PAPERS

Conservation of the *cag* pathogenicity island is associated with *vacA* alleles and gastroduodenal disease in South African *Helicobacter pylori* isolates

M Kidd, A J Lastovica, J C Atherton, J A Louw

Abstract

***Background*—The development of clinically significant disease in South Africa is associated with the vacuolating cytotoxin gene (*vacA*) s1 genotype but not with the presence of the cytotoxin associated gene *cagA*. *cagA* occurs in >95% of South African isolates and is a variable marker for the entire *cag* pathogenicity island (PAI).**

***Aim*—To characterise the *cag*PAI in South African isolates and to investigate if structural variants of this multigene locus were associated with variations in *vacA* status and clinical outcome.**

***Patients and methods*—We studied 109 *Helicobacter pylori* strains (36 from patients with peptic ulceration, 26 with gastric adenocarcinoma, and 47 with no pathology other than gastritis) for differences in selected genes of the *cag*PAI and alleles of *vacA* by polymerase chain reaction.**

***Results*—All strains were *cagA*$^+$. Sixty five (60%) strains had an intact contiguous *cag*PAI; 78% of peptic ulcer isolates, 73% of gastric adenocarcinoma isolates, but only 40% of gastritis alone isolates ($p< 0.01$). The entire *cagII* region was undetectable in 23% of gastritis alone isolates but in only 8% of peptic ulceration isolates ($p<0.05$). The *vacA* signal sequence and mid region demonstrated a strong relationship between the virulence associated *vacA* s1 ($p<0.005$) and *vacA* m1 ($p=0.05$) alleles and an intact *cag*PAI.**

***Conclusion*—Although a complete *cag*PAI was a feature of most infected individuals, deletions in the 5' region of this genetic locus were associated with gastritis alone and with the non-cytotoxic s2/m2 *vacA* genotype.**

(*Gut* 2001;**49**:11–17)

GI Clinic and Department of Medicine, University of Cape Town and Groote Schuur Hospital, Cape Town, South Africa
M Kidd
J A Louw

Department of Medical Microbiology, University of Cape Town and Red Cross Hospital, Cape Town, South Africa
A J Lastovica

Division of Gastroenterology and Institute of Infections and Immunity, University of Nottingham, Nottingham, UK
J C Atherton

Correspondence to:
Dr JA Louw, GI Clinic, E23, Groote Schuur Hospital, Observatory, 7925, Cape Town, South Africa.
jalouw@curie.uct.ac.za.

Accepted for publication
17 January 2001

Keywords: *cagA*; gastric cancer; *Helicobacter pylori*; pathogenicity; peptic ulcer; vacuolating cytotoxin

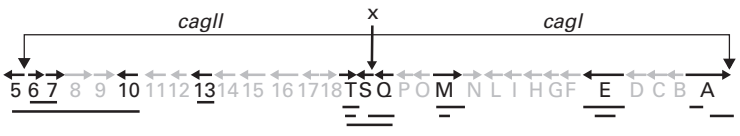

Figure 1 Schematic diagram of the cag pathogenicity island with the putative junction site between cagQ–S where IS605 may be located (vertical arrow, x) noted in NCTC11638. Genes and positions are from GenBank accession number AC000108 (cagII) and U60176 (cagI). Polymerase chain reaction amplimers analysed in this study are denoted by bold horizontal lines beneath the genes.

Helicobacter pylori is the cause of chronic gastritis and is involved in the pathogenesis of peptic ulceration, primary gastric lymphoma, and gastric adenocarcinoma.[1–3] The development of clinically significant disease appears however to depend on a number of factors, including the virulence of the infecting strain, susceptibility of the host, and environmental cofactors.[4]

The majority of vacuolating cytotoxin (VacA) producing isolates[5 6] possess the gene *cagA*, a marker for the *cag* pathogenicity island (*cag*PAI), and production of the CagA protein is associated with clinically significant gastroduodenal disease.[7–10] *cag*$^+$ strains encode a type IV bacterial secretion system which mediates translocation of CagA into host cells where it is phosphorylated and activates a number of intracellular pathways to induce cytoskeletal changes.[11–14] Disruption of many *cag*PAI genes in *cag*$^+$ strains abolishes these effects.

The presence of the *cag*PAI appears to be associated with duodenal ulceration in European and Asian populations[15 16] but patients with peptic ulcer disease sometimes harbour *H pylori* strains with partial or complete deletions of the *cag*PAI[15] and a significant number (53–82%) of isolates from patients without ulcers have intact *cag*PAIs.[15–17] Little is known about the structure of the *cag*PAI in patients with gastric adenocarcinoma or its structure in African populations.

South African *H pylori* isolates from Cape Town are characterised by the universal presence of *cagA*[18] but there is no information on the structure of the *cag*PAI. However, South African isolates have differences in *vacA*, the gene encoding the vacuolating cytotoxin.[18 19] In the current study, we investigated the presence and structural organisation of key regions of the *cag*PAI. The regions targeted and tested using specific polymerase chain reaction (PCR) assays are the functionally important *cagA*, *cagE*, and *cagM* genes in the *cagI* region,[20 21] and three *cagII* marker genes, *cagT* and *cag6–7*[21] (fig 1). Strains which were PCR

Abbreviations used in this paper: PAI, pathogenicity island; PCR, polymerase chain reaction.

Table 1 Clinical profiles

	Gastritis alone (n=37)	*PUD (n=30)*	*GCa (n=19)*
Age (y) (median (range))	39 (14–69)	34 (20–69)	64 (42–82)*
Sex ratio (M:F)	24:13	21:9	14:5
Ethnicity: B:C:W	15:16:6	1:28:1	2:15:2

PUD, peptic ulcer disease; GCa, gastric cancer.
B, Black; C, Cape coloured; W, White.
*p<0.001 *v* gastritis alone and PUD.

negative for these *cagII* marker genes were analysed for the entire *cag5–10* region (which includes the *vir*D4 homologue, *cag10*) as well as for the downstream gene *cag13*. In addition, whether the *cagI* and *cagII* portions of the PAI were contiguous was tested by PCR amplifying their putative junction using primers derived from each side. We hypothesised that *H pylori* isolates from different clinical diseases would exhibit variability in their *cag*PAI and that there would be an association between the type of *cag*PAI and type of *vacA* allele.

Materials and methods

H PYLORI STRAINS AND DNA

A total of 109 *H pylori* isolates were examined from 86 patients. Demographic data and endoscopic and histological diagnoses were recorded for all patients (table 1). Sixty seven (78%) patients had a single strain isolated. Nineteen (22%) patients had two or three non-identical isolates, as assessed by different random amplified polymorphic DNA (RAPD) PCR fingerprint patterns. Fifty six of the strains had previously had their *vacA* types determined in an earlier study.[18] Reference strains 26695 isolated from a patient in the UK with gastritis[22] and J99 isolated from a US patient with duodenal ulcer disease[23] were used as positive controls.

PCR AMPLIFICATION AND DETECTION OF AMPLIFIED DNA PRODUCTS

Oligonucleotide primers used for PCR are listed in table 2. Two sets of primers were used for each gene examined. PCR amplification was performed as previously described.[18 24] Following initial denaturation at 94°C for one minute, each reaction consisted of 35 cycles of denaturation at 94°C for one minute, annealing and extension at an appropriate temperature for 2–4 minutes, and final extension at 72°C for 10 minutes. Annealing temperatures were set at 50°C for cag2/4 and picBF/R; 53°C for cag5/6, cag7/12, cag8/9, cag10/11, cag13/14, cag13F/R, cagTF/R, cagTF2/QR, LECF/cag10R, and cagEF/R; and 55°C for F1/B1, cagMF/R, and cag67F/R.[15 16 25 26] Each PCR mixture (20 µl) was subjected to gel electrophoresis on 1% agarose gels and either a 100 bp or 0.12–23.1 kbp DNA ladder (Roche Diagnostics, Johannesburg, South Africa) was used as a size marker. Long distance PCR was performed with the Expand PCR System (Roche Diagnostics), as recommended by the supplier.[27]

NUCLEOTIDE SEQUENCING

The 3' region of *cagA*, which adequately discriminates between "Asian" and "non-Asian" strains,[28] was sequenced in 12 South African isolates (two Black, seven Cape coloured, three White). PCR products (primer set cag2/4)[25] were gel extracted (Qiaex II gel extraction kit; Qiagen, Cape Town, South Africa) and sequenced on an ABI Prism 377 automated sequencer (ABI, Foster City, California, USA) using the ABI Prism BigDye terminator cycle sequencing reagent kit with AmpliTaq DNA polymerase FS (PE Biosystems, Johannesburg, South Africa), as

Table 2 Primers used to identify cagPAI genes

Gene	*HP No†*	*Primer*	*Primer sequence*	*Location*/+*	*Size (bp)*	*Ref*
cagA 5'	HP547	F1	5' GATAACAGGCAAGCTTTTGAGGGA	19119	349	6
		B1	5' CTGCAAAAGATTGTTTGGCAGA	19468		
cagA 3'	HP547	cag2	5' GGAACCCTAGTCGGTAATG	21498	449	25
		cag4	5' ATCTTTGAGCTTGTCTATCG	21947		
cagE	HP544	picBF	5' TGTTTGGTTTCCCTG	16998	1335	26
		picBR	5' ACGCATTCCTTAACG	15663		
cagE	HP544	cagEF	5' TCTATAAAGAGAGAGGTGTT	17170	2719	16
		cagER	5' GGCTAATCTTTGGTAATCAG	14451		
cagM	HP537	cagMF	5' ATGCTTGCAAAAATTGTTTT	7772	1130	16
		cagMR	5' CTATTCAAAGGGATTATTCT	8902		
cagM	HP537	cag5	5' ACAAATACAAAAAAGAAAAAGAGGC	8022	586	15
		cag6	5' ATTTTTCAACAAGTTAGAAAAAGCC	8608		
cagQ⁻	HP535	CAG12	5' GCAATCATTGAGAAGAGTTTTCGC	1945		15
cagS	HP534	CAG7	5' TTTGGTTGGTAATGGTTTTGGTAGC	6268		
cagQ⁻	HP535	cagTF2	5' GTTTGCTCAGTGGTAAGTGA	1167		16
cagT	HP532	cagQR	5' ATGCTTCCTACTAAAACACG	6569		
cagT	HP532	CAG13	5' TCTAAAAAGATTACGCTCATAGGCG	610	590	15
		CAG14	5' CTTTGGCTTGCATGTTCAAGTTGCC	1100		
cagT	HP532	cagTF	5' ATGAAAGTGAGAGCAAGTGT	21480	842	16
		cagTR	5' TCACTTACCACTGAGCAAAC	22322		
cag13	HP527	cag13F	5' GCTAGAGAAAAGGCTGTTGC	12771	402	16
		cag13R	5' TGGCGTTAATAGTGGCAATA	12369		
cag6⁻	HP520	cag67F	5' ATGGCAACATGGAGATGGTT	4289	885	16
cag7	HP521	cag67R	5' TTAGTTTCCTTTTTTTTCAG	5174		
cag5⁻	HP519	LECF	5' ACATTTTGGCTAAATAAACGCTG	3920	3370	16
cag10	HP524	cag10R	5' TGGGTTCAAGCGAACTGTGA	7290		
tnpA		CAG10	5' ATCAGTCCAAAAAGTTTTTTCTTTCC	4270	344	15
		CAG11	5' TAAGGGGGTATATTTCAACCAACCG	4614		
tnpB		CAG8	5' CGCTCTCCCTAAATTCAAAGAGGGC	5079	569	15
		CAG9	5' AGCTAGGGAAAAATCTGTCTATGCC	5648		

†HP numbers from *Helicobacter pylori* strain 26695 (GenBank accession number AE000511).
*/+Location, position on *H pylori* (GenBank accession number U60176 or AC000108).

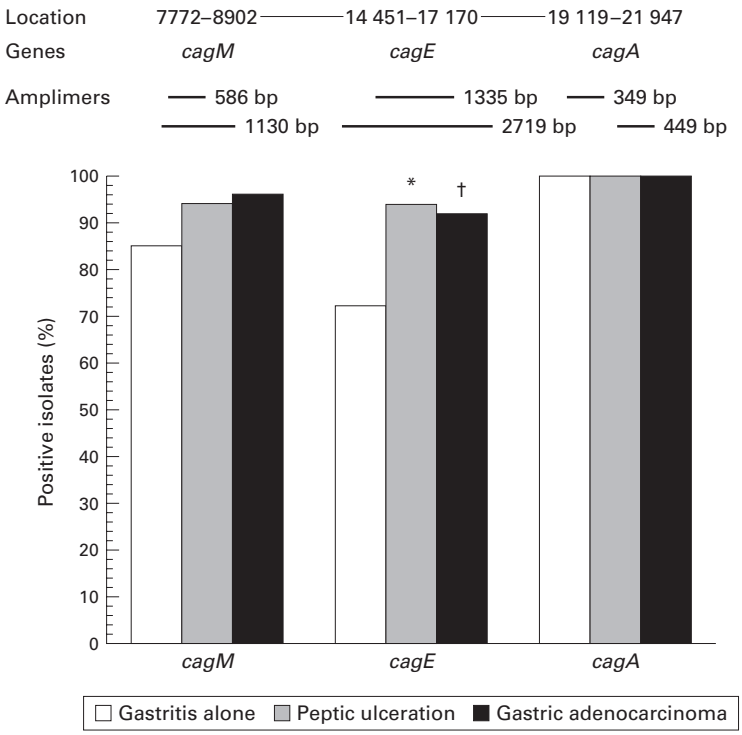

*Figure 2 Schematic diagram of cagI with the polymerase chain reaction amplimers of the three genes analysed (top). Genes and positions are from GenBank accession number U60176. Prevalence of target genes in the different disease groups is shown below. *p<0.009, †p<0.05 versus gastritis alone; Fisher's test.*

described previously.[29] PCR and direct sequencing were performed at least twice to determine DNA sequences for each strain.

COMPUTER ANALYSIS

DNA and protein sequences were analysed using the National Center for Biotechnology Information (NCBI) server (USA), and Internet based searches were preformed at NCBI, the Institute for Genome Research (TIGR, Maryland, USA) and Astra-Zeneca (Boston, USA).

STATISTICAL METHODS

Data (n=109) were analysed using the χ^2 test or Fisher's exact test, as appropriate. Probability levels <0.05 were considered statistically significant.

Results

CLINICAL PROFILES

All samples from patients with clinically significant disease (peptic ulcer disease, gastric adenocarcinoma) were obtained as part of the clinical protocols. Use of non-steroidal anti-inflammatory drugs was an exclusion criterion in all contributing studies. All 30 patients with peptic ulcer disease had active duodenal ulceration at the time of endoscopy. Of the 19 patients with gastric adenocarcinoma, 10 had carcinoma in the antrum, two in the antrum and body, and two in the body alone. In five cases the primary site could not be determined. Tumours were classed histologically as mixed in two cases, poorly differentiated in three, and unknown in six, while four were intestinal and four were diffuse. Patients with no clinically significant upper gastrointestinal disease (gastritis alone) were drawn from two populations. Twenty six (70%) were asymptomatic volunteers receiving no prescribed medication for dyspepsia and 11 were derived from a group of non-ulcer dyspeptics, known not to be receiving potent acid suppressive therapy. Of the 37 patients with gastritis alone, 29 had histological biopsies from the antrum and corpus, allowing for classification of their pattern of gastritis. Of these, 19 (66%) had pangastritis and four (14%) had antral predominant gastritis. Intestinal metaplasia was present in five patients. None of these patients had significant gastroduodenal pathology.

cagA SEQUENCING

Analysis of *cagA* in 12 South African isolates demonstrated no significantly greater differences in homology between isolates from different ethnic groups (83–94% identity) than between isolates from the same ethnic group. *cagA* in all strains demonstrated significantly closer homology with *cagA* from the European strains (HP0547 and JHP495; 73–94% identity) than with *cagA* from the Japanese strains JK25 (GenBank accession number AF043487; 52–57% identity), JK253 (GenBank accession number AF043489; 33–44%), and JK271 (GenBank accession number AF043488; 38–43%). These results confirm that South African *cagA* in this study was "non-Asian" in origin and strongly suggests that these *cag*$^+$ strains share more sequence identity with European[15 17] than with Japanese strains.[16]

PCR ANALYSIS OF THE *cagI* REGION

H pylori isolates were analysed with two different primer sets for the *cagA* (HP547), *cagE* (HP544), and *cagM* (HP537) genes of the *cagI* region. Agreement between primer sets for *cagA* was 97%, *cagE* 100%, and *cagM* 98%. Isolates were considered positive for a gene if one primer set gave successful PCR amplification. The reference strains 26695 and J99 were positive for each primer combination used. Overall, 88 (81%) of 109 strains had all the *cagI* marker genes. All 109 of the strains were *cagA*$^+$ using primers F1/B1 and cag2/4 (fig 2). The gene *cagE* was present in significantly more strains isolated from patients with peptic ulceration (34 of 36 (94%); p<0.009) and gastric adenocarcinoma (24 of 26 (92%); p<0.04) than strains isolated from patients with gastritis alone (34 of 47 (72%)). The gene *cagM* was present in 85% of isolates from patients with gastritis alone, 94% of peptic ulcer disease isolates, and 96% of gastric adenocarcinoma isolates (NS). There were no statistically significant differences in the presence of *cagE* or *cagM* either between ethnic groups or between gastritis alone patients with or without intestinal metaplasia (data not shown).

These results demonstrate that functionally important elements of the *cagI* region (*cagA*–*cagM*) were present in the majority of strains, irrespective of disease status. However, *cagE* (which has previously been shown to be essential for CagA translocation and phosphorylation[14]) was absent in 25% of isolates (from all ethnic groups) from patients with no clinically significant disease.

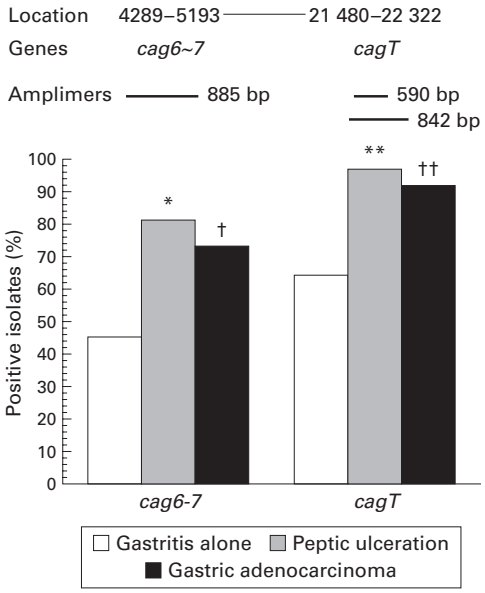

*Figure 3 Schematic diagram of cagII with the polymerase chain reaction amplimers of the three genes analysed (top). Genes and positions are from GenBank accession number AC000108. Prevalence of target genes in the different disease groups is shown below. *p<0.0002, **p<0.0009, †p<0.007, ††p<0.05 versus gastritis alone; Fisher's test.*

PCR ANALYSIS OF THE *cagII* REGION

H pylori isolates were analysed with two different primer sets for *cagT* (HP532), which is present at the 3' end of the *cagII* region, and with one primer set for *cag6–7* (HP520–521) present at the 5' end of this region. Agreement between primer sets for *cagT* was 87%. Overall, 66 (61%) of 109 strains were both *cagT*$^+$ and *cag6–7*$^+$. Thirty five (97%) of 36 isolates from patients with peptic ulceration were *cagT*$^+$ compared with 24 (92%) of 26 isolates from patients with gastric adenocarcinoma and 30 (64%) of 47 isolates from patients with gastritis alone (p<0.0002 *v* peptic ulcer disease; p<0.007 *v* gastric adenocarcinoma) (fig 3). Significantly more isolates from peptic ulcer patients (29 of 36 (81%)) and gastric adenocarcinoma patients (19 of 26 (73%)) were *cag6–7*$^+$ compared with isolates from patients with gastritis alone (21 of 47 (45%); p<0.0009 *v* peptic ulcer disease; p<0.02 *v* gastric adenocarcinoma). Both reference strains produced the expected size *cagT* and *cag6–7* amplicons.

To exclude sequence heterogeneity at primer annealing sites in the 5' region of the *cag*PAI as a reason for the negative PCR results, 40 isolates which were PCR negative for *cag6–7* were also examined for *cag5–10* (HP519–524). This region includes *cag10*, a *vir*D4 homologue and putative toxin, and should result in a PCR product of 3370 bp (predicted from GenBank accession number AC000108). Fourteen (35%) of 40 isolates gave no PCR product, and these are further examined below. The other 26 (65%) of *cag6–7*$^-$ isolates gave PCR amplimers ranging in size from 880 to 2902 bp compared with the expected 3370 bp obtained in all 10 *cag6–7*$^+$ isolates (and in the two reference strains) tested, confirming deletions in the *cagII* region. To define if the deleted *cagII* region extended beyond *cag10*, we next tested whether an additional *cagII* marker gene, *cag13* (HP527), which is between *cag10* and *cagT*, was present in these 26 isolates. Twenty five (96%) of these 26 isolates were *cag13*$^+$ (as were all 10 *cag6–7*$^+$ isolates tested as positive controls, and the two reference strains). Long PCR using primers to examine the entire *cagII* region (*cag5-cagT*) was performed on the one isolate negative for *cag13*. This should result in a PCR product of 18 402 bp (predicted from GenBank accession number AC000108). A product of 8396 bp was amplified. These results suggest that 26 isolates, which were all PCR negative (using two different primer sets) for *cagT*, contain a partial *cagII*.

Thereafter we investigated 14 isolates which appeared to have lost the 5' region of *cagII* (PCR negative for both *cag6–7* and *cag5–10*) using long PCR to examine the entire *cagII* region (*cag5-cagT*). PCR amplicons of 280–3105 bp (compared with a product of 18 402 bp predicted from GenBank accession number AC000108) were generated in 10 of the 14 isolates. These results suggest that these isolates had lost the majority of the *cagII* region. Analysis of the remaining four isolates which gave no PCR amplimers with these primers demonstrated that one isolate had lost all the genes tested to the left of *cagM*; one isolate, all genes tested to the left of *cagE*, and two isolates only contained *cagA*.

Finally, we related these results to clinical status. This showed that a complete intact *cagII* region (*cag6-cagT*) was present in significantly more isolates from patients with peptic ulcer disease (81%) than in patients with gastritis alone (38%; p=0.0001). There were no differences in the distribution of this marker in gastritis alone patients with dyspepsia (36%) or without symptoms (46%). Interestingly, however, the five gastritis alone patients with intestinal metaplasia also had an intact *cag*PAI. Seventy three per cent of isolates from patients with gastric adenocarcinoma had a complete intact *cagII* (p=0.004 *v* gastritis alone). Some elements of the *cagII* region (*cag10-cagT*) were present in all gastric adenocarcinoma isolates whereas 11 isolates from patients with gastritis alone lacked the entire *cagII* region (p<0.006 *v* gastritis alone). Further analysis demonstrated no significant relationship between different ethnic populations and the presence of an intact *cagII* region (χ^2=8.08, p=0.09).

ASSESSMENT OF CONTIGUITY OF *cagI* AND *cagII* REGIONS

The primer combinations of *cag7/12* and *cagTF/QR* test for the presence of the *cagQ~S/T* genes (HP535–532) and also for joined *cagI* and *cagII* regions. Agreement between primer sets for *cagQ~S/T* was 78%. This is lower than for *cagI* but reflects the fact that these primer combinations identify different 5' genes. PCR amplification using these primer sets resulted in the expected amplicon sizes in both 26695 and J99. Seventy five (85%) isolates with all the marker genes for the *cagI* region were *cagQ~S/T*$^+$. Significantly more *cagI* isolates from patients with peptic ulceration (31 of 34 (91%); p<0.02) and gastric adenocarcinoma (23 of 23 (100%); p<0.002)

Table 3 Relationship between cagPAI gene markers and vacA alleles

cag genes detected	*HP numbers*	*s1m1 (n=66)*	*s1m2 (n=17)*	*s2m1 (n=3)*	*s2m2 (n=17)*
All tested genes detected	HP520-HP547$^+$	52 (78%)*	10 (59%)	2 (67%)	2 (12%)
Some cag genes detected	HP524-HP547$^+$	7 (11%)	3 (18%)	1 (33%)	1 (6%)
	HP532-HP547$^+$	2 (3%)	0 (0%)	0 (0%)	0 (0%)
	HP544-HP547$^+$	0 (0%)	1 (6%)	0 (0%)	1 (6%)
	Intervening sequences missing	4 (6%)*	3 (18%)†	0 (0%)	13 (76%)
Only cagA detected	HP547$^+$	1 (2%)	0 (0%)	0 (0%)	0 (0%)

*$p<10^{-5}$ *v* vacA s2m2 (Fisher's test); †$p<0.001$ *v* vacA s2m2 (Fisher's test).

were *cagQ~S/T*$^+$ compared with isolates from patients with gastritis alone (21 of 31 (68%)). Sixty five (98%) of 66 isolates with all the marker genes for the *cagII* region were *cagQ~S/T*$^+$. Overall, sixty five (60%) isolates all had markers for *cagI*, *cagII*, and a contiguous pathogenicity island; 77% of isolates from patients with peptic ulcer disease ($p<0.001$) and 73% of isolates from patients with gastric adenocarcinoma ($p<0.01$) compared with only 18 (40%) isolates from patients with gastritis alone.

Twenty eight isolates did not have an amplifiable *cagQ~S/T* product. One of the isolates was also *cagT*, *cag13*, and *cag6–7* PCR positive. This suggests that this isolate (from a patient with peptic ulcer disease) had a complete but spatially separated pathogenicity island. Sixteen of the remaining 27 *cagQ~S/T*$^-$ isolates (59%) had deletions in the *cagI* region while 26 (96%) had deletions in the *cagII* region. These results suggest that the presence of a *cagQ~S/T* amplimer may be an alternate marker for the *cagII* region.

PREVALENCE OF THE IS605 INSERTION SEQUENCE

While 26695 had an intact *IS605* sequence (*tnpA+tnpB*) detected by PCR and J99 only exhibited *tnpA*, only three (3%) South African isolates were PCR positive for *IS605*. All three had all the marker genes for *cag*PAI and yielded PCR products with the *cag7/12* and *cagTF/QR* primer sets, suggesting that the complete insertion sequence was elsewhere in the genome. Analysis of the *cag*PAI deletion end points in the 28 isolates which did not have an amplifiable *cagQ~S/T* gene product demonstrated that *IS605 (tnpA)* was present in six (27%), suggesting that this insertion sequence may play a role in *cag*PAI disruption in a proportion of isolates.

RELATIONSHIP BETWEEN GENE MARKERS AND *vacA* STATUS

As previously demonstrated,[18 19] *vacA* s1 occurred significantly more often in isolates from patients with peptic ulceration ($p<10^{-5}$) or gastric adenocarcinoma ($p<2\times10^{-4}$) while *vacA* s2 invariably occurred in patients with gastritis alone. In addition, the *vacA* m1 mid region type was present more often in patients with clinically significant disease ($p<0.01$).

When analysing the data by *vacA* status (table 3), there was a significant difference in the distribution of *vacA* allele types between the different *cag*PAI groups ($\chi^2=52.76$; $p<10^{-5}$). A strong association was noted between *vacA* s1m1 and strains containing the complete *cag*PAI. Fifty two (78%) of 66 *vacA* s1m1 strains had all the genetic markers (*cag6-cagA*) compared with two of 18 (12%; $p<10^{-5}$) *vacA* s2m2 strains. Sixteen of 18 *vacA* s2m2 strains (88%) had partial deletions of the *cag*PAI ($p<0.0001$ *v vacA* s1m1; $p<0.005$ *v vacA* s1m2).

There was also a strong association between specific *vacA* alleles and absence of contiguous *cag*PAI. Significantly more *vacA* s2 isolates (13 of 20 (65%)) were *cagQ~S/T* PCR negative compared with 14 of 83 (17%; $p<0.0001$) *vacA* s1 strains, suggesting that the *vacA* s2 allele may be a associated with loss of this region.

Discussion

Our results indicate that the *cag* PAI, using a subset of previously defined functionally important marker genes, appears to be complete and contiguous (*cag6-cagA*) in 60% of South African *H pylori* isolates. However, more than one third of isolates had non-amplifiable gene regions in the island, despite possessing *cagA*. The caveat of over interpretation of PCR results is recognised, as this methodology can fail for a number of reasons, including unexpected sequence divergence at primer annealing sites. Such heterogeneity could conceivably explain our findings although our interpretation of deletion of *cag* regions is much more likely for several reasons. Firstly, we took care to use two established different primer sets per gene site to confirm negative forms. Secondly, these primer combinations have previously been used to identify *cag* genes in European populations,[15 17] and South African *H pylori cag*$^+$ isolates share more sequence homology with European than "Asian" strains (this study and Achtman and colleagues[30]). Thirdly, "missing" gene markers were almost invariably adjacent which could readily be explained by missing regions of *cag* but not by differences in primer annealing.

The insertion sequence *IS605* replaced the deleted regions of the *cagQT* in 25% of strains with deletions. This supports the hypothesis that this DNA element may generate rearrangements and deletions in a small proportion of South African strains to result in subpopulations of organisms with differences in virulence.[20 21] How deletions occur in other strains is unclear. Irrespective of the mode of genetic deletion, strains that carry the *cagA* gene but have internal deletions in the *cag*PAI should probably be classified as functionally *cag*$^-$ rather than *cag*$^+$. This is evidenced by the fact that such strains are often not associated with disease (current study) and that many different artificial mutants in *cag* genes reduce the

ability of *cag*$^+$ strains to induce IL-8 secretion from epithelial cells.[20]

While the majority of isolates (~80%) from patients with peptic ulceration had all the marker genes for a complete *cag*PAI, less than half (~40%) of isolates from patients with gastritis alone had this pathogenic fingerprint. All (100%) tested strains, irrespective of the organisation of the PAI, were *cagA*$^+$. Data on the *cag*PAI from the rest of Africa are scarce but CagA appears to be commonly expressed.[31] Our data suggest dissociation between the presence of *cagA* and other genes in the *cag*PAI island and further that analysis of the PAI may be a prerequisite for investigation of relationships with gastroduodenal disease processes. The presence of an intact *cag*PAI in the majority of South African peptic ulcer disease isolates supports a role for the genes in this island in the pathogenesis of this disease.

Analysis of the distribution of the virulence associated *vacA* alleles demonstrated that subtype s1/m1 was strongly associated with a complete island while subtype s2/m2 was associated with deletions in *cagII*. These findings suggest a functional association between an active s1/m1 vacuolating cytotoxin and an intact *cag* pathogenicity island. This is supported by the observation of a significantly negative relationship between the virulence associated *vacA* s1 allele and deletions in *cagII*.

The association between genes in the *cag*PAI other than *cagA* and gastric adenocarcinoma has not previously been reported. Our results demonstrate significant differences between isolates from patients with gastritis alone and those with this disease. Specifically, an intact frequently contiguous PAI was found more often in patients with cancer than in patients without clinically significant disease. The high prevalence of this marker made subanalysis of the gastric cancer types inappropriate. Interestingly, an intact *cag*PAI was found more often in gastritis alone patients who had intestinal metaplasia. Intestinal metaplasia is known to be a risk factor for gastric adenocarcinoma.

The finding of a conserved pathogenicity element (*cag*PAI) shared by isolates from patients with gastric adenocarcinoma and peptic ulcer disease is of interest given the inverse relationship between these two diseases. That strains with an intact *cag*PAI are present in the older gastric cancer group and the younger peptic ulcer group suggests that the presence of an intact PAI in infecting strains is not due to an age cohort effect. It also suggests that an intact PAI, while being a marker for more severe mucosal damage, may not be a specific marker for either of these diseases. This is entirely compatible with the current understanding of the multifactorial nature of their pathogenesis where the distribution of gastric mucosal damage may be determined by factors such as host genetics and the environment.[4] An alternative or additional hypothesis however is that other bacterial elements may be important. For example, it is possible that while an intact PAI (type IV secretion system) is necessary to deliver CagA into epithelial cells it is the structure of CagA which determines which intracellular pathway (secretory or proliferative) a cell undergoes.[32]

The presence of a contiguous *cag*PAI in 60% of patients suggests that it is conserved in most South Africans as it is in most European and almost all Asian populations.[15–17] The importance of genes in this island (particularly in *cagII*) to the pathogenesis of gastroduodenal disease in South Africa is however suggested by the prevalence of deletions in the 5' region in patients with gastritis alone. In addition, the strong relationship between the virulent *vacA* type s1/m1 and the entire PAI would seem to support the importance of both of these elements to disease pathogenesis although it appears that the relationship with clinically significant disease is stronger for *vacA* alleles than for an intact *cag*PAI. The finding of similar pathogenic elements from strains isolated from patients with either peptic ulcer disease or gastric adenocarcinoma however indicates that further work is required to differentiate the relationship between specific genes, host factors, and disease processes.

This study was partially funded by a David and Freda Becker Trust Award (JAL) and an Abbott-SAGES Award (MK). MK is a recipient of the Claude Harris Leon Foundation Fellowship.

1 Warren J. Unidentified curved bacilli on gastric epithelium in active chronic gastritis. *Lancet* 1983;**i**:1273–5.
2 Wotherspoon A, Ortiz-Hidalgo O, Falzon M, *et al. Helicobacter pylori*-associated gastritis and primary B-cell gastric lymphoma. *Lancet* 1991;**338**:1175–6.
3 Parsonnet J, Friedman G, Vandersteen D, *et al. Helicobacter pylori* infection and the risk of gastric carcinoma. *N Engl J Med* 1991;**325**:1127–31.
4 Atherton J. The clinical relevance of strain types of *Helicobacter pylori*. *Gut* 1997;**40**:701–3.
5 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.
6 Tummuru M, Cover T, Blaser M. 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.
7 Kuipers E, Perez-Perez G, Meuwissen S, *et al. Helicobacter pylori* and atrophic gastritis: importance of *cagA* status. *J Natl Cancer Inst* 1995;**87**:1777–80.
8 Blaser M, Perez-Perez G, Kleanthous H, *et al.* Infection with *Helicobacter pylori* strains possessing *cagA* is associated with an increased risk of developing adenocarcinoma of the stomach. *Cancer Res* 1995;**18**:2111–15.
9 Crabtree J, Taylor J, Wyatt J, *et al.* Mucosal IgA recognition of *Helicobacter pylori* 120 kDa protein, peptic ulceration and gastric pathology. *Lancet* 1991;**338**:332–5.
10 Weel J, van der Hulst R, Gerrits Y, *et al.* The interrelationship between cytotoxin-associated gene A, vacuolating cytotoxin, and *Helicobacter pylori*-related diseases. *J Infect Dis* 1996;**173**:1171–5.
11 Segal E, Cha J, Lo J, *et al.* Altered states: involvement of phosphorylated CagA in the induction of host cellular growth changes by *Helicobacter pylori*. *Proc Natl Acad Sci USA* 1999;**96**:14559–64.
12 Asahi M, Azuma T, Ito S, *et al. Helicobacter pylori* CagA protein can be tyrosine phosphorylated in gastric epithelial cells. *J Exp Med* 2000;**191**:593–602.
13 Stein M, Rappuoli R, Covacci A**.** Tyrosine phosphorylation of the *Helicbacter pylori* CagA antigen after *cag*-driven host cell translocation. *Proc Natl Acad Sci USA* 2000;**97**:1263–8.
14 Odenbreit S, Puls J, Sedlmaier B, *et al.* Translocation of *Helicobacter pylori* CagA into gastric epithelial cells by type IV secretion. *Science* 2000;**287**:1497–500.
15 Jenks P, Megraud F, Labigne A**.** Clinical outcome after infection with *Helicobacter pylori* does not appear to be reliably predicted by the presence of any of the genes of the *cag* pathogenicity island. *Gut* 1998;**43**:752–8.
16 Maeda S, Yoshida H, Ikenoue T, *et al.* Structure of the *cag* pathogenicity island in Japanese *Helicobacter pylori* isolates. *Gut* 1999;**44**:336–41.
17 Slater E, Owen R, Williams M, *et al.* Conservation of the *cag* pathogenicity island of *Helicobacter pylori*: Associations with vacuolating cytotoxin allele and IS*605* diversity. *Gastroenterology* 1999;**117**:1308–15.
18 Kidd M, Lastovica A, Atherton J, *et al.* Heterogeneity in the *Helicobacter pylori* genes *vacA* and *cagA*: Associated with gastroduodenal disease in South Africa? *Gut* 1999;**45**:499–503.

19 Kidd M, Lastovica A, Atherton J, *et al.* Specific genotypes *of Helicobacter pylori vacA* and *cagA* but not the presence *of cagA*, are associated with gastroduodenal disease in South Africa. *Gastroenterol* 1999;**116**:G0928.
20 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.
21 Akopyants N, Clifton S, Kersulyte D, *et al.* Analyses of the *cag* pathogenicity island of *Helicobacter pylori*. *Mol Microbiol* 1998;**28**:37–53.
22 Tomb J-F, White O, Kerlavage A, *et al.* The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature* 1997;**388**:539–47.
23 Alm R, Ling L-S, Moir D, *et al.* Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori*. *Nature* 1999;**397**:176–80.
24 Letley D, Lastovica A, Louw J, *et al.* Allelic diversity of *Helicobacter pylori* vacuolating cytotoxin gene in South Africa: rarity of the *vac* s1a genotype and natural occurrence of an s2/m1 allele. *J Clin Microbiol* 1999;**37**: 1203–5.
25 Rudi J, Kolb C, Maiwald M, *et al.* Diversity of *Helicobacter pylori vacA* and *cagA* genes and relationship to VacA and CagA protein expression, cytotoxin production and associated diseases. *J Clin Microbiol* 1998;**36**:944–8.
26 Wirth H-P, Yang M, Karita M, *et al.* Expression of the human cell surface glycoconjugates Lewis x and Lewis y by *Helicobacter pylori* isolates is related to *cagA* status. *Infect Immun* 1996;**64**:4598–605.
27 van der Water N, Williams R, Browett P**.** Amplification of a 29.7 kb region of the factor VIII grnr using the expand PCR system. *Biochemica* 1996;**2**:11–12.
28 Yamaoka Y, Osato M, Sepulveda A, *et al.* Molecular epidemiology of *Helicobacter pylori*: separation of *H. pylori* from East Asian and non-Asian countries. *Epidemiol Infect* 2000;**124**:91–6.
29 Warnich L, Kotze M, Groenewald I, *et al.* Identification of three mutations and associated haplotypes in the protoporphyrinogen oxidase gene in South African families with variegate porphyria. *Hum Mol Genet* 1996;**5**:981–4.
30 Achtman M, Azuma T, Berg D, *et al.* Recombination and clonal groupings within *Helicobacter pylori* from different geographical regions. *Mol Microbiol* 1999;**32**:459–70.
31 Mandour El-Mahdi A, Pratchett S, Char S, *et al.* Does CagA contribute to ulcer pathogenesis in a developing country such as Sudan? *Eur J Gastroenterol Hepatol* 1998;**10**:313–16.
32 Covacci A, Rappuoli R**.** Tyrosine-phosphorylated bacterial proteins: Trojan horses for the host cell. *J Exp Med* 2000;**191**:587–92.