized among persons carrying *cag*-island<sup>+</sup> *H. pylori* strains. First, to determine immune responses to *H. pylori* group antigens, we examined serum IgG responses in 195 patients to *H. pylori* whole cell lysates. As expected, the 114 *H. pylori*<sup>+</sup> patients showed significantly greater recognition of *H. pylori* group antigens (mean ODU  $\pm$  SD =  $4.16 \pm 2.36$ ) than the 81 *H. pylori*<sup>-</sup> patients ( $0.37 \pm 0.27$ ). To ensure a test population capable of mounting strong immune responses to CagY, from among the 114 *H. pylori*<sup>+</sup> patients, we selected 38 persons whose sera showed significantly ( $P < 0.0001$ ) greater IgG recognition ( $6.03 \pm 1.68$ ) of the *H. pylori* antigens. Next, to confirm the antigenicity of the CagY preparation in the ELISA for-

mat,  $\alpha$ -CagY rabbit immune serum was tested. As expected, this serum recognized the recombinant CagY MRR fragment to a substantially greater extent than the control rabbit serum (Fig. 7 C). These ELISA results are consistent with previous immunofluorescence experiments in which the  $\alpha$ -CagY immune rabbit serum recognized the surface-exposed CagY region on intact *H. pylori* cells (38), indicating that the native CagY structure is intrinsically antigenic. However, CagY-specific serum IgG responses in humans were low, and not significantly different in persons colonized with *cag*<sup>+</sup> or *cag*<sup>-</sup> strains (Fig. 7 C). As a control, and as expected, the 27 persons carrying *cag*<sup>+</sup> strains had significantly ( $P < 0.001$ ) higher IgG responses to a recombinant CagA antigen (18) than sera from 11 patients carrying *cagA*<sup>-</sup> strains (Fig. 7 C), indicating that the CagA protein is antigenic and recognized during colonization. Of 27 sera examined from persons with *cag*<sup>+</sup> strains, only five (18.5%) had  $\alpha$ -CagY serum IgG responses  $>$ mean + 3 SD for persons with *cag*<sup>-</sup> strains (Fig. 7 A). These five patients also had significantly higher IgG responses ( $7.1 \pm 1.3$  vs.  $5.5 \pm 1.5$ ;  $P = 0.049$ ) to *H. pylori* whole cell antigens than the 22 other *cagA*<sup>+</sup> patients, but showed no significant differences in CagA-specific IgG responses. In total, the results of these experiments show that among naturally colonized human hosts who mount strong immune responses to both *H. pylori* and CagA anti-



gens, antibody recognition of the surface-exposed CagY protein and its MRR is minimal or absent despite years or decades of CagY expression by *H. pylori*.

## Discussion

The ability of *H. pylori* to persist within hosts that recognize its presence (10, 47) indicates important adaptations, undoubtedly involving programmed cross-talk between microbe and host (48). However, that common mechanisms for adapting to changing environments, including stringent response and two-component regulatory systems, are underrepresented in *H. pylori* (49), suggests the existence of alternate mechanisms for gene regulation. Recombination involving direct DNA repeat sequences, for example, flanking the *hpyII* restriction modification system (45), is one mechanism that affects gene content and thus phenotype. That a large number of direct repeat sequences are present in *H. pylori* *cagY* in a highly skewed distribution and that rearrangements involving these repeats invariably yield in-frame products is a paradigm for use of repetitive DNA to alter amino acid content of an expressed protein (50). The RecA independence of *cagY* rearrangement is consistent with the recent observation that intragenomic recombination in *E. coli* between direct repeats <400 bp does not require RecA function (30). Although the data presented in this study indicate the role of intragenomic (DNA) recombination in influencing CagY sequence diversity, they do not exclude roles at the transcriptional or translational levels.

*H. pylori* populations within individual hosts are diverse (31, 32, 46) with recombination between repeats leading to varied genotypes (45, 51). The presence of multiple *cagY* alleles in a single host indicates generation of phenotypic variation (Figs. 4 and 6), which provides substrate for host selection based on differential fitness (52). Analysis of CagY protein expression indicates that *H. pylori* cell populations that contain dominant *cagY* alleles also include multiple variants. Comparison of 84–183LP and in vitro-passaged 84–183HP CagY products suggests that development of *cagY* allele variants is a spontaneous event that leads to heterogeneous populations. Within a host, the existence of heterogeneous CagY populations provides substrate for selection of the allele with greatest fitness for that in vivo context. The natural competence of *H. pylori* (53, 54), which provides opportunities for both interstrain (55, 56) and intrastrain (45) recombination between *cagY* alleles, yields an even larger repertoire of phenotypes that are substrate for host selection. However, although there is potential for enormous size variation of the MRR due to deletion or duplication, finding that PCR products from all 107 examined strains range between 2 and 3 kb indicates strong selection for that size range. These findings are consistent with constraints involved in formation of pili that function in type IV systems (25, 26), in which the CagY surface-exposed regions must extend beyond the bacterial envelope and yet not interfere with translocation of the substrate macromolecule(s). The presence of deletions and duplica-

tions, and intergenomic recombination as well (37, 43, 53), may generate an extremely large number of *cagY* alleles. *H. pylori* cells with alleles that are too long or short would be selected against.

Antigenic variation is well recognized in bacterial pili, structures highly subject to host immune surveillance (57, 58). By homology with other genes encoding VirB10-like sequences, *cagY* is a chimera, in which the 3' VHR has classical type IV secretion system function (12, 19–22), and the unique upstream sequences have alternative function. Variations in the 5' G+C content suggests that the current form of *cagY* developed after *H. pylori* had become a distinct species. Confocal laser scanning microscopy and electron microscopy indicate that CagY sheaths the needle-like structure of the type IV secretion system (38) involved in CagA injection into host cells (21), thereby making it susceptible to host immune attack. Despite the presence of the needle-like type IV secretion system at one bacterial pole, CagY was identified on the surface of the entire bacterial cell, suggesting that its role is not limited to the type IV secretion system (38). The extraordinary number of DNA repeats within the *cagY* FRR and MRR facilitates sequence variation through intragenomic recombination at a frequency markedly higher than the expected rate of DNA mutation or alteration in *H. pylori* housekeeping genes (38, 55), providing a mechanism for rapid changes in amino acid composition, and thus epitopes, of CagY surface-exposed regions. Our findings indicating both potential and actual extensive strain-specific protein variation, as well as the observed absent or minimal host responses to the *cagY* product, provide evidence that FRR- and MRR-based antigenic variation reduces CagY vulnerability to host immune responses. Alternatively, the repetitive DNA sequences within CagY may facilitate binding of the type IV secretion apparatus to cellular receptors, provide mechanical support, or protect the apparatus from the surrounding acidic environment.

Bacterial type III or IV secretion systems that inject their substrates into host cells typically include components exposed to host immunity (25, 26). The decades-long, naturally occurring *H. pylori* colonization of its host (59), with ongoing immune recognition (16), provides particularly strong selection for antigenic variation of such a critical constituent. The experimental data from mice, rhesus monkeys, and humans indicate that over relatively short intervals (weeks to months), *H. pylori* cells with *cagY* variants created by intragenomic recombination replaced previous populations. The high mutation rates seen in *H. pylori* (60) may generate heterogeneous CagY populations to evade host immune responses. This hypothesis is consistent with immunoblots of CagY antigenic expression (Fig. 6), which indicate that *H. pylori* cell populations vary in several CagY isoforms and in their host recognition. The ability of the rabbit to mount an immune response to the recombinant CagY protein, whereas the natural host mounts minimal or no antibody responses to native CagY, might be because the rabbit is hyperimmunized with a high concentration of

the CagY recombinant protein as a single protein, delivered parenterally, as opposed to crossing the mucosa.

The general lack of human antibody response to CagY, despite colonization that has been ongoing for years to decades, suggests that *cagY* variation could be occurring so rapidly that the host immune system is not capable of mounting an effective response. Alternatively, lack of detectable host antibodies may reflect the power of even low levels of antibodies to act as selecting agents. *H. pylori* adaptation to this phenomenon would also involve generation of variants, but in this model, host responses to particular epitopes are lost over time and most would not be detectable on an ongoing basis. During the long course of colonization, hosts develop robust immune responses to numerous *H. pylori* proteins. However, because these antigens are usually internal to either the bacterial (e.g., urease) or host (e.g., CagA) cells, their immune recognition should not facilitate *H. pylori* clearance. Other *H. pylori* surface constituents, such as BabA and BabB, also have diversity regions (61), and the outermost lipopolysaccharide constituents, the Lewis antigens present on host cells (62) may also be involved in immune evasion. That several other putative *H. pylori* surface proteins also contain repetitive DNA sequences (unpublished data), suggests that CagY immune evasion might be a model for a more general phenomenon in *H. pylori*. The selection inferred from observing a highly evolved system of immune evasion in *H. pylori* implies that functional host immunity exists against organisms living in the lumen of the human stomach. That *Plasmodium falciparum*, an important cause of human malaria, contains Csp, a surface-exposed protein containing direct repeat sequences that undergo high frequency ( $10^{-3}$ – $10^{-2}$ ) intragenomic recombination (63), suggests that the observed CagY deletions and duplications might be representative of immune evasion mechanisms that span a broad range of microbial species.

In conclusion, our findings demonstrate that *H. pylori* uses extensive DNA rearrangement to alter amino acid composition of the surface-exposed region of a required constituent of its type IV secretion system, which effectively evades the host immune response. That several other bacterial species with relatively small (<2 Mb) genomes also possess extensive and nonrandomly distributed DNA repeats (64), suggests that recombination between repetitive DNA sequences (50, 65) might be a common mechanism in such organisms to regulate gene content. The localization of the *cagY* repeats indicates the evolution of programmed rearrangements, whereas the remarkably extensive possibilities suggest a stochastic process. This combination is a powerful mechanism to effect variation of critical host-exposed structures.

This investigation was supported by National Service Awards 5 T32 AI07180-20 and 5 T32 CA09385, and RO1 GM63270 from the National Institutes of Health, the Medical Research Service of the Department of Veterans Affairs, and by SFB576, project B1.

Submitted: 11 March 2003

Revised: 23 July 2003

Accepted: 22 September 2003

## References

1. Finlay, B.B., and S. Falkow. 1997. Common themes in microbial pathogenicity revisited. *Microbiol. Mol. Biol. Rev.* 61: 136–169.
2. Woolhouse, M.E., L.H. Taylor, and D.T. Haydon. 2001. Population biology of multihost pathogens. *Science*. 292: 1109–1112.
3. Krinos, C.M., M.J. Coyne, K.G. Weinacht, A.O. Tzianabos, D.L. Kasper, and L.E. Comstock. 2001. Extensive surface diversity of a commensal microorganism by multiple DNA inversions. *Nature*. 414:555–558.
4. Dworkin, J., and M.J. Blaser. 1997. Nested DNA inversion as a paradigm of programmed gene rearrangement. *Proc. Natl. Acad. Sci. USA*. 94:985–990.
5. McCulloch, R., G. Rudenko, and P. Borst. 1997. Gene conversions mediating antigenic variation in *Trypanosoma brucei* can occur in variant surface glycoprotein expression sites lacking 70-base-pair repeat sequences. *Mol. Cell. Biol.* 17: 833–843.
6. Stevenson, B., S. Casjens, and P. Rosa. 1998. Evidence of past recombination events among the genes encoding the Erp antigens of *Borrelia burgdorferi*. *Microbiol.* 144:1869–1879.
7. Biswas, B., R. Vemulapalli, and S.K. Dutta. 1996. Molecular basis for antigenic variation of a protective strain-specific antigen of *Ehrlichia risticii*. *Infect. Immun.* 66:3682–3688.
8. Madoff, L.C., J.L. Michel, E.W. Gong, D.E. Kling, and D.L. Kasper. 1996. Group B streptococci escape host immunity by deletion of tandem repeat elements of the alpha C protein. *Proc. Natl. Acad. Sci. USA*. 93:4136–4141.
9. Peek, R.M., Jr., and M.J. Blaser. 2002. *Helicobacter pylori* and gastrointestinal tract adenocarcinomas. *Nat. Rev. Cancer*. 2:28–37.
10. Blanchard, T.G., S.J. Czinn, and J.G. Nedrud. 1999. Host response and vaccine development to *Helicobacter pylori* infection. *Curr. Top. Microbiol. Immunol.* 241:183–214.
11. Perez-Perez, G.I., A. Salomaa, T.U. Kosunen, B. Daverman, H. Rautelin, A. Aromaa, P. Knekt, and M.J. Blaser. 2002. Evidence that *cagA*(+) *Helicobacter pylori* strains are disappearing more rapidly than *cagA*(-) strains. *Gut*. 50:4295–4298.
12. Censini, S., C. Lange, Z. Xiang, J.E. Crabtree, P. Ghiara, M. Borodovsky, R. Rappuoli, and A. Covacci. 1996. *cag*, a pathogenicity island of *Helicobacter pylori*, encodes type I-specific and disease-associated virulence factors. *Proc. Natl. Acad. Sci. USA*. 93:14648–14653.
13. Akopyants, N.S., S.W. Clifton, D. Kersulyte, J.E. Crabtree, B.E. Youree, C.A. Reece, N.O. Bukanov, E.S. Drazek, B.A. Roe, and D.E. Berg. 1998. Analyses of the *cag* pathogenicity island of *Helicobacter pylori*. *Mol. Microbiol.* 28:37–53.
14. Tummuru, M.K., S.A. Sharma, and M.J. Blaser. 1995. *Helicobacter pylori picB*, a homologue of the *Bordetella pertussis* toxin secretion protein, is required for induction of IL-8 in gastric epithelial cells. *Mol. Microbiol.* 18:867–876.
15. Peek, R.M. Jr., G.G. Miller, K.T. Tham, G.I. Perez-Perez, X. Zhao, J.C. Atherton, and M.J. Blaser. 1995. Heightened inflammatory response and cytokine expression in vivo to *cagA*<sup>+</sup> *Helicobacter pylori* strains. *Lab Invest.* 73:760–770.
16. Crabtree, J.E. 1998. Role of cytokines in pathogenesis of *Helicobacter pylori*-induced mucosal damage. *Dig. Dis. Sci.* 43: 46–55.
17. Lindholm, C., M. Quiding-Jarbrink, H. Lonroth, A. Hamlet, and A.M. Svennerholm. 1998. Local cytokine response in *Helicobacter pylori*-infected subjects. *Infect. Immun.* 66:5964–5971.

18. Blaser, M.J., G.I. Perez-Perez, H. Kleanthous, T.L. Cover, R.M. Peek, P.H. Chyou, G.N. Stemmermann, and A. Nomura. 1995. Infection with *Helicobacter pylori* strains possessing *cagA* is associated with an increased risk of developing adenocarcinoma of the stomach. *Cancer Res.* 55:2111–2115.
19. Alm, R.A., L.S. Ling, D.T. Moir, B.L. King, E.D. Brown, P.C. Doig, D.R. Smith, B. Noonan, B.C. Guild, B.L. deJonge, et al. 1999. Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen. *Helicobacter pylori*. *Nature.* 397:176–180.
20. Segal, E.D., J. Cha, J. Lo, S. Falkow, and L.S. Tompkins. 1999. Altered states: involvement of phosphorylated CagA in the induction of host cellular growth changes by *Helicobacter pylori*. *Proc. Natl. Acad. Sci. USA.* 96:14559–14564.
21. Odenbreit, S., J. Puls, B. Sedlmaier, E. Gerland, W. Fisher, and R. Haas. 2000. Translocation of *Helicobacter pylori* CagA into gastric epithelial cells by type IV secretion. *Science.* 287:1497–1500.
22. Stein, M., R. Rappuoli, and A. Covacci. 2000. Tyrosine phosphorylation of the *Helicobacter pylori* CagA antigen after *cag*-driven host cell translocation. *Proc. Natl. Acad. Sci. USA.* 97:1263–1268.
23. Higashi, H., R. Tsutsumi, S. Muto, T. Sugiyama, T. Azuma, M. Asaka, and M. Hatakeyama. 2002. SHP-2 tyrosine phosphatase as an intracellular target of *Helicobacter pylori* CagA protein. *Science.* 295:683–686.
24. Selbach, M., S. Moese, C.R. Hauck, T.F. Meyer, and S. Backert. 2002. Src is the kinase of the *Helicobacter pylori* CagA protein in vitro and in vivo. *J. Biol. Chem.* 277:6775–6778.
25. Galan, J.E., and A. Collmer. 1999. Type III secretion machines: bacterial devices for protein delivery into host cells. *Science.* 284:1322–1328.
26. Christie, P.J., and J.P. Vogel. 2000. Bacterial type IV secretion: conjugation systems adapted to deliver effector molecules to host cells. *Trends Microbiol.* 8:354–360.
27. Tomb, J.F., O. White, A.R. Kerlavage, R.A. Clayton, G.G. Sutton, R.D. Fleischman, K.A. Ketchum, H.P. Klenk, S. Gill, B.A. Dougherty, et al. 1997. The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature.* 388:539–547.
28. Fischer, W., J. Püls, R. Buhrdorf, B. Gebert, S. Odebreit, and R. Haas. 2001. Systematic mutagenesis of the *Helicobacter pylori* *cag* pathogenicity island: essential genes for CagA translocation in host cells and induction of interleukin-8. *Mol. Microbiol.* 42:1337–1348.
29. Liu, G., T.K. McDaniel, S. Falkow, and S. Karlin. 1999. Sequence anomalies in the Cag7 gene of the *Helicobacter pylori* pathogenicity island. *Proc. Natl. Acad. Sci. USA.* 96:7011–7016.
30. Lovett, S.T., R.L. Hurley, V.A. Sutura, Jr., R.H. Aubuchon, and M.A. Lebedeva. 2002. Crossing over between regions of limited homology in *Escherichia coli*. RecA-dependent and RecA-independent pathways. *Genetics.* 160:851–859.
31. van der Ende, A., E.A.J. Rauws, M. Feller, C.J.J. Mulder, G.N.J. Tytgat, and J. Dankert. 1996. Heterogeneous *Helicobacter pylori* isolates from members of a family with a history of peptic ulcer disease. *Gastroenterol.* 111:638–647.
32. Kuipers, E.J., D.A. Israel, J.G. Kusters, M.M. Gerrits, J. Weel, A. van der Ende, R.W. van Der Hulst, H.P. Wirth, J. Hook-Nikanne, S.A. Thompson, et al. 2000. Quasispecies development of *Helicobacter pylori* observed in paired isolates obtained years apart from the same host. *J. Infect. Dis.* 181:273–282.
33. Dubois, A., D.E. Berg, E.T. Incecik, N. Fiala, L.M. Heman-Ackah, J. Del Valle, M. Yang, H.P. Wirth, G.I. Perez-Perez, and M.J. Blaser. 1999. Host specificity of *Helicobacter pylori* strains and host responses in experimentally challenged non-human primates. *Gastroenterology.* 16:90–96.
34. Wilson, K. 1987. Preparation of genomic DNA from bacteria. In *Current Protocols in Molecular Biology*. F.A. Ausubel, R. Brent, R.E. Kingston, D.D. Moore, and J.G. Seidman, editors. John Wiley & Sons, New York. 2.4.1–2.4.2.
35. Israel, D.A., N. Salama, C.N. Arnold, S.F. Moss, T. Ando, H.P. Wirth, K.T. Tham, M. Camorlinga, M.J. Blaser, S. Falkow, et al. 2001. *Helicobacter pylori* strain-specific differences in genetic content, identified by microarray, influence host inflammatory responses. *J. Clin. Invest.* 107:611–620.
36. Sozzi, M., M. Crosatti, S.K. Kim, J. Romero, and M.J. Blaser. 2001. Heterogeneity of *Helicobacter pylori* *cag* genotypes in experimentally infected mice. *FEMS Microbiol. Lett.* 203:109–114.
37. Kersulyte, D., H. Chalkauskas, and D.E. Berg. 1999. Emergence of recombinant strains of *Helicobacter pylori* during human infection. *Mol. Microbiol.* 31:31–43.
38. Rohde, M., J. Püls, R. Buhrdorf, W. Fischer, and R. Haas. 2003. A novel sheathed surface organelle of the *Helicobacter pylori* *cag* type IV secretion system. *Mol. Microbiol.* 49:219–234.
39. Peek, A.S., V. Souza, L.E. Eguiarte, and B.S. Gaut. 2001. The interaction of protein structure, selection, and recombination on the evolution of the type-1 fimbrial major subunit (*fimA*) from *Escherichia coli*. *J. Mol. Evol.* 52:193–204.
40. Baga, M., S. Normark, J. Hardy, P. O'Hanley, D. Lark, O. Olsson, G. Schoolnik, and S. Falkow. 1984. Nucleotide sequence of the *papA* gene encoding the Pap pilus subunit of human uropathogenic *Escherichia coli*. *J. Bacteriol.* 157:330–333.
41. Schmoll, T., H. Hoschutzky, J. Morschhauser, F. Lottspeich, K. Jann, and J. Hacker. 1989. Analysis of genes coding for the sialic acid-binding adhesin and two other minor fimbrial subunits of the S-fimbrial adhesin determinant of *Escherichia coli*. *Mol. Microbiol.* 3:1735–1744.
42. Bessen, D.E., and S.K. Hollingshead. 1994. Allelic polymorphism of *emm* loci provides evidence for horizontal gene spread in group A streptococci. *Proc. Natl. Acad. Sci. USA.* 91:3280–3284.
43. Achtman, M., T. Azuma, D.E. Berg, Y. Ito, G. Morelli, Z.J. Pan, S. Suerbaum, S.A. Thompson, A. van der Ende, and L.J. van Doorn. 1999. Recombination and clonal groupings within *Helicobacter pylori* from different geographical regions. *Mol. Microbiol.* 32:459–470.
44. Bzimek, M., C.J. Saveson, V.V. Feschenko, and S.T. Lovett. 1999. Slipped misalignment mechanisms of deletion formation: in vivo susceptibility to nucleases. *J. Bacteriol.* 181:477–482.
45. Aras, R.A., T. Takata, T. Ando, A. van der Ende, and M.J. Blaser. 2001. Regulation of the *HpyII* restriction-modification system of *Helicobacter pylori* by gene deletion and horizontal reconstitution. *Mol. Microbiol.* 42:369–382.
46. Israel, D.A., N. Salama, U. Krishna, U.M. Rieger, J.C. Atherton, S. Falkow, and R.M. Peek, Jr. 2001. *Helicobacter pylori* genetic diversity within the gastric niche of a single human host. *Proc. Natl. Acad. Sci. USA.* 98:14625–14630.
47. Blaser, M.J., and D. Kirschner. 1999. Dynamics of *Helicobacter pylori* colonization in relation to the host response. *Proc. Natl. Acad. Sci. USA.* 96:8359–8364.

48. Hooper, L.V., and J.I. Gordon. 2001. Commensal host-bacterial relationships in the gut. *Science*. 292:1115–1118.
49. Alm, R.A., and T. Trust. 1999. Analysis of the genetic diversity of *Helicobacter pylori*: the tale of two genomes. *J. Mol. Med.* 77:834–847.
50. Achaz, G., E.P. Rocha, P. Netter, and E. Coissac. 2002. Origin and fate of repeats in bacteria. *Nucleic Acids Res.* 30:2987–2994.
51. Bjorkholm, B., A. Lundin, A. Sillen, K. Guillemin, N. Salama, C. Rubio, J.I. Gordon, P. Falk, and L. Engstrand. 2001. Comparison of genetic divergence and fitness between two subclones of *Helicobacter pylori*. *Infect. Immun.* 69:7832–7838.
52. Webb, G.F., and M.J. Blaser. 2002. Dynamics of bacterial phenotype selection in a colonized host. *Proc. Natl. Acad. Sci. USA*. 99:3135–3140.
53. Nedenskov-Sorensen, P., G. Bukholm, and K. Bovre. 1990. Natural competence for genetic transformation in *Campylobacter pylori*. *J. Infect. Dis.* 161:365–366.
54. Tsuda, M., M. Karita, and T. Nakazawa. 1993. Genetic transformation in *Helicobacter pylori*. *Microbiol. Immunol.* 37: 85–89.
55. Suerbaum, S., J.M. Smith, K. Bapumia, G. Morelli, N.H. Smith, E. Kunstmann, I. Dyrek, and M. Achtman. 1998. Free recombination within *Helicobacter pylori*. *Proc. Natl. Acad. Sci. USA*. 95:12619–12624.
56. Kuipers, E.J., D.A. Israel, J.G. Kusters, and M.J. Blaser. 1998. Evidence for a conjugation-like mechanism of DNA transfer in *Helicobacter pylori*. *J. Bacteriol.* 180:2901–2905.
57. Seifert, H.S. 1996. Questions about gonococcal pilus phase- and antigenic variation. *Mol. Microbiol.* 21:433–440.
58. Allred, D.R., J.M.R. Carlton, R.L. Satcher, J.A. Long, W.C. Brown, P.E. Patterson, R.M. O'Connor, and S.E. Stroup. 2000. The *ves* multigene family of *B. bovis* encodes components of rapid antigenic variation at the infected erythrocyte surface. *Mol. Cell.* 5:153–162.
59. Ghose, C., G.I. Perez-Perez, M.G. Dominguez-Bello, D.T. Pride, C.M. Bravi, and M.J. Blaser. 2002. East Asian genotypes of *Helicobacter pylori* strains in Amerindians provide evidence for its ancient human carriage. *Proc. Natl. Acad. Sci. USA*. 99:15107–15111.
60. Bjorkholm, B., M. Sjolund, P.G. Falk, O.G. Berg, L. Engstrand, and D.I. Andersson. 2001. Mutation frequency and biological cost of antibiotic resistance in *Helicobacter pylori*. *Proc. Natl. Acad. Sci. USA*. 98:14607–14612.
61. Pride, D.T., R.J. Meinersmann, and M.J. Blaser. 2001. Allelic variation within *Helicobacter pylori babA* and *babB*. *Infect. Immun.* 69:1160–1171.
62. Wirth, H.P., M. Yang, M. Karita, and M.J. Blaser. 1996. Expression of the human cell surface glycoconjugates Lewis x and Lewis y by *Helicobacter pylori* isolates is related to *cagA* status. *Infect. Immun.* 64:4598–4605.
63. Ayala, F.J., and S.M. Rich. 2000. Genetic variation and the recent worldwide expansion of *Plasmodium falciparum*. *Gene*. 261:161–170.
64. Aras, R.A., J. Kang, A.I. Tschumi, and M.J. Blaser. Extensive repetitive DNA facilitates prokaryotic genome plasticity. *Proc. Natl. Acad. Sci. USA*. In press.
65. Rocha, E.P., A. Danchin, and A. Viari. 1999. Functional and evolutionary roles of long repeats in prokaryotes. *Res. Microbiol.* 150:725–733.