LEABHARLANN CHOLÁISTE NA TRÍONÓIDE, BAILE ÁTHA CLIATH Ollscoil Átha Cliath

TRINITY COLLEGE LIBRARY DUBLIN The University of Dublin

Terms and Conditions of Use of Digitised Theses from Trinity College Library Dublin

Copyright statement

All material supplied by Trinity College Library is protected by copyright (under the Copyright and Related Rights Act, 2000 as amended) and other relevant Intellectual Property Rights. By accessing and using a Digitised Thesis from Trinity College Library you acknowledge that all Intellectual Property Rights in any Works supplied are the sole and exclusive property of the copyright and/or other IPR holder. Specific copyright holders may not be explicitly identified. Use of materials from other sources within a thesis should not be construed as a claim over them.

A non-exclusive, non-transferable licence is hereby granted to those using or reproducing, in whole or in part, the material for valid purposes, providing the copyright owners are acknowledged using the normal conventions. Where specific permission to use material is required, this is identified and such permission must be sought from the copyright holder or agency cited.

Liability statement

By using a Digitised Thesis, I accept that Trinity College Dublin bears no legal responsibility for the accuracy, legality or comprehensiveness of materials contained within the thesis, and that Trinity College Dublin accepts no liability for indirect, consequential, or incidental, damages or losses arising from use of the thesis for whatever reason. Information located in a thesis may be subject to specific use constraints, details of which may not be explicitly described. It is the responsibility of potential and actual users to be aware of such constraints and to abide by them. By making use of material from a digitised thesis, you accept these copyright and disclaimer provisions. Where it is brought to the attention of Trinity College Library that there may be a breach of copyright or other restraint, it is the policy to withdraw or take down access to a thesis while the issue is being resolved.

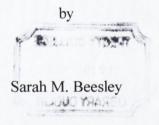
Access Agreement

By using a Digitised Thesis from Trinity College Library you are bound by the following Terms & Conditions. Please read them carefully.

I have read and I understand the following statement: All material supplied via a Digitised Thesis from Trinity College Library is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of a thesis is not permitted, except that material may be duplicated by you for your research use or for educational purposes in electronic or print form providing the copyright owners are acknowledged using the normal conventions. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone. This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

Microevolution in Helicobacter pylori

A thesis submitted for the degree of Doctor of Philosophy



Moyne Institute of Preventive Medicine,

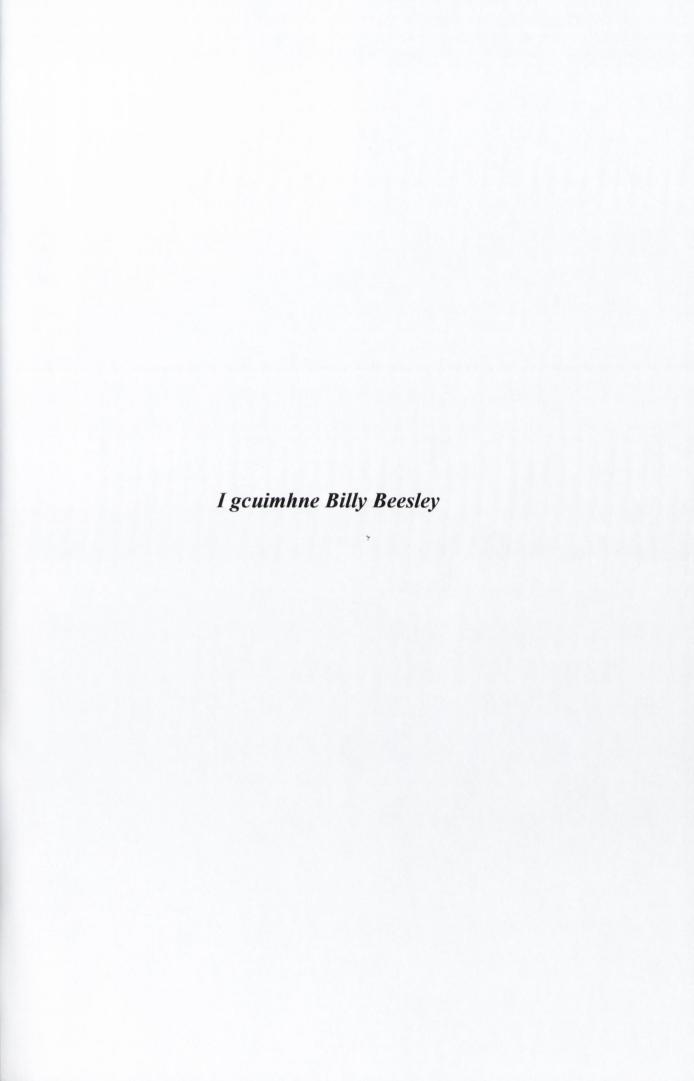
Department of Microbiology,

University of Dublin,

Trinity College

March 2002





DECLARATION

I hereby declare that this thesis has not been previously submitted for a Degree at this or any other University and that it represents my own unaided work except where duly acknowledged in the text.

I agree that this thesis may be lent or copied at the discretion of The Librarian, Trinity College, Dublin.

Sarah Beesley.
Sarah M. Beesley

Contents

	Page
Summary	I
Acknowledgements	III
List of Tables	V
List of Figures	VI
Publications and Conference Communications	X
Chapter 1: General Introduction	1
Chapter 2: Materials and Methods	44
Chapter 3: Diversity in the 3'-region of the cagA gene in paired H. pylori -	
genotypic and phenotypic analysis	66
Chapter 4: 'Microevolution' in paired isolates of <i>H. pylori</i> revealed by	
PCR-subtractive hybridisation	92
Chapter 5: Characterisation of genomic rearrangements at loci jhp0212/HP0227	
(hopM) and jph1261/HP1342 (hopN) in H. pylori strains and	
phenotypic analysis of paired isolates	119
Chapter 6: Genotypic analysis of familial H. pylori isolates and	
an appraisal of the occurrence of clonal variation in isolates	
from different individuals	150
Chapter 7: Analysis of the putative molecular mechanism for <i>H. pylori</i> LPS	
Lewis antigen phenotypic variation	164
Chapter 8: General discussion	183
Bibliography	192

Summary

Helicobacter pylori chronically colonises the human gastric mucosa. It is a major cause of chronic active gastritis and peptic ulcer disease and is associated with the development of gastric neoplasia. The population structure of *H. pylori* is characterised by a high level of genetic diversity. This diversity arises from increased mutation and frequent recombination in the genome of the organism. *H. pylori* is naturally competent for transformation which facilitates acquisition of DNA by horizontal transfer. The non-clonal nature of the *H. pylori* population structure means that distinct isolates are recovered from each colonised individual and, generally, no clonal relationship can be discerned between epidemiologically unrelated isolates. Furthermore, isolates taken from the gastric niche of a single individual have also been shown to exhibit genomic 'microevolution', manifest as small differences in DNA fingerprint profiles between strains when discriminatory fingerprinting methods were utilised. In the present studies chromosomal rearrangement between paired isolates of *H. pylori* was characterised with respect to defined loci and by whole-genome examination.

The *cagA* gene of *H. pylori* was shown to undergo *in vivo* alteration in the copy number of repeat sequences located at the 3'-end of the gene. Approximately 45% of the *H. pylori* paired isolates examined displayed this diversity and variation in repeat copy number could be correlated with the expression of CagA proteins of variable size. Among Irish *H. pylori* isolates, greater than 80% of strains have the same *cagA* subgenotype which consists of one full repeat of 102 bp followed by a 45-bp consensus region. Where differences were detected between paired isolates it was of the order of one repeat unit in the majority of instances. One set of paired isolates was found where a difference of six repeat units existed, one isolate had one repeat and seven were found in the second isolate. The presence of seven 3'-region repeats seemed to correlate with the expression of a CagA protein that was unstable. Evidence to suggest that CagA might be a phase-variable antigen was also obtained.

The PCR-subtractive hybridisation technique was used to analyse microevolution in the chromosomes of *H. pylori* paired isolates which had been shown by oligofingerprinting to exhibit fingerprint profile diversity. Four paired isolates sets were analysed directly by PCR-subtractive hybridisation. The genes in which rearrangements were apparent were housekeeping genes, genes that encoded proteins that interact with DNA, and outer membrane

protein-encoding genes. Rearrangements in outer membrane protein-encoding genes were detected with greatest prevalence. This suggests a possible mechanism by which *H. pylori* might undergo adaptive mutation. The DNA sequences involved in two rearrangement events were characterised in two sets of paired isolates. A putatively horizontally transferred genomic fragment was detected upstream of the *H. pylori hopN* gene in one set of paired strains while insertion of the insertion sequence element IS*Hp608* was detected downstream of the *hopN* gene in a second strain pair. Phenotypic differences were also detected between paired isolates in which genomic microevolution had been analysed. Variability in growth rates was apparent as was expression of a heterogeneous profile of outer membrane proteins.

H. pylori isolates from multiply colonised members of two families were obtained and subjected to oligofingerprinting analysis. In both families, evidence for intrafamilial transmission of a clone of *H. pylori* was detected. PCR-subtractive hybridisation was used to analyse microevolution in the genome of clonal isolates taken from different hosts. Rearrangement of genes that encoded proteins with housekeeping functions was found.

The putative molecular mechanism of *H. pylori* LPS Lewis antigen phenotypic variation was analysed in two sets of paired isolates. Between the isolates in each pair, a variable Lewis phenotype (Lewis^x and Lewis^y) was expressed as determined by Western immunolotting. Analysis of the *H. pylori fucT2* gene indicated that slipped-strand mispairing at a homopolymeric tract in the gene might be responsible for the phenotypic change observed by causing translational frameshifting. The data were somewhat anomalous, however, as genotypic information could not always be correlated with the observed phenotypes.

Acknowledgements

I thank Prof. Cyril Smyth for giving me the opportunity to carry out postgraduate research in his laboratory and for his constant support, guidance and attention to detail in the correction of the manuscript. Academic matters aside, Cyril's *joie de vivre* is truly an inspiration and helped create a wonderfully jolly atmosphere in which to work.

I thank also Prof. Colm Ó'Moráin of the Meath/Adelaide Hospital for providing biopsies and hugely generous financial assistance. I acknowledge the personal stipend and project funding provided by Enterprise Ireland and the generous support of the Society for General Microbiology for attendance at conferences.

Thank you to all members of the Path. Lab present and past, it has been a great pleasure to work with you. I especially thank Jane for her enthusiasm and interest and for being a constant source of encouragement throughout. Thanks to William and Ross for much guidance in the early days and Denise for providing clinical isolates and ongoing support. Thanks to Fred for many comic interludes and for being a knight in shining lab-coat when it came to crashed computers!! I know *H. pylori* will always have a place close to your heart, just as it does mine! Davida, flying the Gram-positive flag, thanks for your sunny greetings and encouragement in recent months. Thank you also to David Marshall for his research enthusiasm and many invaluable suggestions and for passing on much *H. pylori* know-how!

To Henry Kearns and everyone in the Prep. Room, how lucky we all are to work where such a resource is available. Thank you to all for your great efficiency in providing media and glassware of the highest quality and for your constant willingness to help.

Thank you to Prof. Charles Dorman for the invitation to attend the group meetings of his laboratory and to all for many useful contributions. To the many individuals in the Moyne Institute who were often approached for technical advice, thank you for your willingness to engage in the barter of reagents/enzymes etc. and for giving so generously of your time and expertise, it was of invaluable help! I especially thank Padraig Deighan for his assistance with outer membrane protein isolations.

I have been blessed with wonderful friends within and outside the Moyne Institute who have always been tremendously supportive and enthusiastic and who provided of late, much needed encouragement. I first thank Susan, Tara and Emmanuelle for being, quite simply, the best there is. Everybody should have friends like you. To housemates past and present, Catriona, Liz, Michelle, Susan, Ellie, Denise and Ali, thank you all for the day-to-day chats and laughs and for being such a great source of comfort and a marvellous collective support network. I would have been lost without you all. Thank you also Suzanne, Una, Siobhan, Louise and Sinead for being such caring friends.

To my parents Bill and Tess, every educational opportunity I could have wished for you afforded me without question or hesitation and I know your sacrifices were many - 'thank you' really seems far too small an offering. Bill, the work described in this thesis represents the 99% perspiration you believed was at the foundation of any achievement, and of which you so often reminded me by way of encouraging my efforts. Thank you for always taking such a keen interest in science and in my science especially, and for believing in me more than I believe in myself. For the same reasons, thank you Tess and also for being both mother and friend, I am extraordinarily lucky. To my brothers and minders, Art, Liam and Jim, I know you are all dying to read my thesis so you can squabble amongst yourselves over who gets to go first!!! Thanks for everything boys.

Finally, for stalwart support, unwavering optimism and above all for friendship, thank you my Colin.

List of Tables

	Inserted after
pecies members of the <i>Helicobacter</i> genus and natural	
	pp. 6
	7
	pp. 7
	pp. 14 pp. 37
rights of diversity in there, pytori genome.	pp. 37
etails of matched and familial <i>H. pylori</i> strains used	
roughout these studies.	pp. 48
equences of oligonucleotide primers used in PCR,	
APD-PCR and PCR-Subtractive Hybridisation	pp. 54
ings of DCD and dusts obtained using animous Cool and	
	pp. 74
	pp. 74
	pp. 75
	рр. 73
. pylori paired isolates.	pp. 77
	pp. 101
	nn 102
	pp. 102
	pp. 102
S	PP
CR-Subtractive Hybridisation data obtained from the	
nalysis of familial <i>H. pylori</i> isolates MI 431 and MI 577A	pp. 157
ummany of the phonetymes of the two U milesi naised	
	pp. 173
	dvantages and disadvantages associated with various nimal models of <i>Helicobacter</i> infection. Latative virulence factors of <i>Helicobacter pylori</i> . Latative virulence factors of <i>Helicobacter pylori</i> strains used definition of the studies. Latative virulence factors of <i>Helicobacter pylori</i> strains used definition of the capantal strains. Latative virulence factors of <i>Helicobacter pylori</i> strains used definition of the capantal strains of the capantal strains. Latative virulence factors of <i>Helicobacter pylori</i> paired isolates. Latative virulence factors of <i>Helicobacter pylori</i> paired isolates. Latative virulence factors of <i>Helicobacter pylori</i> paired isolate sets. Latative virulence factors of <i>Helicobacter pylori</i> paired isolate sets. Latative virulence factors of <i>Helicobacter pylori</i> paired isolate sets. Latative virulence factors of <i>Helicobacter pylori</i> paired isolate sets. Latative virulence factors of <i>Helicobacter pylori</i> paired isolate sets. Latative virulence factors of <i>Helicobacter pylori</i> paired isolate sets. Latative virulence factors of <i>Helicobacter pylori</i> paired isolate sets. Latative virulence factors of <i>Helicobacter pylori</i> paired isolate sets. Latative virulence factors of <i>Helicobacter pylori</i> set of the capantal se

List of Figures

	\mathbf{I}	nserted after
Chapter 1 Fig. 1.1.	Flow diagram of <i>H. pylori</i> -associated disease.	pp. 8
Chapter 2		
Fig. 2.1.	Flow diagram showing an overview of the PCR-	
Fig. 2.2.	subtractive hybridisation procedure. Schematic flow diagram outlining in greater detail the	pp. 54
	molecular events that take during PCR-subtractive hybridisation.	pp. 54
Chapter 3		
Fig. 3.1.	Southern analysis of <i>Hin</i> dIII-digested genomic DNA from isolate pairs MI 512/MI 541 and MI 561/MI 562 with <i>cagA</i> gene probes (clone 1 and pMC3)	pp. 72
Fig. 3.2.	Southern hybridisation of <i>Hin</i> dIII-restricted total genomic DNA of isolate pair MI 517/MI 566 probed with the clone 1 <i>cagA</i> probe and of <i>Hin</i> dIII, <i>Eco</i> RI and <i>Hae</i> III-restricted gneomic DNA of isolate pair MI 683A/MI 683C probed	
Fig. 3.3.	with the pMC3 <i>cagA</i> probe	pp. 72
F: 2.4	and MI 561/MI 562 probed with the pMC3 cagA gene probe	pp. 72
Fig. 3.4.	Oligofingerprint blot of <i>Hin</i> dIII-restricted total genomic DNA from paired <i>H. pylori</i> isolates.	pp. 73
Fig. 3.5.	Agarose gel electrophoresis of PCR amplification products of	pp. 75
Fig. 3.6.	the <i>cagA</i> gene from isolates MI 512 and MI 541	pp. 74
	isolates.	pp. 74
Fig. 3.7.	Pair-wise alignments of deduced amino acid sequences from the 3'-ends of the <i>cagA</i> genes of isolate pairs MI 512/MI 541,	
E:- 20	MI 517/MI 566 and MI 561/MI 562.	pp. 75
Fig. 3.8.	Schematic representation of the amino acid sequences deduced for the 3'-end of the <i>cagA</i> gene from isolate pairs MI 512/MI 54:	1
	MI 517/MI 566 and MI 561/MI 562.	
Fig. 3.9.	Agarose gel electrophoresis of single and multiple PCR products amplified from individual strains of <i>H. pylori</i> from paired isolate	3
E' 2.10	sets.	pp. 75
Fig. 3.10.	Agarose gel (1% w/v) electrophoresis showing PCR products generated from amplification of the <i>cagA</i> gene using primers Cag1 and Cag2 from <i>H. pylori</i> isolates MI 500A and MI 500C	
	and from single colonies of isolate MI 500C.	. pp. 76
Fig. 3.11.	Western immunoblots probed with anti-CagA polyclonal serum of <i>H. pylori</i> total cell lysates prepared in the presence of	FF.
	protease inhibitors.	pp. 76

Fig. 3.12.	Western immunoblot of total cell lysates of <i>H. pylori</i> strains MI 571, MI 506 and MI 512 prepared from broth culture samples		
Fig. 3.13.	at time points over the growth cycle	pp.	77
Fig. 3.14.	Schematic representation of the possible mechanism of	pp.	81
	rearrangement at the 3'-end of the <i>cagA</i> gene by homologous recombination between mis-aligned repeats during chromosome replication.	pp.	84
	replication.	pp.	01
Chapter 4			
Fig. 4.1.	Oligofingerprint and RAPD-PCR analysis of paired <i>H. pylori</i> isolates.	pp.	98
Fig. 4.2.	RFLP analysis of four paired isolates and two other		
	epidemiologically unrelated isolates using three different DIG-		00
Fig. 4.3.	labelled gene probes	pp.	99
116. 4.5.	Southern hybridisation of secondary PCR amplimers with <i>cagA</i>		
	and flaA gene probes.	pp.	100
Fig. 4.4.	Flow diagram depicting stages in the identification process of		
	DNA fragments that differ or partially differ between <i>H. pylori</i>		100
Fig. 4.5.	paired isolates using PCR-subtractive hybridisation	pp.	100
1.8	probes isolated by PCR-subtractive hybridisation.	pp.	101
Fig. 4.6.	Examples of <i>H. pylori</i> isolate sets from the Moyne Institute		
	collection identified by Southern hybridisation as having		100
Fig. 4.7.	undergone rearrangement at the indicated loci	pp.	102
1 ig. 4.7.	rearrangement by cross-over between directly repeated sequences		
	leading to a deletion.		111
Fig. 4.8.	Schematic representation of the possible cause of the hybridisation		
	pattern observed for clones 3 and 9 isolated by PCR-subtractive		112
	hybridisation of isolates MI 355 and MI 356.	pp.	113
Chapter 5			
Fig. 5.1.	Southern hybridisation analysis with full-length and cleaved		
	clone 2 probe from PCR-subtractive hybridisation of isolate		105
Fig. 5.2.	pair MI 458/MI 459 with <i>Hin</i> dIII digests of these strains Schematic diagram to account for both a putative mutational	pp.	125
11g. 3.2.	and recombinational mechanism of rearrangement between		
	isolate pair MI 458/MI 459 at the <i>hopM/hopN</i> alleles	pp.	126
Fig. 5.3.	PCR-amplification of the chromosomal loci corresponding to		
	alleles HP0227/jhp0212 (hopM) and HP1342/jhp1261 (hopN) in		
	H. pylori isolates where rearrangement at these ORFs was not apparent and analysis by Southern hybridisation of the PCR		
	products obtained.	pp.	126
Fig. 5.4.	Agarose gel electrophoresis of PCR products obtained from	11.	
	amplification of loci HP0227/jhp0212 (hopM) and		
	HP1342/jhp1261(<i>hopN</i>) from <i>H. pylori</i> strains MI 458 and		107
	MI 459 and MI 545A and MI 545C.	pp.	12/

Fig. 5.4.C.	Agarose gel electrophoresis of RAPD-PCR amplimers obtained using three combinations of primers for <i>H. pylori</i>
	strains MI 545A and MI 545Cpp. 128
Fig. 5.5.	Schematic figure showing the <i>hopM</i> - and <i>hopN</i> -encoding alleles,
	flanking genes and location of primer binding sites with agarose gel electrophoresis of PCR products generated with primers specific for
	clone 2 and for the ORFs flanking the HP0227/jhp0212 (hopM)
	and HP1342/jhp1261(<i>hopN</i>) loci. pp. 129
Fig. 5.6.	Schematic representation of the structure of the chromosomal
	regions surrounding the HP0227/jhp0212 (hopM) and
	HP1342/jhp1261 (<i>hopN</i>) loci in isolate pairs MI 458/MI 459 and MI 545A/MI 545C deduced from analysis of PCR products and
	DNA sequencing. pp. 130
Fig. 5.7.	Schematic representation of the putative mechanism of rearrangement
	that took place at the <i>hopN</i> locus between isolate MI 459 and a
Fig. 5.8.	horizontally transferred fragment of DNA
1 ig. 5.6.	region between HP1342/jhp1261 (hopN) and HP1343/jhp1262 where
	the ISHp608 element was found to be inserted in H. pylori strain
	MI 545A and multiple sequence alignment of this intergenic region
Fig. 5.9.	from strains J99, MI 545A and MI 545C
11g. 3.9.	as measured by the OD of the bacterial culture at 600 _{nm}
Fig. 5.10.	SDS-PAGE analysis of outer membrane proteins isolated
	from <i>H. pylori</i> strains MI 458, MI 459, MI 571 and MI 575 pp. 133
Chantan 6	
Chapter 6 Fig. 6.1.A	Oligofingerprint analysis of 12 <i>H. pylori</i> isolates from six members
	of Family 1
Fig. 6.1.B.	RAPD-PCR analysis of 12 <i>H. pylori</i> isolates obtained from six
Fig. 6.2	members of Family 1. pp. 154
Fig. 6.2.	Oligofingerprinting and RAPD-PCR analysis of seven <i>H. pylori</i> isolates from four family members of Family 2 pp. 154
Fig. 6.3.	Similarity coefficient matrices of isolates from Family 1 and
	Family 2
Fig. 6.4.	Southern blot analysis of the clones isolated by PCR-subtractive
	hybridisation of clonally related familial <i>H. pylori</i> isolates MI 431/MI 5877A and MI 218A/MI 221C
Fig. 6.5.	Southern blot analysis of nine <i>H. pylori</i> isolates (representing
	each family member) from Family 1 with a cagA gene probe pp. 157
Cl. 7	
Chapter 7 Fig. 7.1.	Western immunoblot analysis of <i>H. pylori</i> isolates with anti-Le ^x
116. 7.11	and anti-Le ^v monoclonal antibodies
Fig. 7.2.	Analysis of <i>H. pylori</i> isolate pairs MI 561/MI 562 and MI
F: 7.2	222A/MI 222C by oligofingerprinting. pp. 171
Fig. 7.3. Fig. 7.4.A.	Schematic structures of the Lewis X and Lewis Y antigens pp. 172 Alignment of the DNA sequences determined for a region of the
1 ig. 7.4.A.	fucT2 gene from H. pylori isolates MI 222A and MI 222C pp. 172

Fig. 7.4.B.	Amino acid sequence predicted from the translation of the
	fucT2 DNA sequence determined for isolate MI 222A pp. 172
Fig. 7.5.	(A) Alignment of the partial <i>fucT2</i> gene sequences determined
	for isolates MI 561 and MI 562 and (B) alignment of a short region
	of the fucT2 gene of isolate MI 561 determined from two rounds
	of sequencing. pp. 172
Fig. 7.5.C.	Alignment of the DNA sequences obtained for the <i>fucT2</i> gene from
	isolate MI 562 determined at three separate sequencing rounds after
	PCR on three distinct occasions
Chapter 8	
Fig. 8.1.	Summary of findings from studies of microevolution in <i>H. pylori</i>
	paired isolates showing the sources of mutation, the barriers or
	stimulators thereof, and the various consequences for the <i>H. pylori</i>
	genome arising from <i>in vivo</i> diversity

Publications

- D.G. Marshall, W.G. Dundon, **S.M. Beesley** and C.J. Smyth. (1998). *Helicobacter pylori* a conundrum of genetic diversity. *Microbiology* **144**, 2925-2939.
- **S.M.** Beesley, D.G. Marshall, C.A. Ó'Moráin and C.J. Smyth. Characterisation of microevolution in the genome of *Helicobacter pylori* using PCR-subtractive hybridisation. Manuscript in preparation.

Conference Communications

- **S.M.** Beesley, C.A. Ó'Moráin and C.J. Smyth. Phenotypic analysis and highly discriminatory genotypic fingerprinting of *Helicobacter pylori* isolates from Irish families. In *Digestive Disease Week*, Orlando FL, U.S.A. May 1999. Abstract #1404.
- **S.M.** Beesley, C.A. Ó'Moráin and C.J. Smyth. Analysis of 'Microevolution' in *Helicobacter pylori* using PCR-subtractive hybridisation. In 100th General Meeting of the American Society for Microbiology, Los Angeles CA, U.S.A. May 2000. Abstract D-256.
- **S.M. Beesley**, C.A. Ó'Moráin and C.J. Smyth. 'Microevolution' in *Helicobacter pylori* investigated using PCR-subtractive hybridisation. In *Irish Society of Gastroenterology and Midland Gastroenterology Society Millennium Joint Meeting*, Killarney, June 2000. *Irish Journal of Medical Science* **169**, **Suppl. 3**, 20 (abstract).
- **S.M.** Beesley, C.A. Ó'Moráin and C.J. Smyth. 'Microevolution' in *Helicobacter pylori* investigated using PCR-subtractive hybridisation. In *XIIIth International Workshop on Gastroduodenal Pathology and* Helicobacter pylori, Rome, Italy, October 2000. *Gut* 47, **Suppl. 1**, 1/10 (abstract).

Chapter 1 General Introduction

1.1.	Discovery and i	solation of Helicobacter pylori	4
1.2.	Taxonomy		4
1.3.	Bacteriology		4
1.4.	Epidemiology.		5
1.5.	Other members	of the Helicobacter genus.	6
			6
		ctrum and clinical manifestations of H. pylori	
		mans.	7
		3 gastritis	8
	* 1	ulcer disease	9
		. Duodenal ulcer.	9
		2. Gastric ulcer.	9
		e atrophy and intestinal metaplasia	10
		c cancer.	10
		. Adenocarcinoma.	11
		2. MALT lymphoma.	11
		ori infection and non-ulcer dyspepsia.	11
1.8		sation can be advantageous to the host.	12
		ainst H. pylori infection.	12
1.7.		ement of PUD and drug resistance.	12
		I model experiments and human trials.	13
		icating issues in the development of an H. pylori vaccine	14
1 10		determinants of H. pylori.	14
1.10		ence factors present in all isolates of H. pylori.	14
	1.10.1.1	Flagella.	14
	1.10.1.1.		15
	1.10.1.2.	H. pylori urease.	16
		H. pylori neutrophil-activating protein (HP-NAP)	17
	1.10.1.4.	γ-Glutamyltransferase.	18
	1.10.1.5.	Iron-acquisition proteins.	
	1.10.1.6.	Adhesion.	19
	1.10.1.7.	Heat-shock protein homologues.	20
		ence factors present in a subset of H. pylori isolates	20
		Vacuolating cytotoxin (VacA).	20
		Cytotoxin-associated antigen (CagA).	21
	1.10.2.3.	Bab adhesin.	22
		ging H. pylori virulence factors.	22
	1.10.3.1.	Superoxide dismutase and catalase.	22
	1.10.3.2.	H. pylori pore-forming cytolysin – TlyA	23
	1.10.3.3.	iceA locus.	23
1.11		sity in <i>H. pylori</i>	25
		nosomal diversity revealed by molecular typing	
		es – historical perspective	25
		tic diversity in virulence-associated loci – the pre-genome era	25
	1.11.2.1.	The vacA gene.	26
	1.11.2.2.	Diversity in CagA expression, the cagA gene, and the cagPAI	26
	1.11.2.3.	The <i>fla</i> and <i>ure</i> genes.	27
	1.11.2.4.	The bab genes.	27
	1.11.3. Gene	etic diversity in H. pylori – the genomic/proteomic era	28
	1.11.3.1.	Whole-genome microarray analysis.	28
	1.11.3.2.	Proteomics.	29

1.12.	Origins of diversity in the H. pylori genome.	30
	1.12.1. Nucleotide mutation and recombination	30
	1.12.2. The population structure of H. pylori	30
	1.12.3. Quasi-species	31
1.13.	Mechanisms of prokaryotic chromosomal rearrangement.	32
	1.13.1. Tandem duplications, deletions and inversions	32
	1.13.2. Horizontal gene transfer.	33
	1.13.3. Local sequence change.	33
1.14.	Rearrangements and mechanisms of diversity generation in	
	the H. pylori genome inferred from genome sequence analysis	34
	1.14.1. Inversions, transpositions and reciprocal exchange	34
	1.14.2. Gene transfer by 'insertion with long target duplication'	34
	1.14.3. Acquisition of exogeneous DNA – horizontal gene transfer	35
	1.14.4. Insertion sequence (IS) elements.	36
	1.14.5. Phase-variable genes	36
1.15.	Natural transformation and competence in <i>H. pylori</i> as contributors	
	to genomic diversity.	37
	1.15.1. Natural transformability of H. pylori – in vitro and in vivo evidence	37
	1.15.2. The competence machinery	38
	1.15.3. Barriers to transformation and the net contribution of competence	
	to diversity	38
1.16.	Type II restriction-modification systems in <i>H. pylori</i> as barriers to	
	strain-specific transformation.	40
1.17.	Microevolution of the H. pylori chromosome in vivo.	41
1 10	Aims	12

1.1. Discovery and isolation of Helicobacter pylori.

The presence of spiral microorganisms in the stomachs of humans and other animals had been described several times by histologists in the late 19th and early 20th centuries (Salomon, 1896; Krienitz, 1906; Doenges, 1938). Despite this, no role for an infectious agent in the pathology of human gastric disease was envisaged by the medical profession, largely arising from the failure of efforts to culture these organisms from the stomach and the belief that the observed organisms could have been *post mortem* contaminants (Palmer, 1954). However, interest in these apparently unculturable and presumed insignificant organisms was rekindled by Warren and Marshall (1983), who observed a correlation between the presence of these spirochaetes and polymorphonuclear leucocytes. This later led to the culture of a *Campylobacter*-like organism (CLO) under microaerobic conditions (Marshall and Warren, 1984). Soon thereafter a firm link between the presence of these CLOs and gastroduodenal disease (active type B gastritis and duodenal ulceration) in humans was established by several investigators, helped in part by the fulfilment of Koch's postulates for this organism by deliberate ingestion of a suspension of the bacterium (Marshall, 1985).

1.2. Taxonomy.

The Gram-negative, microaerophilic, curved bacillus was initially called *Campylobacter pyloridis* (Marshall and Warren, 1984) but underwent several name changes before being classified into a new genus, termed *Helicobacter*, because of its helical appearance *in vivo* and a propensity to appear rod-shaped (bakter meaning staff) *in vitro* (Goodwin, 1989). Its dissimilarity to the *Campylobacter* type species *C. fetus* in a number of respects, most notably the presence of multiple sheathed flagella, prompted the reclassification of what had by then come to be called *C. pylori* into this new genus (McNulty, 1999). Collectively, members of the genus *Helicobacter* can be distinguished from other Gram-negative curved rods on the basis of analysis of enzymatic activities, fatty acid profiles, nucleic acid hybridisation profiles, growth characteristics and 16S rRNA gene sequences (Versalovic and Fox, 2001).

1.3. Bacteriology.

Typically, *H. pylori* is between 2.5 and 5 μm in length and 0.5 to 1.0 μm in diameter. Although classically described as a spiral organism, this cellular morphology is not generally observed during *in vitro* culture on solid medium. Furthermore, in older cultures, *H. pylori* can be seen to undergo a morphological change from a bacillary to a coccoid form which is

associated with loss of culturability, although the bacterium remains metabolically active (Dunn *et al.*, 1997). The organism is flagellated, normally with four to six flagella being present. The flagella are sheathed and some have terminal bulbs whose function is unknown. *H. pylori* produces oxidase, catalase and urease but does not ferment or oxidise carbohydrates and thus is inactive in most conventional laboratory biochemical tests (Owen, 1998). The genome of the bacterium is a single circular molecule of variable size and is comparatively small at approximately 1.6 Mb on average. The complete genome sequences of two *H. pylori* strains (26695 and J99) have been determined (Tomb *et al.*, 1997; Alm *et al.*, 1999). The G+C% composition of the genome ranges from 34–39 mol% with an average of 35.2 mol%. Plasmids are found in approximately 40% of isolates, although the type strain NCTC11637 is plasmid-free. *H. pylori* plasmids are not known to encode any recognised virulence factors (Owen, 1998).

1.4. Epidemiology.

The prevalence of *H. pylori* infection in populations varies considerably between countries. Globally, it is estimated that 50% of the world's population is infected with H. pylori. In developing countries, H. pylori can colonise up to 80% of the population. The prevalence of infection is lower in developed countries with 25-50% of the population reported to be colonised (Dunn et al., 1997; Mitchell, 1999). The differences in the prevalence of infection between developed and developing countries relate to the rate of acquisition of H. pylori during childhood. In developed countries, in children younger than 10 years, the prevalence of infection is 0-5% by comparison with 13-60% for children in developing countries. Over this age, the increase in prevalence of infection is of the order of 0.5-2.0% per annum. This increase in the prevalence of infection with age had initially been interpreted as acquisition of H. pylori over time. Alternatively, the increasing prevalence of H. pylori occurring from younger to older subjects might reflect the passage of distinct cohorts through the population, i.e., the aging of the colonised population (Mitchell, 1999). If infection had been acquired by older age groups in childhood, an apparent increase in prevalence with age might be observed in developed countries, in particular where gradual improvements in medical care or living conditions are decreasing levels of H. pylori infection associated with younger age groups. No substantial environmental reservoir of H. pylori has been identified, with the source of infection and routes of natural transmission largely unaccounted for (Dunn et al., 1997). Oral carriage of H. pylori has been reported (Namavar et al., 1995) and animals are suspected to act

as reservoirs. Both faecal-oral and oral-oral routes of transmission have been proposed (Dunn *et al.*, 1997).

1.5. Other members of the Helicobacter genus.

It is now 13 years since the genus *Helicobacter* was first proposed, formally distinguishing organisms classified in this taxon from other Gram-negative curved rods, e.g., the *Campylobacter* and *Arcobacter* genera (Goodwin *et al.*, 1989). At present the genus contains 20 formally named species (Table 1.1.) which have been isolated from the gastrointestinal tracts of mammals and birds (Versalovic and Fox, 2001). *H. pylori* remains the best known and most intensively studied species because of its importance as a pathogen of humans. Five other species have been isolated from human hosts. *H. cinaedi*, *H. fennelliae*, *H. pullorum* and *H. canis* can colonise the intestine and are so-called enteric *Helicobacters*, while *H. westmeadii* is a recently isolated helicobacter from blood samples of AIDS patients. The species in the enteric group are also found in non-human hosts but where this is the case strains have distinct host-related genotypes (Versalovic and Fox, 2001).

1.6. Animal models of H. pylori infection.

In elucidating the pathogenic mechanisms of an infectious agent, the availability of an animal model of infection that mimics human disease is an invaluable tool. Not only can information be gleaned about the importance and roles of pathogenicity determinants produced by the organism and the mechanisms of host-bacterium interactions, but the screening and efficacy of potential vaccines and therapeutic agents can be undertaken. Experimental infection with Helicobacter spp. in nude mice, primates, gnotobiotic pigs, cats, dogs, ferrets and gerbils has been achieved, but realistically smaller hosts are more convenient for handling and care purposes and would be desirable for widespread use. However, of the larger animals, the germfree piglet infected with human-derived H. pylori has proven to be a useful model host for colonisation and vaccine studies (Akopyants et al., 1995; Lee, 1998; Eaton, 1999). Natural infection of ferrets with H. mustelae also has utility as a model of human infection (Fox et al., 1990). As with numerous other bacterial infections, a convenient mouse model of H. pylori infection would be preferable. H. pylori can be adapted to colonise the murine stomach (Marchetti et al., 1995) and H. felis, which naturally colonises the stomach of cats has been shown to colonise the gastric mucosa of mice (Eaton, 1999). As a model of H. pylori-related disease, the mouse has been used to assess treatment regimens, to investigate bacterial colonisation factors and to examine the interaction of *H. pylori* with normal flora (Eaton, 1999).

Table 1.1. Species members of the Helicobacter genus and natural animal hosts^a

Recognised species of Helicobacter genus ^b	Natural host
H. acinonyx	Cheetah
H. bilis	Mouse
H. canadensis	Human
H. canis	Dog
H. cholecystus	Hamster
H. felis	Cat
H. fennelliae	Human
H. hepaticus	Mouse
H. muridarum	Rodents
H. mustelae	Ferret
H. nemestrinae	Monkey
H. pametensis	Tern
H. pullorum	Chicken
H. pylori	Human
H. rodentium	Rat
H. salmonis	Cat
H. suncus	Shrew
H. trogontum	Rat
H. typhlonicus	Mouse
H. westmeadii	Human

^aFrom Versalovic and Fox (2001).

^bNote, *H. heilmannii* (formerly, '*Gastrospirillum hominis*') observed in human gastric biopsies is not formally recognised since significant nucleic acid variation was found among various human isolates (Versalovic and Fox, 2001). Such isolates are generally referred to in the literature as '*H. heilmannii*'-like organisms.

However, where H. pylori is concerned, no animal model exists that exactly mimics the diseases found in the human stomach, although colonisation patterns of Helicobacter spp. in experimental infections do approximate those seen in humans. Where gastritis can be induced in animals, the inflammatory responses elicited by the infection resemble those that occur in children, without the neutrophil infiltration seen in infected adults. Moreover, models of ulcerogenesis and carcinoma that do exist cannot be reproduced easily (Lee, 1998). Animal models of Helicobacter gastritis were reviewed by Eaton (1999) who detailed the difficulties associated with the use of animal experimentation for the study of H. pylori-related disease. For example, within the various animal models that have been developed, there is variability in responsiveness to infection with Helicobacter spp. Although this presents problems when it comes to the interpretation of the various studies, such variability emphasises the critical role of the host in determining the outcome of H. pylori-related disease and additionally, as pointed out by Eaton (1999), hinders the evaluation of bacterial virulence factors in vivo. At present, experimental infection of the Mongolian gerbil with H. pylori is the most promising animal model that exists with severe active and chronic gastritis observed, consistent development of gastric ulcers, and critically adenocarcinoma induction with infection alone (Lee, 1999). The standardisation of animal models, especially with respect to the interpretation of histological samples, would greatly advance the usefulness of animal studies in the investigation of many aspects of H. pylori biology, including host responses to infection. The advantages and disadvantages of various animal models of *Helicobacter* infection are summarised in Table 1.2.

1.7. The disease spectrum and clinical manifestations of H. pylori infection of humans.

The involvement of *H. pylori* in gastroduodenal diseases that range from being relatively innocuous to having the gravest of severity is one of the most fascinating aspects of the biology of this organism. Although most individuals colonised by *H. pylori* develop chronic gastric inflammation, it is a condition that generally has few or no symptomatic associations (Kuipers and Blaser, 1998). The progression of *H. pylori*-associated disease from this asymptomatic stage to more serious disease forms is an indolent process, consistent with the persistence of the organism in the stomach environment for decades or longer. Although colonisation is prolonged, to the extent that *H. pylori* colonisation has been termed a 'slow' bacterial infection (Blaser, 1993), immune responses to *H. pylori* are detectable within weeks of its acquisition (Blaser and Parsonnet, 1994). Indeed, Warren's observations that spiral bacteria chiefly overlaid inflamed tissue were responsible for the eventual culture of *H. pylori* from the gastric environment using *Campylobacter*-specific methods and provided the first clue that *H. pylori*

Table 1.2. Advantages and disadvantages associated with various animal models of Helicobacter infection^a

Animal	Colonised by	Advantages	Disadvantages
Primates	H. pylori	Closest animal species to human, endoscopy possible,	Expensive, colonised by endemic strains, presence
		gastric physiology similar to human	of H. heilmannii-like bacteria
Gnotobiotic	H. pylori	Colonisation pattern similar to human, gastric	Chronic gastritis only, expensive short-term
piglets		physiology similar to human, ulcers observed	experiments only
Ferrets	H. mustelae	Natural infection, useful for vaccine studies, gastric	Pattern of colonisation varies from human,
		physiology similar to human	predominantly chronic gastritis only
Cats and dogs	H. pylori	Gastric physiology similar to human	Gnotobiotic and specific-pathogen-free animals
	H. felis		expensive, H. heilmannii-like bacteria present in
			normal cats
Mice	H. pylori	Economical, good for testing vaccines/antimicrobials,	Does not mimic human pathology
	H. felis	good colonising strains of H. pylori available (e.g.,	
	H. heilmannii ^b	Sydney strain)	
		Colonisation by isogenic mutants easily tested,	
		immunological reagents available, transgenic/knockout	
		strains available	
Gerbils	H. pylori	Chronic/active gastritis, gastric ulcers,	Lack of immunological reagents, no
		adenocarcinoma induced with infection alone	transgenic/knockout strains

^aAdapted from Lee (1999).

^bSee Table 1.1.

might play an aetiological role in the pathogenesis of gastritis and more severe forms of gastroduodenal pathology (Warren and Marshall, 1983). A flow-diagram that summarises current views on the progression of *H. pylori*-associated disease from the initial infection of the normal gastric mucosa through gastritis to peptic ulcer disease and gastric cancer is shown in Fig. 1.1.

1.7.1. Type B gastritis.

Type B gastritis is the most commonly occurring form of the group of disorders termed gastritis. It is present in individuals world-wide and increases in prevalence with age. Type B gastritis refers to inflammation of the gastric mucosa in the antrum of the stomach, although the gastric fundus can also be affected (Blaser, 1990). In the aftermath of the discovery of *H. pylori* by Warren and Marshall (1983) and the observation that the organism was chiefly associated with areas of inflamed tissue, further studies demonstrated that patients who did not have gastritis were not colonised with *H. pylori* (Marshall and Warren, 1984) and that *H. pylori* infection could be detected in up to 100% of instances of type B gastritis (Blaser, 1990). Other lines of evidence, including the findings that the clearing of infection by antibiotic intervention also clears the gastritis, that ingestion voluntarily of *H. pylori* resulted in histological evidence of type B gastritis and that experimental animal challenge simulated human infection and resulted in gastritis, further strengthened the case for the direct involvement of *H. pylori* in the pathogenesis of this disorder (Blaser, 1990).

Type B gastritis, also referred to as chronic active gastritis, is characterised histologically by the degeneration of mucosal epithelial cells, by mucosal infiltration by chronic inflammatory cells (lymphocytes, plasma cells and eosinophils) and in adults, by an 'active' infiltrate of neutrophils (Bodger and Crabtree, 1998; Ernst and Gold, 2000). The induction of inflammation by *H. pylori* is thought to happen mechanistically in two ways. Firstly, by direct interaction with surface epithelial cells the organism might directly damage cells or cause the liberation of proinflammatory chemokines derived from the epithelium. Secondly, host non-specific and specific immunity might be stimulated in the underlying mucosa by bacterium-derived molecules. Cytokine release would then ensue resulting in an inflammatory infiltrate (Bodger and Crabtree, 1998).

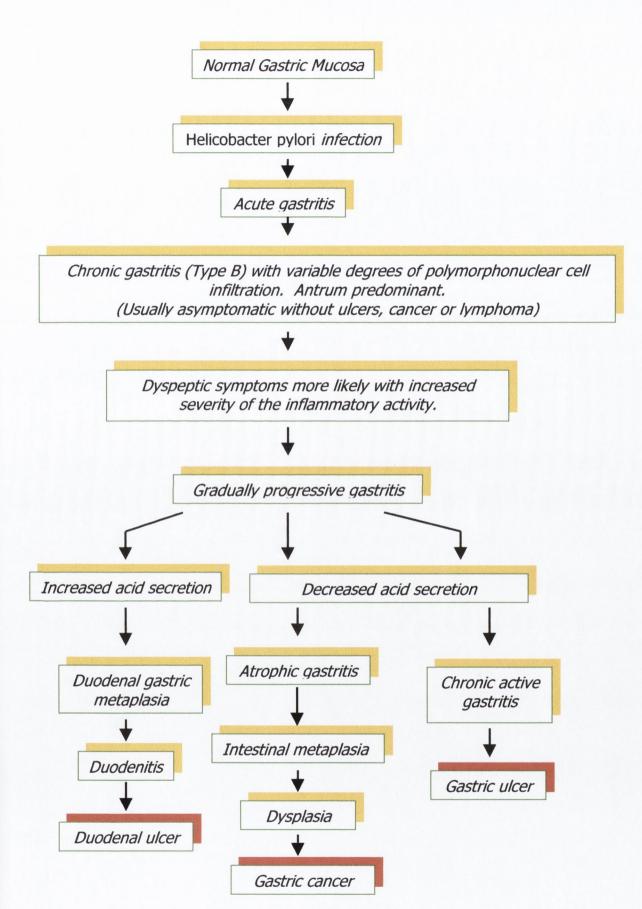


Fig. 1.1. Flow diagram of H. pylori-associated disease.

1.7.2. Peptic ulcer disease.

With the observation that chronic type B gastritis was present in all individuals who also had gastroduodenal ulceration came the realisation that infection with *H. pylori* is also a significant cause of peptic ulcer disease. Between 90–95% of duodenal ulcers and 70–75% of gastric ulcers are attributable to infection with *H. pylori* (Ernst and Gold, 2000), although in populations where the prevalence of *H. pylori* is low this may be an overestimation (Williams and Pounder, 1999). It is estimated that approximately 10% of individuals who are colonised with *H. pylori* will develop peptic ulcer disease (Kuipers and Blaser, 1998; Williams and Pounder, 1999). Eradication of *H. pylori* infection with antimicrobials markedly reduces the recurrence rate of duodenal and gastric ulcers and provides compelling evidence that *H. pylori* plays a causative role in the development of ulceration. Where there is no microbial involvement in ulcerogenesis, non-steroidal anti-inflammatory drug therapy, aspirin intake and Crohn's disease have been implicated in the development of ulcer lesions (Kuipers and Blaser, 1998).

1.7.2.1. Duodenal ulcer.

A duodenal ulcer (DU) is a breach in the duodenal epithelium associated with acute and chronic inflammation. DUs occur more commonly than gastric ulcers (GU) in most countries, although instances where the reverse is found have also been reported (Calam, 1998). If *H. pylori* causes infection of gastric-type epithelium, why is it that most cases of peptic ulcer disease occur in the duodenal region? It is believed that *H. pylori* overlies gastric-type epithelium normally present in the stomach or metaplastic tissue in the duodenum (Blaser, 1990). Gastric metaplasia into the duodenum is found in all patients with DU and is believed to occur in response to injury of pluripotential epithelial stem cells that differentiate into gastric-type epithelium instead of intestinal absorptive-type epithelium (Blaser, 1990). The development of high acid levels has been shown to induce gastric metaplasia and, hence, the pathogenesis of DU disease involves both *H. pylori* infection and gastric acid. Metaplasia of gastric epithelium in the duodenum because of high levels of acid secretion, can become colonised by *H. pylori* in individuals whose stomachs are already colonised by the bacterium and here cause active duodenitis which leads ultimately to ulceration (Blaser, 1990; Calam, 1996).

1.7.2.2. Gastric ulcer.

Similarly to DU disease, GUs exist as lesions in the gastric epithelium associated with inflamed tissue. As alluded to above, gastric ulcers occur less commonly than duodenal ulcers in most

populations and are rarely found in the paediatric population (Ernst and Gold, 2000). Indeed, epidemiological information seems to indicate that GUs occur more commonly among patients who become colonised at an early age while DU disease appears to be associated with infections that take place later in life (Kuipers and Blaser, 1998). Again, the recurrence rate of GU approaches zero if the *H. pylori* infection has been eradicated with antibacterials.

1.7.3. Gastric atrophy and intestinal metaplasia.

The persistent nature of the infection caused by H. pylori, in which a host inflammatory response is induced from the outset, contributes to the complex pathogenesis of this organism (Blaser and Parsonnet, 1994). The chronic, active type B gastritis that predominates in the H. pylori-infected population can, over three or four decades, progress towards increasingly degenerative changes and start to affect gastric glandular structures leading to atrophic gastritis (wasting or destruction of gastric glands). Atrophy of this nature in the stomach is generally followed by the replacement of glandular tissue with intestinal metaplastic cells. These are significant stages in the development of H. pylori-associated disease as they are considered to be pre-malignant lesions in the development of adenocarcinoma of the stomach. Interestingly, however, attempts to locate H. pylori in a cancerous stomach are usually fruitless owing to the replacement of gastric-type epithelium preferentially inhabited by H. pylori with an atrophic mucosa and intestinal metaplasia (Blaser and Parsonnet, 1994; Williams and Pounder, 1999). Atrophy and intestinal metaplasia are estimated to occur eventually in 50% of individuals in the H. pylori-colonised population and the increased risk for the development of gastric cancer associated with these pathologies is estimated to be 5- to 90-fold. Furthermore, these conditions are encountered rarely in the absence of *H. pylori* infection (Kuipers and Blaser, 1998).

1.7.4. Gastric cancer.

In 1994 the International Agency for Research on Cancer (IARC), a department of the World Health Organisation, classified *H. pylori* as a Grade I carcinogen. This classification copper fastened the evidence for the association of *H. pylori* infection with gastric cancer and that the organism can be considered a definite carcinogen. The lifetime risk for the development of gastric cancer among *H. pylori*-positive patients is estimated at 1% (Kuipers and Blaser, 1998), although the prevalence of gastric cancer is more common in nations with emerging economies or among individuals of disadvantaged socioeconomic status in industrialised countries (Ernst and Gold, 2000).

1.7.4.1. Adenocarcinoma.

Gastric adenocarcinoma is the fourteenth most common cause of death worldwide and is predicted by the year 2010 to become the eighth leading cause of death in the world. Of deaths arising from cancer, it is the second commonest fatal malignancy (Williams and Pounder, 1999). *H. pylori* infection is accepted as the most important risk factor for gastric cancer based on several lines of evidence, including studies that associated epidemiological features of *H. pylori* infection with gastric cancer and studies of *H. pylori* infection in patients with cancer (Ernst and Gold, 2000). However, the mechanisms by which the bacterial infection leads to malignancy have not been definitively elucidated (El-Omar and Malfertheiner, 2001).

It has been proposed that persistent oxidative stress attributable to the inflammatory response that occurs during gastritis is responsible for the malignant transformation (Correa, 1992). Oxidative DNA damage that escapes repair within the host cell is considered a highly likely mechanism of inducing carcinoma. Moreover, oxidative stress is involved in the regulation of expression of several genes that are responsible for epithelial cell turnover (Ernst and Gold, 2000). Much interest of late has focused on the role of apoptosis and cell proliferation in carcinogenesis. Apoptosis is thought likely to play a key role in malignancy by increasing cell proliferation or by resulting in gastric atrophy, although apoptosis itself is not a predictor of atrophy. Candidate apoptosis-inducing *H. pylori* factors include the vacuolating cytotoxin and lipopolysaccharide (El-Omar and Malfertheiner, 2001).

1.7.4.2. Mucosa-associated lymphoid tissue (MALT) lymphoma.

The second type of malignancy that can arise from *H. pylori* infection of the stomach is MALT lymphoma. The presence of lymphocytes and lymphoid follicles in the gastric epithelium accompanies *H. pylori* colonisation of the stomach. A causal relationship was speculated, therefore, between *H. pylori* infection and the occurrence of gastric lymphoma. The majority of patients with MALT lymphomas were infected with *H. pylori* and the presence of the bacterium was found to present an increased risk for the development of this type of cancer. Significantly, eradication of the organism resulted in complete remission of the lymphomas in the majority of treated individuals (Kuipers and Blaser, 1998).

1.7.5. H. pylori infection and non-ulcer dyspepsia.

Non-ulcer dyspepsia (NUD) is defined as chronic or recurrent pain or discomfort centred in the upper abdomen in the absence of peptic ulceration (Kalantar et al., 1999). Reports about the

relationship between *H. pylori* infection and dyspepsia are inconsistent. Although *H. pylori* infection is present in 30–60% of individuals with NUD in Western countries, attempts to define a causal role for the organism in eliciting the symptoms of this disorder have not been conclusive. Indeed, the treatment of *H. pylori* infection in cases of NUD is controversial as other causes of dyspepsia in uninfected individuals could also be the causes in many infected individuals (Goodwin, 1997). As noted by Kalantar *et al.* (1999), one of the major questions, therefore, is whether the overlap of *H. pylori* and dyspepsia is causal or coincidental as both conditions are common and could overlap by chance. Nonetheless, management of the disorder in the clinical setting involves the eradication of *H. pylori* infection which has been shown to have beneficial effects on dyspeptic symptoms compared with a placebo (Delaney and O'Morain, 2001).

1.8. H. pylori colonisation can be advantageous to the host.

Although studies on *H. pylori* have been primarily concerned with its molecular bacteriology and clinical investigations to decipher its disease mechanisms, the literature is littered with reports that indicate that *H. pylori* colonisation may be beneficial and that attempts to eradicate it could, in certain instances, prove harmful. It has been proposed that gastric resistance to colonisation by other pathogens is imparted by the mere presence of *H. pylori* in the gastric niche, akin to the colonisation resistance gained from our normal flora at other body sites (Falk *et al.*, 2000). Furthermore, although not associated with *H. pylori* infection, it has been reported from epidemiological studies that the incidence of gastro-oesophageal reflux disease (GERD) increases where *H. pylori* has been eradicated (Labenz *et al.*, 1997; Williams and Pounder, 1999). GERD is a risk factor for oesophageal metaplasia and adenocarcinoma. This correlation presents perhaps the most pressing concern for clinicians. As stated by Williams and Pounder (1999), a dilemma has arisen – 'Should we eradicate *H. pylori* to prevent cancer of the distal stomach, or should we leave it and hence avoid cancer of the proximal stomach or distal oesophagus?'. As yet, the possible mutual benefit of *H. pylori* colonisation of the human stomach is not well defined.

1.9. Vaccination against H. pylori infection.

1.9.1. Management of PUD and drug resistance.

Management of peptic ulcer disease in the clinical setting involves the prescription of various antibiotic triple-therapy regimens with a view to complete eradication of the infection. Although this has proven to be a highly successful therapeutic option, problems, not altogether

unforeseen, have arisen in association with the treatment of patient populations with antimicrobials. Drug-resistant strains of *H. pylori*, e.g., metronidazole-resistant and clarithromycin-resistant isolates, have emerged and are now encountered more commonly by practitioners. Incomplete eradication of the colonising bacteria from the stomach because of drug resistance results in time in recrudescence of the infection and the requirement for the prescription of an alternative multidrug regimen. Such treatment failures are not ideal when considering the desire for patient compliance and expense. Thus, vaccination is viewed as a future alternative to antibiotic therapy by having the advantage of protecting against reinfection while clearing the organism. The interest in developing a therapeutic vaccine against *H. pylori* is set against a backdrop of predictions that in the next 20 years gastric cancer will have become the 8th leading cause of death of any type worldwide.

1.9.2. Animal model experiments and human trials.

In the case of *H. pylori*, two animal models have been developed (see Section 1.6) which mimic human infection sufficiently well to allow useful conclusions to be made about the likely efficacy of prospective prophylactic or therapeutic immunisation. Indeed, using both the ferret and mouse models, mucosal administration of inactivated whole-cell preparations of *H. pylori* accompanied by a mucosal adjuvant conferred protection against challenge with wild-type *H. pylori* or *H. felis* organisms (Chen *et al.*, 1992; Cuenca *et al.*, 1996). Therapeutic immunisation has also been investigated using the mouse model of infection and has been shown to be efficacious in maintaining the clearance of infection (Doidge *et al.*, 1994; Ghiara *et al.*, 1997).

The development of vaccines against *H. pylori* by commercial interests has also taken place. One such trial involved the oral immunisation of *H. pylori*-infected volunteers with *Helicobacter* urease plus *E. coli* heat-labile toxin as a mucosal adjuvant. The results were not promising, however, as the adjuvant induced diarrhoea in the volunteers (Michetti *et al.*, 1999). A recent report has outlined a trial in human volunteers for the first time of a *Helicobacter* whole-cell vaccine (Kotloff *et al.*, 2001). The formalin-inactivated oral whole-cell vaccine was administered with or without mutant *E. coli* heat-labile toxin as the mucosal adjuvant. Although preliminary, the data indicate that a mucosal (IgA) and systemic immune response to *H. pylori* antigens can be stimulated using an inactivated whole-cell vaccine. Mucosal immunity is known, however, not to be protective against *H. pylori* infection. The possibility of using live *Salmonella* vector vaccines has also been investigated in the mouse model and in human volunteers using attenuated *S. typhimurium* and *S. typhi* strains, respectively, expressing *H.*

pylori urease subunits. However, the mucosal immune response to the *Helicobacter* urease in these human trials was poor (Sutton, 2001).

1.9.3. Complicating issues in the development of an H. pylori vaccine.

Studies have shown somewhat surprisingly that effective immunisation against *H. pylori* is independent of an antibody response and dependent on CD4⁺ T-helper cells whose role in mediating protective immunity is unclear. It is also known that both Th1 and Th2 immune responses may be protective against *H. pylori* infection but how the T-helper cell response effects a reduction in bacterial load in the stomach remains a quandary (Sutton, 2001). There is also evidence to suggest that the genetic background of the host is important in the efficacy of therapeutic vaccination against *H. pylori* (Kim *et al.*, 2001). Clearly, potential exists for the development of an *H. pylori* vaccine but, as yet, the ultimate goal of achieving complete immunity has not been realised. Vaccine development remains a challenging arena in *H. pylori* research that will undoubtedly prove fruitful once the ideal set of bacterial components are identified and an adequate immune response achieved.

1.10. Pathogenicity determinants of *H. pylori*.

H. pylori elaborates several factors that are involved or putatively involved in the pathogenesis of gastroduodenal disease. These factors and their putative roles in contributing to *H. pylori*-related disease are summarised in Table 1.3.

1.10.1. Virulence factors present in all isolates of H. pylori.

1.10.1.1. Flagella.

In order to colonise the viscous mucosal layer which overlies the gastric epithelium and to persist there for a prolonged period, it is necessary for *H. pylori* to be motile. Motility is conferred by the presence of a unipolar bundle of sheathed flagella which generally number between 4 and 6. The sheath is believed to protect the acid-labile flagellar filament from gastric acidity (Clyne *et al.*, 2000). These flagella constitute an important virulence factor of *H. pylori* as non-motile mutants are defective in colonisation in the gnotobiotic pig model of disease (Eaton *et al.*, 1996). A study by Clyne *et al.* (2000) that investigated the adherence properties of flagellum-negative mutants of *H. pylori* and *H. mustelae* to human and ferret gastric epithelial cells, respectively, revealed that the flagella of these helicobacters do not play a role in their adherence to mucosal epithelium. Yet, adherence of *H. pylori* to the gastric mucosa is a prerequisite for persistence of the infection.

Table 1.3. Putative virulence factors of *Helicobacter pylori*.

Putative virulence	Gene(s)	Function	Reference(s)
factor			
FlaA, FlaB	flaA, flaB	Motility (flagella)	Eaton et al. (1996)
Urease	ureA-ureI	Gastric acid	Mobley et al. (1988)
		neutralisation	
HP-NAP	napA	Activation of neutrophils	Dundon et al. (2001)
		through the production of	
		oxygen radicals	
γ-Glutamyltransferase	ggt	Putative role in	Chevalier et al. (1999)
		colonisation	McGovern et al. (2001)
Iron-acquisition	feo, fec,	Iron assimilation	McGee & Mobley (1999)
proteins	ceuE, pfr, lbp		
AlpA/AlpB	alpAB	Adhesins	Odenbreit et al. (1999)
HopZ	hopZ		Peck et al. (1999)
Heat shock proteins	hspA, hspB	Molecular chaperonin	Dunn et al. (1997)
		Nickel uptake for urease	
VacA	vacA	Vacuolating cytotoxin	Montecucco et al. (1999)
CagA	cagA	Highly immunogenic,	Covacci et al. (1993)
		becomes phosphorylated	Stein et al. (2000)
		upon translocation to	
		host cells	
Bab	babA2, babB	Adhesin (binds Le ^b)	Ilver et al. (1998)
Superoxide dismutase	sodA	Resistance to killing by	McGee & Mobley,
and catalase	katA	phagocytes	(1999)
TlyA	tlyA	Pore-forming cytolysin	Martino et al. (2000)
IceA	iceA	Unknown function,	Peek et al. (1998)
		induced by contact with	
		epithelial cells	

Over 40 genes are involved in the regulation, secretion and assembly of the flagellar structure and are distributed throughout the chromosome (Allan et al., 2000). The proteins which comprise the flagellar filament are designated FlaA and FlaB, the former protein being the major subtype present (Spohn and Scarlato, 1999). Synthesis of H. pylori flagella appears to be controlled by a transcriptional hierarchy in which the master regulator FlgR (flagellum regulator) is transcribed, in turn activating transcription of the genes encoding structural elements of the basal body-hook complex and the *flaB* gene while also repressing transcription of the *flaA* gene. This proposed model system is reminiscent, therefore, of the flagellar biosynthesis systems present in Enterobacteriaceae and the Caulobacter genus (Spohn and Scarlato, 1999). Furthermore, the biosynthesis of FlaA has been found to be regulated at the transcriptional level in response to the presence of the early components of the flagellar assembly pathway (Allan et al., 2000). The absence of these components results in reduced levels of flaA mRNA by an unknown mechanism. Flagellar assembly and hence motility in H. pylori has also been proposed to be regulated by slipped-strand mispairing-mediated mutagenesis in a poly-cytidine tract present in the fliP gene encoding the basal body protein FliP (Josenhans et al., 2000).

1.10.1.2. H. pylori urease.

One of the characteristic features of H. pylori is the production of high levels of urease. The enzyme causes the hydrolysis of urea to CO_2 and NH_3 and is necessary for $in\ vivo$ growth of the organism but not for $in\ vitro$ culture (Eaton $et\ al.$, 1991). It is considered to be an important H. $pylori\ virulence$ factor for this reason and is also the most potent enzyme of its kind among bacterial species, binding its substrate with extremely high affinity (the K_m for urea is 0.8 mM). The high specific activity of the enzyme renders it highly efficient at binding the limited amounts of urea available in the stomach (Mobley $et\ al.$, 1988).

At a mechanistic level, much has been elucidated recently about how urease imparts an acidresistant phenotype to *H. pylori*. In its simplest terms, this is achieved through the hydrolysis of
urea present in gastric juice, thus protecting the organism from stomach acid with a so-called
'cloud of ammonia'. However, the process of urea hydrolysis, which is necessary to alkalify the
environment surrounding the infecting organisms, must be regulated because at later stages of
infection, when the organism is present within the mucous layer and thus protected from acid,
continued urea hydrolysis would be damaging and potentially lethal. The pH rise caused by an
excess of ammonia within and surrounding the bacterium would reach levels injurious to and

uninhabitable by *H. pylori*, respectively (Mobley, 2000). Indeed, this was verified experimentally with the demonstration that in the presence of urea *H. pylori* requires an acidic environment to survive and that *H. pylori* does not survive above pH 4.0 in the presence of urea because of the alkaline pH generated and not because of ammonia toxicity (Clyne *et al.*, 1995). Clearly, *H. pylori* acid resistance is highly dependent on and sensitive to the pH of the surrounding milieu.

The functional characterisation of the protein encoded by the *ureI* gene of the urease gene cluster has illuminated the acid-tolerance phenomenon of H. pylori. The functional part of the H. pylori urease is composed of two structural subunits encoded by the ureA and ureB genes. These genes are present in a cluster that contains seven genes, ureABIEFGH. The cluster is believed to be transcribed as two operons, ureAB and ureIEFGH (Akada et al., 2000). The genes ureE, ureF, ureG and ureH function as accessory genes required for nickel incorporation to produce the catalytically active metalloenzyme. The gene designated ureI has been shown to encode an inner membrane protein with six transmembrane domains which is activated at acidic pH to passively transport urea into the cytoplasm. The gene is not present in non-gastric Helicobacter species (Scott et al., 2000; Weeks et al., 2000). Furthermore, the ureI gene had previously been shown not to be involved in the formation of active urease but its absence vitiated acid survival in vitro (Skouloubris et al., 1998). Additionally, H. pylori acid resistance was demonstrated to be dependent on Urel and also the relative impermeability to acid of the bacterial inner membrane (Rektorschek et al., 2000). Thus, the acid-dependent activity of UreI determines the rate of urea entry into the cytoplasm whereupon urea activates cytoplasmic The ammonia generated by urea hydrolysis buffers the periplasm in an acidic urease. environment.

1.10.1.3. *H. pylori* neutrophil-activating protein (HP-NAP).

Much interest has been generated in recent times about the HP-NAP protein (*H. pylori* neutrophil-activating protein) and its potential efficacy as a vaccine candidate. Based on sequence similarities to bacterioferritins, the HP-NAP protein is likely to have had an ancestral role in iron acquisition but it has been put forward that this function has been superceded, as the name of the protein would suggest, by a neutrophil-activating function during *H. pylori* infection (Dundon *et al.*, 2001). Neutrophils form an important component of the inflammatory infiltrate of the gastric mucosa during *H. pylori* infection and the extent to which the mucosa is damaged can be correlated with the degree of neutrophil infiltration (Warren and Marshall,

1983). A 150-kDa oligomeric protein was isolated from *H. pylori* extracts that could promote neutrophil adhesion to endothelial cells and induce neutrophils to produce reactive oxygen radicals that may contribute to damage of the gastric mucosa. HP-NAP can also act as a chemotactic factor for human neutrophils and monocytes and is also highly immunogenic (Satin *et al.*, 2000; Dundon *et al.*, 2001). Indeed, mice administered recombinant HP-NAP were protected against challenge with *H. pylori* (Satin *et al.*, 2000). Its vaccine candidature resides in this feature of the molecule in conjunction with the high conservation of the protein among geographically distinct *H. pylori* strains (Dundon *et al.*, 2001). In terms of disease pathogenesis, reactive oxygen radicals have host-cell-damaging properties and activated neutrophils, by altering tight junctions and basal membranes, could serve via HP-NAP to release nutrients into the mucosal milieu and enhance the growth of *H. pylori* (Blaser, 1993). It has also been postulated that the main biological activity of HP-NAP has yet to be elucidated and that nutrient release from the mucosa may be a side activity to an unknown primary function (Montecucco *et al.*, 1999).

1.10.1.4. γ-Glutamyltransferase.

The glutamyl transpeptidase (GGT) enzyme has recently emerged as a virulence factor of *H. pylori*. In mammalian tissues it functions in transpeptidation reactions and glutathione synthesis. In reports by Chevalier *et al.* (1999) and McGovern *et al.* (2001) it was established that the enzyme is not necessary for the viability of *H. pylori* as *ggt*-negative mutants could grow normally *in vitro*. However, when *in vivo* studies were carried out, conflicting data were reported by both groups. The former group ascertained that GGT was necessary for colonisation of the gastric mucosa because a *ggt*-negative *H. pylori* strain was unable to cause infection of the mucosa in a murine model of disease. At variance to this, McGovern *et al.* (2001), using both a gnotobiotic pig model and a mouse model, reported that GGT resulted in decreased bacterial loads but that *ggt*-deficient isogenic mutants did cause successful mucosal infections in both models. Thus, it was concluded by McGovern *et al.* (2001) that GGT activity is advantageous in establishing colonisation and is, hence, considered a virulence factor.

It must be borne in mind, as pointed out by Graham and Yamaoka (2000), that to be considered a true virulence factor, the candidate molecule must pass the tests of biological feasibility, i.e., its mechanism of action must explain why it might be one. This is difficult to rationalise in the case of the GGT enzyme. The inconsistency in data reported between the two groups is also

worrisome despite acknowledged differences in experimental design and execution. However, the *in vivo* evidence presented by Chevalier *et al.* (1999) and McGovern *et al.* (2001) does make for a compelling argument.

1.10.1.5. Iron-acquisition proteins.

The ability to acquire iron from the environment is an essential physiological feature of *H. pylori* and indeed bacteria in general. Iron becomes incorporated into haem groups in cytochromes and is required as a co-factor for many essential bacterial enzymes. Metabolism would cease if iron was not available. Iron also acts as a stimulus for the differential expression of genes (besides genes involved in iron uptake) and iron-regulated genes can commonly include those encoding important virulence determinants (Fassbinder *et al.*, 2000). Iron is poorly soluble in biological fluids and is scarce, therefore, in the stomach environment inhabited by *H. pylori*. The organism, thus, has systems in place to assimilate iron, including the production of a lactoferrin-binding protein, a surface-bound iron-binding protein and secreted siderophores (McGee and Mobley, 1999). *H. pylori* also produces a novel non-haem ferritin known as Pfr that is involved in the formation of iron-containing subcellular inclusions for the storage of non-haem iron (Frazier *et al.*, 1993; Bereswill *et al.*, 1998).

In *H. pylori* strain 26695 sequenced by Tomb *et al.* (1997), homologues to uptake systems for ferrous iron (*feo*), ferric iron (*ceu*) and iron-citrate complexes (*fec*) were identified. The ferric uptake regulator (Fur) also has an orthologue in strain 26695 (HP1027). In other organisms, when iron levels are high, Fur acts as a transcriptional repressor of genes involved in iron acquisition. Recently, the Fur protein of *H. pylori* has been shown to be a DNA-binding protein that binds to sequences that resemble the consensus *E. coli* Fur box sequences (Delany *et al.*, 2001). Szczebara *et al.* (1999) determined that transcription of the *vacA* gene was up-regulated in response to iron limitation in a seemingly Fur-independent manner. Moreover, several *H. pylori* outer membrane proteins (Omps) have been identified whose synthesis was induced in iron-restrictive conditions (Worst *et al.*, 1995). In the same study, haemin-agarose affinity chromatography was used to isolate three haem-binding *H. pylori* Omps thought likely to be involved in the uptake of haem from the host. Taken together, the ability of *H. pylori* to acquire iron from the human stomach for use as an essential metabolite and a regulator of gene expression, in particular the expression of virulence genes that are members of the Fur regulon, is likely to be of importance to the bacterium in the pathogenesis of infection.

1.10.1.6. Adhesion.

The literature asserts that *H. pylori* is free-living in the mucosal layer overlying the gastric epithelium but that, in any population of cells, some bacteria will be adherent to the cells of this mucosa. Thus, *H. pylori* must produce adhesins that allow the bacterial cells to form biological interactions with molecules expressed on host gastric epithelial cell surfaces. As these adhesins are responsible for helping *H. pylori* maintain colonisation of an environment where epithelial cells turnover constantly and where ingested material and cellular debris pass continually by the action of peristalsis, they constitute a class of molecule critical to colonisation and therefore to virulence.

Recently, a locus has been identified that was found to encode H. pylori outer membrane proteins (Omps) essential for the adhesion of H. pylori to human gastric epithelial tissue (Odenbreit $et\ al.$, 1999). The alpAB locus is organised as an operon with AlpA specifying a protein with a functional lipoprotein signal sequence (adherence associated lipoprotein A). AlpB has strong amino acid identity to AlpA and at its N-terminus has a putative standard signal sequence. At their C-termini, classic porin-like β -barrel sequences are present with 14 transmembrane amphipathic β -strands predicted. Using isogenic mutant strains, these proteins were found to mediate specific adherence of H. pylori to gastric epithelial tissue sections. However, Odenbreit $et\ al.$ (1999) were unable to establish whether AlpA and AlpB themselves represented the adhesin molecules or whether these proteins were in fact necessary for correct surface display of the adhesin. The importance of these proteins in colonisation has been established nonetheless.

At around the same time that the characterisation of the *alpAB* locus took place, a member of the *H. pylori* Hop protein family (HopZ) was analysed and found to have a role in adhesion of the organism (Peck *et al.*, 1999). The *hop* genes that encode the proteins of the Hop family are members of a supergene family in *H. pylori* of which there are 32 members and to which the *alpA* and *alpB* genes also belong (Tomb *et al.*, 1997; Alm *et al.*, 1999). Peck *et al.* (1999) determined that HopZ was located at the surface of *H. pylori* and that *hopZ* isogenic mutants displayed a reduction in adherence to AGS cells. A function for this Omp in adherence is proposed and prompts speculation as to the function of the as yet uncharacterised members of the Hop family. Could the characterised porins have moieties that might also mediate adhesion and exist *in vivo* as multifunctional proteins?

1.10.1.7. Heat-shock protein homologues.

The *H. pylori* GroES and GroEL heat-shock protein homologues HspA and HspB, respectively, are encoded by the genes *hspA* and *hspB* genes which form a bicistronic operon. The HspB chaperone has a proposed specialised function in the stabilisation and/or export of the urease enzyme. The HspA protein contains a unique nickel-binding site at its carboxy terminus. HspA is thought to act in the integration of nickel into the functional urease molecule. Co-expression of HspA and HspB together with the urease gene cluster increases the activity of urease in functional complementation studies (Dunn *et al.*, 1997).

1.10.2. Virulence factors present in a subset of H. pylori isolates.

1.10.2.1. Vacuolating cytotoxin (VacA).

Approximately 60% of *H. pylori* isolates were recognised as being capable of inducing the formation of intracellular vacuoles in eukaryotic cells *in vitro* by virtue of their production of an extracellular protein, the so-called vacuolating cytotoxin (VacA). Peptic ulcer patients had higher antibody titres to the toxin than did patients with gastritis (Cover *et al.*, 1990). Similarly, a higher percentage of strains isolated from individuals with polymorphonuclear leukocyte infiltration of the gastric mucosa were cytotoxin-positive than those obtained from patients in which inflammatory cell infiltration was not evident (78% versus 33%, respectively) (Phadnis *et al.*, 1994).

The construction of isogenic mutants in the *vacA* gene independently by several groups confirmed that the VacA antigen was responsible for eukaryotic cell vacuolation (Cover *et al.*, 1994; Phadnis *et al.*, 1994; Schmitt and Haas, 1994). Indeed, oral administraion of purified VacA to mice induced gastric epithelial erosions that resembled those seen in humans with peptic ulcer disease (Telford *et al.*, 1994). Thus, qualitatively based on its effect on eukaryotic cells, its inherent immunogenicity and the correlation of VacA+ *H. pylori* with inflammation, the *H. pylori* vacuolating cytotoxin appeared to be a factor of marked significance in the pathogenesis of this organism. Interestingly, although *H. pylori* isolates phenotypically negative for cytotoxin production accounted for 40% of isolates in the *H. pylori* population, these isolates were found to have the *vacA* gene (Phadnis *et al.*, 1994; Schmitt and Haas, 1994). All *H. pylori* isolates are, therefore, genotypically positive for the *vacA* gene but with diversity existing between isolates in the expression of the cytotoxin.

Molecular characterisation of the *vacA* gene elucidated that it encodes a precursor protein of ~139 kDa. This precursor molecule then undergoes specific proteolytic cleavage of a signal sequence 33 amino acids in length and the removal also of a domain of 50 kDa located at the carboxy terminus of the protein (Telford *et al.*, 1994; Montecucco *et al.*, 2001). The signal sequence and C-terminal domain are thought to be responsible for translocation of the active cytotoxin across the inner and outer membranes, respectively, of the Gram-negative bacterial envelope (Phadnis *et al.*, 1994; Schmitt and Haas, 1994). Although the deduced protein sequence of VacA shows no significant similarity to those of any known proteins, it is recognised as being an A-B type toxin with separate domains responsible for cell binding and toxic activity. The cytosolic activity of VacA is mostly localised to the amino-terminal (p37) region of the protein and requires the entire, highly conserved N-terminus of the toxin and a small portion of the p58 moiety for full vacuolation activity (de Bernard *et al.*, 1998; Ye *et al.*, 2000; Montecucco *et al.*, 2001).

Mature VacA has been found as an oligomeric protein composed of identical monomeric subunits that adopt a ring structure, although in this form it has very low activity (Reyrat et al., 1999; Montecucco et al., 2000). Greatly heightened activity of VacA occurs upon exposure to acidic pH which also results in the dissociation of the toxin into monomers. In this apparently activated state, VacA becomes capable of interacting with membranes (de Bernard et al., 1995). The intracellular target of VacA still awaits identification but the toxic effects of VacA on eukaryotic cells are known to be elicited through disruption of the endocytic pathway, the likely induction of permeability changes in gastric epithelial cells (by extension of in vitro observations in polarised epithelial cell monolayers), and the proposed formation of anion-selective channels in lipid bilayers (Czajkowsky et al., 1999; Reyrat et al., 1999).

1.10.2.2. Cytotoxin-associated antigen (CagA).

Recognised as an immunogenic and high molecular weight protein produced by cytotoxic strains of *H. pylori* (Covacci *et al.*, 1993), the CagA antigen became one of the most intensively studied virulence factors produced by this organism. The *cagA* gene is located in the '*cag* pathogenicity island' (*cagPAI*) whose genes are believed to specify a novel type IV secretion apparatus (Odenbreit *et al.*, 2000). The recently defined role for the CagA antigen, namely, its delivery into epithelial cells by the PAI-encoded secretion system, whereupon it hijacks cell signalling (Stein *et al.*, 2000), has provided answers to long-standing questions about the nature of its involvement in bacterium-host interactions. This will provide a basis for studying its

mechanism of host-cell translocation and the tight association of the CagA⁺ phenotype with severe disease outcomes. A more thorough discussion of the *cagA* gene and the *cagPAI* takes place in Chapter 3.

1.10.2.3. Bab adhesin.

H. pylori is capable of binding to host glycoconjugates containing Lewis b (Le^b) structures (Borén et al., 1993). A protein expressed by H. pylori that mediated adhesion of the bacterium to the blood group antigen Le^b was identified by Ilver et al. (1998) using a novel technique called 'retagging'. The protein, designated Bab for blood group antigen-binding adhesin has a molecular mass of 75 kDa and is expressed on the H. pylori cell surface but not on the flagellar sheath. Its activity was detected in 66% of the strains examined in the study by Ilver and coworkers (1998). Three alleles can exist at the bab locus namely, babA1, babA2 and babB. Only the latter two alleles encode functional proteins. Furthermore, an epidemiological association between the CagA antigen and the ability of strains to bind Le^b was noted.

In order to assess the clinical relevance of the Bab adhesin, Gerhard *et al.* (1999) examined the presence and transcription of the *babA2*, *cagA* and *vacA* alleles in clinical isolates and correlated the data obtained to specific *H. pylori*-related disease. Their findings revealed that the *babA2* allele was a good marker for the presence of duodenal ulcer disease and also gastric adenocarcinoma, whereas the VacA/CagA-positive phenotypes of 'type I' strains were only reliably predictive of duodenal ulcer. It has been hypothesised that the tight adherence of *H. pylori* to gastric epithelial cells, mediated in part by the Bab adhesin, might allow the roles of VacA and CagA in virulence to be more effectively played out. However, these hypotheses and the apparent high clinical relevance of the adhesin are set against a backdrop where the Le^b blood group antigen is widely distributed on epithelial cells with which *H. pylori* does not interact and where not all *H. pylori* strains bind the Le^b antigen, as noted by McGee and Mobley (1999).

1.10.3. Emerging H. pylori virulence factors.

1.10.3.1. Superoxide dismutase and catalase.

The enzymes superoxide dismutase (SodB) and catalase (KatA) are bacterial factors that detoxify the damaging oxygen metabolites produced when phagocytic cells undergo an oxidative burst. This takes place upon phagocytosis of foreign material by these cells. Reactive oxygen metabolites, e.g., hydrogen peroxide (H_2O_2), superoxide anions (O_2^-) and the hydroxyl

radical (·OH), are toxic to micro-organisms and have also been shown to induce injury to the gastric mucosa (McGee and Mobley, 1999). Ramarao *et al.* (2000) have reported that *H. pylori* can actively adhere to polymorphonuclear neutrophils, block its own phagocytosis by these cells and in doing so cause these professional phagocytes to release toxic oxygen metabolites extracellularly. The catalase enzyme appears to be responsible for the survival of the bacterium at the phagocyte cell surface (Ramarao *et al.*, 2000). Thus, the persistence and severity of *H. pylori* infection can be exacerbated by this mechanism of avoiding phagocytic killing, with oxygen metabolites having the potential to induce host tissue damage. These enzymes are not yet considered virulence factors because a *sodB* mutant has never been analysed and the importance of catalase is unclear based on conflicting evidence from different studies (McGee and Mobley, 1999).

1.10.3.2. *H. pylori* pore-forming cytolysin – TlyA.

Analysis of the genome sequence of *H. pylori* strain 26695 led to the identification of two loci with putative haemolysin-encoding DNA sequences. One of these loci, HP1086, possessed similarity to that encoding a pore-forming cytolysin (TlyA) from other bacterial species. Characterisation of the *H. pylori* recombinant TlyA orthologue by expression of the protein in a heterologous non-haemolytic host revealed that it too encoded a protein with haemolytic activity that appeared to act via pore formation. Further, a *H. pylori tlyA* mutant had reduced *in vitro* haemolytic activity (Martino *et al.*, 2001). Although Martino *et al.* (2001) reported that *H. pylori* demonstrates haemolytic activity when cultured on unlysed blood agar plates, it is not a feature that has been noted in the culture of Irish clinical strains (S. Beesley, unpublished data). Nonetheless, their study does present evidence that TlyA is a likely pore-forming cytolysin. Furthermore, experimental infection in the mouse model revealed that this protein is an important *H. pylori* colonisation factor and thus a putative pathogenicity determinant. The significance of haemolysis in *in vivo* pathogenesis of *H. pylori* infection is likely to be hotly debated. Nonetheless, it is suggested from results presented by Martino *et al.* (2001) that this property of TlyA might be secondary to an adhesive property possessed by the molecule.

1.10.3.3. iceA locus.

In a search for *H. pylori* genes that are possibly induced by contact with epithelial cells, Peek *et al.* (1998) identified a locus they designated as *ice* (induced by contact with epithelium). The gene is found as two alleles, *iceA1* and *iceA2* but only *iceA1* expression increased following adherence to epithelium. The *iceA1* and *iceA2* alleles are unrelated to each other. The role of

the *iceA* gene in human infection still remains to be established, although the *iceA1* allele encodes a gene product with 65% identity to the restriction endonuclease *Nla*III of *Neisseria lactamica*, suggesting a putative function (Xu *et al.*, 1997). Xu and Blaser (2001) have further ascertained that *iceA1*- and *iceA2*-positive strains regulate *hpyIM* (the gene located immediately downstream of *iceA1/2*) transcription differently with different promoters for the *hpyIM* gene present in *iceA* strains of each subtype. The *iceA1* strains exhibited a higher level of transcription of the *hpyIM* gene than did *iceA2* strains, although the contribution of trans-acting factors to these findings cannot be discounted.

In the initial studies of this locus, strains of H. pylori in which the iceA1 allele was present were shown to be strongly associated with duodenal ulcer disease, while iceA2 strains were present at higher frequency in individuals with gastritis (Peek et al., 1998; van Doorn, L.-J. et al., 1998). This finding prompted other studies of this apparent association in an attempt to ascertain if H. pylori virulence factors including the iceA gene might be predictive of clinical outcome. The findings of Peek and co-workers (1998) were not universal. No association was found between the iceA genotype of a strain and clinical outcome in a study of Portuguese patients (Figueiredo et al., 2001). In concurrence with the latter study and also refuting the claim by Peek et al. (1998) that genotyping of iceA and cagA allowed for the effective identification of patients with ulcers, a study by Yamaoka et al. (1999b) in which a panel of over 400 strains from different geographical origins were analysed found that, although iceA status shows considerable geographical differences, the iceA allele along with the vacA and cagA alleles was not associated with clinical outcome and hence of no disease-predictive value. Indian strains with an iceA2 genotype were found to be what was described as 'weakly disease associated' (Mukhopadhyay et al., 2000). Interestingly, in the strain population analysed in the latter study a novel deletion, never before detected in strains from other geographical locations, was identified in the *iceA1* gene in 20% of the isolates.

Without a verified function for its encoded gene product and given the widely conflicting data concerning its disease-specific association, there is insufficient evidence to consider the *iceA1* gene as a pathogenicity determinant of *H. pylori*. Nonetheless, questions have been raised for future research especially in relation to understanding the function of the *iceA1* gene, i.e., does it encode a restriction endonuclease and what, if any, is the link between the ability of *iceA1* strains to better transcribe the *hpyIM* gene and virulence?

1.11. Genetic diversity in H. pylori.

1.11.1. Chromosomal diversity revealed by molecular typing studies – historical perspective.

The genetic heterogeneity present between isolates in the H. pylori population exists at an unprecedented level compared to other bacterial species. It is such that all unrelated isolates can be distinguished from each other by a variety of typing and molecular fingerprinting methods and means that epidemiological associations between isolates are rarely discernible. diversity of the H. pylori genome first became apparent when research on this organism was in its infancy. In reports by Langenberg et al. (1986) and Majewski and Goodwin (1988), the HindIII restriction patterns of all the isolates examined were found to be unique. Similarly, using restriction endonuclease analysis of chromosomal DNA and plasmid profiling (Simor et al., 1990; Oudbier et al., 1990), the same conclusions were drawn. A variety of high-resolution molecular typing methods have also been employed to address this question, including PCRbased restriction fragment length polymorphism (RFLP) analysis of the flagellin and urease genes (Akopyanz et al., 1992b; Foxall et al., 1992; Shortridge et al., 1997), random amplified polymorphic DNA (RAPD) analysis (Akopyanz et al., 1992a), oligofingerprinting (Marshall et al., 1996), repetitive extragenic palindromic (REP) PCR (van Doorn N.E.M. et al., 1998), ribotyping (Owen et al., 1992) and pulsed-field gel electrophoresis (Taylor et al., 1992). This body of work asserts unquestionably that the genome of H. pylori displays a diversity never before encountered among bacterial pathogens and has stimulated enormous interest in its biological significance and the mechanisms of its generation.

1.11.2. Genetic diversity in virulence-associated loci – the pre-genome era.

The association of *H. pylori* infection with the pathogenesis of a spectrum of diseases has been recognised since its original isolation but is still incompletely understood. The disparate outcome of infection in different individuals suggested the existence of strains of *H. pylori* with enhanced virulence potential and the possibility of the presence in some strains of virulence-associated loci (Atherton, 1997). Yet, the contribution of host and environmental factors to the outcome of *H. pylori*-associated disease cannot be discounted. Investigations to probe the diversity of the *H. pylori* genome based on molecular typing studies have brought to light that not only does this bacterium exhibit a heterogeneity that extends throughout the chromosome, (Section 1.11.1) but also, that variability exists between *H. pylori* isolates with respect to (i) the carriage of specific genes by sub-groups of isolates and (ii) the diversity at the level of gene structure and expression of specific genes.

1.11.2.1. The *vacA* gene.

Inter-strain diversity in VacA expression exists (Section 1.10.2.1). Active toxin is only produced by approximately 50% of isolates despite the ubiquitous presence of the gene (Cover et al., 1990). In addition to this, vacA alleles also exhibit a type of genotypic variation known as mosaicism, with conserved regions and highly divergent regions occurring within these alleles (Atherton et al., 1995). The mosaic structure of the vacA gene is believed to be evidence that recombination has occurred between strains in vivo. The divergent regions comprise the middle domain and the signal sequence of the gene. Four signal sequence types have now been identified, namely, s1a, s1b, s1c and s2. Mid-region types m1 and m2 predominate, although vacA alleles where further subdivisions of these mid-region types was merited have been identified in a minority of isolates (Marshall et al., 1998; van Doorn et al., 1998). A large body of evidence indicates that there is a strong correlation between the vacA genotype of a strain and Strains of signal sequence/mid-region combination type s1/m1 were disease outcome. identified by Atherton et al. (1995) as producing high levels of the vacuolating cytotoxin in vitro and were associated with increased gastric epithelial damage, enhanced gastric inflammation and duodenal ulceration (Atherton et al., 1997). Although geographic differences can be detected, it is generally accepted that H. pylori strains with vacA m1 alleles are more toxigenic than those with m2 alleles which are non-toxigenic or only mildly so. Furthermore, s1a/m1 alleles produce a more potent toxin than do s1b/m1 alleles. In vivo evidence indicated that the vacA signal sequence type was associated with levels of mucosal inflammatory cell infiltrate; s1a alleles were associated with more inflammation than were s1b or s2 type alleles and indeed, s1a strains were most commonly isolated from ulcer patients (Atherton, 1998; Marshall et al., 1998).

1.11.2.2. Diversity in CagA expression, the *cagA* gene, and the *cagPAI*.

As was found with the VacA cytotoxin, inter-strain diversity in expression of the CagA antigen was also detected. However, where CagA⁻ strains were found, it was because they lacked the *cagA* gene. *H. pylori* strains can arbitrarily be divided into six groups, based on whether they possess the *cag* genetic locus and express the CagA protein and active VacA toxin (Xiang *et al.*, 1995). However, most isolates are classifiable as either Type I (CagA⁺/VacA⁺), or Type II (CagA⁻/VacA⁻). Isolates of Type I are more commonly isolated from patients with peptic ulcer disease and gastric cancer. However, in countries where the incidence of *H. pylori* infection is high, the majority of strains are *cagA*⁺ irrespective of clinical presentation (Miehlke *et al.*, 1996; Pan *et al.*, 1997). The *cagA* gene itself has a variable 3'-region which can be correlated with the

size of the CagA antigen expressed by particular isolates. Subtypes A–D have been deduced for the *cagA* 3'-end from Japanese isolates (Yamaoka *et al.*, 1998) with one of these subtypes (Type C) found to be associated with more severe degrees of gastric atrophy, higher levels of anti-CagA antibody and gastric cancer. Variants of the *cagA* 3'-end are thought to arise in the stomach (Yamaoka *et al.*, 1999a). Furthermore, the *cagPAI* has been shown to vary in structure between isolates and can be disrupted by the presence of chromosomal DNA or the IS element IS605 (Censini *et al.*, 1996), although overall, the gene content and arrangement of the PAI are well conserved between isolates (Akopyants *et al.*, 1998a).

1.11.2.3. The fla and ure genes.

The roles played by the flagella and urease in the initial colonisation of the gastric mucosa by *H. pylori* have already been discussed. The genes which encode the flagellin proteins, *flaA* and *flaB*, can vary extensively in sequence between isolates, as evidenced by RFLPs when PCR products were digested with restriction endonucleases (Ohta-Tada *et al.*, 1997). Moreover, Forbes *et al.* (1995) presented evidence that the sequence diversity found in the *flaA* and *flaB* genes had resulted from the reassortment of sequences between strains in the bacterial population. Extensive restriction site diversity in the urease structural genes *ureA*, *ureB* and *glmM* (formerly *ureC*) is also found, a fact that has been exploited extensively for the molecular typing of strains (Foxall *et al.*, 1992; Shortridge *et al.*, 1997).

1.11.2.4. The bab genes.

Three *bab* alleles were identified (*babA1*, *babA2* and *babB*) which encode the putative virulence adhesin Bab that mediates binding to the fucosylated Le^b histo-blood group antigen (Section 1.10.2.3) (Ilver *et al.*, 1998). The alleles *babA1* and *babA2* are identical with the exception of a 10-bp deletion of a repeat motif in the signal peptide sequence of the *babA1* gene. This results in ablation of the translational initiation codon and suggests the presence of a hot spot for phenotypic variation in the *bab* gene family. The *babA1* allele is therefore not functional. The *babB* allele encodes a functional protein that has regions of heterogeneity and similarity with the BabA2 protein. The *bab* genes are members of the extensive Hop gene family identified by Tomb *et al.* (1997) that encode outer-membrane proteins and between them show extensive amino-acid sequence identity in the N-terminal and C-terminal domains. The potential for recombinational events between domains of similarity in the genes that encode these proteins is suggested and might lead to mosaicism (Tomb *et al.*, 1997; Ilver *et al.*, 1998).

1.11.3. Genetic diversity in H. pylori – the genomic and proteomic era.

The availability of complete genome sequence information for an organism is the forerunner to using microarray technology to carry out genome-scale analyses. The quantity of information that can be generated using these sophisticated methodologies that provide for the simultaneous interrogation of the full complement of ORFs in an organism, combined with the speed at which the data is available, underlie the usefulness of these technological advances. Furthermore, in principle the inspection of the profile of proteins expressed by a particular organism from a clinical setting has more meaningful information to contribute than does genome analysis alone. Nonetheless, the identification of putatively significant proteins using a proteomics approach is considerably more useful if the genes which encode these proteins can also be identified. Thus, genomics and proteomics are complementary techniques.

Where *H. pylori* is concerned, genomic and proteomic studies have been reported which attempted to exploit the availability of complete *H. pylori* genome sequences to gain new perspectives on the genetic diversity of *H. pylori* and on aspects of pathogenicity. Moreover, considered as an organism that co-evolves with its host over years of chronic infection, the analysis of the *H. pylori* proteome is seen as an approach whereby the dynamics of the host-bacterium relationship and the process of co-evolution can be better understood (Björkholm *et al.*, 2001b; Engstrand *et al.*, 2001).

1.11.3.1. Whole-genome microarray analysis.

The genetic diversity that exists between *H. pylori* strains of differing virulence potential was investigated by analysing the genomic content of 15 *H. pylori* strains using a whole-genome microarray (Salama *et al.*, 2000). A minimal functional core of 1,281 genes was defined for this organism which included genes that encoded metabolic, biosynthetic, regulatory and cellular functions. It was found that up to 22% of genes were dispensable, among which were included genes that encoded outer-membrane proteins, R-M systems and transposases, and genes involved in lipopolysaccharide biosynthesis. Similarly, using microarray technology strain-specific differences in gene content between two *H. pylori* strains of different clinical background (one gastric ulcer strain and one duodenal ulcer strain) were identified (Israel *et al.*, 2001a). A deletion of a region of the *cag*PAI between these two strains was identified and again, R-M system genes featured among the strain-specific genes that were found. It was also possible to ascertain that the presence of an intact *cag*PAI was influential in determining the host inflammatory response to infection *in vivo* and *in vitro*.

1.11.3.2. Proteomics.

As noted by Chakravarti *et al.* (2001), the advances in genomic and bioinformatical methods paralleled a renewed interest in post-genomic high-throughput protein analysis. Employing two-dimensional polyacrylamide gel electrophoresis (2-D PAGE) as a method of analysing a bacterial proteome, along with the availability of protein analytical methods such as mass spectrometry, means that strain-specific differences in protein content between isolates can be identified. Moreover, the proteins expressed in a cell under a given set of conditions can be separated and analysed. High variability of the *H. pylori* proteome was identified in an extensive study of *H. pylori* strains 26695, J99 and SS1 (Jungblut *et al.*, 2000). However, it was pointed out by these authors that proteins with a single amino acid change resulting in a net change of only one charge are shifted in a 2D-PAGE profile. This study attempted the systematic identification of *H. pylori* proteins predicted from genome sequencing studies and confirmed the expression of 27 predicted conserved hypothetical ORFs. Further, 152 proteins in total were identified which included nine pathogenicity factors and 28 antigens.

When 2-D PAGE was employed to test whether *H. pylori* strains from individuals with different clinical symptoms could be grouped according to protein patterns, the results were interesting, if largely inconclusive. Despite extensive variation in spot pattern, nine putative disease-specific proteins were identified, which had, however, no similarities to typical virulence factors (Enroth *et al.*, 2000). In terms of cost and the amount of labour required, 2-D PAGE is considered unsuitable for routine diagnostic purposes (Enroth *et al.*, 2000). The effect of pH on the *H. pylori* proteome has also been analysed by 2-D PAGE and protein N-terminal sequencing (Jungblut *et al.*, 2000; Slonczewski *et al.*, 2000) and indicates the efficacy of these techniques for an appraisal of the effect of environmental changes on protein profile.

The studies briefly described above are a testimony to the great promise resulting from determining complete genome sequences and represent very useful practical applications of such technologies. A functional genomics approach will doubtless form a critical avenue of investigation in further deciphering mechanisms of *H. pylori*-host interactions and in identifying potentially useful candidate antigens for vaccine design.

1.12. Origins of diversity in the *H. pylori* genome.

1.12.1. Nucleotide mutation and recombination.

Genome-wide molecular typing studies and the analysis of specific loci bore witness to H. pylori being an organism of tremendous genetic diversity in that every infected individual is apparently colonised by a unique strain. Furthermore, as detailed above, diversity at the virulence-associated loci vacA and cagA in particular has a significant bearing on the severity of disease caused by a particular strain. By what mechanism does this diversity arise?

PCR-based RFLP analysis of the *ure* and *fla* genes gave, perhaps, the first indication that unique gene sequences were present in each isolate. Kansau *et al.* (1996) confirmed this by examining the DNA sequence of a 294-bp fragment of the *glmM* gene from 29 isolates and found a unique sequence in each strain. Additionally, the presence of a mosaic gene structure in *vacA* alleles was considered evidence that these mosaic alleles had arisen by recombination between strains (Atherton, 1998), a phenomenon that was also echoed in the regions of heterogeneity and similarity that exist between the *babA* and *babB* alleles (Ilver *et al.*, 1998). Additionally, as alluded to previously, Forbes *et al.* (1995) found evidence for the reassortment of sequences in the *flaA* and *flaB* genes between strains. Thus, nucleotide mutation and recombination both seemed to be involved in the genesis of the genetic heterogeneity in *H. pylori*.

1.12.2. The population structure of H. pylori.

These examples of gene diversity referred to above related to a small fraction of the entire complement of genes present on the *H. pylori* chromosome. The diversity detected at these specific loci could feasibly have arisen as a result of selective pressure on their encoded gene products because of their association with virulence and host immune responses to them. To understand more fully the origins of the genomic diversity exhibited by *H. pylori*, it became necessary to determine whether the rest of the *H. pylori* chromosome exhibited the same level of variability as was detected for these virulence-associated genes and whether any detected diversity was applicable to all isolates. Diversity also poses a broader question – what is the population structure of *H. pylori*?

Population genetics is the study of the natural variability of bacterial populations. Different bacterial species can vary in their population structure and three types have been described, namely, clonal, pannictic and epidemic (Maynard-Smith *et al.*, 1993). These distinct types of population structure relate to associations between genes at different loci. It is possible to

determine the frequency of recombination in natural populations using multilocus enzyme electrophoresis (MLEE) to estimate levels of allelic variation in multiple chromosomal genes. A clonal population structure describes a situation where the rate of recombination of large genome segments is not great enough to randomise genomes or break up clonal associations (Maynard-Smith *et al.*, 1993). A panmictic population structure shows evidence of free recombination and frequent horizontal exchange in which clonality has been disrupted. In a panmictic population, alleles at different loci are said to be in a state of linkage equilibrium and, thus, phylogenetic trees would vary depending on the locus under examination. This type of population structure is exemplified by *N. gonorrhoeae*. Epidemic populations refer to the rapid expansion of clones within a recombining species, e.g., the epidemic clones of *N. meningitidis*.

Employing MLEE, the extent of allelic variation in six H. pylori genes which encoded metabolic house-keeping enzymes was assessed, thereby avoiding the detection of diversity that could have arisen from selective pressure (Go et al., 1996). Based on such analysis H. pylori has a panmictic population structure. The mean genetic diversity calculated for the experimental sample exceeded the level of diversity recorded in all other bacterial species Since then, free recombination in the H. pylori population has been studied by MLEE. demonstrated using the homoplasy test and compatibility matrices by analysing the sequences of the vacA, flaA and flaB genes from H. pylori strains isolated in Germany, Canada and South Africa (Suerbaum et al., 1998). Again, comparisons with sequences from E. coli, N. meningitidis and Drosophila melanogaster showed that recombination in H. pylori is much more frequent than in these organisms. Thus, within the H. pylori population, attempting to compare the genetic relatedness of epidemiologically unrelated isolates is of little use, even although the population structure of the organism can display evidence of clonality over short time periods after natural transmission (e.g., between family members) (Marshall et al., 1998; Suerbaum et al., 1998). It is now accepted that the genotypic diversity displayed by H. pylori arises from a combination of a high mutation rate and frequent recombination. As noted by Suerbaum et al. (1998), recombination does not create sequence polymorphisms but can increase the number of unique sequences present by effecting the shuffling of polymorphic nucleotides or even importing nucleotide diversity from other species.

1.12.3. Quasi-species.

The possibility that within *H. pylori* there may exist subspecies or what has also been termed a species-complex or quasi-species, where a single strain may represent a cluster of closely

related organisms, has been put forward because of the low level of clonality observed between unrelated isolates of this organism (Censini *et al.*, 1996; Hazell *et al.*, 1997; Covacci and Rappuoli, 1998). Indeed, the existence of intraspecies variation (genes found in the chromosome of some strains but not in others) within *H. pylori* as a manifestation of the non-clonal population structure of this organism has led to the concept of a 'species genome' for organisms such as *H. pylori*, *N. meningitidis*, *E. coli* and *Salmonella enterica* (Lan and Reeves, 2000).

It is known that between strains of these organisms a substantial proportion of chromosomal genes are not found among all members of the species. These genes can be considered auxiliary genes to a core set of genes that are found in almost 100% of species members. The 'species genome' concept strives, therefore, to be more inclusive in that the definition of the 'species genome' will take in both core and auxiliary genes and comprise all the genes found in the species. Thus, the 'species genome' for an organism will give a more accurate and in-depth appraisal of the genetic potential of a bacterial species and should, in the genome era, be a concept that is attainable for many pathogens.

1.13. Mechanisms of prokaryotic chromosomal rearrangement.

1.13.1. Tandem duplications, deletions and inversions.

Rearrangements that occur within bacterial chromosomes can take place by a variety of mechanisms that are RecA-dependent and involve repeated DNA sequences that provide homologous regions at which the occurrence of recombination can take place. Chromosome replication is the primer for the occurrence of deletions, tandem duplications and inversions by unequal crossing-over between misaligned repeated sequences (Roth *et al.*, 1996; Weinstock and Lupski, 1998). The types of repeated sequences encountered in bacterial genomes include the multigene families (e.g., ribosomal RNA genes), insertion sequence (IS) elements, transposons, shorter REP (repetitive extragenic palindromic) sequences or, indeed, short repetitive sequences of any type. Deletions and tandem duplications take place between repeated sequences which exist in the same orientation. Inversions of chromosomal regions can also take place by recombination between oppositely oriented repeated sequences by a process that is considered to be more demanding than that which is required to achieve tandem duplication or deletion (Roth *et al.*, 1996). Deletions, the segregation of duplications and inversions, can arise also from intrachromosomal exchanges.

1.13.2. Horizontal gene transfer.

The phenomenon of horizontal gene transfer (HGT) or the acquisition of foreign DNA sequences has been described as a strategy whereby microbes can share the evolutionary success of others (Arber, 2000). Further, de la Cruz and Davies (2000) have noted that HGT has now become the substance of evolution and that it is no longer confined to being of interest for practical applications alone. Investigated extensively in the course of studies that examined the dissemination of antibiotic resistance genes in bacterial populations, the horizontal transfer of DNA can confer significant new phenotypes in a single step of acquisition. The transfer on mobile genetic elements (e.g., plasmids, bacteriophages, transposons and gene cassettes) of antibiotic resistance genes, entire biodegradation pathways and pathogenicity determinants can be considered as emergency responses by bacteria to strong selection pressures. Additionally, in instances where a large proportion of the genetic information that specifies the virulence determinants of a particular organism is located on extrachromosomal elements, the occurrence of HGT can be viewed as a speciation event. Thus, the significance of the involvement of HGT in the evolution of bacterial pathogenesis cannot be overstated (de la Cruz and Davies, 2000).

There are three main mechanisms by which naturally occurring interspecies or intraspecies gene exchange may occur in bacteria, namely, natural transformation, conjugation and bacteriophage transduction. The frequency with which HGT takes place depends on the degree to which the barriers to genetic exchange possessed by bacteria slow down DNA transmission. These barriers include transfer barriers, e.g., the host ranges of genetic exchange vectors and different microhabitats, and establishment barriers, e.g., bacterial restriction-modification systems that act to control the genetic isolation of bacterial species by limiting gene exchange. However, sequence divergence is reported to be the greatest limiting factor to successful gene transfer and the establishment of the acquired genetic information (Matic *et al.*, 1996).

1.13.3. Local sequence change.

Chromosomal rearrangements and systems for horizontal DNA transfer are mediated by genetic functions. However, point mutations (local sequence changes) arise because of the intrinsic structural instability of nucleotides that can be directly mutable or result in infidelity upon DNA replication (Arber, 2000). Point mutations theoretically lead to a stepwise improvement of biological function on which natural selection steadily exerts pressure (Arber, 2000; Schloter, 2000). However, the mutational generation of a novel biological function happens rarely.

1.14. Rearrangements and mechanisms of diversity generation in the *H. pylori* genome inferred from genome sequence analysis.

1.14.1. Inversions, transpositions and reciprocal exchange.

The genomic organisational comparison of the two sequenced *H. pylori* strains, 26695 and J99, required artificial inversions and transpositions in order to align homologous regions of the two genomes (Alm *et al.*, 1999). The chromosomal segments involved ranged in size from 1 kb to 83 kb and, significantly, most of the artificial end-points created were associated with IS elements, repeated genes or sequences, and DNA restriction modification (R-M) systems. Inversion of a genomic fragment also implicated recombination between inverted repeated sequences that encoded the carboxy-terminal domains of two members of the multigene family of outer-membrane protein encoding alleles (Alm *et al.*, 1999; 2000). Moreover, evidence for reciprocal exchange between the genes for the Lewis-antigen binding adhesins *babA* and *babB*, was also detected (Alm *et al.*, 1999). Thus, the occurrence of ancestral chromosomal rearrangement events in the genomes of progenitors of strains 26695 and J99 yield the organisational differences that exist between them and implicate these mechanisms of rearrangement in the likely continuous evolution of the genome of this bacterium over a prolonged time-scale.

1.14.2. Gene transfer by 'insertion with long target duplication'.

In intricate work reported by Nobusato *et al.* (2000b), a novel mechanism of gene mobility termed 'insertion with long target duplication' was identified by genome comparison of strains 26695 and J99. It can be described as a process whereby a part of one genome is duplicated in direct orientation in the other genome with these repeats now flanking a DNA segment not present in the first genome. The duplicated target sequence is much longer (between \sim 40–500 bp) than that which becomes duplicated upon insertion of transposable elements but the overall mechanism of insertion is formally similar. This process is not limited to the acquisition of restriction and/or modification system genes, but these were found adjacent to or within the insertion in the majority of examples (Nobusato *et al.*, 2000b). Indeed, as pointed out by these authors, the *cagPAI* becomes inserted with a duplication of a 31-bp long sequence in the genomes of *cagA*⁺ strains (Censini *et al.*, 1996). Examples of this mechanism of insertion (with long target duplication) were identified in the genomes of *N. meningitidis* and *N. gonorrhoeae* (Nobusato *et al.*, 2000b). With respect to the acquisition of R-M systems by *H. pylori*, their probable horizontal transfer had been deduced (Nobusato *et al.*, 2000a) and it was hypothesised

that this novel mechanism of gene mobility resulted from attack by restriction enzymes on the chromosome in the absence of sufficient methylation activity (Nobusato *et al.*, 2000b).

1.14.3. Acquisition of exogeneous DNA – horizontal gene transfer (HGT).

In the years prior to the availability of genome sequence information for *H. pylori*, the contribution of horizontal gene transfer (HGT) to the evolution of pathogenesis in this organism was recognised with the identification of the *cag*PAI (Censini *et al.*, 1996). Furthermore, analysis of the G+C% content and sequence data suggested the presence of sequences related to eukaryotes, archaea and other prokaryotes. For example, analysis of the *H. pylori* major sigma factor (RpoD) revealed that the amino acid sequence of RpoD showed strongest similarity to those of Gram-positive bacteria, e.g., *Staphylococcus aureus*, *Listeria monocytogenes* and *Enterococcus faecalis* (Marshall *et al.*, 1998).

When the genome sequence of strain 26695 was published, it became clear that inter-species horizontal transfer played a significant role in the evolutionary development of H. pylori. Many of the proteins described in this Gram-negative bacterium show greatest amino-acid sequence similarity to proteins from the different major taxonomic groups (the non-Proteobacteria). Indeed, Goldman and Kranz (1998) published evidence for the transfer of an entire biosynthetic pathway for c-type cytochromes. A type II pathway of cytochrome biogenesis is characteristic of Gram-positive bacteria but was detected in H. pylori, indicating transfer of genetic information from different phylogenetic groups. Moreover, a total of five chromosomal regions were identified in strain 26695 that had a significantly different G+C% compared to the mean of 39 mol% for the rest of the genome (Tomb et al., 1997). Two of these regions contain at least one copy of the IS605 insertion sequence, flanked by a 5S rDNA sequence and a 521-bp repeat. The other regions include the cagPAI, a region associated with β and β' RNA polymerases and elongation factor EF-G, and a region associated with two restriction/modification systems. Similar analysis of strain J99 led to the designation of one of these regions of lower G+C% as the 'plasticity zone', as it contained 46% and 48% of the strain-specific genes of strains 26695 and J99, respectively. The plasticity zone constitutes a contiguous region in strain J99 but is split into two domains in strain 26695 by ~600 kb of DNA (Alm et al., 1999).

1.14.4. Insertion sequence (IS) elements.

The identification of the *cag*PAI was accompanied by the recognition in *H. pylori* for the first time of an insertion sequence element. The element, IS605, was found to disrupt the PAI and to split it into right and left segments (Censini *et al.*, 1996). *H. pylori* strain 26695 was found to have five full-length copies of IS605 and eight partial copies. Five partial copies of the element were found in strain J99. Additionally, a distinct element, IS606, was identified by genome sequencing of strain 26695 (Tomb *et al.*, 1997) in which two partial and two full-length copies of this element were found. Strain J99 possesses one complete IS606 element and four partial copies, two of which are variable in length (Alm *et al.*, 1999). Both IS605 and IS606 encode two divergently transcribed transposases (TnpA and TnpB). The two elements share less than 50% nucleotide sequence identity and their encoded transposases have approximately 29% amino-acid identity.

The *H. pylori* IS605 transposable element family has been extended and now has two more members designated IS607 (Kersulyte *et al.*, 2000) and the recently identified IS*Hp608* (Kersulyte *et al.*, 2002). IS606, IS607 and IS*Hp608* are related to IS605 on the basis of protein level similarities. Thus, by acting as repetitive DNA, the presence of IS elements in a genome might serve as substrates for recombination and thereby effect rearrangement of the genome and contribute to the evolutionary development of the bacterium. The transposition of the IS element intrachromosomally with the attendant potential for the modification of gene expression contributes significantly to a genome in a state of continuous flux.

1.14.5. Phase-variable genes.

Genome sequence scrutiny led to the identification of homopolymeric tracts and dinucleotide repeats in promoter regions or coding sequences of numerous *H. pylori* genes. Apart from those genes that encode putative proteins of unknown function, the putative proteins encoded by the phase-variable genes are thought to be involved in lipopolysaccharide (LPS) biosynthesis and DNA restriction-modification systems and to be cell-surface associated (Tomb *et al.*, 1997; Saunders *et al.*, 1998; Alm *et al.*, 1999). It is thought likely that the presence of these simple nucleotide repeats facilitates the phase variation of the genes they are associated with by slipped-strand mispairing. This mechanism involves the displacement of strands of duplex DNA during replication followed by mispairing of complementary bases at sites of short tandem repeats. The insertion or deletion of nucleotides or repeat units can result and consequently an ORF may be moved in or out of frame with respect to an upstream translation

initiation codon (Levinson and Gutman, 1987). Indeed, mechanistically, this has been proven for two genes involved in LPS biosynthesis in H. pylori namely, the $\alpha(1,2)$ fucosyltransferase gene (Wang $et\ al.$, 1999a) and both the $\alpha(1,3)$ fucosyltransferase genes (Appelmelk $et\ al.$, 1999), resulting in the phase-variable expression of Lewis antigens expressed on the O-antigen chain of H. $pylori\ LPS$.

H. pylori has a paucity of regulatory proteins. Thirty-two gene products have been classified as possibly having regulatory functions. This constitutes half the number reported for H. influenzae which has a genome of comparable size to H. pylori and less than a quarter of those predicted for E. coli (Scarlato et al., 2001). Thus, regulation of gene expression might be simpler in H. pylori than in these other organisms. However, the putative differential expression of the biologically relevant cell-surface protein encoding genes and LPS biosynthesis genes allowed by slipped-strand mispairing may be a mechanism evolved by H. pylori to exert a level of regulatory control over colonisation and persistence, for example (Alm and Trust, 1999). Table 1.4 summarises the various diversities that exist in H. pylori genomes and the methods that led to their identification.

1.15. Natural transformation and competence in *H. pylori* as contributors to genomic diversity.

1.15.1. Natural transformability of H. pylori - in vitro and in vivo evidence.

The ability of a bacterium to acquire genetic information horizontally implies its competence for natural transformation. Early *in vitro* studies revealed that most strains of *H. pylori* are naturally competent for DNA uptake (Nedenskov-Sørensen *et al.*, 1990; Wang *et al.*, 1993b), a phenomenon that has been exploited widely for experimental mutagenesis manipulations. As detailed above, genome sequence analysis testified to the important role played by HGT in shaping the genetic make-up of *H. pylori* and in the evolution of virulence in this bacterium. Thus, over short time-periods genetic competence might be inextricably linked with the disease process. Indeed, the occurrence of HGT involving the *cag*PAI has recently been demonstrated *in vivo* (Kersulyte *et al.*, 1999), indicating that the process of natural transformation can occur under the living conditions of the bacterium and can allow the acquisition or loss of advantageous virulence traits during infection while simultaneously contributing to genetic diversity (Israel *et al.*, 2001b).

Table 1.4. Origins of diversity in the *H. pylori* genome.

Type of diversity	Method of identification	Reference(s) Forbes et al. (1995); Shortridge et al. (1997)	
Local sequence change	PCR-RFLPs in ure and fla genes		
(point mutation)	PCR product sequencing	Kansau et al. (1996)	
	MLEE	Go et al. (1996)	
	Homoplasy test	Suerbaum et al. (1998)	
Horizontal gene transfer	cagPAI identification	Censini et al. (1996)	
	'Plasticity zone' identification	Tomb et al. (1997); Alm et al. (1999)	
	Loss of cagPAI in vivo	Kersulyte et al. (1999)	
	Genes transferred from other taxonomic groups	Marshall et al. (1998), Section 1.14.3.	
	Linkage equilibrium of alleles	Suerbaum et al. (1998)	
Insertion sequence transposition	IS605	Censini et al. (1996)	
or act as repeated DNA leading	IS606	Tomb et al. (1997)	
to recombinations	IS607	Kersulyte et al. (2000)	
	IS <i>Hp608</i>	Kersulyte et al. (2002)	
Allelic recombination	Mosaic gene structures in vacA and bab alleles	Atherton et al. (1995); Ilver et al. (1998)	
Insertion with long target Genome sequence analysis duplication		Nobusato et al. (2000b)	
Phase-variable genes arising	Genome sequence analysis	Tomb et al. (1997); Saunders et al. (1998)	
from slipped-strand mispairing Experimental characterisation		Appelmelk et al. (1999); Wang et al. (1999)	

1.15.2. The competence machinery.

The components of the DNA transformation machinery in *H. pylori* and the molecular mechanisms involved have been explored extensively by Hofreuter and colleagues (2001) who have recently shown that natural transformation competence in this organism is encoded by the genes of the *comB* locus that specify a type IV secretion system. The *comB* locus consists of four tandemly arranged genes with partially overlapping ORFs which constitute a single transcriptional unit and were designated *orf2*, *comB1*, *comB2* and *comB3* when first identified (Hofreuter *et al.*, 1998). These ORFs were given the alternative names *comB7*, *comB8*, *comB9* and *comB10*, respectively, in the more recent work because of the amino acid sequence identity present between these gene products and those of the type IV secretion apparatus of *Agrobacterium tumefaciens* and the proteins VirB7, VirB8, VirB9 and VirB10 specifically.

From the *comB* locus, the proteins ComB8, ComB9 and ComB10 are indispensable for transformation competence. The putative lipoprotein specified by the *comB7* gene is not essential but is thought to stabilise the apparatus and modulate transformation efficiency. The available evidence indicates that ComB8, ComB9 and ComB10 are membrane-associated proteins which, along with accessory protein functions encoded by the VirB4-homologous ATPase (ORF HP0017 or ComB4), the *comH* gene (Smeets *et al.*, 2000) and the ORF at locus HP0333 (Ando *et al.*, 1999), are involved in DNA uptake in *H. pylori* by a second type IV transport system. As noted by Hofreuter *et al.* (2001), pathogenic *H. pylori* strains possess two independent type IV transport systems, namely, the *cag*PAI system involved in protein export and this second system for natural transformation.

1.15.3. Barriers to transformation and the net contribution of competence to diversity.

Although the evidence is substantial that the genomic structure of *H. pylori* has been modified by the occurrence of rearrangement within the genome and is likely to be heavily influenced by the continuous horizontal transfer of genetic information, it is noteworthy that *H. pylori* appears not to possess uptake sequences that can, in other Gram-negative bacteria, enhance the binding and uptake of homospecific DNA during transformation (Saunders *et al.*, 1999). Numerous copies of these uptake-specific sequences are present in the genomes of *H. influenzae* and *N. meningitidis* (1465 uptake sequences were identified in the former organism) (Marshall *et al.*, 1998). However, in analysis of the genome sequence of *H. pylori* strain 26695 no such sequences were identified. This scenario is postulated to reflect the restricted environmental

niche inhabited by *H. pylori* in which the opportunity to acquire DNA from a heterologous species rarely arises (Saunders *et al.*, 1999).

The absence of uptake-specific sequences in *H. pylori* is set against a backdrop of the discovery that the complete genome sequences of this organism do not contain orthologues of several enzymes including MutH and MutL from the DNA mismatch repair system, although the methyltransferase encoding *mutS* and *uvrD* genes are present (Tomb *et al.*, 1997; Alm *et al.*, 1999). In bacteria, the occurrence of recombination between DNA substrates that are partially homologous depends on the extent of DNA sequence identity that exists between them. This is monitored by the enzymes of a system known as the mismatch repair system. Mismatch-binding components, MutS and MutL, of these systems strongly inhibit interspecies recombination by blocking RecA-mediated strand transfer between DNA substrates that may only be weakly or moderately diverged (Matic *et al.*, 1996).

As an opposing force to the inhibition of recombination at diverged sequences by bacterial mismatch repair systems, bacteria also possess what is known as an SOS system that acts as an inducible positive regulator of interspecies recombination between similar sequences. The physiological response of the SOS system is induced by exposure of bacterial cells to conditions that might damage DNA or might interfere with replication. Induction of the SOS system enhances the capacity for DNA repair and increases genetic variation and as such improves the survival of bacteria under stressful conditions (Matic *et al.*, 1996).

Thus, the opposing activities of mismatch repair and SOS systems determine the rate at which sequence divergence is accumulated and, therefore, the extent of genetic isolation of a species. The absence of MutH and MutL orthologues from the *H. pylori* genome sequence suggests that *H. pylori* has somewhat less stringent control of replicative errors (Alm and Trust, 1999) and, furthermore, that neither interspecies transfer nor the establishment of acquired genetic information may be impeded by the components of this machinery. There were no identifiable components of an SOS system present in the genome sequences either.

Thus, the classical prokaryotic barrier to chromosomal gene transfer, the mismatch repair system, is likely at best to be only semi-functional in *H. pylori*. This, however, may be of valuable consequence to a bacterium under certain environmental conditions because of the apparent absence of an inducible SOS system which would stimulate chromosomal

establishment of heterologous DNA under adverse conditions. Moreover, the absence in *H. pylori* of sequences that enhance homospecific DNA uptake during natural transformation does not appear to be an impediment to the occurrence of horizontal transfer in this organism.

What therefore constitutes the greatest physiological barrier to DNA uptake in *H. pylori* and, in so doing, protects the genetic isolation of this species despite a population structure that indicates a significant propensity to undergo recombination? Although the environmental niche inhabited by *H. pylori* affords a barrier to some degree, it is thought that endogenous restriction modification systems present the strongest barrier against the adulteration of the genome by both foreign and heterologous *H. pylori* DNA (Israel, 2001). Taken together, the natural competence of *H. pylori* is most likely liable to contribute to the diversity of the species when DNA transfer from a homologous strain is involved over the time-scale of natural infection.

1.16. Type II restriction-modification systems in *H. pylori* as barriers to strain-specific transformation.

Restriction-modification (R-M) systems of bacteria can be classified as type I, type II or type III based on several factors, including enzyme subunit composition, co-factor requirements, the structure of recognition sequences and the reaction products generated (Wilson and Murray, 1991). Type II R-M systems, which in *H. pylori* represent the majority, consist of two enzymic activities of opposing function, the methyltransferase and the restriction endonuclease. The methyltransferase is responsible for the methylation of endogenous DNA at the endonuclease recognition site, thereby protecting it from endonucleolytic cleavage by the cognate restriction enzyme. Unmodified foreign DNA is rapidly degraded by the restriction endonuclease.

Analysis of the sequenced genomes of *H. pylori* strains 26695 and J99 (Tomb *et al.*, 1997; Alm *et al.*, 1999) revealed the presence of more than 20 putative R-M systems in each strain comprising 4% of each genome. Sequence comparison of the two genomes also revealed that 15–20% of strain-specific genes in strains 26695 and J99 encoded enzymes of R-M systems. Approximately three-quarters of these R-M systems in each strain were of type II. Furthermore, comparison of strains J166 and 26695 by PCR-subtractive hybridisation isolated 18 J166-specific clones of which seven displayed DNA sequence identity to genes of R-M systems (Akopyants *et al.*, 1998b). It was apparent, therefore, that R-M system genes contributed to interstrain diversity and comprised a large part of the complement of strain-specific genes possessed by *H. pylori* isolates. Biochemical analysis of strains 26695 and J99 to assess the

functionality of a subset of the R-M systems in each strain revealed that, in spite of having 90% of R-M system genes in common, different sets of genes were active in each strain. Strain-specific R-M system genes were, however, fully functional (Lin *et al.*, 2001). Other studies in which different isolates were examined for the presence and specificity of their various R-M systems also indicated that each strain analysed had a varying complement of restriction endonucleases (Xu *et al.*, 2000a; Xu *et al.*, 2000b).

Thus, *H. pylori* isolates are believed to be rich in R-M systems which are highly diversified among strains. Instances where the methyltransferase gene was active but where the restriction enzyme gene was inactive were also detected (Lin *et al.*, 2001; Virkute *et al.*, 2001). It follows, therefore, that the possession by individual *H. pylori* strains of variable complements of these R-M systems and their disparate functionality between distinct populations (i.e., individual stomachs) hinders the chromosomal establishment of horizontally acquired DNA, especially where it has been acquired from a non-homologous *H. pylori* strain that is likely to be of different modification status or where it is from a different species of organism. The modification status of transforming DNA and the expression of R-M systems determine the natural transformation efficiency of *H. pylori*.

1.17. Microevolution of the H. pylori chromosome in vivo.

Microevolution is described as the occurrence of changes in one or a few chromosomal loci within a clonal population (Hobbs et al., 1994). Employing various high resolution DNA e.g., restriction endonuclease analysis (Oudbier et al., 1990), methods. oligofingerprinting (Marshall, 1996), and RAPD-PCR (Kersulyte et al., 1999; Kuipers et al., 2000; Israel et al., 2001b), microevolution has been observed between the genomes of H. pylori isolates and is believed to be evidence that recombinant subtypes of H. pylori arise during In these examples, microevolution was detected as a difference in genomic fingerprint profiles between related strains. Such differences were characterised in the studies by Kersulyte et al. (1999) and Israel et al. (2001b) as involving the cagPAI, Omp-encoding genes, and genes from the plasticity zone, to mention but some. There is no doubt that genomic rearrangement and the horizontal acquisition of genetic information made an important contribution to the evolutionary development of H. pylori over an evolutionary time-scale. However, over the time-scale of natural infection (years or decades), it is interesting that the genome of H. pylori should exhibit a propensity to undergo recombination. In many other bacterial species, e.g., N. meningitidis, N. gonorrhoeae, E. coli, Streptococcus pneumoniae and H. influenzae, microevolution is known to be responsible for the rapid emergence of variants with novel virulence and resistance properties (Ziebuhr et al., 1999). Genome plasticity, thus, is a mechanism that can be exploited for the adaptation of the bacterium to changing environmental conditions and therefore constitutes an important physiological property of certain organisms (Ziebuhr et al., 1999).

Similarly, the observation of microevolution in *H. pylori* was suspected to reflect bacterial adaptation to niche conditions that may not be uniform throughout the stomach, e.g., the pH level within different regions of the stomach is known to vary (Marshall *et al.*, 1998). Moreover, microevolution might reflect adaptation to selection pressures arising from a changing spectrum of immune and inflammatory responses and various other physiological changes that infection brings about (Kersulyte *et al.*, 1999). As described in Section 1.14., the presence in the *H. pylori* genome of repetitive DNA, e.g., IS elements, and a multigene-family encoding *H. pylori* Omps, combined with the natural competence of the organism for transformation, would easily facilitate intra-genomic recombination and the acquisition of exogenous DNA, respectively, thus allowing continuous microevolution of the genome during infection.

1.18. Aims.

The analyses described in the present studies aimed to characterise the microevolution of *H. pylori* with respect to defined loci and by whole-genome examination and consisted of the following investigations:

- Genotypic and phenotypic analysis of diversity at the 3'-end of the cagA gene between H.
 pylori paired isolates, i.e. matched isolates recovered from the same or related (close contact family members) individuals that have been shown to have identical or very similar oligonucleotide fingerprint profiles.
- The use of PCR-subtractive hybridisation to isolate DNA fragments that differed or partially differed between the chromosomes of *H. pylori* paired strains as a method of analysing microevolution in the entire genomes of epidemiologically related isolates.
- Characterisation of the molecular basis for distinct recombinations identified by PCRsubtractive hybridisation and an analysis of phenotypic variation between paired isolates.

- Examination of clonally related familial isolates by PCR-subtractive hybridisation to assess
 the microevolution that might occur upon transmission of an *H. pylori* isolate to the stomach
 of a different host.
- Analysis of the putative molecular mechanism for *H. pylori* Lewis antigen phenotypic variation between paired isolates.

Chapter 2
Materials and Methods

2.1.	Helicob	pacter pylori Bacteriological Techniques.	47
	2.1.1.	Primary isolation of H. pylori from gastric biopsies and storage of isolates.	47
	2.1.2.	Culture of H. pylori on solid medium	47
	2.1.3.	Culture of H. pylori in liquid medium	47
	2.1.4.	Handling and nomenclature of H. pylori isolates used in the	
		present studies	48
2.2.	E. coli	Strains.	48
2.3.	Molecu	ılar Techniques.	49
2.4.	Agaros	se Gel Electrophoresis.	49
2.5.	Prepar	ration and Transformation of Competent Bacteria.	49
2.6.	DNA I	solation.	50
	2.6.1.	Extraction of genomic DNA from H. pylori.	50
	2.6.2.	Small-scale preparation of plasmid DNA by alkaline lysis	51
	2.6.3.	Rapid screening of transformants.	52
2.7.	Southe	rn Hybridisation.	52
	2.7.1.	Capillary blotting.	52
	2.7.2.	Probe hybridisation.	53
2.8.	Polymo	erase Chain Reaction and probes for Southern Hybridisation	53
	2.8.1.	Amplification of H. pylori genes and genomic fragments by PCR	53
	2.8.2.	Gene probes for Southern hybridisation.	54
2.9.	PCR-S	ubtractive Hybridisation.	54
	2.9.1.	Isolation of bacterial genomic DNA.	54
	2.9.2.	Restriction enzyme digestion.	55
	2.9.3.	Clean-up of restriction endonuclease reaction and preparation of	
		DNA for adaptor ligation.	55
	2.9.4.	Ligation of adaptor oligonucleotides to restriction endonuclease-cleaved	
		tester DNA.	55
	2.9.5.	First hybridisation.	56
	2.9.6.	Second hybridisation.	56
	2.9.7.	Primary PCR amplification.	57
	2.9.8.	Secondary PCR amplification	58
	2.9.9.	Analysis of amplification products from primary and secondary	
		PCR reactions.	58
	2.9.10	. Construction of subtracted DNA library	59

	2.9.11.	Screening of subtracted DNA library	59
	2.9.12.	Labelling of amplified insert DNA and probing restricted genomic DNA	59
	2.9.13.	DNA sequencing.	60
2.10.	Protei	n Analysis.	60
	2.10.1.	H. pylori cell lysate preparation.	60
	2.10.2.	Proteinase K treatment of H. pylori for phenotypic analysis for the	
		expression of $Lewis^x$ and $Lewis^y$ on the lipopolysaccharide (LPS) molecule.	.61
	2.10.3.	Preparation and analysis of H. pylori outer membrane fractions	61
	2.10.4.	$So dium\ dode cyl-sulphate\ polyacrylamide\ gel\ electrophores is\ (SDS-PAGE).$	62
	2.10.5.	Transfer of separated proteins to nitrocellulose membranes	62
	2.10.6.	Western immunoblotting.	62
2.11.	H. pyl	ori Colony Blotting.	63
2.12.	Finger	rprinting of H. pylori Isolates.	64
	2.12.1.	Oligofingerprinting.	64
	2.12.2.	RAPD-PCR fingerprinting	65
	2.12.3.	Similarity analysis of fingerprint profiles	65

2.1. Helicobacter pylori bacteriological techniques.

2.1.1 Primary isolation of H. pylori from gastric biopsies and storage of isolates.

Strains of *Helicobacter pylori* were isolated from pinch biopsy specimens obtained from patients undergoing gastroduodenal endoscopy at the Adelaide/Meath Hospital, Dublin. The tissue sample obtained from endoscopy, with the blood vessels facing upwards, was streaked using a sterile inoculating loop onto the surface of two 90 mm diameter petri dishes (Sterilin Ltd., Feltham, UK) each containing ~20 ml Campylobacter Agar Base (Oxoid, Basingstoke, UK) enriched with lysed or defibrinated horse blood (7% v/v). The plates were then incubated in a microaerobic atmosphere at 37 °C for 5 d after which time they were examined for the growth of colonies resembling those produced by *H. pylori*. Successful culture of *H. pylori* from gastric biopsies was confirmed by subjecting presumptive *H. pylori* colonies from primary isolation plates to Gram stain and tests for oxidase, catalase and urease. After further subculturing, *H. pylori* cells were collected from the surface of 48-h lawned agar plates and suspended in the cryo-fluid contained in 'Protect' tubes (Protect Technical Service Consultants, Lancs., UK) and stored at –70 °C.

2.1.2. Culture of H. pylori on solid medium.

H. pylori was routinely cultured on Campylobacter Agar Base (Oxoid) containing 7% (v/v) lysed or defibrinated horse blood and the antibiotics vancomycin (dissolved in sterile distilled water) and amphotericin B (dissolved in DMSO) at concentrations of 10 and 5 μg/ml, respectively. A 10-μl aliquot of partially thawed cryofluid containing viable H. pylori cells was inoculated onto the agar plate and streaked out using a sterile inoculating loop. Agar plates were incubated at 37 °C in a microaerobic atmosphere (5% (v/v) O₂, 10% (v/v) CO₂), created using a gas-generating kit for campylobacters (Oxoid) in an anaerobic jar. A tissue dampened with sterile water was placed at the bottom of the jar to humidify the atmosphere. Plates were typically incubated for a period of 48–72 h.

2.1.3. Culture of H. pylori in liquid medium.

Culture of *H. pylori* in liquid medium was carried out in Brucella Broth (Oxoid) supplemented with foetal calf serum (10% v/v) (GibcoBRL, Life Technologies, Germany) and the antibiotics vancomycin (10 μg/ml), amphotericin B (5 μg/ml), trimethoprim (5 μg/ml) and cefsulodin (5 μg/ml). Cell culture flasks (25 cm² canted neck flask, 0.2 μm vented cap) (Corning Costar Corp., MA, USA) containing 5 ml broth were inoculated with

H. pylori cells from young plate cultures (24–36 h). Inoculated flasks were placed in an anaerobic jar containing a Gas-Pak for campylobacters and catalyst and then inserted into a 2-litre port in an orbital shaker and incubated at 37 °C and 150 r.p.m.

2.1.4. Handling and nomenclature of H. pylori isolates used in the present studies.

Table 2.1 details the matched and familial *H. pylori* isolates that were used in these analyses and named in the following experimental chapters. The strains utilised in these studies that were present in the Moyne Institute *H. pylori* strain collection prior to the commencement of these investigations were subject to single-colony-purification at the beginning of this work and are those strains depicted in blue coloured font in Table 2.1. Matched isolates that were cultured from biopsy tissue during the course of these investigations (as described in Section 2.1.1) were also subject to single-colony purification from primary isolation plates prior to storage. Such isolates are depicted in red coloured font in Table 2.1. For particular isolates, growth was consistently observed to be poor. In these instances sweeps were taken from plates for storage. This however, is not considered good practise for a bacterium as genotypically diverse as *H. pylori*. Such strains are marked with an asterix in Table 2.1. The remaining isolates in Table 2.1 were isolated in the Adelaide/Meath Hospital and were stored as pooled colonies upon receipt and were later subject to single-colony purification.

All *H. pylori* strains isolated in the course of these studies and received from elsewhere were named according to pre-existing nomenclature (Marshall, 1996), i.e., preceded by the letters MI (for Moyne Institute). In the case of newly isolated *H. pylori* strains, the number that follows the letter designation MI, consisted of the last three or four digits of the hospital number accompanying the isolate. Where the letter A or C follows this number, this designation indicates that the isolate in question was taken from the antrum (A) or the corpus (C) of that individual's stomach. This was usually used in instances where two isolates were obtained from an individual, one from each of these distinct sample sites.

2.2. E. coli strains.

E. coli XL-1 Blue (Stratagene, CA, USA) and TOP10 (Invitrogen, CA, USA) were used as recipient strains for transformation. The genotypes of these strains are as follows: XL-1 Blue – supE44 hsdR17 recA1 endA1 gyrA46 thi relA1 lac F'[proAB+ lacIq lacZΔM15 Tn10 (Tet]], and TOP10 – F- mcrA Δ(mrr-hsdRMS-mcrBC) Φ80lacZΔM15 ΔlacX74 recA1 deoR araD139

Table 2.1. Details of matched and familial H. pylori strains used throughout these studies.

H. pylori strain	Date of isolation	Single/Sequential ^a	Clinical origin ^b	Time between samplings ^c	Sample site ^d	Antimicrobial treatment ^e
MI 512	pre '96				A	
MI 541	pre '96	Sequential	DU	14 m	A	DLan
MI 506	pre '96	204			A	
MI 520	pre '96	Sequential	DU	3 m	A	DFT
MI 571	pre '96	1	GI, DI		A	
MI 575	pre '96	Sequential	DI	6 m	A	DFT
MI 517	pre '96	•	GI		A	DFT
MI 566	pre '96	Sequential	DU	14 m	A	DFT/Lan
MI 355	pre '96		ASY		A	
MI 356	pre '96	Single	ASY		A	
MI 458	pre '96				A	
MI 459	pre '96	Single	n/k		A	
MI 683A	June '99	8			A	
MI 683C	June '99	Single	DU		C	
MI 212A	March '99	0			A	
MI 212C	March '99	Single	DU		C	
MI 862A*	April '99				A	
MI 862C	April '99	Single	DU		C	
MI 334A	Feb '99				A	
MI 334C*	Feb '99	Single	DU		C	
MI 545A	Feb '99				A	
MI 545C*	Feb '99	Single	DU		C	
MI 393A	Feb '99				A	
MI 393C	Feb '99	Single	DU		C	
MI 688A	Aug '98				A	
MI 688C	Aug '98	Single	DU		C	
MI 500A	Dec '98				A	
MI 500C*	Dec '98	Single	DU		C	
MI 745A	Dec '98				A	

MI 745C	Dec '98	Single	DU	C
MI 886A	June '99			A
MI 886C	June '99	Single	DU	C
E				
Familial H. pylor	i isolates			
Family 1				
MI 431	April '97		DU	C
MI 577A	May '98			A
MI 577C	May '98	Single	DU	C
MI 9613A	May '98			A
MI 9613C	May '98	Single	ASY	C
MI 018	Jan '98			A
MI 613	Jan '98	Single	ASY	C
MI 561	Nov '97			A
MI 562	Nov '97			C
MI 563	Nov '97	Single	ASY	A
MI 610	Jan '98			A
MI 611	Jan '98	Single	ASY	C
Family 2				
MI 218A	Aug '98			A
MI 218C	Aug '98	Single	ASY	C
MI 605A	May '98			A
MI 605C	May '98	Single	ASY	C
MI 222A	Aug '98			A
MI 222C	Aug '98	Single	ASY	C
MI 221C	Aug '98	Single	ASY	C

^a'Sequential' refers to when matched isolates were obtained at two endoscopic sessions, and 'single' refers to when strains were obtained on one sampling occasion.

^bClinical origin; DU, duodenal ulcer; GI, gastritis; DI, duodenitis; ASY, asymptomatic; n/k, not known.

^cTime elapsed to nearest month (m) from previous biopsy sampling.

^dSite of stomach from which biopsy was taken; 'A' refers to an antral biopsy; 'C' refers to a corpus biopsy.

^eD, Denol (bismuth chelate); F, Flagyl (metronidazole); T, tetracycline; Lan, Zoton (Lansoprazole).

^{*}See Section 2.1.4.

 Δ (ara-leu)7697 galU galK rpsL (Str^R) endA1 nupG. E. coli strains were cultured in L-broth and on L-agar (Oxoid) containing appropriate selective antibiotics. Blue-white selection on agar of E. coli transformants harbouring plasmids that contained a portion of the lacZ gene was allowed by the inclusion of 5-bromo-4-chloro-3-indolyl- β -D-galactoside (X-Gal) (Melford Laboratories Ltd., Suffolk, UK) at a concentration of 16 μ g/ml.

2.3. Molecular techniques.

Routine molecular techniques, e.g., restriction endonuclease digestion, DNA ligation, clean-up of PCR reactions, phenol-chloroform extraction of DNA and DNA precipitation were all carried out according to standard protocols (Ausubel *et al.*, 1987; Sambrook *et al.*, 1989) and the specific instructions supplied with restriction and modification enzymes by the manufacturers. Commercially purchased kits were also used for some applications (e.g. for the isolation of high quality plasmid DNA) according to the accompanying specifications.

2.4. Agarose gel electrophoresis.

DNA was electrophoresed through agarose gels (Promega, WI, USA) according to standard protocols (Sambrook *et al.*, 1989) using Tris–Borate–EDTA buffer at a 0.5 × final concentration [1 1 10 × stock contains 890 mM Tris base, 890 mM boric acid, 40 ml 0.5 M EDTA, pH 8.0] or 1 × Tris–Acetate–EDTA buffer [1 litre 50 × stock contains, 2 M Tris base, 57.1 ml glacial acetic acid, 0.1 M Na₂EDTA.2H₂O, pH ~8.5] where appropriate. Agarose gels were routinely cast at a 1% (w/v) concentration in buffer including ethidium bromide (0.5 μg/ml) from a 20 mg/ml stock. Electrophoresed DNA was visualised under ultra-violet (302 nm) light by exposure of the gel on a Chromato-Vue UV transilluminator, (UVP Inc., CA, USA). Horizon gel tanks (Gibco BRL, Life Technologies, Germany) of various sizes were routinely used.

2.5. Preparation and transformation of competent bacteria.

Calcium-chloride-competent *E. coli* cells were prepared according to the method of Tang *et al.* (1994). L-broth (2 ml) was inoculated with *E. coli* strain XL1-Blue and incubated for \sim 16 h at 37 °C. After overnight incubation, 500 μ l of this culture was inoculated into 100 ml prewarmed L-broth and incubated at 37 °C with shaking until an OD_{600nm} of 0.4 was reached. After brief centrifugation (\sim 2000 g for 3 min) in the 216 rotor of the IEC Centra-GP8 Bench Centrifuge (IEC, MA, USA) the cell pellets were resuspended in 10 ml of an ice-cold

solution containing 80 mM CaCl₂ and 50 mM MgCl₂ and stored on ice for 10 min. This treatment was repeated twice. After the last centrifugation step, the pellet was resuspended in 1 ml of ice-cold 0.1 M CaCl₂ and 1 ml 50% (v/v) glycerol and stored as 80-μl aliquots at – 70 °C.

For transformation, 80 µl of a thawed competent cell suspension was mixed with an appropriate volume of ligation reaction mix and kept on ice for 30 min. Heat-shock at 42 °C for 40 s was carried out followed by cooling of the mixture on ice for 2 min. A 400 µl aliquot of SOC medium [2% (w/v) tryptone, 0.5% (w/v) yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl₂, 10 mM MgSO₄, 20 mM glucose] was then added and the mixture incubated for 30 min at 37 °C. Aliquots of the transformed mixture were plated on L-agar containing antibiotics appropriate to the selectable marker encoded by the vector.

2.6. DNA isolation.

2.6.1. Extraction of genomic DNA from H. pylori.

H. pylori strains from which genomic DNA was to be isolated were lawned on the surface of Campylobacter Agar plates containing lysed or defibrinated horse blood and antibiotics. After incubation for 72 h at 37 °C, genomic DNA was extracted using a modification of the method of Ausubel et al. (1994). The cells were harvested into 1 ml phosphate buffered saline (PBS) [137 mM NaCl, 2.7 mM KCl, 4.3 mM Na₂HPO₄.7H₂O, 1.4 mM KH₂PO₄, pH ~7.3] using sterile cotton swabs. Cells were pelleted by centrifugation at maximum speed (~11,000 g) in a bench-top centrifuge (Sorvall MicroSpin 24) for 15 s. After washing in PBS, genomic DNA was extracted by resuspension of the washed pellet in 567 μl TE buffer [10 mM Tris-HCl, pH 8.0, containing 1 mM EDTA], 30 μl 10 % (w/v) SDS and 3 μl proteinase K (20 mg/ml) followed by incubation of the resuspended cells in a water bath at 37 °C for 1 h. After this incubation, 100 μl 5 M-NaCl was added to the lysed cell suspension followed by thorough mixing. An 80-μl aliquot of CTAB/NaCl solution [10% (w/v) CTAB in 0.7 M NaCl] was then added, followed by thorough mixing of the suspension and the mixture incubated for 10 min at 65 °C.

To obtain the nucleic acid from the cell lysate an equal volume (\sim 750 μ l) of chloroform was then added. Organic and aqueous phases were then mixed by vigorous shaking. The organic

phase containing precipitated polysaccharide and protein was separated from the nucleic-acid-containing aqueous phase by centrifugation at 11,000~g in a bench-top centrifuge for 8 min. The upper aqueous phase was removed to a clean 1.5 ml Eppendorf tube by pipetting small aliquots. Two further extractions of this aqueous phase were then carried out with phenol–chloroform in a similar manner to that described for the chloroform extraction step. Precipitation of the DNA from this solution was achieved by adding 0.6 volume of isopropanol to the fully extracted aqueous phase and allowing the mixture to stand at room temperature for 10 min. The DNA was pelleted by centrifugation at 11,000~g for 10~min. The pellet was washed with 1~ml 70% (v/v) ethanol followed by a further 5-min centrifugation at 11,000~g. The ethanol was poured off carefully and the isolated DNA was routinely dried at room temperature and resuspended in $50~\mu l$ TE buffer. The DNA concentration was determined by agarose gel electrophoresis against purified bacteriophage lambda DNA of known concentration.

2.6.2. Small-scale preparation of plasmid DNA by alkaline lysis.

E. coli strains harbouring plasmids containing cloned H. pylori chromosomal fragments were cultured in L-broth containing antibiotics appropriate to the selectable marker on the plasmid. Typically, 2-ml L-broths were inoculated and incubated at 37 °C and 250 r.p.m. for 16 h. After overnight incubation, 1.5 ml of the culture was poured into a microfuge tube and centrifuged at 11,000 g for 30 s. The supernatant spent medium was then decanted.

The plasmid DNA was then extracted using a modification of the method of Birnboim and Doly (1979) as described by Sambrook *et al.* (1989). The bacterial pellet was resuspended in 100 μl GET buffer [50 mM glucose, 25 mM Tris-Cl (pH 8.0), 10 mM EDTA (pH 8.0)] by vigorous vortexing. After resuspension, 200 μl of a freshly prepared solution containing 0.2 M NaOH and 1% SDS was added. The contents of the tube were mixed by rapid inversion and then held on ice. A 150-μl aliquot of 'Solution III' [5 M potassium acetate (60 ml), glacial acetic acid (11.5 ml) and H₂O (28.5 ml)] was added and the contents of the tube mixed by vortexing briefly. The tubes were stored on ice for 5 min. Centrifugation at 11,000 g for 5 min then took place. The supernatant fraction was then transferred to a fresh tube. Plasmid DNA was precipitated by the addition of 2 volumes of absolute ethanol followed by vortexing. The mixture was held at room temperature for 2 min followed by centrifugation at 11,000 g for 5 min. After removal of the supernatant fluid the pellet was rinsed with 1 ml

70% (v/v) ethanol and again centrifuged briefly. The supernatant fluid was again removed and the pellet allowed to dry. The extracted plasmid DNA was dissolved in 50 μ l TE buffer (pH 8.0) containing DNAase-free RNAase (20 μ g/ml).

When insert DNA was to be sequenced, plasmid DNA of high purity was isolated using mini-prep kits, typically the High Pure Plasmid Isolation Kit supplied by Roche Molecular Biochemicals (East Sussex, UK).

2.6.3. Rapid screening of transformants.

Rapid screening of undigested plasmid DNA from transformant colonies taken directly from agar plates was carried out according the method described by Le Gouill and Dery (1991).

2.7. Southern hybridisation.

2.7.1. Capillary blotting.

Agarose gels, through which DNA had been electrophoresed in order to transfer it to nylon membranes for Southern hybridisation experiments, were treated as follows prior to transfer. The positions of the DNA fragments in a molecular weight ladder electrophoresed alongside the DNA to be probed were marked with a needle when visualising with ultra-violet light. The gel was then soaked in 0.2 M HCl for 10 min (depurination step that aids the transfer of high molecular weight DNA). This treatment was followed by the denaturation of the DNA contained in the gel by submerging the gel in denaturing solution [1.5 M NaCl, 0.5 M NaOH] for 45 min with gentle shaking. The gel was rinsed briefly in distilled H₂O to enhance neutralisation. Finally, the gel was soaked in neutralisation solution [1.5 M NaCl, 1 M Tris-HCl (pH 7.4)] for 45 min while maintaining gentle shaking.

Transfer to positively charged nylon membrane (Hybond-N+, Amersham, Buckinghamshire, UK) by capillary blotting using 20 × SSC [3 M NaCl, 0.3 M sodium citrate] was then carried out as previously described (Southern, 1975), typically by turning the gel upside down so that the DNA electrophoresed through the bottommost part of the gel has only to move a short distance before binding to the positively charged nylon. Blotting took place overnight, after which time the transfer system was disassembled and the positions of the molecular weight ladder marked on the nylon membrane while the dehydrated gel was still sandwiched to the membrane. The transferred DNA was fixed to the membrane by heating at 80 °C for 2 h.

2.7.2. Probe hybridisation.

DNA probes labelled with Digoxigenin (DIG) were used throughout these studies. Probes were labelled by random primed labelling or PCR DIG-labelling in PCR reactions which incorporated DIG-dUTP into the synthesised DNA. Hybridisation of the probe to the membrane was carried out in standard hybridisation buffer [5 x SSC, 0.1% (w/v) Nlauroylsarcosine, 0.02% (w/v) SDS, 1% (w/v) blocking reagent (Roche Molecular Biochemicals)] when high stringency temperatures were utilised (65-68 °C). Following prehybridisation for 2 h, hybridisation took place overnight at the same temperature and in the same buffer used for prehybridisation. Detection of the chemiluminescent signal from labelled hybridised probes was carried out as previously described (Boehringer Mannheim, GmbH, 1995) with the following modifications. The membrane was washed twice in 2 × SSC-0.1% (w/v) SDS for 10 min each, followed by two washes of 20 min each at 65-68 °C in $0.5 \times SSC-0.1\%$ (w/v) SDS. The membrane was then washed in maleic acid buffer [0.1] M Maleic acid, 0.15 M NaCl (pH 7.5)] containing 0.3% (v/v) Tween 20 for 5 min followed by blocking for 30 min in a 1% (w/v) solution of blocking reagent (diluted in maleic acid buffer). Anti-DIG Fab fragments were then diluted 1:10,000 in the blocking solution and the filter allowed to incubate for 30 min in this. Unbound antibody was washed from the membrane in maleic acid buffer containing Tween 20 followed by detection of the chemiluminescent signal with CSPD or CDP-Star chemiluminescent substrates (Roche Molecular Biochemicals) diluted in detection buffer [100 mM Tris-HCl (pH 9.5), 100 mM NaCl]. Signals were detected upon exposure of the filter to X-ray film (RP-X1 Medical X-Ray Film, Med Ray Photosystems Ltd., Dublin).

2.8. Polymerase chain reaction (PCR) and probes for Southern hybridisation.

2.8.1. Amplification of H. pylori genes and genomic fragments by PCR.

PCR amplifications were carried out in a PTC-100 Programmable Thermal Cycler (M.J. Research, MA, USA). *Taq* DNA polymerase was used for most applications with the 10 × buffer and MgCl₂ supplied with the enzyme by Promega. The Expand™ (Roche Molecular Biochemicals) thermostable polymerase was used for amplification of longer templates (Section 3.2.4). Reactions were carried out according to standard protocols (Ausubel *et al.*, 1987; Sambrook *et al.*, 1989) typically in a final volume of 25 µl containing 1 µl of DNA template (~50–100 ng/µl), 3 mM MgCl₂, 10 pmol of each primer, 2.5 units of *Taq* polymerase, 0.4 mM (each) dATP, dCTP, dGTP, dTTP (Promega) in 1 × reaction buffer.

Primer annealing temperatures were determined according to primer melting temperature. A standard set of thermal cycling parameters consisted of 1 cycle of 94 °C (5 min) followed by 30 cycles of 94 °C (1 min), 55 °C (1 min), 72 °C (2 min), and finally 72 °C (10 min). The DNA sequences of oligonucleotide primers used in PCR are shown in Table 2.2.

2.8.2. Gene probes for Southern hybridisation.

Probes for the *vacA* and *flaA* genes used in Southern hybridisation were amplified by PCR using the primers shown in Table 2.2. Two probes for the *cagA* gene were utilised and were derived from the plasmids pMC3 (Tummuru *et al.*, 1993) and 'clone 1' (Covacci *et al.*, 1993) which both contained a cloned *cagA* gene.

2.9. PCR-Subtractive Hybridisation.

PCR-subtractive hybridisation was carried out on four sets of paired isolates obtained from individuals either simultaneously or consecutively, and, on two sets of isolates obtained from each of two individuals within two multiply colonised families. The technique was carried out as outlined by Akopyants *et al.* (1998b) and according to the specifications in the user manual which accompanies the PCR-Select Bacterial Genome Subtraction Kit (Clontech, CA, USA) with several modifications as detailed below. A schematic overview of the technique is shown in Fig. 2.1 with a more detailed account of the various DNA species recovered at each step appearing in Fig. 2.2.

2.9.1. Isolation of bacterial genomic DNA.

H. pylori genomic DNA for PCR-subtractive hybridisation experiments was isolated as described in Section 2.6.1 according to the method of Ausubel et al. (1994) and with the outlined modifications. As high quality DNA is imperative for successful subtraction, organic extraction steps were performed with the utmost of care in order to prevent phenolic contamination of the isolated DNA. The concentration and quality of the extracted DNA were assessed by electrophoresis of small aliquots (typically 2 μl) alongside aliquots of lambda DNA of known concentration. Successful subtraction also depends strongly on having accurate concentrations of the two DNA populations in use so determinations were made with care.

Table 2.2. Sequences of oligonucleotide primers used in PCR, RAPD-PCR and PCR-Subtractive Hybridisation.

Primer name	Sequence (5' to 3')	Gene ^a	Strand	Position (bp) ^b	GenBank Accession No.
Cln2#1	GTTCAAGGCAGGCATCC	hopM/hopN	+/_c	7792-7808	AE00635
Cln2#2	CCTGGGAATGTAGTGG	hopM/hopN	+/_c	8085-8100	AE00635
1260	CCAGACGAAACCGTGCAT	jhp1260/HP1341	+	6983-7000	AE00635
1262	GCCTTGCTCATAGAATGCTTG	jhp1262/HP1343	-	9498-9518	AE00635
211	CGGTAATGCCGGTGAATAG	jhp0211/HP0226	+	7142-7160	AE001459
213	GAATGTTGCTCAGCCATTC	jhp0213/HP0228	-	9782-9801	AE001459
flaF	GCCCATTTTCATGCTCTT	flaA	+	1780-1798	X60746
flaR	ACCCAACACAATAACCGC	flaA	-	10-28	X60746
cagAF1	TAAGCAACTCCATAGACCACT	cagA	+	9-29	X70039
cagAR1	GCTTGAGATAACCCATTACCG	cagA	-	3081-3101	X70039
cagAF2	ATCTCAATCAAGCGGTATCAG	cagA	+	2891-2911	X70039
cagAR2	ATAGGCGTTTTTGATAGCGG	cagA	_	5791-5810	X70039
cagAF	GATAACAGGCAAGCTTTTGAAG	cagA	+	691-712	X70039
Cag1	ACCCTAGTCGGTAATGGGTTA	$cagA^{\mathrm{d}}$	+	3073-3093	X70039
Cag2	GTAATTGTCTAGTTTCGC	$cagA^{\mathrm{d}}$	-	3586-3603	X70039
FucT6	TTGCGCTGGTATTCTTCC	fucT2	_	574-591	AF076779
FucT7	GAACACTCACACGCGTC	fucT2	+	1-17	AF076779
1,2FucR	CCCACTCCTTACAAAGGATATTTC	fucT2	_	965-989	AF076779

VA1-F	ATGGAAATACAACAAACACAC	vacA ^e	+	797-818	
VAG-R	GCGTCTAAATAATTCCAAGG	$vacA^{\mathrm{f}}$	-	2620-2640	
	RAPD-PCR primers ^e				
1254	CCGCAGCCAA				
D11344	AGTGAATTCGCGGTGAGATGCCA				
1281	AACGCGCAAC				
1283	GCGATCCCCA				
D8635	GAGCGGCCAAAGGGAGCAGAC				
	Oligonucleotides and primers used in	PCR-Subtractive Hyb	oridisation ^h		
Adaptor-1	CTAATACGACTCACTATAGGGCTCGAC				
Adaptor-2R	CTAATACGACTCACTATAGGGCAGCG				
PCR Primer-1	CTAATACGACTCACTATAGGGC				
Nested Primer-1	TCGAGCGGCCGCCCGGGCAGGT				
Nested Primer-2R	AGCGTGGTCGCGGCCGAGGT				
aRefers to the gene to	which the primer hinds and is shown as gene	name or locus number	from H mile	wi strains 26605 an	4 10

U05676 U05676

^aRefers to the gene to which the primer binds and is shown as gene name or locus number from *H. pylori* strains 26695 and J99.

^bBase pair position of oligonucleotide with reference to given GenBank sequence accession number.

[°]Served as forward and reverse primers depending on which of primers 1260, 1262, 211 and 213 they were used with (see Fig. 5.5).

^dFrom Yamaoka et al. (1998).

^eFrom Atherton et al. (1995).

From Atherton et al. (1999).

^gPrimer nomenclature as used by Akopyanz et al. (1992a)

^hFrom Clontech PCR-Select Bacterial Genome Subtraction Kit User Manual (1997)

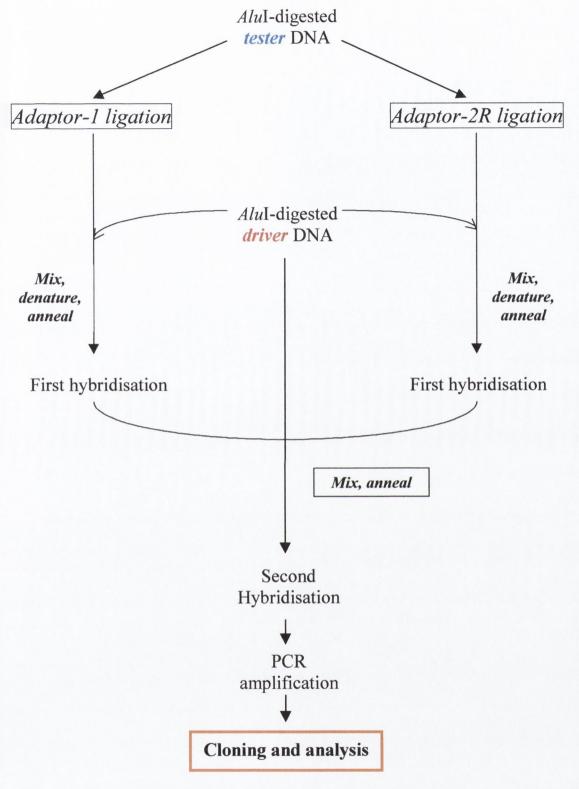


Fig. 2.1. Flow diagram showing an overview of the PCR-subtractive hybridisation procedure. Copied from Clontech PCR-Select Bacterial Genome Subtraction Kit User Manual (1997).

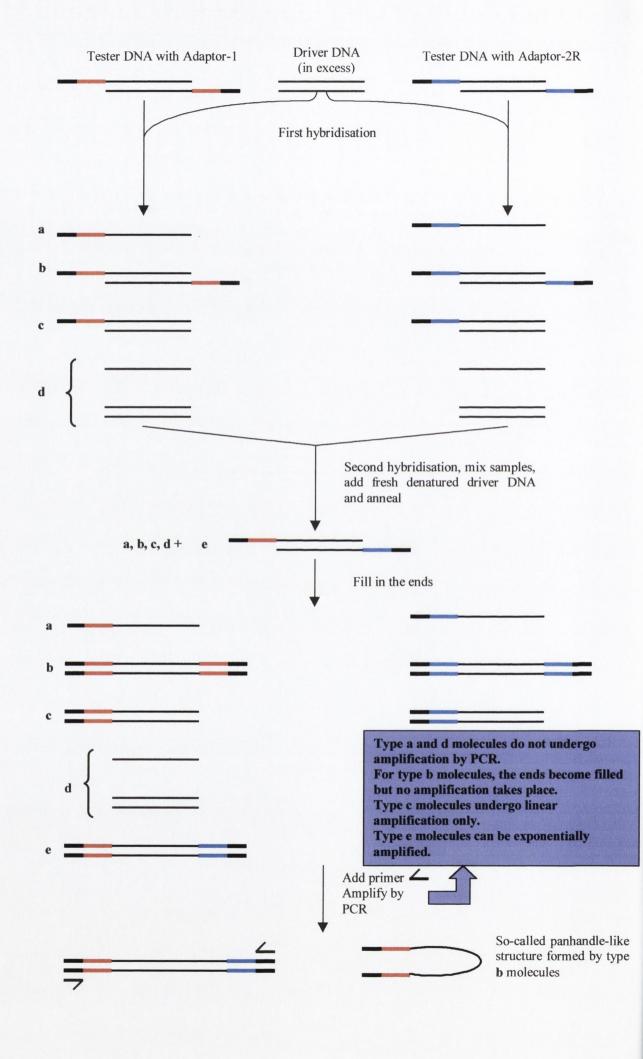


Fig. 2.2. Schematic flow diagram outlining in greater detail the molecular events that take place during PCR-subtractive hybridisation.

- ❖ After DNA has been isolated from the two strains being compared and digested with *Alu*1, the tester DNA is divided into two pools followed by the ligation of each one of these with a different adaptor (adaptor-1 or adaptor-2R).
- The first hybridisation step is then carried out as outlined in text generating type a, b, c and d molecules in each sample. The single stranded type a molecules represent tester-specific sequences that have become enriched in the mixture. DNA fragments that are not tester specific form type c molecules with the corresponding sequences from the driver population.
- During the second hybridisation, mixing of the two primary hybridisation samples without denaturation takes place ensuring that only subtracted single stranded tester DNA (type a) molecules from both mixtures can reassociate and form new type e hybrids. The type e molecules represent double stranded tester-specific sequences with different ends. An aliquot of freshly denatured driver DNA is added in order to further enrich for tester-specific sequences.
- ❖ *Taq* polymerase is added to fill in the ends. The type *e* molecules now have different primer annealing sites on each of their 5′- and 3′- ends.
- ❖ PCR is then carried out to amplify the tester-specific sequences. As outlined in the shaded box, type a and d molecules cannot be amplified because they are missing primer annealing sites. Type b molecules are not amplified because they form a panhandle-like structure upon denaturation during PCR because of the same adaptor sequences present on both ends of the fragment. Type c molecules can only be amplified linearly because they have only one primer annealing site. It is only type e molecules that can undergo exponential amplification because they have different adaptors at either end. A second round of PCR is carried out using nested primers which further reduces any background PCR products and enriches for tester-specific sequences.

[Fig. 2.2. and legend copied with modification from the Clontech PCR-Select Bacterial Genome Subtraction Kit User Manual (1997)].

2.9.2. Restriction enzyme digestion.

The two DNA populations (tester DNA, derived from the strain in which unique DNA sequences are being sought, and driver DNA, derived from the strain whose genome is being subtracted from the tester pool) were digested with the restriction endonuclease AluI (Roche Molecular Biochemicals). This restriction endonuclease generates short, blunt-ended restriction fragments necessary for optimal subtraction and adaptor oligonucleotide ligation. The reaction was carried out in a volume of 50 μ l using 2 μ g genomic DNA and 15 units of the endonuclease. Reaction components were mixed by vortexing briefly followed by centrifugation at 11,000 g for 10 s. Incubation at 37 °C for 16 h then followed.

2.9.3. Clean-up of restriction endonuclease reaction and preparation of DNA for adaptor ligation.

Gel electrophoresis of a 5-µl aliquot of the digestion reaction mixture was used to assess the efficiency of restriction cleavage of the DNA. Of 0.2 M EDTA, 2.5 µl was added to terminate the restriction reaction. To ease pipetting during organic extraction, 300 µl sterile distilled H₂O was added to the restriction reaction mixture bringing the total volume in the tube to ~350 µl. An equal volume of phenol-chloroform (1:1) (350 µl) was added and the tube vortexed vigorously. The phases were separated by centrifugation at 15,000 r.p.m. for 10 min. The upper aqueous phase was removed to a fresh tube by careful pipetting of small aliquots. An equal volume (~350 µl) of chloroform was added followed by mixing and centrifugation as before. After removal of the aqueous phase the DNA was precipitated by the addition of 0.5 volume of 4 M ammonium acetate and 2.5 volumes (of the total resulting volume) of 95% (v/v) ethanol. Centrifugation of the tubes for 20 min at 11,000 g followed thorough vortexing. The pellet was often barely visible so extreme care was exercised in the removal of the supernatant fraction. The pellet was rinsed in 200 µl 80% ethanol and centrifuged at 11,000 g for 5 min. Again, the supernatant fraction was removed carefully and the pellet allowed to air dry. Pellets of AluI-cleaved tester and driver DNA pools were dissolved in 6.5 µl sterile distilled H₂O to give an approximate final concentration of 300 ng/ul.

2.9.4. Ligation of adaptor oligonucleotides to restriction endonuclease-cleaved tester DNA.

Before adaptor oligonucleotides can be ligated to cleaved tester genomic DNA it is necessary to dilute the pool of tester restriction fragments in order that, when driver DNA is being

added during the subtractive hybridisation step, it exists in significant excess to the tester molecules. Thus, 1.2 μl *Alu*I-cleaved tester DNA was diluted with 1.8 μl sterile distilled H₂0. This diluted tester DNA was then ready for ligation to each of the two sets of adaptor oligonucleotides (Adaptor-1 and Adaptor-2R, Table 2.2). Ligation reactions were carried out in ε final volume of 10 μl consisting of the following components: 4 μl sterile H₂O, 2 μl 5 × ligation buffer (supplied with the enzyme), 1 μl T4-DNA ligase (400 units/μl) (New England Biclabs, MA, USA), 1 μl diluted tester DNA and 2 μl of either Adaptor-1 or Adaptor-2R (from 10 μM stocks). The tubes were vortexed, centrifuged briefly and incubated at 16 °C overnight. The ligation reaction incubation was usually performed in the block of a thermal cycler held at 16 °C in order to maintain a constant uniform temperature during the important adaptor ligation step. After overnight incubation the reactions were stopped by the addition of 1 μl 0.2 M EDTA and heating at 72 °C for 5 min to inactivate the ligase.

2.9.5. First hybridisation.

This step involved the addition of an excess of driver DNA to the Adaptor-1- and Adaptor-2R-ligated tester DNAs. The samples were then heat-denatured and allowed to anneal. Two PCR-tubes were set up containing the following reagents: 2 µl *Alu*I-digested driver DNA (Section 2.9.3) in both tubes, 1 µl Adaptor-1-ligated tester to one tube and 1 µl Adaptor-2R-ligated tester to the second tube, 1 µl 4 × hybridisation buffer [200 mM HEPES-HCI (pH 8.0), 2 M NaCl, 0.8 mM EDTA (pH 8.0)] to both tubes. A drop of mineral oil was added to both samples and the tubes centrifuged briefly. Denaturation and annealing were carried out in a thermal cycler by heating at 98 °C for 1.5 min. The temperature was then lowered to 65 °C and the tubes held at this temperature for 1.5 h.

2.9.6. Second hybridisation.

During this step, the two samples from the first hybridisation are mixed together and combined with an additional aliquot of freshly denatured driver DNA. In doing so, the subtracted sample is further enriched for tester-specific sequences and, importantly, hybrid tester-specific molecules with different adaptors on each end are formed. These new hybrids correspond to the tester-specific DNAs that remain after removal of all the sequences that are common between the tester and driver DNA populations. The significance of the presence on each end of the tester-specific hybrids of different adaptors lies in the PCR amplification

steps when the different adaptors enhance the specificity of amplification by serving as binding sites for nested primers.

In order to denature the driver DNA, the $4 \times$ hybridisation buffer was diluted with an equal volume of distilled H₂O to make 2 × hybridisation buffer. Of the 2 × hybridisation buffer, 1 μl was added to 1 μl driver DNA (again from Section 2.9.3), the sample overlaid with mineral oil and then denatured by heating in a thermal cycler at 98 °C for 1.5 min. The samples from the first hybridisation were removed to a heating block set at 65 °C prior to the denaturation of the driver DNA. Next, the freshly denatured aliquot of driver DNA was mixed with each of the two hybridisation samples from the first hybridisation step. It is imperative that the two hybridisation samples only mix together in the presence of driver DNA. This was achieved as follows. A Gilson P200 (Gilson Pipetman®, WI, USA) pipette was set to pipette a small volume, e.g., 40 µl. The first hybridisation sample was pipetted into the tip and the tip removed from the tube. A small amount of air was drawn into the pipette tip to create an air space below the sample. This step is repeated with the tube containing the freshly denatured driver DNA, now separated from the first hybridisation sample by a pocket of air. Finally, the entire mixture was transferred to the tube containing the second hybridisation sample and mixed by pipetting. This tube, containing all of the components for the next cycle of annealing, was heated at 65 °C for 14 h. A 200-µl aliquot of dilution buffer [20 mM HEPES-HCl (pH 8.3), 50 mM NaCl, 0.2 mM EDTA] was added to the now complete subtractive hybridisation reaction and the mixture heated in a thermal cycler at 65 °C for 7 min as a final step to eliminate non-specific hybridisation.

2.9.7. Primary PCR amplification.

The primary PCR step amplifies tester-specific DNA hybrids formed during the preceding second subtractive hybridisation step. Other single- and double-stranded molecular species will exist in the tube, e.g., single-stranded tester molecules with one adaptor ligated, single-stranded driver molecules, tester-driver hybrids and double-stranded tester molecules with the same adaptors at each end (Fig. 2.2). However, only double-stranded DNAs with different adaptors at each end are exponentially amplified during primary PCR.

Primary PCR was carried out in a final volume of 25 μ l which included the following components: sterile water, $10 \times PCR$ reaction buffer, 3.5 mM MgCl₂, 1 μ l 10 mM dNTP mix,

1 μl from a 10 μM stock of PCR Primer-1 (see Table 2.2. for sequence), 0.5 units *Taq* polymerase (Promega) (added after preincubation of the other reaction components at 75 °C for 1 min) and 1 μl of the diluted subtracted sample (see Section 2.9.6.) as template. The cycling parameters employed for primary PCR were 75 °C for 1 min (hot-start) followed by the addition of *Taq* DNA polymerase, 72 °C for 5 min (fills in the missing strands of the adaptors and creates binding sites for PCR Primer-1), 94 °C for 1 min and then 30 cycles of 94 °C for 30 s, 66 °C for 30 s and 72 °C for 1.5 min. When cycling was completed, 1 μl of each primary PCR mixture was diluted with 39 μl sterile H₂O. Of this diluted sample 1 μl was then used as template for secondary PCR.

2.9.8. Secondary PCR amplification.

This amplification step employs nested primers, which further reduce background and non-specific amplification, and enriches for tester-specific sequences. The sequences of the nested primers are given in Table 2.2. Similarly to primary PCR amplification, secondary PCR was carried out in a total volume of 25 µl with the same components as detailed above for primary amplification but with the inclusion of 1 µl of a 10 µM stock of each nested primer (nested Primer-1 and nested Primer-2R). The cycling parameters included the hot-start at 75 °C for 1 min as before after which *Taq* polymerase was added. This was followed by a 1 min incubation at 94 °C and 16 cycles of 94 °C for 30 s, 68 °C for 30 s and 72 °C for 1.5 min.

2.9.9. Analysis of amplification products from primary and secondary PCR reactions.

The success of PCR-subtractive hybridisation was assessed by electrophoresis of 7-µl aliquots from both the primary and secondary PCR amplifications on 2% (w/v) agarose gels. Primary PCR products generally appeared as a smear on the gel from 0.5–2 kb, occasionally with some distinct bands. The pool of amplicons that are generated by the secondary PCR should contain fewer bands than those obtained in primary PCR, with the prominent bands in primary PCR now more so than before in the secondary PCR.

2.9.10. Construction of a subtracted DNA library.

The amplicons obtained from secondary PCR with nested primers were cloned into the T-vector pCRII-TOPO (Invitrogen) followed by transformation by heat-shocking into chemically competent *E. coli* TOP10 cells provided with this vector. Colonies containing

plasmid DNA with the cloned insert were selected for by blue–white selection on L-agar containing ampicillin (50 μ g/ml) and 40 μ l X-Gal (40 mg/ml) per plate. White colonies were picked for further analysis.

2.9.11. Screening of subtracted DNA library.

For each set of paired isolates being analysed by PCR-subtractive hybridisation, 48 white colonies were picked initially for screening. Plasmid DNA was isolated using the small-scale alkaline lysis procedure as described in Section 2.6.2. The presence of insert DNA in the recovered plasmid was confirmed by re-amplification of the insert from the plasmid using nested primers and secondary PCR conditions. The resulting PCR products were electrophoresed in two identical gels and transferred by capillary blotting to nylon filters for Southern hybridisation (Section 2.7) using as probes DIG-labelled total genomic DNA from each of the two isolates in a particular subtractive hybridisation experiment. Of genomic DNA, 500 ng was DIG-labelled by random-priming using Klenow enzyme and DIG-dUTP (Roche Molecular Biochemicals). Hybridisation was carried out at 68 °C in standard hybridisation buffer (see Section 2.7.2). After stringency washing, further washing and signal detection was carried out as described under Section 2.7.2. Cloned amplicons identified using this approach as possibly being unique to one strain from a pair or partially different between the two were analysed further.

2.9.12. Labelling of amplified insert DNA and probing restricted genomic DNA.

The DNA of putative unique or partially different amplicons was excised from TAE-agarose gels using a clean scalpel blade. The gel slice containing the DNA was then subjected to 'clean-up' using a Geneclean kit (Anachem, Luton, UK) which removes the agarose from the DNA by melting the gel slice in a buffer containing sodium iodide. Silica-coated glass beads that bind the DNA, now in solution, were then added. After washing, the DNA was eluted from the glass beads into a small volume of H₂O. The extracted DNA was then labelled with DIG-dUTP by random-primed labelling as described above (Section 2.9.11). These probes were used in hybridisation with a Southern blot of *Hind*III-restricted genomic DNA from each strain in paired sets in order to examine directly the chromosomes of these pairs to determine if the presence of genomic fragments isolated by PCR-subtractive hybridisation differed between the two.

2.9.13. DNA sequencing.

Sequencing of clones containing DNA that showed evidence from Southern blot analysis as possibly having undergone rearrangement between paired isolates was carried out using the commercial sequencing service at MWG-Biotech (Milton Keynes, UK). Plasmid DNA for sequencing was isolated using the High Pure Plasmid Isolation Kit (Roche Molecular Biochemicals), typically from culture volumes of 3 ml which yield ~20 µg of plasmid DNA. *Taq* DNA polymerase was used to generate all the PCR products throughout these studies that were subject to DNA sequence analysis, and PCR products were routinely sequenced from a single strand. This practise was deemed adequate for the present studies given that gross genome rearrangements (as opposed to single base changes) were of interest for most of the investigations described save for those of Chapter 7. The potential limitations of this experimental strategy are discussed in relevant chapters. Sequences were identified using the BLAST algorithm at the NCBI and the *H. pylori* genome sequence databases of strain 26695 at TIGR [http://www.tigr.org/tdb/CMR/ghp/htmls/Splash-Page.html] and of strain J99 at AstraZeneca Pharmaceuticals [http://www.astra-boston.com/hpylori].

2.10. Protein analysis.

2.10.1. H. pylori cell lysate preparation.

H. pylori cells were harvested for cell lysate preparation from 48–72 h lawned plate cultures using sterile cotton swabs into 1 ml PBS. The cells were washed once in 1 ml PBS containing protease inhibitors (CompleteTM, Roche Molecular Biochemicals) followed by a final resuspension of the washed cells in 200 μl PBS with protease inhibitors. The turbidity of each suspension was adjusted to an OD_{600nm} of 4.0. For lysate preparation, equal volumes of adjusted cell suspension (typically 20 μl) and Laemmli (1970) buffer [40 mM Tris-HCl, pH 6.8, 20% (w/v) glycerol, 4% (w/v) SDS, 10% (v/v) β-mercaptoethanol, 0.025% (w/v) bromophenol blue] were mixed, boiled for 2 min, and centrifuged for 1 min at 11,000 g in a bench-top centrifuge to pellet large cellular debris. Lysates were then ready for application to the wells (20-μl aliquots) of SDS-PAGE gels for the separation of cellular protein molecules. Unused lysates were stored at –20 °C.

2.10.2. Proteinase K treatment of H. pylori for phenotypic analysis for the expression of Lewis^X and Lewis^Y on the lipopolysaccharide (LPS) molecule.

Treatment of *H. pylori* whole cell lysates with proteinase K was necessary prior to electrophoresis of the samples for Western blotting and probing with anti-Lewis^x and anti-Lewis^y monoclonal antibodies. Following the adjustment of the turbidity of each suspension to an OD_{600nm} of 4.0, the cells were solubilised in 200 μl of final sample buffer (Laemmli, 1970). Samples were boiled for 10 min and then allowed to cool. Of a stock proteinase-K solution (2 mg/ml), 40 μl was added to each sample and incubated at 60 °C for 1 h. The enzyme was then denatured by heating to 100 °C for 5 min (Preston and Penner, 1987).

2.10.3. Preparation and analysis of H. pylori outer membrane fractions.

The outer membranes of H. pylori were prepared by differential extraction with N-laurylsarcosine (SarkosylTM), using a modified procedure to those described by Nikaido (1994) as detailed in Deighan $et\ al.$ (2000). The extraction procedure is based on the ability of this detergent to disaggregate and solubilise proteins and other components of the cytoplasmic membrane in a bacterial lysate as distinct from the LPS-based outer membrane. Sarkosyl has a similar charge density per surface area as LPS, i.e., mimics LPS, and thus the outer membrane is not therefore disaggregated by this detergent.

H. pylori was cultured in liquid medium as described in Section 2.1.3. in 25 cm² cell culture flasks until maximum growth had been achieved. Typically, this took approximately 72 h. A Sarkosyl-insoluble fraction containing the outer membranes was prepared as follows. Cells were harvested by centrifugation and resuspended in 5 ml PBS. Cells from an aliquot of this suspension equivalent to 20 ml of $OD_{600mm} = 0.6$ were harvested, washed and resuspended in 0.6 ml ice-cold sonication buffer [10% (w/v) sucrose, 50 mM Tris-HCl pH 7.5, 100 mM NaCl, 1 mM EDTA, 5 mM dithiothreitol] and incubated with lysozyme (200 μg/ml) on ice for 20 min. Complete lysis was achieved by sonication and cellular debris cleared by centrifugation at ~5,000 g for 5 min. The supernatant fraction was then incubated with 0.5% (w/v) Sarkosyl for 30 min at room temperature. The insoluble outer membrane fraction was harvested by centrifugation at 16,000 g for 30 min and resuspended in 100 μl 1 × final sample buffer. Samples (15 μl) were boiled and then separated by SDS-PAGE (as described in the next section) at 170 V for 2.5 h. Gels were stained with GelCode Blue Stain Reagent (Pierce, IL, USA).

2.10.4. Sodium dodecyl-sulphate polyacrylamide gel electrohoresis (SDS-PAGE).

The discontinuous SDS system of Laemmli (1970) was utilised for SDS-PAGE employing the Bio-Rad "Mini PROTEAN II" casting plates and gel tank. Gels were electrophoresed at a constant voltage of 170 V for 1 h. Polypeptides were resolved through separating gels consisting of 10% or 12% (w/v) polyacrylamide (made from a 30% stock of acrylamide mix [29% (w/v) acrylamide, 1% (w/v) N,N'-methylenebisacrylamide]), 375 mM Tris-HCl (pH 8.8), 0.1% (w/v) SDS, 0.1% (w/v) ammonium persulphate and 0.04% (v/v) N,N,N',N'-Stacking gels, consisting of 5% (w/v) tetramethylethylenediamine (TEMED). polyacrylamide, 125 mM Tris-HCl (pH 6.8), 0.1% (w/v) SDS, 0.1% (w/v) ammonium persulphate and 0.001% (v/v) TEMED, were also employed in the electrophoresis of H. pylori cell lysates. Tris-glycine electrophoresis buffer [25 mM Tris, 250 mM glycine (pH 8.3), 0.1% (w/v) SDS] was added to the gel tank before the samples were applied to the wells. The proteins separated in the gel were stained with Coomassie Brilliant Blue [250 mg Coomassie brilliant blue, 90 ml methanol:H₂O (1:1 v/v), 10 ml glacial acetic acid] for a minimum of 4 h. Excess dye was removed from the gel by soaking in destain solution (methanol-acetic acid solution described above but omitting the dye) for several hours with frequent changing of the destain solution.

2.10.5. Transfer of separated proteins to nitrocellulose membranes.

When the polypeptides separated by SDS-PAGE were to be analysed by Western immunoblotting with antiserum, the electrophoresis apparatus was disassembled and the separating gel immersed in transfer buffer [25 mM Tris-HCl (pH 8.3), 192 mM glycine, 20% (v/v) methanol] to equilibrate for 30 min with gentle shaking. The separated polypeptides were transferred electrophoretically onto nitrocellulose membranes (BA85, 0.45-µm pore size, Schleicher and Schuell) at a constant voltage of 19 V for 1 h using a semi-dry transfer cell (Bio-Rad). After transfer was completed, any proteins remaining in the gel were stained with Coomassie brilliant blue.

2.10.6. Western immunoblotting.

Following immobilisation of the separated proteins on nitrocellulose filters, the filters were first blocked and then probed with either polyclonal serum or monoclonal antibodies to the proteins of interest. Blocking typically took place for 1 h at room temperature, or at 4 °C overnight, in blocking solution consisting of 5% (w/v) dried milk (Marvel, Chivers Ireland

Ltd., Dublin), 0.15% (v/v) Tween 20 in PBS with gentle shaking. Blocked filters were washed for 15 min in PBS containing 0.1% Tween 20. Incubation with primary antibody then followed. The immunoglobulins were either murine monoclonal anti-Lewis^x or anti-Lewis^x antibodies (Signet laboratories, MA, USA), or polyclonal serum raised against an electroeluted 120-kDa CagA protein (Marshall, 1996). The anti-CagA primary antibody was diluted 1:500 in a solution of 0.5% (w/v) Marvel in PBS, and the anti-Lewis^x and anti-Lewis^x monoclonal antibodies were diluted 1:1000 in a 1.0% (w/v) blocking solution.

Primary antibodies were incubated with the membrane for 1 h at room temperature (generally carried out in a sealed bag with all air bubbles removed for close contact between the diluted antibody and the filter). The antibody solution was decanted after 1 h and the filter washed for a total of 30 min with approximately six changes of PBS. Secondary antibodies conjugated to horse-radish peroxidase were diluted 1:1000 in 0.5% (w/v) Marvel in PBS (rabbit anti-mouse IgG as secondary to the murine monoclonal antibodies, or goat anti-rabbit IgG as secondary to the anti-CagA serum, both purchased from Dako Ltd., Cambridgeshire, UK). Incubation of the filter with secondary antibody was carried out as described for the primary antibody incubation. Washing of the filter after incubation with the secondary antibody took place for 2×5 min in PBS followed by one 15-min wash in PBS containing 0.15% (v/v) Tween 20 and was completed with 3×5 min washes in PBS. Detection of the chemiluminescent signal was carried out with the WestPico chemiluminescent substrate (Pierce) on filters exposed to X-Ray film.

2.11. Colony blotting of *H. pylori* with anti-Le^x and anti-Le^y monoclonal antibodies.

When *H. pylori* was to be cultured on solid medium in order to carry out colony blotting, a single colony of a strain of interest was picked from a plate and resuspended in 20 µl sterile PBS. This was pipetted onto the surface of a new culture plate and lawned with a sterile spreader. The plates were incubated for 3–4 days in order to allow colonies of a reasonable size to form. After this time, the colonies were transferred from the surface of the plate to a nitrocellulose filter by pressing the filter to the plate. The filter was sealed and put at –20 °C overnight to allow the colonies to stick. As described above (Section 2.10.6), the filters were blocked, washed and incubated with primary (anti-Lewis^x and anti-Lewis^y) and secondary (rabbit anti-mouse IgG) antibodies as was carried out for Western immunoblotting. These filters however, were not sealed in a bag for antibody incubation steps as this tended to

detach the colonies from the surface of the membrane. For similar reasons, washing steps were carried out extremely gently. The binding of anti-Lewis^x and anti-Lewis^y antibodies to the *H. pylori* cell surface was detected colourimetrically as follows: after the final washing step the blot was placed in H_2O and then transferred to 18 ml 50 mM Tris, pH 7.4–8.0 to which was added 2 ml of 4-chloro-1-naphthol (Sigma-Aldrich Ireland Ltd., Tallaght, Dublin, Ireland) (from a 6 mg/ml stock prepared in methanol and stored at -20 °C) followed by 10 μ l H_2O_2 . Colour development was allowed to take place and was stopped after sufficient time by rinsing the blot in H_2O .

2.12. Fingerprinting of *H. pylori* isolates.

2.12.1. Oligofingerprinting.

H. pylori isolates were analysed by the oligofingerprinting technique described for this organism by Marshall et al. (1996). H. pylori genomic DNA was prepared for hybridisation to oligonucleotide probes by digestion of approximately 10 µg DNA to completion with the restriction endonuclease HindIII (Roche Molecular Biochemicals) for 16 h using the conditions specified by the manufacturer. The restricted DNA was electrophoresed through 0.9% (w/v) TAE agarose gels in a Horizon 20.25 gel tank (Gibco BRL, Life Technologies, Germany) followed by Southern transfer to nylon membranes. Oligonucleotides (GGAT)₄, (GACA)₄ and (GTG)₅ (MWG-Biotech, Milton Keynes, UK) were labelled at their 3'-ends with DIG using a DIG oligonucleotide 3'-end labelling kit (Roche Molecular Biochemicals). Nylon filters containing transferred restricted H. pylori genomic DNA were prehybridised in 'High-SDS buffer' [7% (w/v) SDS, 50% (v/v) formamide, 5 × SSC, 2% (w/v) blocking reagent, 50 mM sodium phosphate buffer (pH 7.0), 0.1% (w/v) N-lauroylsarcosine] or DIG-Easy Hyb (Roche Molecular Biochemicals) for 2 h at 45 °C. Labelled oligonucleotide probes were then diluted in fresh prehybridisation solution and hybridisation was continued for approximately 16 h. Detection of the chemiluminescent signal from the hybridised oligonucleotide probe was then carried out by washing the filter twice for 10 min each in 6 × SSC-0.1% (w/v) SDS followed by a further two washes for 20 min each with 2 × SSC-0.1% (w/v) SDS. All stringency washings took place at room temperature. Remaining washing and detection was carried out according to instructions in the DIG System User's Guide for Filter Hybridisation (1995) using the CDP-Star chemiluminescent substrate.

2.12.2. Random amplified polymorphic DNA (RAPD)-PCR fingerprinting.

H. pylori isolates were fingerprinted by RAPD-PCR (Akopyanz et al., 1992a) using various pair-wise combinations of RAPD primers, the sequences of which are given in Table 2.2. RAPD-PCR reactions took place in a final volume of 25 μl that included 10 × reaction buffer, 3 mM MgCl₂, 10 pmol each primer, 0.4 mM each of dATP, dGTP, dCTP, dTTP, 2.5 units of Taq DNA polymerase (Promega) and 1 μl genomic DNA as template (~50–100 ng/μl). The cycling parameters routinely employed were as follows: 4 cycles of 94 °C for 5 min, 40 °C for 5 min, 72 °C for 5 min (low stringency synthesis steps) followed immediately by amplification at higher stringency for 30 cycles of 94 °C for 1 min, 55 °C for 1 min, 72 °C for 2 min. A final 10-min extension at 72 °C was carried out to finish incomplete elongation of products. RAPD-PCR products were electrophoresed through 1.8% (w/v) agarose gels containing ethidium bromide (0.5 μg/ml) and fingerprints visualised by under ultra-violet (302 nm) light by exposure of the gel on a Chromato-Vue UV transilluminator, (UVP Inc., CA, USA).

2.12.3. Similarity analysis of oligofingerprint profiles.

The degree of relatedness exhibited between familial *H. pylori* isolates (carried out in Chapter 6) was manually assessed by the pair-wise analysis of oligofingerprint profiles obtained after hybridisation of *Hin*dIII-cleaved total genomic DNA with the oligonucleotide (GTG)₅, and the application of a simple coefficient of similarity (S_D) (Dice, 1945) calculated with the following formula:

 $S_D = 2 \times \text{(number of matched bands between profiles)} / \text{(total number of bands in both profiles)}$

The % similarity (% S_D) is $100 \times S_D$.

Where a $%S_D$ of $\geq 90\%$ was determined between epidemiologically related isolates these strains were deemed to be clonal. This 10% variation in $%S_D$ was allowed in order to take into account band variability between fingerprint profiles that is likely to have arisen because of microevolutionary changes in the genomes of these strains i.e., *de novo* mutation or gross genome rearrangements.

Chapter 3

Diversity in the 3'-region of the cagA gene in paired H. pylori isolates - genotypic and phenotypic analysis

3.1.	Introd	uction.	68
3.2.	Result	s	72
	<i>3.2.1. 3.2.2.</i>	Genotypic analysis of paired H. pylori isolates using a cagA gene probe Confirmation that observed putative rearrangements did not result	72
	3.2.2.	from point mutations occurring within the cagA gene	72
	3.2.3.	Fingerprinting of paired isolates with cagA gene polymorphisms	72
	<i>3.2.4.</i>		12
	5.2.7.	rearrangement that had taken place in the cagA gene	73
	3.2.5.	DNA sequence determination of PCR products.	74
		Analysis of paired and individual H. pylori isolates using primers	
	2.2.0.	Cag1 and Cag2	75
	3.2.7.	Multiple products obtained by PCR using primers Cag1 and Cag2	
		result from a heterogeneous culture	75
	3.2.8.	Analysis by Western immunoblotting of the sizes of the CagA	
		protein expressed by paired isolates.	76
	3.2.9.	Is the expression of CagA growth-phase dependent?	77
3.3.	Discus		
	3.3.1.	The cagA gene exhibits diversity at its 3'-end	78
	3.3.2.	Variability between two paired strains in the 5'-region of the cagA gene	
	3.3.3.		
	3.3.4.	Analysis of the 3'-region of the cagA gene from six Irish isolates	80
	3.3.5.		
		between Irish and Italian strains	81
	3.3.6.	Prevalence of cagA 3'-end repeat copy number variation in paired Irish	
		isolates.	82
	3.3.7.	Prevalence of particular cagA subtypes in the Irish collection	83
	3.3.8.	Strain MI 562 – the cagA gene and CagA protein	83
	3.3.9.	Homologous recombination as a putative mechanism of rearrangement	
		at the cagA gene 3'-end	84
	3.3.10). Model to explain the structural subgenotypes of the cagA gene found,	
		bearing in mind a putative homologous recombination mechanism	
		of rearrangement	85
	3.3.11	. Analysis of paired H. pylori isolates by Western blotting with	
		polyclonal anti-CagA serum.	86
		P. Could CagA be a phase-variable antigen?	88
	3.3.13	3. The deduced amino acid sequence of the CagA antigen of isolate MI 562	
		evidences the occurrence of frameshifting	89
	3.3.14	! Significance of rearrangement at the 3'-end of the cagA gene	89

3.1. Introduction.

The recognition of *H. pylori* as an important pathogen of humans prompted an immediate interest in the possible disease-causing factors possessed by this unusual bacterium. How was *H. pylori* equipped to establish colonisation in the stomach environment and indeed persist and cause disease there? When considering the issue of the pathogenesis and virulence of an organism, one enters complex terrain in which an elaborate interplay of factors, co-ordinately and temporally regulated, combine to elicit disease symptoms, not discounting the contribution of host and environmental influences. Virulence, therefore, is multifactorial and rarely completely understood even for some of the most well characterised pathogens. The diversity present at many levels in *H. pylori* further complicates a complete definition of virulence for this organism. The likely outcome of infection with a strain of a particular genotype is accordingly difficult to predict, and especially so when one considers the diversity that can exist in what is perhaps its single most important virulence determinant, the *cag* locus. Diversity between isolates at this locus is present at genotypic, phenotypic and structural levels and is responsible for influencing directly the severity of *H. pylori*-related gastroduodenal disease.

A highly immunogenic, surface-expressed protein was identified in broth culture supernatants of cytotoxic *H. pylori* strains (Cover *et al.*, 1990; Covacci *et al.*, 1993; Tummuru *et al.*, 1993). The protein was designated CagA (cytotoxin-associated gene A) to account for the apparent link that existed between cytotoxicity and the expression of this protein in most of the isolates examined (Covacci *et al.*, 1993). Indeed, antibody titres against CagA correlated with disease severity, thereby strongly implicating the protein in pathogenesis (Covacci *et al.*, 1993; Xiang *et al.*, 1993). VacA and CagA appeared to be coexpressed and it seemed that the two proteins had something of an inextricable role in *H. pylori*-related gastric pathology (Xiang *et al.*, 1995).

While the role of the VacA cytotoxin in disease aetiology was fairly clear-cut (Phadnis *et al.*, 1994; Schmitt and Haas, 1994), the scenario for CagA was apparently more complex (Ghiara *et al.*, 1995). Not all isolates carry the *cagA* gene (Xiang *et al.*, 1995; Cover *et al.*, 1995). In most developed countries, 70–80% of people infected with *H. pylori* carry strains that express CagA (Weel *et al.*, 1996) and 88–100% of these individuals will have presented with peptic ulcer disease. This compares with the presence of the allele in only 50–60% of isolates from patients with superficial gastritis alone. Isolates can be grouped into two broad

types based on the expression of VacA and CagA proteins. Isolates that produce the CagA antigen and the cytotoxin are classified as Type I strains (CagA+/VacA+). Bacteria of Type II are genotypically negative for the *cagA* gene and in addition do not express the cytotoxin (*cagA*-/VacA-). Type I strains were typically found in patients suffering with more severe disease symptoms such as gastroduodenal ulceration and gastric adenocarcinoma (Xiang *et al.*, 1995). Thus, the CagA antigen became a marker for strains of more enhanced virulence.

Type I and Type II strains were found to differ markedly in the chromosomal region upstream of the *cagA* gene. Not only are Type II strains devoid of the *cagA* allele, but they also lack an entire 40-kb segment of DNA, which contains up to 31 genes, that is present in Type I strains as a chromosomal insertion in the glutamate racemase gene. The '*cag*-locus' is commonly known as the *cag*-pathogenicity island (*cagPAI*) with the *cagA* gene located at its right-most extremity (Censini *et al.*, 1996; Akopyants *et al.*, 1998a).

The structure of the *cag*PAI has also been shown to vary between isolates. It has been found variously as either a single uninterrupted unit, split into two regions (Cag I and Cag II) by a large piece of chromosomal DNA or the novel insertion sequence IS605, or partially deleted (Censini *et al.*, 1996; Akopyants *et al.*, 1998a; Occhialini *et al.*, 2001). As well as being a marker for strains which were likely to be involved in causing more severe pathology, the *cagA* gene was also a marker for the entire *cagPAI* region present in isolates of Type I. *cagPAI*s which were partially deleted in structure were observed in one study to be associated with milder inflammatory infiltration in the antrum of the stomach and milder neutrophil infiltration in the corpus and antrum than occurred with strains with fully intact pathogenicity islands (Ikenoue *et al.*, 2001). However, it is generally perceived that the structure of the *cagPAI* has little or no modulatory impact on the outcome of *H. pylori*-related disease and that PAIs that are complete or interrupted have no difference in their ability to induce interleukin-8 (IL-8) secretion from host gastric epithelial cells (Audibert *et al.*, 2001).

What of the functions of both the CagA antigen and the genes of the *cag*PAI? Evidence suggests that the role of the proteins encoded by the genes that comprise the *cag*PAI are of pivotal importance in interactions with host cells. The ability of CagA⁺ strains but not *cagA*⁻ *H. pylori* strains to induce the secretion of IL-8 from gastric epithelial cell lines was described by Crabtree *et al.* (1994, 1995a). Molecular dissection of the *cag*PAI by Censini *et*

al. (1996) employing transposon mutagenesis was carried out to establish if the genes of the cag locus were responsible for the induction of this proinflammatory mediator. Their analysis revealed that individual alleles from a cluster in the right-most part of the PAI (Cag I region) were responsible for IL-8 induction but that the cagA gene itself had no involvement in this process. Crabtree et al. (1995b) presented similar findings. The CagA protein has been found to be exported from the bacterial cell and inserted into host gastric epithelial cells whereupon it becomes phosphorylated on a tyrosine residue. The cagPAI was believed, based on sequence similarities, to encode a Helicobacter-specific type IV secretion apparatus that was involved in the secretion of IL-8 from host cells. Overwhelming evidence exists to support the export of CagA using the secretion machinery encoded by the cagPAI (Odenbreit et al., 2000; Backert et al., 2000; Segal et al., 1999; Stein et al., 2000).

Thus, the *cag*PAI, through its ability to induce the secretion of mediators of inflammation, was seen as being of major importance in the pathogenesis of *H. pylori*-related disease. Factors which heighten gastric inflammation are likely to increase the risk of severe disease outcomes such as gastroduodenal ulcer and gastric adenocarcinoma. In addition, the secretion of CagA from the bacterial cell and its transport into cells of the host's gastric epithelium is likely to play an as yet undefined role in interactions of the bacterium with the host. In this new location, CagA is likely to have a marked impact on pathogenesis through its insertion into and subversion of eukaryotic signal transduction pathways. The *cag*PAI is, therefore, based on these two counts, considered to be the single most important delineating factor in determining disease outcome for a particular strain.

H. pylori is an organism characterised by heterogeneity and diversity. The ancestral inheritance of the cagPAI modified the virulence of H. pylori by marking the differentiation of a more virulent type of bacterium within the Helicobacter genus (Censini et al., 1996; Covacci et al., 1999). Yet, further layering of complexity within the cag locus in structural and genotypic terms is likely to have potential modulatory effects on the virulence of a strain. Such is the number of factors which act in concert to bring about a particular pathology, with CagA acting as a major player in this process, it seems plausible that diversities between strains within the cag locus could have subtle but as yet latent influences on virulence modulation.

Another layer of complexity is added to the spectrum when one considers the inherent diversity of the cagA gene. The CagA protein varies in size between isolates and has been found as a molecule of between 120 and 140 kDa. In the characterisation of the cagA gene that was carried out by Covacci et al. (1993), it was found that the size variation in the protein was attributable to the existence of cagA genes of variable size because of duplications of regions within the gene. These regions consisted of repeat sequences and were located at the 3'-end of the gene. Indeed, Yamaoka et al. (1998) identified four types of gene structure for cagA in H. pylori isolates depending on the type and number of repeats present. Thus, the cagA gene can undergo recombination, with heterogeneity existing in the cagA gene sequences of strains recovered from different individuals. It was also observed that isolates with variable 3'-region structure elicited different titres of anti-CagA antibodies, suggesting that immune responses of variable strength and duration depend on the structure of the 3'-region of the cagA gene. This factor is likely to be of considerable importance when one considers the key role played by the inflammatory response of the host in the development of gastric pathology. Yamaoka et al. (1999a) also observed, where isolates had more than three repeat regions in the 3'-end of the gene, that such strains were associated with enhanced histological injury and with reduced survival in acidic conditions.

In the present study, epidemiologically related *H. pylori* isolates, which had highly similar fingerprint profiles, have been analysed for the possible occurrence of rearrangements within the *cagA* gene. The occurrence of microevolution between the strains in these isolate pairs had been observed using the discriminatory oligofingerprinting typing method (Marshall, 1996). This was postulated to be a mechanism whereby the genomes of these *H. pylori* isolates had undergone host-specific adaptation. It was of interest to ascertain if rearrangement within the *cagA* gene might be involved in the generation of microevolution in *H. pylori*, possibly as a reflection of selection pressure being applied to the CagA antigen because of its high immunogenicity. Analysis of *H. pylori* isolates of proven clonality such as these allows a direct appraisal of the propensity of the *cagA* gene to undergo rearrangement *in vivo* during the lengthy period of gastric colonisation.

3.2. Results.

3.2.1. Genotypic analysis of paired H. pylori isolates using a cagA gene probe.

A panel of nine paired *H. pylori* isolates from the Moyne Institute collection was analysed initially by probing Southern blots of *Hin*dIII-restricted genomic DNA with each of two *cagA* gene probes. The probes used were the entire *cagA* gene from the clone described by Covacci *et al.* (1993), and the shorter 1.2 kb-*Hin*dIII fragment from the pMC3 plasmid into which a *cagA* gene had been cloned (Tummuru *et al.*, 1993). The latter probe was used because of the presence on the Covacci clone of additional sequences 5' and 3' to the *cagA*-coding sequence. The patterns of restriction fragments hybridising to these probes were seen to differ between two of these paired isolate sets (MI 512/MI 541 and MI 561/MI 562) with both probes, suggesting that rearrangement had taken place within the *cagA* gene during infection (Fig. 3.1). Using these probes on additional isolates, namely MI 517/MI 566 and MI 683A/MI 683C, revealed the same dissimilar pattern of fragments hybridising to the probes between these related isolates when genomic DNA was restricted with *Hin*dIII (Fig. 3.2.A and Fig. 3.2.B I).

3.2.2. Confirmation that observed putative rearrangements did not result from point mutations occurring within the cagA gene.

To confirm that what were perceived as rearrangements within the *cagA* gene because of the pattern of probe hybridisation described in Section 3.2.1 were indeed such and not due to point mutations occurring within this gene, the genomic DNA from these isolates was also restricted with the endonucleases *EcoRI* and *HaeIII*. Following electrophoresis of the restricted DNA and Southern blotting, the 1.2 kb-*HindIII* fragment from pMC3 was used to probe the nylon membrane. In the case of isolate pairs MI 512/MI 541, MI 561/MI 562 and MI 683A/MI 683C, the patterns of restriction fragments hybridising to the probe were also seen to differ when their DNAs were cleaved with *EcoRI* (Fig. 3.2.B II and Fig. 3.3.A II and B II) and similarly for isolate pair MI 683A/MI 683C when the enzyme *HaeIII* was used to restrict the DNA (Fig. 3.2.B III). These results were taken as evidence that *bona fide* rearrangement and not point mutations had taken place within the *cagA* gene between related isolates.

3.2.3. Fingerprinting of paired isolates with cagA gene polymorphisms.

Although it had been determined that rearrangement had taken place within the cagA gene of related isolates (i.e., isolates obtained from the same individual), it was necessary to

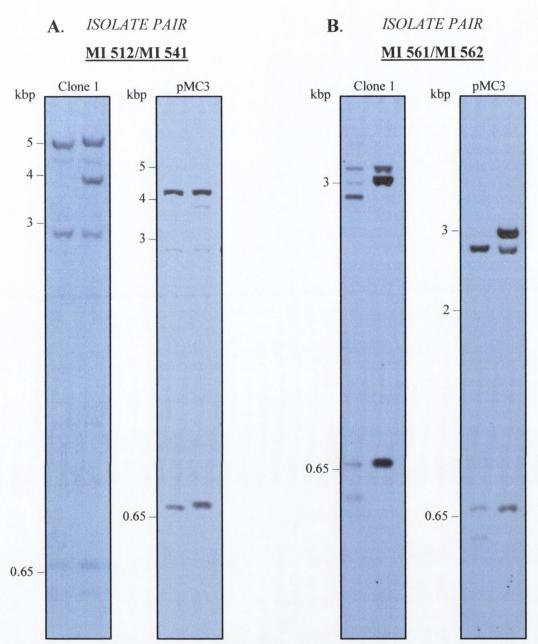


Fig. 3.1. Southern analysis of *Hin*dIII-digested genomic DNA from isolate pairs MI 512/MI 541 and MI 561/MI 562 with *cagA* gene probes (clone 1 and pMC3).

(A) HindIII-digested DNA from isolate MI 512 was electrophoresed in the left lane of each individual panel and that from isolate MI 541 was electrophoresed in the right lane of each panel. (B) HindIII-digested DNA from isolate MI 561 was electrophoresed in the left lane of each panel and that from isolate MI 562 was electrophoresed in the right lanes. The probe clone 1 refers to the cloned cagA gene obtained from Covacci and co-workers (1993) that also contains DNA sequence 5' and 3' to the coding region of cagA. The probe pMC3 refers to the 1.2-kb HindIII fragment from the cagA gene cloned by Tummuru et al. (1993). The positions of DNA molecular size standards (kbp) are indicated to the left of each panel.

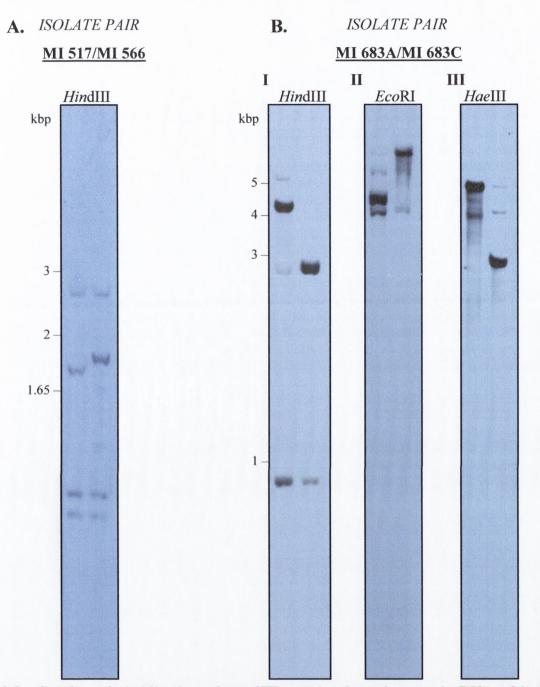


Fig. 3.2. Southern hybridisation of *HindIII*-restricted total genomic DNA of isolate pair MI 517/MI 566 probed with the clone 1 *cagA* probe and of *HindIII*, *EcoRI* and *HaeIII*-restricted genomic DNA of isolate pair MI 683A/MI 683C probed with the pMC3 *cagA* probe.

(A) Restricted DNAs from isolates MI 517 and MI 566 were electrophoresed in the left and right lanes of this panel, respectively. (B) Restricted DNAs from isolates MI 683A and MI 683C were electrophoresed in the left and right lanes of each panel, respectively. The enzymes used to restrict the genomic DNA were (I) *HindIII*, (II) *Eco*RI and (III) *HaeIII* and are indicated at the top of each image. The positions of DNA molecular size standards (kbp) are shown to the left in panels A and B.

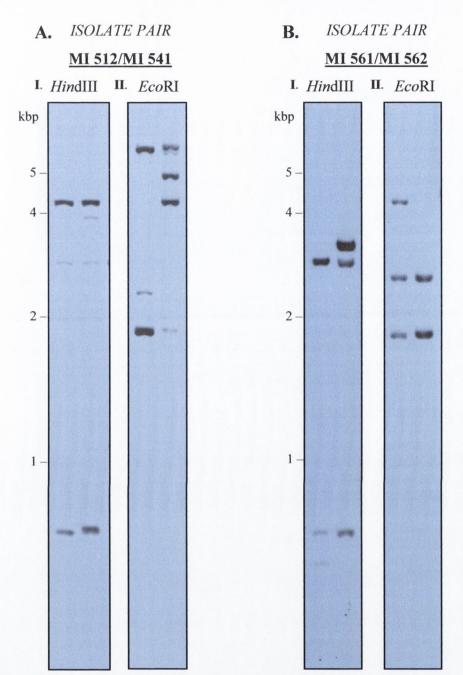


Fig. 3.3. Southern hybridisation analysis of *Hin*dIII- and *Eco*RI-restricted total genomic DNA from isolate pairs MI 512/MI 541 and MI 561/MI 562 probed with the pMC3 *cagA* gene probe.

(A) Restricted DNAs from isolates MI 512 and MI 541 were electrophoresed in the left and right lanes of each panel, respectively. (B) Restricted DNA from isolate MI 561 was electrophoresed in the left lane of each panel and that from isolate MI 562 was electrophoresed in the right lanes. The enzymes used to restrict the DNA in both panels (A) and (B) were lane I - HindIII and lane II - EcoRI (indicated at the top of each panel). The images obtained for the HindIII digested genomic DNA samples were displayed previously in Fig. 3.1. The positions of DNA molecular size standards (kbp) are indicated to the left in each panel.

genotypically fingerprint the isolates using a discriminatory technique in order to ascertain if the isolates were genotypically related, i.e., if they could be considered to have a clonal origin and, in effect, deemed to be the same strain. Isolate pairs MI 512/MI 541, MI 517/MI 566, MI 561/MI 562 and MI 683A/MI 683C were subjected to oligofingerprint analysis using the oligonucleotide (GTG)₅ to determine their degree of genetic relatedness. Examination of the fingerprint profiles generated for these isolates using the (GTG)₅ probe revealed that within paired sets, isolates had a high level of fingerprint identity, although minor differences in band profiles were evident, most notably between the isolates MI 561/MI 562 and MI 683A/MI 683C (Fig. 3.4). The paired isolates in which rearrangements of the *cagA* gene were identified can be regarded as the same. Based on these discriminatory fingerprint analyses the rearrangements can be considered to have taken place *in vivo*, i.e., during infection.

3.2.4. Analysis of paired isolates to decipher the possible mechanism of rearrangement that had taken place in the cagA gene.

PCR was used to attempt to elucidate the mechanism of rearrangement that had taken place in the *cagA* gene between paired isolates. The gene itself is approximately 3441 bp in length, specifying a large protein of 1147 amino acids (Covacci *et al.*, 1993). The cloned *cagA* gene (Covacci *et al.*, 1993) with flanking DNA sequences 5' and 3' to the coding region of the gene is, in total, 5225 bp in length. Two sets of primers were designed (CagAF1/CagAR1 and CagAF2/CagAR2) for use in PCR that amplified across the entire coding region of the gene in two segments (each being >2.5 kb). This was done in order to explore as much as possible of the DNA sequences involved in the rearrangement event that had taken place while mindful of the fact that the diversity present in the gene between isolates involved sequences located within the *cagA* gene based on hybridisation studies (Section 3.2.1). Although the 5'-region of the *cagA* gene is known to be conserved, it was thought useful to include this region of the gene in the PCR analysis of the *cagA* gene in order to investigate if *in vivo*-occurring rearrangements between genotypically related isolates maintained conservation of this section of the gene.

Amplification of regions of DNA of this length requires the use of what is known as long-range PCR, and involves the use of a DNA polymerase capable of maintaining association with the template over a longer distance than would *Taq* polymerase. Surprisingly, for

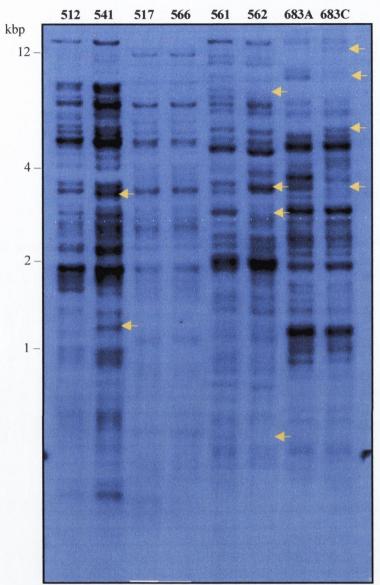


Fig. 3.4. Oligofingerprint blot of *Hin*dIII-restricted total genomic DNA from paired *H. pylori* isolates.

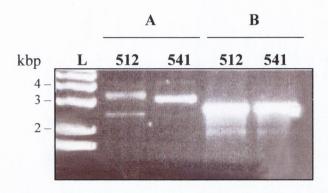
Southern blots of *Hin*dIII-cleaved genomic DNA from isolate pairs MI 512/MI 541, MI 517/MI 566, MI 561/MI 562 and MI 683A/MI 683C were probed with the DIG-labelled oligonucleotide (GTG)₅. Lanes are labelled with the isolate name with the letter designation MI omitted. The most prominent band polymorphisms in the fingerprint profile between paired isolates are indicated with yellow arrows. The hybridisation band profiles range in size from >12 kbp to approximately 500 bp as indicated to the left of the figure.

isolate pair MI 512 and MI 541 there was a slight size difference in the PCR products obtained using primers CagAF1 and CagAR1 that amplify the 5'-most part of the *cagA* gene and sequences upstream thereof (Fig. 3.5.A). The amplification products generated using primers CagAF2 and CagAR2 were uninformative because of their size identity (Fig. 3.5.B). The apparent diversity in the 5'-region of the *cagA* gene present between isolates MI 512 and MI 541 was analysed further by using an alternative forward primer (CagAF), which bound within the *cagA* gene at a location close to the start codon, and the reverse primer CagAR1. The products obtained for both isolates using this primer combination were indistinguishable in size (Fig. 3.5.C), suggesting that the size difference observed initially involved DNA sequences located upstream of the *cagA* gene in the intergenic region between the *cagA* gene and the divergently transcribed *cagB* gene.

The primers Cag1 and Cag2 used in the study by Yamaoka *et al.* (1998) to amplify solely the repeated sequences located at the 3'-end of the gene were next employed in further attempts to elucidate the nature of the rearrangement event that had taken place between paired isolates and to possibly localise the event to a very specific region of the *cagA* gene. Upon amplification, the size differences of the PCR products generated for each isolate in paired sets (MI 512/MI 541 and MI 517/MI 566) using these primers was taken as evidence of the involvement of the 3'-end associated repeats in the rearrangements of the *cagA* gene observed initially by Southern hybridisation (Fig. 3.6). Application of these primers to isolates MI 561/MI 562 and MI 683A/MI 683C (Fig. 3.6) also resulted in PCR products of different sizes between paired isolates (Table 3.1). The sizes of the PCR products generated for three of these paired sets differ by approximately 100 bp between the two, lending further weight to the possibility that repeats which have been reported to be 102 bp long were involved in the generation of *cagA* diversity between paired isolates.

3.2.5. DNA sequence determination of PCR products.

The DNA sequences of the PCR products were determined in order to ascertain if the 3'-end associated repeats were in fact responsible for causing the size differences observed in the PCR products. The PCR products were first cloned into a T-vector (pCR II-TOPO) before being sequenced by the automated custom sequencing service at MWG-Biotech Ltd. The resulting sequence data were analysed using the BLAST algorithm at the NCBI, the CllustalW multiple sequence alignment program, and the ExPASy DNA translation tool.



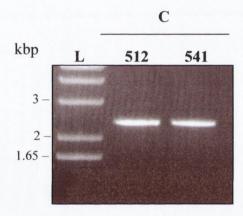
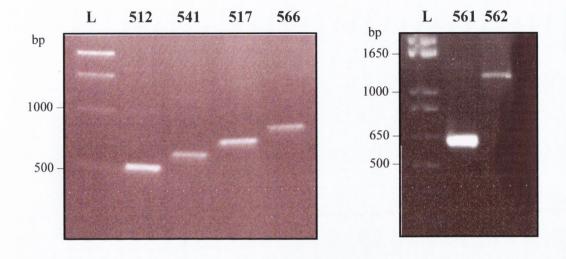


Fig. 3.5. Agarose gel electrophoresis of PCR amplification products of the cagA gene from isolates MI 512 and MI 541.

(A) Agarose gel (1% w/v) electrophoresis of the PCR products generated for the indicated strains using primers CagAF1/CagAR1. (B) Amplification products obtained when the primers CagAF2/CagAR2 were used for isolates MI 512 and MI 541. (C) Amplification products obtained when primers CagF and CagAR1 (both located within the *cagA* gene and amplify most of the first half of the gene apart from the extreme 5'-end) were used. The lanes labelled L show where the DNA size ladder was electrophoresed with sizes (kbp) of the indicated fragments detailed to the left of each panel.



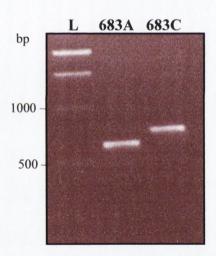


Fig. 3.6. Agarose gel electrophoresis of PCR amplification products obtained using primers Cag1 and Cag2 for paired *H. pylori* isolates. Agarose gels (1% w/v) were loaded with aliquots of PCR reactions carried out using primers Cag1 and Cag2 on the strains indicated. The lanes labelled L are where the DNA molecular size ladder was electrophoresed and the sizes (bp) of

the indicated bands are shown to the left of each panel.

Table 3.1. Sizes of PCR products obtained using primers Cag1 and Cag2 for paired isolate sets.

Isolate Name	Approximate PCR Product Size ^a (bp)	Actual Size Of PCR Product ^b (bp)
MI 512	550	533
MI 541	650	635
MI 517	750	738
MI 566	850	838
MI 561	650	635
MI 562	1200	1244
MI 683A	750	N.D.°
MI 683C	850	N.D.

^a approximate size deduced from gel electrophoresis of PCR products.

^b size determined from PCR product sequences.

^c not determined.

Pairwise ClustalW alignments of the deduced amino acid sequences are shown in Fig 3.7. Sequencing results are shown schematically in Fig 3.8.

3.2.6. Analysis of paired and individual H. pylori isolates using primers Cag1 and Cag2. Paired clinical isolates available in the Moyne Institute collection were subjected to PCR analysis using primers Cag1 and Cag2 to examine these strains for the occurrence of cagA 3'-end repeat unit copy-number variation and to determine if rearrangements of this nature within the cagA gene were a prevalent phenomenon. PCR followed by gel electrophoresis of the products for an additional 29 sets of paired isolates revealed 11 other examples (making 15 pairs in total out of the complete set of 33 paired isolates examined) where each isolate from a pair gave different sized PCR products. Where different, the sizes of the PCR products obtained typically differed by approximately 100 or 200 bp, an indication that the likely difference in the 3'-region of the cagA gene between these isolates was one or two repeat units of 102 bp.

Additionally, single isolates obtained from individuals were also examined which, along with the data obtained for paired isolates sets, were used in order to determine the likely structure of the 3'-region of the cagA gene most commonly occurring among Irish H. pylori isolates. This was found to be a PCR product of 650 bp in size, that, by comparison with the literature and the sequence determined for isolates MI 541 and MI 561, is likely to consist of one 102 bp region and a 45 bp region. Agarose gels showing representative PCR products from paired isolates with variable cagA 3'-ends and identical cagA 3'-ends are shown in Fig. 3.9. Table 3.2 gives the complete set of data for the isolates examined in the present study.

3.2.7. Multiple products obtained by PCR using primers Cag1 and Cag2 result from a heterogeneous culture.

In a number of instances, the amplification of the repeat region of the *cagA* allele by PCR resulted in multiple products (e.g., in isolates MI 334A/ 334C and MI 545A/MI 545C, Fig. 3.9). The multiplicity of products was thought likely to be an indication that the individual from whom the biopsy was obtained may have been colonised at a particular site of the stomach by *H. pylori* isolates which were genotypically heterogeneous. This could encompass isolates that may have been genetically related (i.e., clonal isolates that may be considered the same strain but which differed genotypically in a minor way) or genotypically

	A		
F	7	1	

512 541	LVGNGLSGIEATALTKNFSDIKKELNEKFKNFNNNNNGLKNSGEPIYAKVNKKKTGQVDS LVGNGLSGIEATALTKNFSDIKKELNEKYKNFNNNNNGLKNSGEPIYAKVNKKKTGQVDS ************************************
512 541	PEEPIYTQVAKKVKAKIDRLNQIASGLGGVGQAVGPEEPIYTQVAKKVKAKIDRLNQIASGLGGVGQAVGFPLKRHDKVDDLSKVGLSASPEPIY ************************************
	FPLKRHDKVDDLSKVGLSRNQE-A-KIDNLNEAVS-AKTCHFGNLDQMIDKATIDDLGGPFPLKRHDKVDDLSKVGLSRNQELAQKIDNLNQAVSEAKTCHFGNLDQMIDK
512 541	LKDSTKKNVMNLYVESEKKVPTSLSAKLDNY LKDSTKKNVMNLYVESAKKVPTSLSAKLDNY ************************************

B.

517 566	TLVGNGLSGIEATALAKNFSDIKKELNEKFKNFNNNNNGLKNSGEPMYAKVNKKKTGQAA TLVGNGLSGIEATALAKNFSDIKKELNEKFKNFNNNNNGLKNSGEPIYAKVNKKKTGQAA ***********************************
517 566	SLEEPIYTQVAKKVNAKIDRLNQIASGLGGVGQAAGSHEEPIYTQVAKKVNAKIDRLNQIASGLGGVGQAAGFP-KRHSKVDDLSKVGLSANHEPI * ***********************************
517 566	FPLKRHSKVDDLSKVGLSANHEPIYATIDDLGGPFPLKRHSKVDDLSKVG YATIDDLGGPFPLKRHSKVDDLSKVELSANHEPIYATIDDLGGPFPLKRHSKVDDLSKVG ************************************
	LSANHEPIYATIDDLGGP FPLKRHSKVDDLSKVELSRNQELAQKIDNLNQAVSEAKAGFF LSANHEPIYATIDDLGGP FPLKRHSKVDDLSKVELSRNQELXQKIDNLNQAVSEAKAGFF ***********************************
517 566	GNLEQTIDKLKDSTKHNSMNLWVESAKKVPASLSAKLDNY GNLEQTIDKLKDSTKXXL-IFGLKVQKSAC-FVSETRQL- ************************************





Fig. 3.7. Pair-wise alignments of deduced amino acid sequences from the 3'-ends of the cagA genes of isolate pairs MI 512/MI 541, MI 517/MI 566 and MI 561/MI 562.

The ExPASy DNA translation tool was used to translate the nucleotide sequences obtained for the 3'-regions of isolate pairs A. MI 512/MI 541, B. MI 517/MI 566, C. MI 561/MI 562 of each region. ClustalW was then used to align the amino acid sequences in a pair-wise fashion. The sequence of the 34 amino acid (102 bp) consensus is shown in blue font with the 'EPIYA' motif from outside and within the repeat region highlighted in red. The 15 amino acid (45 bp) consensus sequence is highlighted in green and follows the 45 amino acid repeats in all strains. It was necessary to translate the nucleotide sequence obtained for isolate MI 562 in two reading frames in order to obtain intelligible amino acid sequence data for this region of the gene.

Nucleotide consensus sequence of 102 bp region:

ttc cct ttg aaa agg cat tct aaa gtt gat ctc agt aag gta ggg ctt tca gct aac cat gaa ccc att tac gct acg att gat gat ctc ggc gga cct

Deduced 34 amino acid consensus sequencne:

F P L K R H S K V D D L S K V G L S A N H F F T V A T I D D L C

Deduced 15 amino acid sequence from the 45 bp consensus region:

DDL

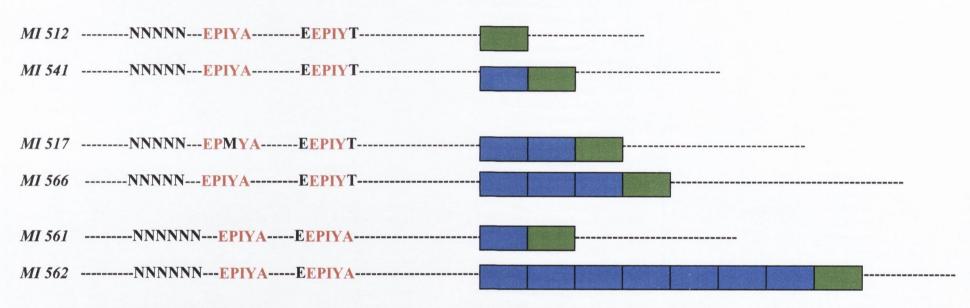


Fig. 3.8. Schematic representation of the amino acid sequences deduced for the 3'-end of the cagA gene from isolate pairs MI 512/MI 541, MI 517/MI 566 and MI 561/MI 562.

The 102 nucleotide consensus sequence is shown at the top of the figure. Beneath this (highlighted in blue) is the deduced 34 amino acid sequence from this 102 bp region. Beneath this again (highlighted in green) is shown the sequence of the 15 amino acids that comprise the 45 nucleotide consensus. Conserved motifs that also occur elsewhere in the protein are shown in red and yellow. A schematic representation of the C-terminus end of the CagA antigen for each isolate includes the conserved asparagine residues followed by the two occurrences of the 'EPIYA' sequence not within the repeat region. The 34 and 15 amino acid consensus sequences are depicted as blue and green coloured boxes, respectively.

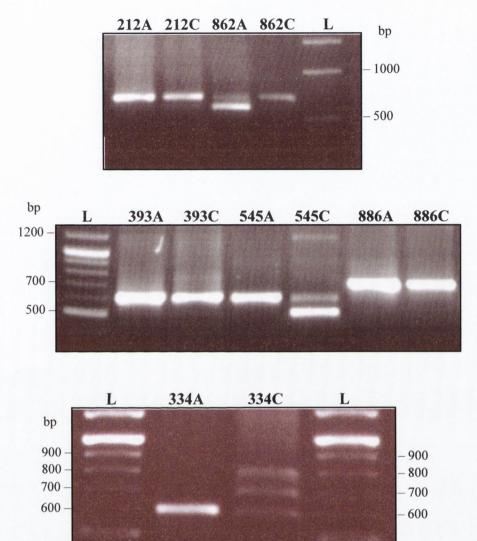


Fig. 3.9. Agarose gel electrophoresis of single and multiple PCR products amplified from individual strains of *H. pylori* from paired isolate sets.

Agarose gels (1% w/v) were loaded with aliquots of PCR reactions carried out on the *H. pylori* isolate pairs indicated at the top of each lane. The gel images show examples of the multiple products obtained from single *H. pylori* isolates, e.g., MI 862C, MI 545C and MI 334C in the above Fig. The lanes labelled L are those in which the DNA molecular size ladder was loaded and the sizes (bp) of the indicated bands are shown to the left and/or right of individual panels.

Table 3.2. Data from the analysis of the 3'-ends of the cagA gene in *H. pylori* paired and single isolates (%).

	<u>Paired is</u>	olates $n = 33$	
Ammon DCD	Identical PCR products ^a 18 (54.54) ^d	Different PCR products	
Approx. PCR product size		$\frac{Single\ products^{b}}{4\ (12.12)^{d}}$	Multiple products ^c 11 (33.3) ^d
550 bp		1*	
650 bp	15 (83.3) ^e	2	10
750 bp	3 (16.6) ^e	2	6
850 bp		2	6
950 bp			2
1200 bp		1	2

Single isolates n = 20

Approx. PCR product size	No. of isolates (%)
<i>550</i> bp	1 (5)
650 bp	16 (80)
750 bp	3 (15)
850 bp	<u> </u>
950 bp	<u> </u>
1200 bp	_

^aPCR products of identical size were amplified for each of the two isolates in each pair.

^bRefers to instances where a single PCR product was obtained for each isolate in a pair but differed in size between the two strains.

^cRefers to instances where multiple PCR products were generated for either isolate from a pair from a single genomic DNA preparation.

^d% value of the total number of paired strains.

^{°%} value of the total number of strains which displayed identical PCR products i.e. 18.

^{*} Figures in bold refer to the number of times PCR products of these approximate sizes were represented in paired isolates that displayed different single or multiple PCR products, e.g., there were six 750-bp PCR products obtained among the total of 11 paired strains from which multiple products were amplified.

unrelated isolates indicating that the individual had a mixed infection. Nonetheless, either scenario would have resulted in multiple products during PCR amplification if, at primary isolation, sweeps were taken from the agar plate for storage as detailed in Section 2.1.4 and Table 2.1. This would have been carried out if the bacteria grew poorly from the biopsy specimen.

To this end, the relevant strains were streaked for single colonies on agar plates and incubated in the usual fashion. PCR was then carried out on single colonies by using as templates aliquots of the suspensions obtained when the colonies were lifted from the surface of the plate on pipette tips and resuspended in sterile water. The cells of the colonies are lysed during PCR, releasing the genomic DNA which then serves as template for the PCR reaction. When this was carried out for isolate MI 500C in which multiple 3'-region bands had been observed, only single PCR products were obtained, indicating that the original culture on which PCR analysis was performed had been heterogeneous (Fig. 3.10).

3.2.8. Analysis by Western immunoblotting of the sizes of the CagA protein expressed by paired isolates.

Western blotting of total cellular lysates of paired isolates using polyclonal anti-CagA serum was carried out in order to determine the sizes of the CagA proteins expressed by strains from paired isolates. Did variation in 3'-end repeat copy number between isolate pairs cause variation in the size of the CagA protein produced by these isolates? Generally, clinical isolates of *H. pylori* express CagA proteins of molecular weights that increase or decrease depending on the number of repeats present within the *cagA* allele of a particular strain.

When cell lysates of isolates MI 512 and MI 541 were analysed by SDS-PAGE and Western immunoblotting, isolate MI 512 gave no signal indicating expression of CagA (Fig. 3.11.A). The remaining paired isolate sets whose *cagA* 3'-region had been subject to DNA sequence analysis were also analysed by Western immunoblotting. Similarly, for isolates MI 517 and MI 566, expression of CagA in these strains was found to be discordant. However, upon repetition, it was observed that an immuno-signal could be obtained for isolates that were originally deemed phenotypically negative for CagA based on the results of previous immunoblotting experiments (Fig. 3.11.B and C). Further preparation of total cell lysates from different cultures of each isolate eventually resulted in a signal being detected for all

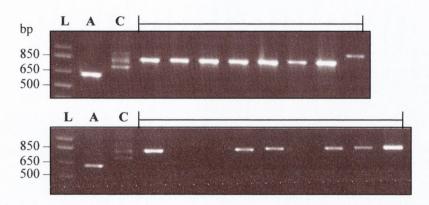


Fig. 3.10. Agarose gel (1% w/v) electrophoresis showing PCR products generated from amplification of the *cagA* gene using primers Cag1 and Cag2 from *H. pylori* isolates MI 500A and MI 500C and from single colonies of isolate MI 500C.

A DNA size ladder was electrophoresed in the lane labelled L. Lanes labelled A and C denote where the amplification products obtained from bulk preparations of total genomic DNA of isolates MI 500A and MI 500C were loaded, respectively. The remaining lanes in both upper and lower panels (indicated with bracketing) show the 14 PCR products obtained when single-colony PCR was carried out on 18 single colonies of isolate MI 500C. The sizes of the bands (bp) in the molecular size ladder are indicated to the left of both panels.

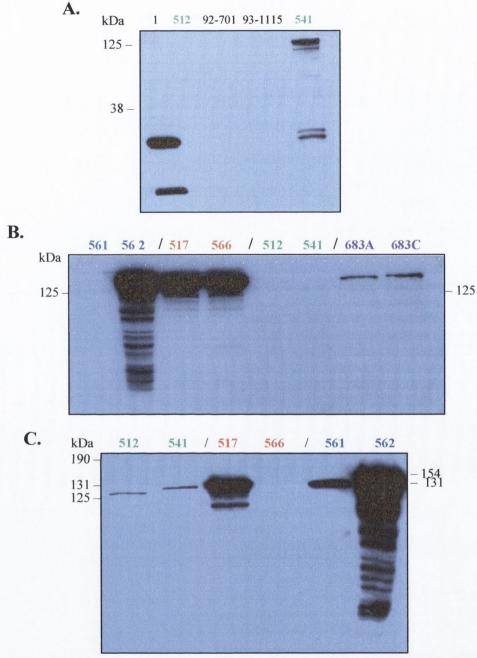


Fig. 3.11. Western immunoblots probed with anti-CagA polyclonal serum of *H. pylori* total cell lysates prepared in the presence of protease inhibitors.

Blot A, lanes correspond to the following: 1. *E. coli* strain XL-1 Blue transformed with a plasmid containing the cloned *cagA* gene. Lanes 2–5, *H. pylori* isolates MI 512, MI 92-701, MI 93-1115 (both *cagA*- isolates) and MI 541 as labelled. Blot B, lanes correspond to the following *H. pylori* isolates: 1, MI 561; 2, MI 562; 3, MI 517; 4, MI 566; 5, MI 512; 6, MI 541; 7, MI 683A; 8, MI 683C as labelled. Blot C, lanes correspond to the following *H. pylori* isolates: 1, MI 512; 2, MI 541; 3, MI 517; 4, MI 566; 5, MI 561; 6, MI 562 as labelled. Lanes where paired isolates were electrophoresed are indicated in coloured font that is comparable between blots. Relative molecular sizes (kDa) are indicated to the left and right of the panels.

strains. In instances where no signal indicating expression of CagA was detected upon Western immunoblotting, this was not considered to arise from inadequate loading of proteinaceous cell lysate material as determined by Coomassie staining of SDS-PAGE gels. Additionally, theses observations were made following repetition of cell lysate preparation and Western analysis for each of these strains on at least four occasions. The deduced sizes of the CagA antigens expressed by these isolates correlated with the number of repeats present in the 3' region of the gene (Table 3.3).

3.2.9. Is the expression of CagA growth-phase dependent?

Many Gram-negative bacteria activate virulence gene expression by a 'quorum sensing' mechanism in response to the concentration of a signalling molecule that reaches a threshold level at high cell densities (Withers *et al.*, 2001). As CagA appeared not to be consistently expressed in different cell lysate preparations for particular isolates, it was thought that the antigen might only be expressed at a certain stage of the *H. pylori* growth cycle. If the *H. pylori* cells from which lysates were prepared had been harvested from agar plates at a point during the growth phase when CagA was not expressed, it would be impossible to detect the antigen by Western immunoblotting. With this in mind, broth cultures were set up of strains MI 512, MI 506 and MI 571 from which samples were taken throughout the growth cycle. Lysates were prepared from all of these samples and Western blots were probed with anti-CagA polyclonal serum. A CagA antigen could be detected at all stages of the growth cycle for these three isolates based on these analyses (Fig. 3.12).

Table 3.3. Deduced sizes of the CagA proteins expressed by *H. pylori* paired isolates.

H. pylori <i>isolate</i>	Number of repeats present at the 3'-end of the cagA gene ^a .	Deduced size of the CagA protein expressed (kDa).
MI 512	15 aa ^b	128
MI 541	1 + 15aa	131
MI 517	2 + 15aa	135
MI 566	3 + 15aa	139
MI 561	1 + 15aa	131
MI 562	7 + 15aa	154
MI 683A	2 + 15aa	135
MI 683C	3 + 15aa	139

^aDetermined from sequencing of PCR products for all isolates apart from MI 683A and MI 683C for which the number of repeats present at the 3'-end of the *cagA* gene was deduced from the sizes of the PCR products obtained for these strains.

^bRefers to the sequence of 15 amino acids that follows the 34 amino acid repeats in each of these strains at the 3'-end of the *cagA* gene. The number of 34 amino acid repeats in each of these strains is indicated by the number shown.

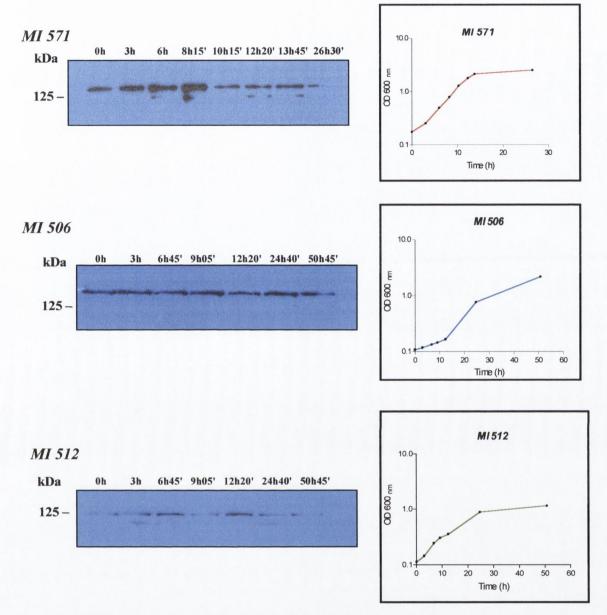


Fig. 3.12. Western immunoblot of total cell lysates of *H. pylori* strains MI 571, MI 506 and MI 512 prepared from broth culture samples at time points over the growth cycle.

Total cell lysates of samples taken at the time points detailed above the various lanes were prepared and probed with polyclonal anti-CagA serum. The position of the 125-kDa molecular size marker is shown to the left of each panel. Note, lanes were not equally loaded. A plot of the growth curves $[OD600_{nm}]$ versus time (h)] for each of the *H. pylori* strains analysed is shown to the right of each relevant panel.

3.3. Discussion.

In general, when considering *H. pylori*, much research interest of late has focused on the interactions the bacterium has with cells of the host, the responses to infection triggered in the eukaryotic cell and the implications for gastric homeostasis that these events elicit. The role played by the CagA antigen and the *cag*PAI in *H. pylori* disease aetiology have of late been shown to have a very intricate involvement in the process. Recent data garnered on the role of CagA provide valuable insights about the pivotal part the protein plays in bacterium—host interactions and justify the apparent shift in direction that has taken place towards a fuller appreciation of the host's involvement in the development of gastroduodenal pathology.

Nonetheless, the *cagA* gene and the protein it encodes, taken as individual molecules, represent interesting entities in the biology of *H. pylori*, their attendant virulence implications notwithstanding. The studies described herein provide information on the occurrence of rearrangements at the 3'-end of the *cagA* gene. A collection of paired and single *H. pylori* isolates was also examined to ascertain the prevalence of rearrangement of this nature in the *H. pylori* population and also to determine the most common *cagA* 3'-end subtype found in Irish strains. Phenotypic analysis of CagA expression in a set of paired *H. pylori* isolates was also undertaken.

3.3.1. The cagA gene exhibits diversity at its 3'-end.

When the likely importance of the antigen found to be co-expressed with the vacuolating cytotoxin and associated with more severe pathologies was realised, DNA sequence analysis of the gene designated *cagA* from different isolates revealed that the 5'-region of the gene was highly conserved between isolates, with variability in the 3'-end of the gene accounting for size differences in the proteins produced. Thus, allelic variation in *cagA* exists and it appeared that the *cagA* gene was capable of undergoing rearrangement within *H. pylori* populations by altering the structure of the 3'-end of the gene and the type and number of repeated sequences located there (Covacci *et al.*, 1993; Yamaoka *et al.*, 1998; Rota *et al.*, 2001). The isolation from single individuals of strains of *H. pylori* containing multiple subtypes of the *cagA* gene has been reported previously (Yamaoka *et al.*, 1999a; Rota *et al.*, 2001). Yamaoka *et al.* (1999a) performed repetitive extragenic palindromic DNA sequence-based PCR (REP-PCR) on such isolates to analyse the DNA fingerprinting pattern of these strains and concluded, based on profile identity between isolates, that the variants had arisen

in the stomach. The evidence presented here corroborates these findings based on the DNA sequence information for the 3'-region of the *cagA* gene from paired isolates (Fig. 3.7) and the discriminatory oligofingerprinting analysis which proves their genetic clonality (Fig. 3.4.). Thus, during the prolonged period of gastric colonisation with *H. pylori*, the *cagA* gene is capable of undergoing rearrangement. The possible reasons for these events are discussed later.

3.3.2. Variability between two paired strains in the 5'-region of the cagA gene.

On the basis of PCR analysis, the finding that there appeared also to exist variability of some type in the extreme 5'-region of cagA or the region upstream thereof (Fig. 3.5.A, B & C) between isolates MI 512 and MI 541 was at variance with the widely held belief that the 5'region of the gene is highly conserved and was somewhat surprising therefore. The cagA gene is located adjacent to the cagB gene in the cagPAI and the genes are divergently transcribed. It is also known that there is considerable overlap in the promoters for both genes with complex mechanisms of transcriptional regulation postulated (Spohn et al., 1997). The diversity involving this end of the cagA gene and of sequences upstream between isolates MI 512 and MI 541 was not investigated further because Southern blot analysis indicated an intragenic variability in the cagA gene between these isolates as with isolate pairs MI 517/MI 566, MI 561/MI 562 and MI 683A/MI 683C. Thus, a possible common mechanism of rearrangement was sought. Nonetheless, the intensity of the PCR product generated using primers CagAF1 and CagAR1 for strain MI 541 compared with strain MI 512 and the slight size difference in these PCR products seems to suggest a definite DNA sequence variability in the region. It may perhaps be an indication of a structural interaction of some description between the DNA sequences of the divergently transcribed and overlapping promoters for the cagA and cagB genes as a result of the effect of regulatory proteins binding to the 436-bp intergenic region with the outcome of a local rearrangement in the DNA sequence. This is highly speculative. However, there are examples from other bacteria of the induction of structural changes in the intergenic regions between divergently transcribed genes upon the binding of regulatory proteins (Rhee et al., 1998). Additionally, effects on the topology of the DNA in the intergenic region because of changes in DNA supercoiling due to transcription from the divergently oriented promoters have also been documented in Vibrio cholerae (Parsot and Mekalanos, 1992).

3.3.3. Distinct subgenotypes of the cagA gene exist in different populations.

Previous studies in which the structure of the 3'-region of the cagA gene was characterised for large numbers of isolates revealed that distinct subgenotypes of cagA existed in H. pylori isolates from different geographical populations. For example, primers used to amplify the 3'-end of the cagA gene from American isolates could not amplify this region of the gene from Korean strains indicating the existence of allelic variants of the gene in different populations (Miehlke et al., 1996). Yamaoka et al. (1998) identified four cagA types with different primary gene structures in isolates from a Japanese population. In addition, these Japanese isolates could readily be distinguished from Western isolates according to the sequence of 15 amino acids (Yamaoka et al., 1999a), a factor thought to be useful for epidemiological studies. In Japanese isolates the amino acid sequence FPLKRH(D/S)KVDDLSKV found in Western isolates was replaced by the sequence KIASAGKGVGGFSGA. The former sequence was present in the 3'-region of the three sets of paired isolates whose DNA sequence was determined in the present study, not a surprising finding given that all isolates were obtained from Irish patients (Fig. 3.7 and 3.8).

The typing system defined in the study of Japanese isolates which enabled cagA genes to be categorised into one of four types (A-D) could not be applied to the strains in the Irish collection analysed here because of the markedly different sequence of the amino acids present in these different sets of H. pylori isolates from different geographic populations. However, the region designated R1 by Yamaoka $et\ al.$ (1998), which is also found in European (Covacci $et\ al.$, 1993) and American (Tummuru $et\ al.$, 1993) isolates, possessing within it a five amino acid sequence EPIYA was present both upstream of the repeat region and within the repeat region in the Irish strains (Fig. 3.7 and 3.8). It is, thus, a highly conserved motif across cagA genes. The polyasparagines (5 or $6 \times Asn$) located upstream of the repeat sequences in the Irish, European and American isolates (Covacci $et\ al.$; 1993; Tummuru $et\ al.$, 1993) were not a feature of the Japanese strains.

3.3.4. Analysis of the 3'-regions of the cagA genes from six Irish isolates.

The structures of the 3'-regions of the *cagA* genes in the six strains whose DNA sequence was determined for this region are shown schematically in Fig. 3.8 from the deduced amino acid sequences of the region. The deduced amino acid sequence for the same region in the isolates analysed by Covacci *et al.* (1993) and Tummuru *et al.* (1993) contains high (>97%) amino

acid sequence identity with the Irish collection. However, despite the amino acid and nucleotide sequence identity that exists between the strains in the Irish collection and the European (Italian) strain, there appears to be a variable juxtaposition of short stretches of amino acids within the entire 34 amino acid (102 bp) repeat region between these isolates of different geographic origin. In the Italian isolate the sequence FPLKRHDKVDDLSKV is located at the end of the 34 amino acid repeat stretch at variance with its location at the beginning in the Irish strains, where it precedes the stretch of 19 amino acids of sequence GSLAN(H/P)EPIYAT(V/I)DDLGGP (Fig. 3.7).

3.3.5. Evidence for subtle differences in the structure of the cagA 3'-end between Irish and Italian strains.

The studies carried out by Covacci et al. (1993) detailed the sequence of strain CCUG 17874 which has apparent similarity to that of strain MI 512 in its possession of one 15 amino acid sequence of the consensus described in the schematic Fig. 3.8. The sequence of a second strain G39 was also determined and was described as being identical to that of CCUG 17874 but also had two repeat stretches of 102 bp inserted at a location in the cagA gene immediately following the 45-bp sequence that encoded the 15 amino acid consensus and which existed alone in strain CCUG 17874. The amino acid sequence deduced for the two pairs of Irish isolates with one or more copies of the 102-bp consensus sequence (isolates MI 517/MI 566 and MI 561/MI 562) consisted of contiguous repeats of 34 amino acids followed by the 15 amino acid region and not preceded by it as in the isolate G39. This is depicted schematically in Fig. 3.13. Hence, fundamental structural differences appear to reside within cagA genes from different populations despite the deduced protein sequences sharing a high degree of amino acid sequence identity. Sequence analysis of a larger number of isolates from an Italian population would be necessary to determine if, in general, cagA genes structurally similar to that possessed by strain G39 are present with high frequency in H. pylori strains from this location. Nonetheless, the possibility that H. pylori isolates with cagA genes of a certain substructure disseminate in different geographical regions is an interesting one with rearrangement within the gene taking place in accordance with a mechanism defined in an ancestral strain. By this reasoning, the cagA genes from the Irish isolates appear to consistently undergo allelic variation such that repeats of 102 bp are always followed by the 45-bp region. Similarly, a bigger study in which sequence analysis of the gene from a larger isolate collection would be carried out would be desirable to clarify this point.

Strain CCUG 17874



<u>Irish strains e.g., MI 517, MI 566,</u> MI 561, MI 562



Fig. 3.13. Comparison of structural differences in Irish and Italian *H. pylori* isolates at the 3'-end of the *cagA* gene.

The 3'-end of the *cagA* gene is depicted schematically for Italian strains CCUG 17874 and G39 (from Covacci *et al.*, 1993) and for Irish strains (e.g. MI 541, MI 517, MI 566, MI 561 and MI 562) [also see Fig. 3.8]. The structural differences present between the strains from these different geographic regions are explained in text and shown above. In the Italian, strains the 15 amino acid consensus sequence (depicted by the green shaded box) precedes the repeats of 34 amino acids (depicted by the blue shaded box) whereas the reverse scenario was found in the Irish strains analysed.

3.3.6. Prevalence of cagA 3'-end repeat copy number variation in paired Irish strains.

The isolates examined in the present study from the Irish collection included 33 paired strains in which two isolates were obtained either sequentially or at a single endoscopic session from individuals. The paired isolates were obtained from biopsy specimens taken from both the antrum and corpus regions of the stomach or from two biopsies taken at different sample sites in the antrum. Interest in paired isolates stemmed from the initial observations coupled with DNA sequence evidence that individuals could be colonised with genotypically identical or highly similar isolates which had undergone allelic variation in *cagA* structure during infection. In order to ascertain how prevalent the occurrence of rearrangement within the *cagA* gene *in vivo* was, as many paired isolates as possible were obtained and analysed.

Of the 33 sets of paired isolates examined, 15 (~45%) pairs yielded PCR products of different sizes using the primers Cag1 and Cag2 (Table 3.2). Although the DNA sequence of these PCR products was not determined, the size difference in the PCR products was taken as evidence of the involvement of the 3'-end associated repeats in the rearrangement event, by comparison with isolates MI 512/MI 541, MI 517/MI 566 and MI 561/MI 562. The possibility remains, however, that, if more biopsy samples from individuals were taken, more colonies of *H. pylori* obtained from these multiple biopsies examined and a larger study population included, the prevalence of isolating allelic variants of *cagA* from single individuals would be greater. Nonetheless, the 45% of Irish paired isolates which displayed genotypic diversity in *cagA* is 28% higher than that reported by Rota *et al.* (2001).

In the study carried out by Rota *et al.* (2001), multiple *cagA* subtypes from single individuals were observed by the generation of multiple amplification products from a single template DNA preparation. In agreement with the observations of Rota and colleagues, this phenomenon was also noticed among Irish strains, i.e., that multiple amplimers were obtained for certain isolates and generally displayed one prominent PCR product accompanied by other less intense bands. As described in Section 3.2.7, this was found to result from a heterogeneous *H. pylori* population existing at the sample site by the separating out of colonies from the original stock of the isolate and carrying out single colony PCR on same. Only one PCR product was obtained for each colony when this was carried out. This is believed to be an indication that the multiple products observed originally resulted either from a mixed *H. pylori* infection at that sample site or the rearrangement of the *cagA* gene in

subclones of the original founder colony. As outlined in Section 3.2.7, in general, single colonies from primary isolation plates were subcultured for storage but in cases where growth of *H. pylori* from the biopsy specimen was poor, sweeps of several colonies were taken. This accounts for the generation of multiple PCR products from a single template stock as reported similarly by Rota *et al.* (2001).

3.3.7. Prevalence of particular cagA subtypes in the Irish collection.

As alluded to previously, the four *cagA* subtypes described by Yamaoka *et al.* (1998) cannot be applied to non-Asian isolates. However, the majority (81%) of Irish strains obtained singly and in pairs contain a *cagA* 3'-repeat region of approximately 650 bp. This region is the same size as the Type A variant of the *cagA* gene described for the Japanese isolates but they only contain in common the EPIYA motif. By comparison with the sequences determined for isolates MI 541 and MI 561, it is likely that the repeat region in these strains contains the 102-bp consensus followed by the 45-bp region. Twenty-two isolates yielded longer amplimers of 750 and 850 bp, either as a single product or multiple PCR products from a heterogeneous population, which by analogy with isolates MI 517 and MI 566 are thought likely to have two and three 102-bp regions, respectively, again followed by the 45-bp region. Large scale sequence analyses would be necessary in order to definitively characterise the complete range of *cagA* structural subtypes present in the *H. pylori* population from Irish patients in order to propose a typing scheme for this geographical area akin to that described by Yamaoka *et al.* (1998) for Japanese isolates.

3.3.8. Strain MI 562 – the cagA gene and CagA protein.

The size of the PCR product obtained for isolate MI 562 (1244 bp) was the largest amplimer generated for isolates in the present study and the largest product amplified using the Cag1 and Cag2 primers in any study which has employed these primers. The largest PCR product reported by Yamaoka *et al.* (1998) was 810 bp long and that reported by Rota *et al.* (2001) was 850 bp long. The product obtained for isolate MI 562 was accordingly something of an unprecedented finding. The DNA sequence of the PCR product revealed a *cagA* 3'-region that contained 7 complete-102 bp regions followed, as in the other sequenced strains, by the 45 bp region. There is also a very faint band of approximately 1200 bp in isolate 545C (Fig. 3.9), an isolate in which two other products were also obtained for these primers. The CagA protein expressed from a *cagA* gene of this length would be expected to be approximately 150

kDa in size. Western blot analysis of total cellular lysates of this strain with anti-CagA polyclonal serum revealed a protein of this relative molecular weight. Although lysates were prepared in the presence of protease inhibitors, there appeared to be considerable degradation of the CagA antigen from isolate MI 562 with a smear of intense signal obtained in the lane where MI 562 was loaded when chemiluminescent detection was complete. Ruling out proteolytic degradation of the antigen because of the presence of protease inhibitors and because lysates of other strains prepared with the same batch of protease inhibitor cocktail remained undegraded, it seems plausible that a protein as large as ~150 kDa that is possibly anomalously big at such a size might be inherently unstable and subject to degradation very readily, perhaps as a result of inefficient folding of the molecule and an unattainable stable tertiary structure. Lysates prepared from different cultures of isolate MI 562 on several different occasions consistently displayed smearing when probed with anti-CagA antibody.

3.3.9. Homologous recombination as a putative mechanism of rearrangement at the cagA gene 3'-end.

A consideration of rearrangement of DNA sequences must address the likely mechanism by which such rearrangement takes place. Rearrangement of the cagA gene results in alteration in copy number of repeat regions at the 3'-end of the gene. It would seem that the most likely mechanism by which this occurs would be by homologous recombination taking place between mis-aligned repeat units of the cagA gene during chromosome replication, the repeat units serving as substrates for RecA (Petes and Hill, 1988). Indeed, a H. pylori recA gene was cloned and sequenced by two groups (Schmitt et al., 1995; Thompson and Blaser, 1995). A double cross-over event taking place at a location where the 3'-region of the cagA gene is juxtaposed between two recently replicated chromosomes would result in both duplication and deletion of a repeat unit with respect to the structure of the original strain. One daughter cell would now have an additional repeat copy in its cagA allele compared to the original strain and the other daughter cell, one repeat unit less than existed before rearrangement took place (Fig. 3.14). In essence, therefore, if this model of homologous recombination is a reflection of events that take place within cells of H. pylori, the fullest description of what takes place during rearrangement of the 3'-end of the cagA gene is tandem duplication and tandem deletion of repeat units.

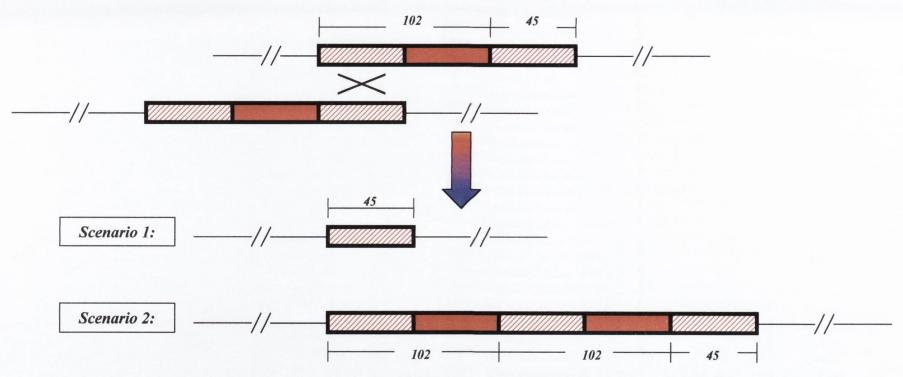


Fig. 3.14. Schematic representation of the possible mechanism of rearrangement at the 3'-end of the cagA gene by homologous recombination between mis-aligned repeats during chromosome replication.

The top half of the figure depicts the replicated region of *cagA* from isolate MI 541 or MI 561 before chromosome segregation. Both these isolates, as detailed in text, contain one 102-bp repeat region followed by a 45-bp consensus at the 3'-end of *cagA*. The first 45 nucleotides of the 102-bp region are identical to the 45-bp consensus and encode the amino acid sequence FPLKRHDKVDDLSKV in Western isolates. Double cross-over at this region would result in daughter cells which, as represented in scenario 1, have lost the copy of the 102-bp region and remain solely with the 45-bp region (as detected in isolate MI 512) or, as in scenario 2, have gained a copy of the 102-bp region (tandem duplication of repeat region). Sequences upstream and downstream of *cagA* are represented by horizontal and diagonal lines. It follows that, if homologous recombination within the *cagA* gene was to occur between *cagA* alleles of the substructure generated in scenario 2, it could also occur between mis-aligned sequences from the latter portion of the 102-bp region, i.e., that which comes after the 45-bp consensus and represented by solid red shading in this figure.

3.3.10. Model to explain the structural subgenotypes of the cagA gene found, bearing in mind a putative homologous recombination mechanism of rearrangement.

Where repeat copy-number variation was identified in the paired Irish isolates analysed herein, the sizes of the PCR products obtained suggested a difference between the two strains of one repeat unit for the majority of pairs. However, the model for rearrangement outlined above intimates that paired isolates which display allelic variation in the cagA gene should at least differ by the order of two repeat units, one daughter cell with an additional repeat unit compared to the original strain and the second daughter cell with one less repeat. How then, does one account for the fact that, where variation exists between paired isolates, the difference is generally of the order of one repeat region? Clearly, in a homogeneous population every cell should be capable of undergoing rearrangement with equal frequency, although not every cell will undergo rearrangement simultaneously or indeed ever. It is tempting to speculate, therefore, where rearrangement in the cagA gene is concerned, that when two genotypically related strains are recovered that have cagA subtypes differing in repeat unit copy number of one, that these two isolates represent the original strain and the daughter strain post-rearrangement of the cagA gene by homologous recombination that contains the gene with one repeat unit deleted. This latter point is thought likely because for the Irish collection of strains, the most prevalent subtype of the cagA gene that was isolated consisted (by comparison with isolates MI 541 and MI 561 as alluded to previously) of one 102-bp repeat region followed by the 45-bp consensus. Thus, this is the shortest structure for the 3'-end of the cagA gene that includes a 102-bp region. Indeed, in the study carried out by Yamaoka et al. (1998), the most prevalent subgenotype of the cagA gene found among Japanese isolates was also the shortest, the type designated Type A in that study.

The apparent instability of the CagA antigen expressed by isolate MI 562 and the infrequency with which cagA subtypes containing multiple 102-bp regions are isolated suggests that it is not favourable to allow tandem repeat duplication to carry on unchecked. It is possible that isolates such as MI 541 and MI 561 and by extension, all isolates from which a \sim 650 bp PCR product was amplified (which probably also contain a single 102-bp region) are most frequently isolated from the H. pylori population for two reasons. First, it might be that cagA genes of this structure express the most stable protein while still maintaining full immunogenicity and virulence functions. Secondly, the maintenance of one 102-bp repeat unit accompanied by the 45-bp sequence allows scope for the generation of diversity in this

region of the gene. From this perhaps 'basic' structure for the region, the formation of any number of *cagA* subtypes is possible. Without the maintenance of a 102-bp region, as exemplified by isolate MI 512, the possibility of generating allelic variants of the *cagA* gene no longer exists. There is, in essence, nowhere to go at this point. Thus, if these hypotheses are valid, it would seem that there are compelling reasons for *H. pylori* isolates to possess a *cagA* gene of the shortest 3'-subgenotype, a situation where allelic diversity and possibly antigenic diversity could be generated if required.

Isolates MI 500C and MI 334C displayed multiple PCR products from cells of a single biopsy specimen (Fig. 3.9 and 3.10). The sizes of these PCR products are thought likely to represent the allelic variants of the *cagA* gene from the original colonising strain and the *cagA* gene from each of the two daughter strains generated after homologous recombination at the 3'-end of *cagA*. All three *cagA* subgenotypes are thought likely to be represented in PCR reactions from these biopsy specimens because of the sweeps taken for subculture and storage from the primary isolation plate and because these specimens are fresh clinical isolates which had been subjected to only a minimum of subculturing. For this reason, the tandemly duplicated and, therefore, larger *cagA* variant resulting from homologous recombination may not yet have been lost from the population.

It is also noteworthy that the manner in which these isolates were stored could feasibly allow the occurrence of cagA 3'-end repeat copy number variation by an alternative mechanism to that which has been described above. As it is evident that these particular strains (MI 500C and MI 334C) contained multiple clones, the possibility cannot be excluded that natural transformation and homologous recombination took place *in vitro* to effect the diversity observed in the cagA gene by transfer of DNA containing the 3' repeat region of the cagA gene.

3.3.11. Analysis of paired H. pylori isolates by Western blotting with polyclonal anti-CagA serum.

Western blot analysis of total cellular lysates of a selection of *H. pylori* isolates was carried out to determine the sizes of the CagA antigens expressed from each isolate in paired sets that had displayed alteration in *cagA* 3'-end repeat unit copy number. Unexpectedly, the studies revealed that different lysate preparations of the same isolate, namely MI 512, MI 541, MI

scluded that this was as a result of inadequate protein loading during electrophoresis. The cells cultured originally for Western blotting were cultured on solid medium which makes it difficult, therefore, to determine the stage of growth that the majority of the cells on the plate might be in. For this reason it seemed plausible that, when isolates were apparently phenotypically negative for CagA as determined by Western analysis, the cells were harvested from plates at a stage that preceded or came after the point where CagA was expressed. To assess the possibility that CagA expression in *H. pylori* might be growth-phase dependent and as such, putatively controlled by a novel transcriptional regulator, broth cultures of *H. pylori* strains MI 512, MI 506 and MI 571 were inoculated and samples taken from the culture throughout the growth cycle. Western blot analysis of lysates of these strains indicated that the antigen was expressed at all phases of growth (Fig. 3.12). Thus, sampling of the entire growth cycle as described excluded the possibility that CagA expression in *H. pylori* is growth-phase dependent.

Why then, even in the presence of protease inhibitors, do certain isolates vary phenotypically with respect to CagA expression? Although no DNA sequence information is known for the 5'-end of the *cagA* gene from any of the isolates MI 512/MI 541, MI 517/MI 566 or MI 561/MI 562, the sequence information available for the 3'-regions of these isolates suggests that the premature location of stop codons can occur within this region of DNA sequence. In isolate MI 512, for example, there appears to be nucleotide substitution near the end of the sequenced stretch of DNA which creates a stop codon at this location (depicted as gaps in the sequence of this strain downstream of the 15 amino acid motif in Fig. 3.7 A). However, it must be borne in mind that the CagA antigen is encoded by an extremely long open-reading frame of 3.4 kb. It is difficult to rationalise how the premature occurrence of a stop codon only a few hundred base pairs from the end of such a long gene would affect protein expression. The efficient translation and folding of a marginally truncated protein is assumed in this scenario. The occurrence of similar nucleotide substitutions in the unsequenced 5'-region of the genes cannot be ruled out of course and would be more likely to have a direct impact on expression of a full-length CagA protein.

3.3.12. Could CagA be a phase-variable antigen?

The above suggestions do not account for the observations that isolates MI 512 and MI 541 appear to vary phenotypically with respect to CagA expression and that this protein can in effect, based on Western immunoblot evidence, be considered to be a candidate phase-variable antigen. This postulate includes similar evidence accumulated for isolate MI 566 and MI 561. Furthermore, genome sequencing of *H. pylori* strains 26695 and J99 (Tomb *et al.*, 1997; Alm *et al.*, 1999) identified that phase-variable elements in *H. pylori* belonged to genes involved in LPS biosynthesis, cell-surface associated proteins and DNA R-M systems. Thus, as a surface-expressed protein with several short homopolymeric tracts of nucleotides in its gene coding sequence, CagA might be a member of the group of phase-variable proteins of this organism. Reversible phase variation of CagA could be what is in evidence in the present study because of slipped-strand mispairing in the homopolymeric regions during DNA replication with the ultimate effect of translational frameshifting.

In the case of isolates MI 517 and MI 566, the latter strain seems to undergo phenotypic variation of CagA. Similarly to isolate MI 512 there is a stop codon located close to the end of the sequenced repeat-containing region of the 3'-end (Fig. 3.7 B). However, in the immediate nucleotides upstream of this codon there appears to have been misreading of the bases located there which may account for the erroneous reading of a stop codon at a location before the end of the gene. Furthermore, for the same reasons discussed above in relation to isolate MI 512, the occurrence of a stop codon at such a location may have a negligible impact on CagA expression.

In the analysis of such sequence data it is important to note that DNA sequence determinations of the *cagA* 3'-end for each of these isolates took place by sequencing PCR products that were generated using the error-prone *Taq* DNA polymerase and furthermore, that these PCR products were sequenced from a single strand only. Thus, the possibility cannot be excluded that the observations discussed above (Sections 3.3.11 and 3.3.12) have arisen as a result of these experimental limitations. The potential error that these limitations introduce is also relevant to the following discussion.

3.3.13. The deduced amino acid sequence of the CagA antigen of isolate MI 562 evidences the occurrence frameshifting.

When considering isolates MI 561 and MI 562 the scenario becomes more complicated still but does provide some evidence for translational frameshifting of the cagA gene. Of this pair, isolate MI 561 varies in the detectability of CagA by Western immunoblotting and the deduced amino acid sequence does imply the insertion of a stop codon two amino acids away from the end of the sequenced repeat region. For reasons outlined in the consideration of isolates MI 512 and MI 566, this is thought likely to be of minor significance in terms of CagA expression. However, isolate MI 562 which consistently gave an immuno-signal by Western blotting required that the sequenced region of the 3'-end of the gene be translated in two different reading frames and spliced together as it were, in order to obtain intelligible sequence for this region. The first 68 amino acids of the region were deduced from the translation of the DNA sequence in the +1 reading frame, while the remainder were deduced from translation in the +3 reading frame. This alone suggests that upstream of the repeat region either translational frameshifting took place as a result of slippage of the DNA polymerase during chromosome replication at regions of homopolymeric tracts or that sequence divergence of the cagA gene in strain MI 562 had taken place such that gain or loss of nucleotides occurred compared to isolate MI 561. Both of these possibilities are represented in the analysis of the nucleotide sequence of the repeat region of MI 562. It is difficult to account for the deduced amino acid sequence of the 3'-end of isolate MI 562, more so when one takes into account that expression of the antigen from this isolate appears to be uniform in cell lysate preparations.

3.3.14. Significance of rearrangement at the 3'-end of the cagA gene.

What is the possible significance of the alteration in the copy number of the 3'-end associated repeats of the CagA antigen during infection? What selective pressure exists within the stomach to render it advantageous for the CagA antigen to display allelic and hence probable antigenic diversity and why, as postulated here, are *cagA* alleles maintained in the majority of isolates with a subgenotype which allows scope for the generation of diversity but which itself might be a basic structural form of the allele?

In the report of the characterisation of the 128-kDa CagA antigen (Covacci et al., 1993), reasons were put forward to explain why heterogeneous CagA antigens are found in *H. pylori*

populations. It was suggested that it could be a mechanism for generating antigenic diversity in the molecule or immunodominant non-protective epitopes that aid in escaping host immunity. Indeed, there are many examples of pathogens that express antigens which contain repeating sequences such as *Streptococcus* spp., *Plasmodium falciparum* and *Clostridium difficile* (Deitsch *et al.*, 1997). Furthermore, the repeats in the alpha C protein of Group B streptococci have been found to undergo alteration in copy number as a means to generate antigenic diversity and that deletion of repeats within the protein negates antibody recognition of the molecule (Gravekamp *et al.*, 1998). However, both Covacci *et al.* (1993) and Yamaoka *et al.* (1998) make the point that rearrangement in the CagA antigen duplicates sequences already present in the protein without generating antigenic diversity and that within the same geographic population (as also observed between Irish isolates) the primary gene sequence of these regions is significantly conserved.

Despite the widely held opinion that repeat duplication in the *cagA* gene does not generate antigenic diversity in the classical definition of the term, the fact that duplication of nucleotide sequences results in the translation of additional amino acids cannot be ignored. All proteins possess structural features required for both stability and functionality. The question of how the CagA antigen accommodates additional amino acids when duplication occurs is an interesting one in terms of protein structure and the issue of antigenic diversity. It seems difficult to rationalise how the presence of new stretches of amino acids could not somehow affect the structure of the protein by perhaps impacting on the way the protein folds and hence by adjusting in some way the conformation of what might be a native CagA molecule. This is especially true if repeat copy number continues to 3, 4, 5 or more. If conformational changes are induced, antigenic diversity could be generated by exposing parts of the protein that might have been sequestered in an antigen of alternative conformation.

Thus, although rearrangement by the duplication of repeated sequences is not perceived to generate antigenic diversity *per se*, alternative, immunogenic, non-protective epitopes of CagA could be revealed because of the presence of newly duplicated repeats. The immune system might thus be distracted from the protective epitopes. A three-dimensional crystal strucutre of CagA combined with knowledge of the precise function of the repeat-containing moieities of the protein would be required in order to further arguments for or against their part in the generation of antigenic diversity in the molecule. Covacci *et al.* (1993) remarked that duplication generates new peaks of hydrophilicity, predicted to be surface exposed and

highly immunogenic. This type of information could dramatically change an argument based on the induction of conformational change in the antigen if the repeats always require to be surface expressed. Nonetheless, speculation is permissible until experimental evidence supporting a single theory becomes available. Additionally, in a less exciting but equally valid possibility, rearrangement at the 3'-end of the *cagA* gene might be neutral, neither conferring any disadvantage nor providing any advantage over the immune system.

It is also worth noting that the recently determined biological function of CagA involves the phosphorylation of a tyrosine residue on the protein when it is inserted into host cells. Within the amino acid sequence of CagA there are three tyrosine phosphorylation motifs, located at the N- and C-termini of the protein and at the beginning of the CagA variable region. That located at the N-terminus is thought most likely to be responsible for the phosphorylation of CagA in host cells. The motif in the variable region, loosely of the consensus GLKNEPIY in Irish isolates, is thought to be functional in *H. pylori* strains but only in approximately 58% of isolates (Evans and Evans, 2001). Repeat copy number variation in the *cagA* gene amplifies a partial element of this motif, i.e., EPIYA. Interesting questions for future research are raised therefore about the activity of the motif in this variable region of the protein and the consequences this might have for the CagA protein *in vivo*.

Chapter 4

'Microevolution' in paired isolates of H. pylori revealed by PCR-

Subtractive Hybridisation

4.1.	Introduction.		94
4.2.	Result	S	98
	4.2.1.	Oligofingerprinting of paired isolates reveals microevolution	98
	4.2.2.	RAPD-PCR fingerprinting confirms high similarity between paired	
		isolates and also reveals microevolution	98
	4.2.3.	Examination of paired isolates by Southern hybridisation and restriction	
		fragment length polymophism (RFLP) analysis as an assessment of strain	
		relatedness	99
	4.2.4.	PCR-subtractive hybridisation and optimisation of the technique	99
	4.2.5.	Application of PCR-subtractive hybridisation to the analysis	
		of H. pylori paired isolates.	100
	4.2.6.	Analysis of the collection of paired isolates for the occurrence	
		of rearrangements at particular loci	101
	4.2.7.	MI 571/MI 565 – Clone 4.	102
	4.2.8.		103
		MI 506/MI 520 – Clone 4.	103
		. MI 355/MI 356 – Clone 3.	104
		. MI 355/MI 356 – Clone 9.	104
	4.2.12.	Membrane protein-encoding gene fragments	104
		4.2.12.1. MI 506/MI 520 – Clone 18.	105
		4.2.12.2. <i>H. pylori</i> Omps.	106
4.3.	Discus	ssion.	107
		Molecular fingerprinting of paired H. pylori isolates	107
		PCR-subtractive hybridisation.	108
	4.3.3.	ORFs identified by PCR-subtractive hybridisation to be involved in the	
		generation of microevolution.	110
		4.3.3.1. MI 571/MI 575 – Clone 4.	110
		4.3.3.2. MI 355/MI 356 – Clone 1	111
		4.3.3.3. MI 506/MI 520 – Clone 4.	112
		4.3.3.4. MI 355/MI 356 – Clone 3.	112
		4.3.3.5. MI 355/MI 356 – Clone 9.	112
		4.3.3.6. MI 506/MI 520 – Clone 18. 4.3.3.7. MI 571/MI 575 – Clone 41.	113
		4.3.3.8. MI 571/MI 575 – Clone 47.	114
		4.3.3.9. MI 458/MI 459 – Clone 2.	114
		4.3.3.10. MI 458/MI 459 – Clone 6.	11:
	131	General remarks.	110
	T.J.T.	GUILUI UI I UIII UII IW	11

4.1. Introduction.

The diversity present at the genomic level in H. pylori was first recognised in the course of investigations to find a suitable typing method for the differentiation of clinical isolates for molecular epidemiological purposes (Langenberg et al., 1986; Majewski and Goodwin, 1988; Akopyanz et al., 1992a and 1992b). Briefly, when subjected to discriminatory molecular fingerprinting analysis, e.g., RAPD-PCR, oligofingerprinting and epidemiologically unrelated isolates (i.e., those obtained from different individuals) were found to have completely different fingerprint profiles, with each infected individual apparently colonised by a unique strain. The genomic diversity present in H. pylori was such that it rendered the efforts to find epidemiological, and hence, disease associations between clinical isolates of little use. However, with such diversity at a level unseen in other bacterial pathogens (Blaser, 1997a; Achtman, 2001), huge interest was generated in this aspect of the biology of H. pylori, its genesis and its implications for disease aetiology (Logan and Berg, 1996; Marshall et al., 1998; Blaser, 1997b).

Genome mapping studies of *H. pylori* using pulse-field gel electrophoresis (PFGE) also evidenced macrodiversity within the species with the strains examined having a disparate physical genome organisation and chromosomal gene order (Taylor *et al.*, 1992; Bukanov and Berg, 1994; Jiang *et al.*, 1996). Thus, as alluded to above, DNA fingerprinting and molecular typing studies, combined with multilocus enzyme electrophoresis data (Go *et al.*, 1996) and DNA sequence analysis of the *glmM* (formerly *ureC*) and *flaA* genes (Fujimoto *et al.*, 1994; Forbes *et al.*, 1995; Shortridge *et al.*, 1997) amongst others, testified to the high level of genetic variability that exists within *H. pylori*. Indeed, the population structure of the species was defined as being non-clonal or panmictic (Go *et al.*, 1996; Suerbaum *et al.*, 1998). DNA sequence analysis of multiple alleles revealed that the pool of alleles present is so large that evidence of clonality between isolates is difficult to discern. Panmixis implies a lack of linkage disequilibrium between alleles and a high frequency of horizontal genetic exchange (Suerbaum *et al.*, 1998).

Although accounting in large part for the perceived genomic diversity of *H. pylori*, the occurrence of *de novo* nucleotide sequence variability only partially explains the allelic variation observed in this organism. Mutation occurs with the same frequency in both *H. pylori* and *Neisseria meningitidis* but unique sequences are observed more frequently in *H. pylori* than they are in *N. meningitidis*. The panmictic or freely recombining nature of the *H.*

pylori population structure allows for the occurrence of recombinations so frequently as to generate a large number of unique sequences more often than would be possible by the sequential accumulation of mutations alone. Recombination per se does not create polymorphic sites, but increases the number of unique sequences generated when polymorphic nucleotides are present (Suerbaum et al., 1998; Suerbaum and Achtman, 1999; Wang et al., 1999b; Suerbaum, 2000).

Collectively, the accumulated data were indicative of a highly diverse species which seemed to exhibit both marked microdiversity and macrodiversity, little or no evidence of clonal descent between isolates and a population structure that displayed a high frequency of horizontal gene transfer as evidence of free recombination between strains. However, the comparison of the completed genome sequences of two unrelated H. pylori strains, 26695 and J99 (Tomb et al., 1997; Alm et al., 1999), somewhat surprisingly revealed a relatively conserved proteome and a less complex gene order and genomic organisation than would These observations indicated, furthermore, that the degree of have been anticipated. macrodiversity present in H. pylori as assessed by PFGE may have been over-estimated (Alm et al., 1999; Alm and Trust, 1999). These interstrain comparisons also illuminated that a major contributory factor to the variability in genomic fingerprint profiles that are generated for different isolates by the nucleotide-based typing techniques RAPD-PCR, REP-PCR and oligofingerprinting was the frequent occurrence of synonymous nucleotide substitutions at the third base position of codons that do not affect the protein sequence encoded. synonymous site nucleotide variation in H. pylori results in conservation of the proteome while simultaneously exaggerating the diversity of the species (as was found similarly with PFGE), given that such substitutions are not manifest at the phenotypic level. Nonetheless, the existence of marked genomic heterogeneity within H. pylori is a recognised characteristic of this bacterium and is unique among pathogens, which generally exhibit a high level of clonality (Achtman, 2001).

Free recombination and frequent horizontal exchange are permitted in *H. pylori* by the natural competence of the organism for transformation (Wang *et al*, 1993b). Indeed, the horizontal transfer of genetic information between co-colonising *H. pylori* strains has been demonstrated experimentally by Kersulyte *et al.* (1999) and represents physical evidence for the reshuffling of sequences that homoplasy testing and compatibility matrix analysis testified to (Suerbaum *et al.*, 1998). The transposition of the novel *H. pylori* insertion sequence elements IS605 and

IS606 also contributes to organisational rearrangements of the cagPAI (Akopyants et al., 1998a) and the mixing of alleles at different loci in the chromosomes of strains 26695 and J99 (Alm and Trust, 1999). They are also likely to be involved in sequence shuffling and the generation of allelic variation during natural infection. Thus, the apparent frequency with which recombination takes place in *H. pylori*, easily facilitated by the natural competence of the organism for exchange of genetic information by transformation, together with the prevalence with which synonymous nucleotide substitutions can be detected, seems to suggest that much of the recombination that occurs in the *H. pylori* species is not brought about by selective pressure. However, a study by Akopyants et al. (1995) in which spontaneous mutations in the *H. pylori* genome were selected for by serial passage in gnotobiotic piglets suggested that host-specific adaptation of H. pylori can occur and that such adaptation selected in this instance for the progressive improvement in fitness of the infecting strain. Indeed, pathogenicity-adaptive or 'pathoadaptive' mutations by loss- or modification-offunction random mutations in chromosomal genes that confer a strong selective advantage upon the bacterial clone in the virulence niche without horizontal transfer are found extensively among both Gram-positive and Gram-negative genera (Sokurenko et al., 1999).

Several reports have noted the recovery from individuals of *H. pylori* isolates showing minor differences in fingerprint profile when two or more biopsy samples were taken and the isolated bacteria analysed by molecular techniques (Langenberg et al., 1986; Majewski and Goodwin, 1988; Oudbier et al., 1990; Kuipers et al., 2000). These differences are not believed to be caused by laboratory manipulation of the isolates and are considered evidence for the occurrence of recombination in the genome of the infecting strain in vivo. The subtypic variants that arise as a result of these undefined DNA rearrangements coexist within the resident H. pylori population of the stomach. With a population structure that displays panmixis, it follows that in vivo-occurring DNA rearrangements in H. pylori suggested by fingerprint profile variation could be direct evidence for DNA exchange between clonally related strains during infection or equally for intra-strain chromosomal recombinations, both of which would generate fingerprint polymorphisms. However, there also exists the possibility that these differences in fingerprint profile are so-called pathoadaptive mutations occurring in the genomes of H. pylori isolates during infection in response to the changes in the micro-environment induced in the virulence niche of this bacterium as a direct result of the progress of infection, its sequelae and the host's immune intervention (Kuipers et al., 1995).

The studies described in the present chapter were aimed to probe the subject of in vivooccurring H. pylori genome recombination. Minor differences in fingerprint profile between clonally related H. pylori isolates were observed when oligofingerprinting analysis was employed (Marshall, 1996). In essence, herein an analysis of four such pairs of clonally related but so-called 'microevolved' paired isolates was conducted using a technique called PCR-subtractive hybridisation (Akopyants et al., 1998b). The DNA fragments which differed or partially differed between microevolved paired isolates were isolated using this technique and enabled the definition of the loci which had undergone rearrangement between H. pylori subtypic variants by DNA sequence analysis of these fragments. The technique is the successor to the powerful yet technically unwieldy subtractive hybridisation and representation difference analysis methodologies (Lisitsyn et al., 1993; Tinsley and Nassif, 1996; Calia et al., 1998) and represents a significant advance in the area of strain-specific DNA fragment isolation. Manipulation of the isolated fragments is markedly easier because they can be amplified readily by PCR and cloned by a simple cloning procedure which utilises a single base-pairing principle involving the overhanging adenine nucleotide left by Taq polymerase and a vector manufactured with thymine nucleotide overhangs. PCR-subtractive hybridisation is ideally suited to the analysis of H. pylori isolates in which microevolution was observed to have occurred by allowing direct isolation and analysis of the DNA fragments responsible for fingerprint polymorphisms.

A more in-depth understanding of the genetic diversity of *H. pylori* is afforded by carrying out analyses such as these, given that this question is addressed at the most appropriate level, i.e., analysis of diversity between clonally related isolates that is believed to have taken place *in vivo*.

4.2. Results.

4.2.1. Oligofingerprinting of paired isolates reveals microevolution.

The use of oligonucleotides (GGAT)₄ and (GTG)₅ as probes under low stringency conditions in Southern hybridisation with HindIII-restricted H. pylori genomic DNA was found to be a useful fingerprinting method for this organism and could readily differentiate isolates (Marshall et al., 1996). When oligonucleotide probes were used to analyse sets of isolates obtained from individuals on one or more sampling occasions, in most instances it was found that these individuals were colonised by a homogeneous H. pylori population with all isolates having identical fingerprints (data not shown). However, it was also observed that several sets of paired isolates had oligofingerprint profiles that were highly similar to each other but differed slightly (Marshall, 1996). The profiles were, nonetheless, of a level of similarity such that the isolates are still considered to be of clonal origin and are in essence subtypic variants. The presence of additional bands and the absence of hybridising bands at the same position between the two profiles or both, are taken as evidence of the occurrence of microevolution between isolates. Fig. 4.1.A shows examples of the fingerprint profiles generated for four paired isolate sets (MI 458/MI 459, MI 571/MI 575, MI 506/MI 520 and MI 355/MI 356) in which microevolution was apparent and are reproduced in the present studies using DIG-labelled (GTG)₅ and (GGAT)₄ oligonucleotide probes.

4.2.2. RAPD-PCR fingerprinting confirms high similarity between paired isolates and also reveals microevolution.

In studies by other groups the commonly used RAPD-PCR fingerprinting technique has also allowed the detection of *H. pylori* isolates that had minor but distinct differences in fingerprint profile. The application of pair-wise combinations of RAPD primers to the analysis of the isolate pairs MI 458/MI 459, MI 571/MI 575, MI 506/MI 520 and MI 355/MI 356 confirmed the high similarity or identity of the isolates that had been observed previously by oligofingerprint analysis. Similarly, band differences observed between the fingerprints of some of the paired isolates using RAPD-PCR were further evidence that recombination had taken place in the genomes of these isolates during infection (Fig. 4.1.B). The identity and therefore clonality of the paired isolates was also confirmed using RAPD-PCR. However, RAPD-PCR also showed that not as much of the genome is analysed by this technique as is with oligofingerprinting, as evidenced by the failure of RAPD-PCR to pick up profile band differences in all three sets of paired isolates examined using this technique.

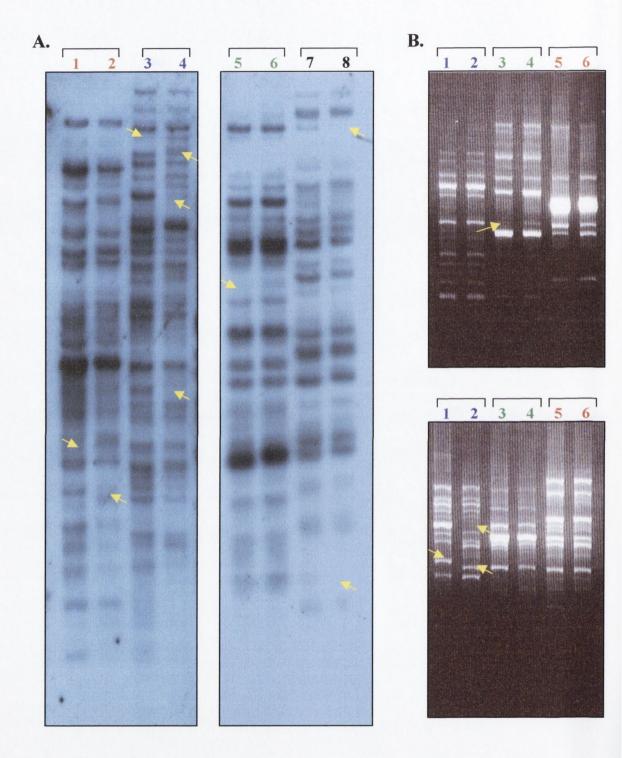


Fig. 4.1. Oligofingerprint and RAPD-PCR analysis of paired H. pylori isolates.

- (A.) Oligofingerprint analysis of four *H. pylori* paired isolate sets with DIG-labelled oligonucleotide probes (GTG)₅ (left panel) and (GGAT)₄ (right panel). Lanes correspond to the following *H. pylori* paired isolate sets: lanes 1 and 2, MI 458 and MI 459, respectively; lanes 3 and 4, MI 571 and MI 575, respectively; lanes 5 and 6, MI 506 and MI 520, respectively; lanes 7 and 8, MI 355 and MI 356, respectively. Yellow arrows point to areas of band polymorphism between isolates in paired sets.
- **(B.)** Agarose (1.8% w/v) gel electrophoresis of amplimers obtained following RAPD-PCR analysis of three *H. pylori* paired isolates sets. The lanes labelled 1-6 in both upper and lower panels correspond to the following *H. pylori* paired isolates: lanes 1 and 2, MI 571 and MI 575, respectively; lanes 3 and 4, MI 506 and MI 520, respectively; lanes 5 and 6, MI 458 and MI 459, respectively. Primer pair D11344 and 1281 was used to generate the profiles in the upper panel and primer pair 1254 and D8635 was used in the reactions electrophoresed in the lower panel (Section 2.12.2; Table 2.2).

In both (A) and (B), brackets join isolates from paired sets and the lane numbers are coloured to correlate the isolates in the profiles of both the oligofingerprint and the RAPD-PCR fingerprint.

4.2.3. Examination of paired isolates by Southern hybridisation and restriction fragment length polymorphism (RFLP) analysis as an assessment of strain relatedness.

HindIII-digested DNA from the four sets of paired isolates analysed by oligofingerprinting and RAPD-PCR were also probed with three DIG-labelled probes to the H. pylori genes vacA, cagA and flaA. As a means of strain typing and differentiation, digested PCR products of the H. pylori ureC gene have been analysed in the past for RFLPs (Foxall et al., 1992; Fujimoto et al., 1994; Shortridge et al., 1997). The analyses performed here were used as a simple method to further ascertain the level of strain relatedness exhibited by the various paired isolates. Although oligofingerprinting and RAPD-PCR are both highly discriminatory fingerprinting methods in that they give a genome-wide fingerprint, it was of interest to determine if results from fingerprinting analysis could be correlated with RFLP data for the four paired isolates examined and ultimately give a more rounded appraisal of the level of clonality present within pairs, without the necessity to perform DNA sequence analysis of, for example, a house-keeping gene in order to construct a relationship dendrogram for the isolates.

The hybridisation profiles obtained for each of the four paired isolates and two other unrelated strains with each of the given probes are shown in Fig. 4.2. Three out of four paired isolate sets, namely, MI 571/MI 575, MI 458/MI 459 and MI 355/MI 356 gave identical hybridisation profiles with all three probes apart from the difference of one minor band with isolate pair MI 571/MI 575 when probing with *vacA*. The isolates MI 506/MI 520, however, gave disparate patterns of hybridisation with all three probes and might, based alone on RFLP analysis of these genes, be considered to lack clonality. The fingerprinting evidence for these isolates presented in Fig. 4.1. suggests the contrary.

4.2.4. PCR-subtractive hybridisation and optimisation of the technique.

The recently described PCR-subtractive hybridisation methodology (Akopyants *et al.*, 1998b) was chosen for analysis of the microevolution observed in *H. pylori* paired isolates MI 458/MI 459, MI 571/MI 575, MI 506/MI 520 and MI 355/MI 356. The technique presented itsell f as being ideally suited to the analysis of microevolution in these isolates as it enables the isolation of genetic differences between two closely related genomes by hybridisation and subtraction of all the DNA that is common between two strains to leave the unique sequences remaining.

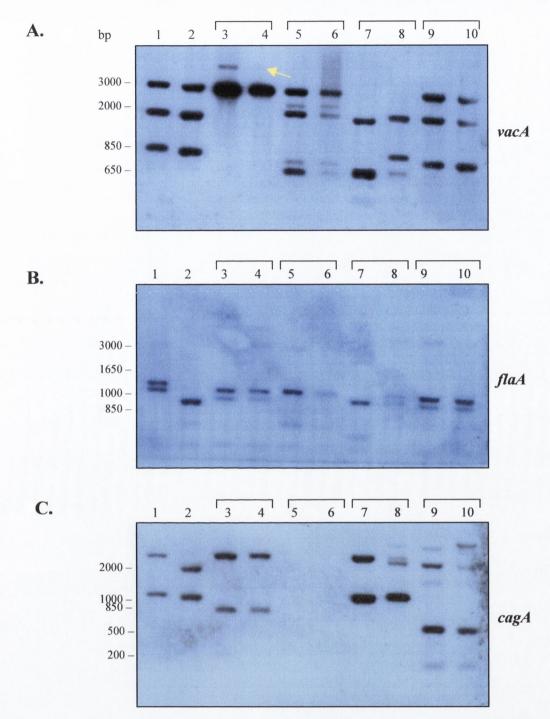


Fig. 4.2. RFLP analysis of four paired isolates and two other epidemiologically unrelated isolates using three different DIG-labelled gene probes.

Panels A, B and C show hybridisation band patterns of *HindIII*-digested genomic DNA when probed with each of the indicated probes: (A) a *vacA* gene probe, (B) a *flaA* gene probe and (C) a *cagA* probe. Lanes: 1, NCTC 11638; 2, MI 497; 3, MI 571; 4. MI 575; 5, MI 458; 6, MI 459; 7, MI 506; 8, MI 520; 9, MI 355; 10, MI 356. Molecular sizes (bp) are indicated. The yellow arrow in panel (A) points to the polymorphic band referred to in text when the *vacA* probe hybridised to strains MI 571 and MI 575. In (A), (B) and (C), brackets join the lanes where isolates from paired sets were electrophoresed.

To optimise and trouble-shoot the protocol for PCR-subtractive hybridisation, it was first applied to the analysis of two unrelated *H. pylori* isolates from the Moyne Institute collection. One strain possessed the cagA gene while the other strain lacked the cagA gene. Optimum conditions for clean-up of digested DNA, adaptor oligonucleotide ligation and hybridisation steps, as well as appropriate PCR cycling parameters, were determined using these two unrelated strains with definite known chromosomal differences. Southern blot analysis of the amplimers obtained upon primary PCR of the unsubtracted tester control sample and the subtracted sample with a labelled cagA gene probe gave hybridisation signals for both samples but the hybridisation was considerably more intense for the subtracted sample, indicating that this isolated collection of unique or partially different genomic fragments contained cagA DNA sequences in accordance with what would have been expected had subtraction worked efficiently. Probing the same PCR amplification products with the ubiquitous flaA gene gave a hybridisation signal with the unsubtracted tester control PCR amplicons that was not present in the collection of amplicons generated during PCR of the subtracted sample (Fig. 4.3). This indicated that PCR-subtractive hybridisation had been optimised to efficiently isolate DNA that was unique to one strain (cagA DNA as exemplified here) and not fragments that were common between the two (*flaA* in this example).

4.2.5. Application of PCR-subtractive hybridisation to the analysis of H. pylori paired isolates.

The strategy adopted in the application of PCR-subtractive hybridisation to the analysis of *H. pylori* paired isolates is shown in Fig. 4.4. For the four sets of paired strains examined (MI 458/MI 459, MI 571/MI 575, MI 506/MI 520 and MI 355/MI 356) a subtracted library was constructed by T-vector cloning of the amplicons present upon secondary PCR. After identification of the colonies containing plasmid vectors with insert DNA by blue/white colony screening, the insert DNA was reamplified from the vector and electrophoresed in each of two identical gels. Identification of the amplicons that were unique or which partially differed between the isolates in paired sets was achieved by probing each of these gels with labelled total genomic DNA from both isolates. Presumptive unique or partially different clones identified in this manner were confirmed to be such by DIG-labelling of specific PCR products in order to probe a Southern blot of *Hind*III-restricted genomic DNA from each strain. A different pattern of hybridising bands between strains in paired isolates was taken as evidence of the occurrence of DNA rearrangement of some description in the genomes of these isolates involving the cloned chromosomal fragment (insert DNA). The chromosomal

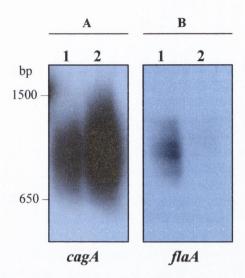
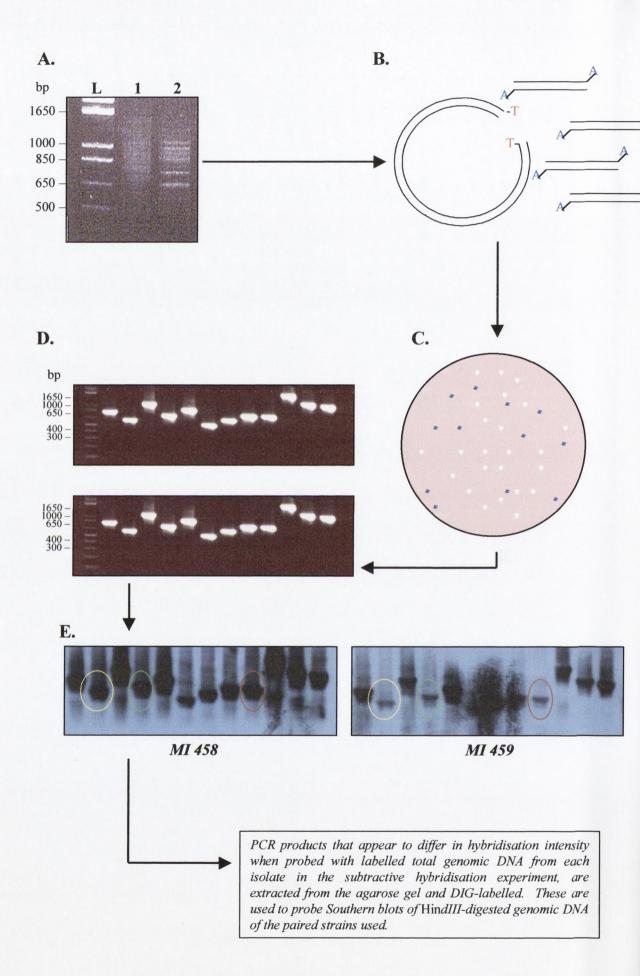


Fig. 4.3. Verification of the efficacy of PCR-subtractive hybridisation by Southern hybridisation of secondary PCR amplimers with *cagA* and *flaA* gene probes.

PCR-subtractive hybridisation was carried out on two *H. pylori* isolates, one $cagA^-$ strain (MI 215) and one $cagA^+$ strain (MI 506). In both panels A and B, lanes labelled 1 and 2 correspond to where secondary amplification products generated from the unsubtracted tester control sample and the subtracted sample were electrophoresed, respectively. Panel A was probed with a DIG-labelled cagA probe as indicated and panel B was probed with a DIG-labelled flaA probe. The size range of the smeared signals obtained is indicated in bp to the left of panel A.



- Fig. 4.4. Flow diagram depicting stages in the identification process of DNA fragments that differ or partially differ between *H. pylori* paired isolates using PCR-subtractive hybridisation.
- (A.) Representation of the appearance upon agarose gel electrophoresis (2% w/v) of the amplicons obtained after secondary (nested) PCR on subtractively hybridised collections of fragments. Lane 1, amplicons obtained from subtractive hybridisation of two unrelated *H. pylori* isolates (MI 506 and MI 575) and lane 2, amplicons obtained from the analysis of isolate pair MI 458/ MI 459 by PCR-subtractive hybridisation. DNA molecular size standards were electrophoresed in the lane labelled L and the sizes (bp) of the indicated bands are shown to the left of the panel.
- (B.) A subtracted library of the collection of amplicons generated by secondary PCR was made by cloning the PCR products into a T-vector in a shot-gun cloning manner. Competent *E. coli* were transformed with the ligation mixture and plated on agar containing appropriate antibiotic selection (ampicillin) and X-Gal to permit blue/white colony screening.
- (C.) Cartoon of agar plate depicting the blue and white colonies obtained after transformation. White colonies containing inserts were picked (~50 for each subtraction experiment). Plasmid DNA was prepared from the transformed cells and nested primers 1 and 2R (Table 2.2) were used to reamplify the insert from the vector.
- **(D.)** Representation of the electrophoresis in duplicate gels of these reamplified insert fragments. The gels were treated and blotted by capillary blotting to nylon membrane.
- (E.) DIG-labelled total genomic DNA from each isolate of a pair (isolates MI 458 and MI 459 in this instance) were used to probe these blotted PCR products as indicated. Although band intensities differ between blots because of labelling of unequal quantities of total genomic DNA, it is possible to discern the clones which represent likely *bona fide* differences between the two strains. They are indicated by the coloured circles.

fragments or ORFs in question were determined by DNA sequence analysis of the insert DNA. The physical evidence by Southern blotting that clones of interest contain genomic DNA fragments that appear to be rearranged between a pair of *H. pylori* isolates is shown in Fig. 4.5. The complete set of data gleaned from PCR-subtractive hybridisation analysis of the four paired isolate sets is shown in Table 4.1.

Nucleotide sequence identities between the clones isolated by PCR-subtractive hybridisation of paired clinical isolates and the sequenced H. pylori strains 26695 and J99 are also shown in Table 4.1 and are given as percentage values. When BLAST analysis was carried out all but two of the subtractively hybridised clones (clone 1 from strain pair MI 355/MI 356 and clone 4 from strain pair MI 506/MI 520) gave high identity scores to ORFs from each of strains 26695 and J99. Indeed, clone 4 from isolate pair MI 506/MI 520 has a G+C% of only 27%, vastly different from the ~39% value calculated for the entire H. pylori genome. This has implications in terms of the likely origins of this DNA fragment and strongly suggests its acquisition by horizontal transfer from a heterologous species. The remainder of the clones encoded a diverse array of putative proteins and included (i) conserved hypothetical proteins with no functional annotation, (ii) housekeeping genes involved in energy metabolism, gene regulation and DNA interactions, and (iii) genes that encoded putative outer membrane proteins (Omps) and an integral membrane protein in one instance. In terms of the hypothesis that H. pylori might undergo microevolution in vivo as an adaptive response to changes in its microenvironment, it is striking that rearrangement was found in several Omp-encoding genes. It seems plausible that these membrane-located molecules might be candidate proteins through which adaptive responses are elicited, perhaps selected for as antigenic variants in response to the mounting by the host of an immune response to the Omp. The possible wider significance of the rearrangements at the loci detected by PCR-subtractive hybridisation is discussed in Section 4.3.

4.2.6. Analysis of the collection of paired isolates for the occurrence of rearrangements at particular loci.

The availability of a large collection of paired clinical isolates allowed the examination of strains obtained from other individuals to determine whether the occurrence of a rearrangement within a particular ORF was a prevalent phenomenon or whether the rearrangement was a rare event occurring only in the set of isolates on which subtractive hybridisation was originally carried out or in a small number of other paired strains. It

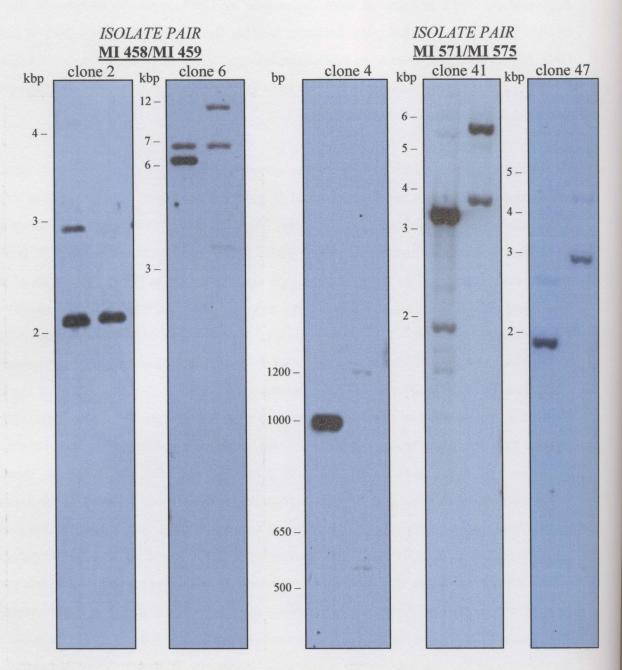


Fig. 4.5. A. Southern hybridisation of paired *H. pylori* isolates with DNA probes isolated by PCR-subtractive hybridisation.

HindIII-restricted genomic DNA of the indicated H. pylori isolate pairs was probed with DIG-labelled probes from the clones (shown at the top of each panel) isolated by PCR-subtractive hybridisation of each given isolate pair. For Southern blotting of HindIII-digested DNA from isolate pair MI 458/MI 459, DNA from strain MI 458 was electrophoresed in the left lane and DNA from strain MI 459 in the right lane. Similarly, for the three right-most blots shown above, HindIII-digested DNA of strain MI 571 was electrophoresed in the left lane and of strain MI 575 in the right lane. For PCR-subtractive hybridisation isolates MI 458 and MI 571 served as the tester strains. Molecular size markers (kbp and bp) are shown to the left of each panel.

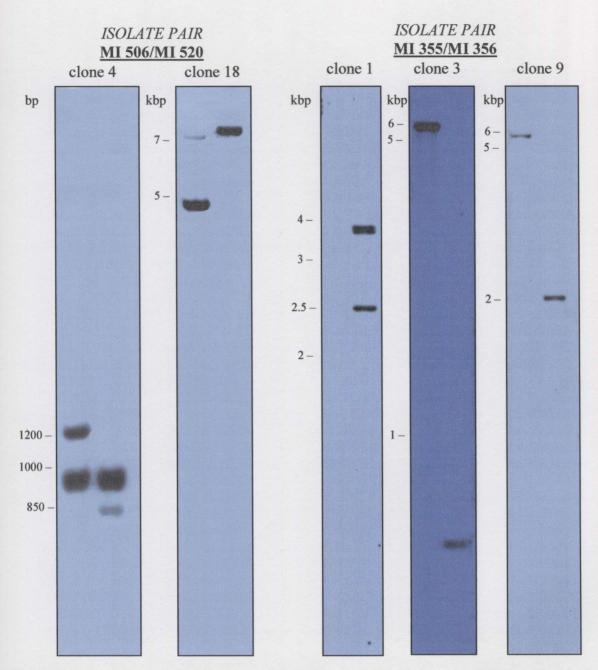


Fig. 4.5. B. Southern hybridisation of paired *H. pylori* isolates with DNA probes isolated by PCR-subtractive hybridisation.

HindIII-restricted genomic DNA of the indicated H. pylori isolate pairs was probed with DIG-labelled probes from the clones (shown at the top of each panel) isolated by PCR-subtractive hybridisation of each given isolate pair. For Southern blotting of HindIII-digested DNA from isolate pair MI 506/MI 520, DNA from strain MI 506 was electrophoresed in the left lane and DNA from strain MI 520 in the right lane. Similarly, for the three right-most blots shown above, HindIII-digested DNA of strain MI 355 was electrophoresed in the left lane and of strain MI 356 in the right lane. For PCR-subtractive hybridisation, isolates MI 506 and MI 356 served as the tester isolate. Molecular size markers (kbp and bp) are shown to the left of each panel.

Table 4.1. Data obtained from PCR-Subtractive Hybridisation analysis of four H. pylori paired isolate sets.

Isolate pair	Clone name	No. of nt sequenced	Corresponding locus in J99	Corresponding locus in 26695	% identity ^a		No. of nt overlap ^b	Function/location of (putative) protein	
					J99	26695		encoded	
MI 458/ MI 459	Clone 2	431	jhp1261 jhp0212	HP1342 HP0227	~73%	~93%	~430	Putative outer membrane protein	
	Clone 6	533	jhp1084	HP1157	~84%	~80%	~530	Putative outer membrane protein	
MI 506/MI 520	Clone 4°	510							
	Clone 18	462	jhp1262	HP1343	~95%	~94%	~459	Putative protein (J99 annotation)	
								Conserved hypothetical integral membrane protein (26695 annotation)	
				HP0548		~99%	~184	DNA helicase	
MI 355/MI 356	Clone 1 ^d	568	jhp1213	HP1293	~76%	~76%		DNA-directed RNA polymerase, α subunit $(rpoA)$	
						~98%	~288	DNA helicase (HP0548) and mini IS605°	
	Clone 3 ^f	840	jhp0041	HP0048	~92%	~95%	~558	Putative transcriptional regulator (<i>hypF</i>) ^g	

			jhp0040	HP0047	~89%	~89%	~229	Hydrogenase expression/formation protein (<i>hypE</i>)
	Clone 9	338	jhp1312	HP1417	~97%	~94%	~336	Conserved hypothetical protein
MI 571/MI 575	Clone 4	521	jhp1364	HP1471	~87%	~85%	~515	Putative Type II restriction enzyme
	Clone 41	552	jhp1008	HP0373	~70%	~85%	~545	Putative outer membrane protein (J99 annotation)
								Conserved hypothetical protein (26695 annotation)
a P. C.	Clone 47	476	jhp1103	HP1177	~82%	~96%	~463	Putative outer membrane protein

^aRefers to percentage identity of the DNA sequence of the particular clone to the corresponding ORFs of strains J99 and 26695.

^bRefers to the number of nucleotides of sequenced DNA that base pair with designated ORFs from strains J99 or 26695. For example, approximately 380 nucleotides of sequenced DNA of Clone 1 from strain pair MI 355/MI 356 contained no homologous sequence in either *H. pylori* strain J99 or 26695 or when an NCBI BLAST search was conducted.

^cThe DNA sequence of Clone 4 from isolate pair MI 506/ MI 520 contains no significant identity or similarity to any ORF sequence from either *H. pylori* strain J99 or 26695. It has a G+C% of 27%.

^d Using BLAST, Clone 1 from this isolate pair gave highest identity scores with different ORFs from strains 26695 and J99.

^e Information obtained from sequence submitted to GenBank by McGee et al. (1999).

^f DNA sequence of this clone spans two ORFs.

^gNow not believed to have a gene regulatory function based on functional characterisation of *hypF* from *R. capsulatus* (see text).

follows that the likely adaptive significance of rearrangement in Omp-encoding loci, for example, is heightened if these events are not peculiar to a single paired isolate set.

The same probes that were employed for the screening of the original strains (Fig. 4.5) were used in Southern hybridisation with *Hin*dIII-cleaved genomic DNA from a panel of paired clinical isolates, with the exception of the three clones isolated by PCR-subtractive hybridisation of isolate pair MI 355/MI 356. This analysis revealed that rearrangement at a particular ORF was not peculiar to a single set of paired isolates, clone 4 from isolate pair MI 506/MI 520 notwithstanding. Several probes hybridised to numerous isolate pairs (Fig. 4.6 and Table 4.2) with a pattern of hybridisation that was taken as evidence of the occurrence of rearrangement within that particular locus in the given strain pair. The details of the Ompencoding loci isolated by PCR-subtractive hybridisation and found among other paired isolate sets are given in Table 4.3.

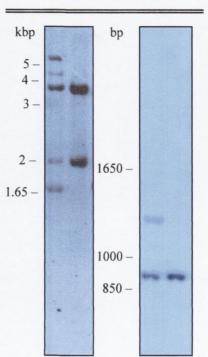
4.2.7. MI 571/MI 575 - Clone 4.

In strain pair MI 571/MI 575 the DNA sequence of the genomic fragment present in clone 4 encodes a putative type II restriction endonuclease gene and is present in both strain 26695 (locus number HP1471) and strain J99 (locus number jhp1364). The restriction/modification (R-M) systems present in H. pylori are extensive and highly diversified among strains (Xu et al., 2000b; Lin et al., 2001). The majority of H. pylori R-M systems, which themselves comprise more than 4% of the genome, are of Type II and consist of two separate enzymes, the restriction endonuclease and the methyltransferase (Lin et al., 2001). A total of 22 R-M systems have been identified among H. pylori strains with 18 distinct specificities (Xu et al., 2000b) and are believed to play an important role in *H. pylori* by preventing the insertion into the chromosome of exogenous DNA that might be acquired through the natural competency of the organism (see Section 1.16). Studies have found that the presence of an R-M system in an isolate does not imply its functionality. Indeed, Lin et al. (2001) indicated that the genes of the particular R-M system encoded at locus HP1471/jhp1364 and respective adjacent genes showed strong similarity to the genes of the BcgI R-M system and that they were inactive in both strains. Furthermore, this R-M system is pointed out to be unusual in that it contains an M gene fused with an R gene, with a separate gene encoding an S subunit of the enzyme which confers target sequence specificity. The system might also be subject to phase variation by virtue of the presence of a run of 14 G-residues in the coding sequence of locus HP1471/jhp1364.

A.
HP1157/jhp1084 (clone 6)
encodes HopL Omp

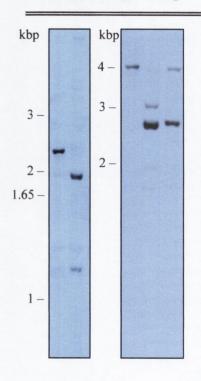
kbp kbp kbp 4 -3 -1.65 -0.85 -0.85 -0.85 - B.

HP1177/jhp1103 (clone 47) encodes HopQ Omp



C.

HP0373/jhp1008 (clone 41) encodes HomC Omp



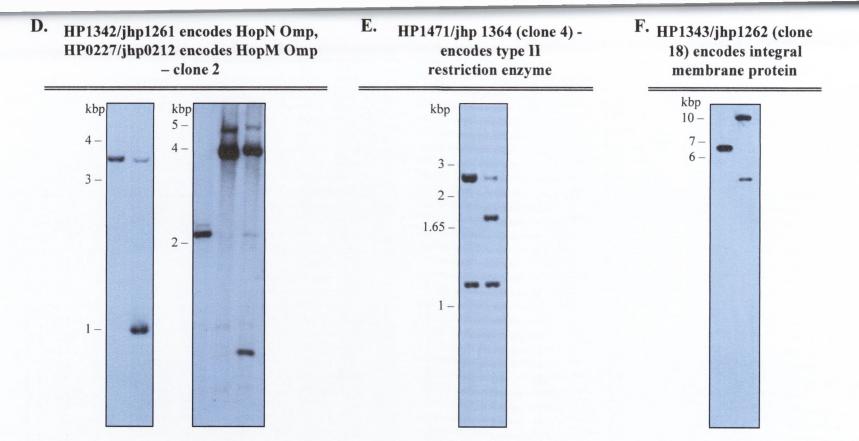


Fig. 4.6. Examples of *H. pylori* isolate sets from the Moyne Institute collection identified by Southern hybridisation as having undergone rearrangement at the indicated loci.

A – E are Southern blots of *Hin*dIII-restricted genomic DNA from *H. pylori* paired isolate sets using DIG-labelled probes of the indicated clones (isolated originally by PCR-subtractive hybridisation). (A.) From left to right, the *H. pylori* strains probed in each panel are as follows: MI 745A/MI 745C, MI 610/MI 611, MI 334A/MI 334C, MI 688A/MI 688C, MI 683A/MI 683C. (B.) The isolates probed in the left panel are MI 545A/MI 545C and in the right panel, MI 862A/MI 862C. (C.) *H. pylori* paired isolates MI 545A/MI 545C (left panel) and isolates MI 561/MI 562 and MI 563 (right panel). (D.) Panels correspond to isolates probed in (C.). (E. & F.) *H. pylori* isolate pair MI 458/MI 459. The approximate position of bands in the DNA molecular size ladder (bp and kbp) are shown to the left of each panel.

Table 4.2. Data obtained from the analysis of the *H. pylori* paired isolate collection for the existence of polymorphisms at the loci identified by PCR-subtractive hybridisation among other paired strains.

Isolate pair where polymorphism was originally identified	Clone name	No. of paired strains examined	Total no. of paired strains showing polymorphism (%) ^a	Names of paired strains in question ^b
MI 458/MI 459	Clone 2	43	3 (6.97)	MI 458/MI 459 MI 545A/MI 545C MI 561/MI 562/MI 563
	Clone 6	25	7 (28)	MI 334A/MI 334C MI 458/MI 459 MI 571/MI 575 MI 610/MI 611 MI 683A/MI 683C MI 688A/MI 688C MI 745A/MI 745C
MI 571/MI 575	Clone 4	18	2 (11.1)	MI 458/MI 459 MI 571/MI 575
	Clone 41	25	3 (12)	MI 545A/MI 545C MI 561/MI 562/MI 563 MI 571/MI 575
	Clone 47	25	3 (12)	MI 545A/MI 545C MI 571/MI 575 MI 862A/MI 862C
MI 506/MI 520	Clone 4	9	1 (11.1)	MI 506/MI 520
	Clone 18	17	2 (11.76)	MI 458/MI 459 MI 506/MI 520
MI 355/MI 356	Clone 3 Clone 9	N.D.°	N.D.	N.D.

^aPercentage values (in brackets) refer to the value obtained when the total number of strains showing polymorphism at a particular locus was expressed as a percentage of the total number of paired isolates examined for such a polymorphism.

^bNames of isolate pairs are depicted in coloured font where paired isolates were identified by Southern blot screening to be polymorphic at two loci.

^cN.D. = not determined.

Table 4.3. Details of Omps encoded by alleles identified by PCR-subtractive hybridisation as having undergone rearrangement between the strains in the given *H. pylori* isolate pairs (all information obtained from Alm *et al.*, 2000).

Strain pair	Clone name	Locus number		Length (aa)		% Size variation	Omp group	C-terminal residues	Gene name
		26695	J99	26695	J99	(no. of aa)			
458/459	Clone 6	1157	1084	1,230	1,237	0.6 (7)	Нор	MGF	hopL
458/459	Clone 2	1342	1261	691	696	0.7 (5)	Нор	FAY	hopN
		0227	0212	691	696	0.7 (5)	Нор	FAY	hopM
571/575	Clone 47	1177	1103	641	643	0.3 (2)	Нор	FAY	hopQ
571/575	Clone 41	0373	1008	700	751	6.8 (51)	Hom	WVF	homC

4.2.8. MI 355/MI 356 - Clone 1.

Approximately 568 bp of DNA sequence was determined for clone 1 isolated by PCR-subtractive hybridisation of isolate pair MI 355/MI 356. BLAST analysis of this sequence against the genomes of strains 26695 and J99 revealed that only one-third of this clone had significant identity to DNA sequences in each of these strains. Furthermore, the highest identity scores in strains 26695 and J99 were obtained from different ORFs (see Table 4.1). The remaining almost 400 bp of sequence from this clone contained no significant identity to chromosomal or ORF sequences in either strain. In strain 26695, 184 bp of clone 1 had 99% identity to the locus HP0548. This locus is specific for strain 26695 with no homologous counterpart present in strain J99. In strain J99, however, 76% identity between clone 1 and locus jhp1213 was obtained, with the same identity reported for the corresponding locus in strain 26695 (HP1293). The HP1293/jhp1213 ORF putatively encodes RpoA, the DNA-directed RNA polymerase α-subunit.

Locus HP0548 is thought likely to encode a DNA helicase. Such enzymes catalyse the unwinding of double-stranded DNA during replication, recombination and repair and are found ubiquitously in bacteria (Lohman, 1992). In a study carried out by McGee *et al.* (1999), the putative DNA helicase-encoding gene designated HP0548 was found to enhance urease production in *E. coli* cells co-transformed with a cloned urease gene and a *H. pylori* DNA library. The HP0548 locus is located at one extremity of the *cag* pathogenicity island (*cag*PAI) immediately upstream of the glutamate racemase gene (HP0549) into which the *cag*PAI is inserted. Both isolates MI 355 and MI 356 are *cagA*⁺, this gene serving as a marker for the presence of the *cag*PAI. As discussed in the preceding chapter, the structure of the *cag*PAI has been reported to vary between isolates (Censini *et al.*, 1996). Both chromosomal DNA and the insertion sequence element IS605 have been found within the *cag*PAI, thereby segmenting it into two, as well as structures consisting of the *cag*PAI unit with partial deletions of some coding sequence. The DNA sequence submitted to GenBank by McGee *et al.* (1999) showed identity between clone 1 and locus HP0548 as well as with what is termed mini-IS605 encoding sequence.

4.2.9. MI 506/MI 520 - Clone 4.

The DNA sequence of this clone contains no significant identity or similarity to any ORF from either strain 26695 or strain J99. However, over lengths of no greater than 29 bp it does

display identities in the region of 93-100% with these strains. In its entirety, clone 4 has a G+C% of only 27% which is at considerable variance to the rest of the *H. pylori* genome, the G+C% content of which is ~39% as referred to above. Screening of a panel of nine paired isolates also revealed that the novel DNA sequence present in clone 4 is not found in any other set of strains apart from MI 506/MI 520. Indeed, when using a probe made from clone 4, a pattern of hybridisation was detected within this pair that was taken as evidence of the occurrence of a rearrangement between isolates MI 506 and MI 520, thus explaining the isolation of clone 4 by PCR-subtractive hybridisation. This resulted in the pattern of hybridisation observed by Southern blotting displayed in Fig. 4.5.B.

4.2.10. MI 355/MI 356 - Clone 3.

This clone consisted of DNA sequence that spanned two adjacent ORFs, namely, the gene designated *hypF* that was originally annotated to encode a putative transcriptional regulator of the hydrogenase operon because of its sequence similarity to the *hypF* gene in *Rhodobacter capsulatus* (Tomb *et al.*, 1997; Alm *et al.*, 1999). The *R. capsulatus hypF* gene has since been shown to be involved in maturation of the hydrogenase subunits by nickel incorporation (Scarlato *et al.*, 2001) and therefore has no function at the transcriptional level. The second ORF spanned by the sequence of clone 3 is involved in hydrogenase expression/formation and is encoded by the gene designated *hypE*.

4.2.11. MI 355/MI 356 - Clone 9.

In sequencing the genomes of strains 26695 and J99 it was not always possible to assign a function to certain putative ORFs because no significant identity existed between the *H. pylori* gene sequence and any annotated gene sequence previously deposited in gene sequences databases. In such instances, these putative protein-coding sequences were designated 'conserved hypothetical proteins' (Tomb *et al.*, 1997; Alm *et al.*, 1999). Luckily perhaps in the course of the present study, clone 9 isolated by PCR-subtractive hybridisation of isolate pair MI 355/MI 356 was the only gene fragment that belonged to an ORF which was classified in both strains 26695 and J99 as encoding a conserved hypothetical protein.

4.2.12. Membrane protein-encoding gene fragments.

The remaining DNA fragments isolated by PCR-subtractive hybridisation in isolate pairs MI 571/MI 575, MI 458/ MI 459 and MI 506/MI 520 all consisted of coding sequence of ORFs that putatively encode membrane proteins. Clone 18 from isolate pair MI 506/MI 520

encodes a protein likely to be located in the inner membrane. The remaining DNA fragments are from genes which code for outer membrane proteins (Omps).

4.2.12.1. MI 506/MI 520 - Clone 18.

The DNA fragment designated clone 18 was isolated by PCR-subtractive hybridisation of isolate pair MI 506/MI 520. It contains sequence with high identity to locus HP1343 from strain 26695 which corresponds to locus jhp1262 in strain J99. Rearrangement at the same locus was also detected between isolate pair MI 458/MI 459 (Fig. 4.6). The protein encoded at this locus is thought to be located in the inner membrane according to the annotation made by TIGR (Table 4.1).

Proteins of the inner membrane in Gram-negative bacteria serve a pivotal role in the cell by acting as key players in the translocation of molecules into and out of the intracellular environment. For example, protein export is required for surface expression of virulence factors, nutrient acquisition and release of effector proteins (Ilver *et al.*, 2001). As part of the general secretion pathway, in concert with periplasmic and outer membrane proteins, they mediate the translocation of toxins and hydrolytic enzymes out of the cell (Sandkvist, 2001). The import from the extracellular environment of nutrients and other metabolic molecules necessitates the negotiation of the inner membrane and is facilitated in certain instances by the proteins located there. Moreover, as members of what are known as two-component signal transduction systems, inner membrane proteins acting as environmental sensors convey this signal to response proteins in the cytoplasm where an intracellular phosphorylation cascade is triggered (Hoch, 2000) and a variety of adaptive cellular responses elicited at many levels.

In *H. pylori* specifically, UreI is the only inner membrane protein that has been characterised in any detail. This protein regulates cytoplasmic urease activity by acting as a urea-specific transporter that translocates the urease substrate only under conditions of low pH (Rektorschek *et al.*, 2000; Weeks *et al.*, 2000). The ability of *H. pylori* to hydrolyse urease is intimately connected with its pathogenicity and thus also is UreI. The CagA antigen, perhaps the single most important pathogenicity determinant of *H. pylori*, is translocated out of the cell and into close contact with cells of the host gastric epithelium by a type IV secretion apparatus (Backert *et al.*, 2000). Proteins located in the inner membrane have an important role to play, therefore, in the interaction between CagA and host.

4.2.12.2. *H. pylori* Omps.

Three of the four Omp-encoding loci found by PCR-subtractive hybridisation to have undergone rearrangement between paired isolates encode Hop proteins (clones 2 and 6 from isolate pair MI 458/MI 459 and clone 47 from isolate pair MI 571/MI 575), while the fourth locus is from what is called the Hom (*H. pylori* outer membrane proteins) family of proteins (clone 41 from isolate pair MI 571/MI 575). This family is considerably smaller than that which comprises the Hop and Hor proteins (see below) and has three members in strain 26695 and four in strain J99, with one strain-specific member present in this strain. Proteins of the Hom family have conserved N- and C-termini with variability existing in the central domain (Alm *et al.*, 2000).

Annotation of the genome sequences of *H. pylori* strains 26695 (Tomb *et al.*, 1997) and J99 (Alm *et al.*, 1999) allowed the identification of a large family of Omp-encoding loci. Prior to the availability of complete genome sequences, a family of five Omps that shared N-terminal sequence similarity had been identified (HopA to HopE) and were shown to function as porins (Doig and Trust, 1994; Doig *et al.*, 1995). These were the original members of the major *H. pylori* Omp family that now includes an additional 16 Hop proteins and 12 Hor (Hop related) proteins (33 proteins in total). Although the HopA–E proteins were identified by N-terminal sequence homology, it is now apparent that, among the Hop/Hor proteins, the majority of the sequence similarity exists at their C-terminal ends (Alm *et al.*, 2000). This amino acid sequence similarity is presumably reflected to some extent in the gene sequences that encode these proteins.

4.3. Discussion.

4.3.1. Molecular fingerprinting of paired H. pylori isolates.

Short oligonucleotide probes containing eukaryotic microsatellite sequences were first utilised as a molecular fingerprinting tool to differentiate eukaryotic micro-organisms, including species of the yeast genus Candida, and the prokaryote Mycobacterium tuberculosis (Sullivan et al., 1995; Thierry et al., 1995). When applied to HindIII-restricted genomic DNA of H. pylori, such probes yielded reproducible and highly discriminatory DNA fingerprints of the isolates examined (Marshall et al., 1996). The existence of paired H. pylori isolates with minor differences in fingerprint profile when subjected to molecular typing has been observed by several groups (Oudbier et al., 1990; Kersulyte et al., 1999; Kuipers et al., 2000; Israel et al., 2001). Using oligofingerprint analysis of paired H. pylori isolates, similar findings were made and were referred to as microevolution (Marshall, 1996). Both mutation and a high horizontal exchange frequency are believed to be responsible for the genomic diversity exhibited by H. pylori (Suerbaum et al., 1998) but the origins of the diversity observed between epidemiologically related isolates have not been well characterised. subtypic variants are considered most likely to arise as a result of genomic rearrangements taking place within or between H. pylori isolates during the prolonged period of colonisation of the gastric environment by this organism.

Oligofingerprint analysis of the paired isolates MI 458/MI 459, MI 571/MI 575, MI 506/MI 520 and MI 355/MI 356 using DIG-labelled oligonucleotide probes (GGAT)₄ and (GTG)₅ and RAPD-PCR fingerprinting of the same strains (Fig. 4.1.A and B) testified to the clonality of the isolates in paired sets, with minor band differences indicating that genetic rearrangement of some description had taken place between them as was found previously (Marshall, 1996). Oligofingerprint profiles were consistently found to be reproducible as were RAPD-PCR profiles, however, the clarity of profiles obtained for RAPD-PCR depended strongly on the quality and quantity of template DNA used. Additionally, *in vitro* culture was not found to result in genetic changes in the genome of *H. pylori* as determined by RAPD-PCR analysis of genomic DNA isolated from a strain that had been multiply passaged in liquid medium (data not shown). Similarly, it has previously been shown using oligofingerprinting that *in vitro* culture did not result in the generation of genotypic diversity in *H. pylori* strains (Marshall, 1996). Interestingly, when *Hind*III-restricted genomic DNA of the same isolates was probed with three *H. pylori* gene probes (*vacA*, *cagA* and *flaA*), RFLPs were observed between the isolates in strain pair MI 506/MI 520 with all probes used. This was a somewhat surprising

observation given that both oligofingerprinting and RAPD-PCR indicated that, although mnor fingerprint differences existed between them, these two strains were almost identical. It also points to the care that must be exercised in the interpretation of RFLP data as a method of strain typing. Had RFLP analysis alone been carried out, isolates MI 506 and MI 520 would have been classified as genotypically unrelated and this patient deemed to have a mxed infection. For the purposes of the present study, because oligofingerprinting is a genome-wide fingerprinting method, this pair are considered as epidemiologically related isolates that have undergone microevolution and in which polymorphisms exist at the cagA, vacA and flaA loci. It also indicates that fingerprinting/typing methods that are not confined to the analysis of one gene might give a more accurate appraisal of strain relatedness. By the sane token, if RFLP analysis is utilised as a method of strain typing, the use of gene probes to krown virulence genes, as has been described herein, might detect genotypically evidence for the occurrence of selective pressure on the proteins encoded by these genes (i.e., a rearrangement within the gene becoming amplified in the population as a reflection of selective pressure on the protein). Thus, probing Southern blots with probes made to housekeeping genes might have been a more suitable approach.

4.3.2. PCR-subtractive hybridisation.

The existence of isolates between which in vivo genomic rearrangement has taken place might be considered a snap-shot in the development of genetic diversity in H. pylori as these events unloubtedly contribute to the diversity seen in the species as a whole. More importantly, hovever, a unique ability to define precisely the nature and significance of the genetic changes that take place in vivo in H. pylori is afforded because the polymorphisms exist between strains that are clonal in origin. In this way, it is possible to decipher whether the genetic changes that have occurred in vivo are likely to be pathoadaptive mutations selected for by the changes that take place in the microenvironment inhabited by H. pylori as disease progresses or whether they reflect the panmictic, freely recombining nature of the species and have, therefore, little adaptive consequence. Thus, towards a better understanding of the way H. pylori behaves in the stomach, its high genomic diversity and the nature of the putative adaptive changes made in the genomes of isolates during infection, the present studies sought to dentify the loci or chromosomal DNA fragments involved in the generation of subtypic variants. Insights about the specific nature of these changes might also help elucidate how H. pybri fine-tunes its virulence and explain the more unusual aspects of its disease pathogenesis.

PCR-subtractive hybridisation was employed to isolate the DNA fragments which differed between microevolved pairs. The technique was designed for the isolation of genetic differences between two closely related genomes and was used to analyse two unrelated *H. pylori* isolates in its original application (Akopyants *et al.*, 1998b). It is the successor to the original subtractive hybridisation technique (Straus and Ausubel, 1990) and indeed the more recent 'Representational Difference Analysis' methodologies. It has several advantages over both pre-existing techniques including the requirement for only very small quantities of genomic DNA from each strain. Moreover, complicated adaptor removal and readdition with successive rounds of PCR amplification are not necessary nor is the physical removal of tester/driver complexes. Furthermore, DNA fragments that partially differ between tester and driver strains can be recovered and there is a built-in ability to amplify by PCR the fragments remaining after the two genomes have been subtractively hybridised (Akopyants *et al.*, 1998b).

PCR-subtractive hybridisation was employed in the present studies to directly analyse four sets of paired *H. pylori* isolates. The DNA fragments thus isolated were then used to screen a large panel of paired clinical strains for the presence of rearrangements within the same ORFs in other epidemiologically paired isolates (Table 4.1; Figs. 4.5.A and B and Fig. 4.6). PCR-subtractive hybridisation and Southern blotting allowed the identification of 10 genomic fragments among the four paired isolate sets examined that appeared to be rearranged between pairs. The annotated genome sequences of *H. pylori* strains 26695 and J99 were used to identify the loci involved and the putative proteins encoded by these genes. The potential implications and possible adaptive significance of rearrangement at each of these identified fragments will now be discussed in turn, however, the potential limitations of the PCR-subtractive hybridisation methodology will first be scrutinised.

The analysis of rearranged DNA structures in Fig. 4.5 and Fig. 4.6 suggest that most of these are likely to result in hybridisation and subtraction during PCR-subtractive hybridisation. Thus, using this methodology, these gene fragments have become amplified but not necessarily as a result of selection and enrichment. PCR-subtractive hybridisation is expected to identify DNA sequences that are missing entirely in the driver strain, or, sequence that is sufficiently divergent along its whole length to preclude hybridisation in the reasonably stringent conditions utilised.

In order to prove that the methodology has indeed identified rearranged genes or gene fragments, two possibilities are proposed. First, if primers internal to the rearranged fragments were used to amplify genomic DNA from both strains and these amplimers were then used to probe both strains in Southern hybridisation experiments, an identical pattern of hybridising bands would be expected if PCR-subtractive hybridisation was successful. Secondly, the PCR could be used to amplify similarly sized fragments from genes not identified using this methodology and again use the obtained amplimers in Southern hybridisation. These would also be expected to give identical Southern blot profiles and would, in such a scenario, indicate that PCR-subtractive hybridisation had worked efficiently.

4.3.3. ORFs identified by PCR-subtractive hybridisation to be involved in the generation of microevolution.

4.3.3.1. MI 571/MI 575 - Clone 4.

The genes of *H. pylori* R-M system homologues were found to have a G+C% content that was lower than that of other chromosomal genes and some of them displayed a different codon use bias to average genes (Nobusato *et al.*, 2000a) suggesting that *H. pylori* R-M systems were acquired by horizontal transfer (Nobusato *et al.*, 2000a; Xu *et al.*, 2000a). Indeed, a model of R-M system acquisition by horizontal gene transfer and mutational inactivation of the system followed by complete loss of these inactive systems has been proposed (Lin *et al.*, 2001).

The finding of rearrangement within a locus encoding a Type II restriction endonuclease gene in each of two paired isolate sets (MI 571/MI 575 and MI 458/MI 459, Figs. 4.5 and 4.6 and Table 4.2) might indicate that, if the rearrangement event is adaptive in purpose, it is a rearrangement to a functional or non-functional R-M system. However, there are no studies that report non-functional R-M system loci mutating to become active systems. Without characterisation of the rearrangement event at locus HP1471/jhp1364, such hypotheses are purely speculative, and may *in vivo* not be a plausible explanation for what has taken place in these paired *H. pylori* isolates. Indeed, one envisages the involvement of promoter sequences in dictating if a gene is functional or not, but there is also the possibility that nucleotide substitution or mutation of the coding region might alter enzyme functionality by affecting residues of the active site.

Within the coding sequence of the ORFs HP1471/jhp 1364, there exists a repeated sequence of 59 bp in length, separated by 584 and 560 bp of DNA in strains 26695 and J99,

respectively. It is also possible that local recombination by cross-over between these two direct repeats, as depicted schematically in Fig. 4.7, is responsible for the rearrangement observed, with deletion of the intervening region of chromosomal sequence occurring and one repeat copy remaining (Petes and Hill, 1988). This entire 59-bp sequence is not located elsewhere in the genomes of strains 26695 or J99 and contains no significant identity to any DNA sequence deposited in the database at the NCBI. This suggests that ancestral duplication of this sequence took place and that these sequences may now be used as substrates for intrachromosomal recombination.

4.3.3.2. MI 355/MI 356 - Clone 1.

In PCR-subtractive hybridisation of isolate pair MI 355/MI 356, strain MI 356 served as the 'tester' strain, the strain in which unique DNA was being sought and from whose genome that of strain MI 355 was being subtracted. In all four sets of paired isolates analysed by PCRsubtractive hybridisation, it was mostly DNA fragments that only partially differed between the two strains that were isolated. Bona fide unique DNA, i.e., DNA only present in one isolate from a pair, was isolated less frequently. However, the DNA fragment in clone 1 of which two-thirds consisted of novel DNA sequence as well as coding sequence from the cagPAI (HP0548) and also mini-IS605, was found by Southern blotting to be one such example and to be unique to isolate MI 356 (Fig. 4.5.B). Over its entire 568 bp, the G+C% of clone 1 is 38%, the same approximately as that calculated for the entire H. pylori genome. Thus, there is no reason to suspect that the novel DNA sequence contained in clone 1 is not of Helicobacter origin. Taken together, the presence of clone 1 DNA solely in isolate MI 356 and the association of the assigned sequence of the clone with sequences located in the cagPAI (HP0548) and with IS605, as well as the existence of a large amount of novel DNA adjacent to these, suggests the involvement of the IS element in the loss of the DNA region from strain MI 355.

Rearrangement at locus HP0548 is less easily explained, such as in the role found by McGee et al. (1999) in the enhancement of urease activity by the putatively encoded DNA helicase. Urease production in *H. pylori* is regulated at the level of transcription in response to pH and other factors (Akada et al., 2000; van Vliet et al., 2001). The probable modulation of the expression of the helicase function as a result of the rearrangement observed herein and the downstream impact this has on urease activity was probably of consequence in the pathogenesis of disease caused by isolates MI 355 and MI 356. However, it is more difficult

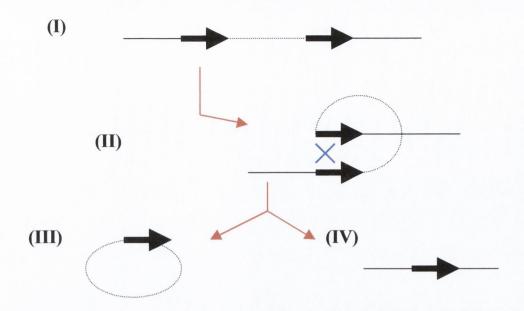


Fig. 4.7. Schematic representation of the occurrence of a chromosomal rearrangement by cross-over between directly repeated sequences leading to a deletion.

(I) The black arrows represent the 59-bp directly repeated sequences in locus HP1471/jhp1364. (II) Intrachromosomal cross-over between these repeats could lead to deletion of the intervening sequence (depicted as a dotted line) from the chromosome, with the sequence located there now on an excised circular molecule with one copy of the repeat (III). (IV) The chromosome now contains only one repeat copy (Adapted from Petes and Hill, 1988).

to rationalise how this rearrangement might be widely relevant when urease activity is in question.

4.3.3.3. MI 506/MI 520 - Clone 4.

As detailed in the results section, the DNA sequence of clone 4 has a G+C% of 27% and is not significantly identical to any ORF in either strain 26695 or strain J99. These pieces of information strongly suggest that this DNA fragment was acquired by horizontal transfer from a heterologous species, became inserted into the genome of the founder strain from which the subtypic variants MI 506/MI 520 arose and there underwent recombination, perhaps utilising the aforementioned short regions of sequence identity located around the *H. pylori* chromosome (Section 4.2.9) as substrates for recombination or indeed integration of the sequence following horizontal transfer of the fragment in the first instance. Interaction between small (15 bp or less) dispersed repeats causing rearrangement events has been detected in *E. coli* (Petes and Hill, 1988) and thus lends weight to the possibility that in *H. pylori* recombination might be permissible over regions of identity from 19–29 bp.

4.3.3.4. MI 355/MI 356 - Clone 3.

Bacteria that produce hydrogenase can oxidise hydrogen gas to produce energy, with the hydrogenase enzyme serving to catalyse the oxidation step. The association of HypE and HypF with energy metabolism/electron transport suggests that a genomic rearrangement event in which they are involved might not be selected for *in vivo* by the sequelae of prolonged colonisation by *H. pylori*. Their putative cellular functions notwithstanding, the intracellular location of these proteins makes discerning the possibility of their being subject to adaptive mutation difficult.

4.3.3.5. MI 355/MI 356 - Clone 9.

The third clone (designated clone 9) isolated by PCR-subtractive hybridisation of isolates MI 355/MI 356 them consisted of sequence that showed high identity to that annotated for a conserved hypothetical protein in both strains 26695 and J99 (loci HP1417 and jhp1312, respectively). Without the availability of functional information for this ORF it is impossible to speculate about how or why rearrangement might have occurred *in vivo* at this locus. It is interesting to note, however, that the *Hin*dIII-restriction fragment to which the clone 9 probe hybridises in strain MI 355 is the same as the restriction fragment to which the clone 3 (discussed above) probe hybridises (Fig. 4.5.B). It is, therefore, tempting to speculate that

between isolates MI 355 and MI 356 rearrangement at the loci encoded by clone 3 and clone 9 occurred from, or to, the same chromosomal location.

It is important to point out, however, that the pattern of hybridisation observed by Southern blotting with probes for clone 3 and clone 9 (Fig. 4.5.B) can also be explained by the occurrence of a point mutation which either creates or inactivates a *HindIII*-restriction site, thereby allowing the probe to hybridise to restriction fragments that are shorter or longer, respectively as depicted schematically in Fig. 4.8. Although the occurrence of rearrangement at these loci is an infinitely more interesting possibility than the accounting for the Southern blot evidence by spontaneous point mutation, the latter cannot be ignored, especially given the propensity of the *H. pylori* genome to undergo mutation (Wang *et al.*, 1999b).

4.3.3.6. MI 506/MI 520 - Clone 18.

Although lacking functional characterisation and with its inner membrane protein annotation a putative one, is it possible that rearrangement at locus HP1343/jhp1262 (and by extension, at loci encoding inner membrane proteins in general) is a mechanism whereby *H. pylori* exerts an additional dimension of control over the molecular translocation properties of this group of proteins? Given that rearrangement at this locus was found in only two out of 17 paired strains examined and that, in the entire analysis of microevolution by PCR-subtractive hybridisation herein, this was the only locus encoding an inner membrane protein that was detected in which rearrangement was apparent, it seems unlikely that rearrangement occurred at this locus in isolate pairs MI 506/MI 520 and MI 458/MI 459 because of adaptive pressure.

It is interesting to note that the finding of what is perceived as rearrangement at locus HP1343/jhp1262 in isolate pair MI 458/MI 459 was not detected by PCR-subtractive hybridisation of these strains, but was instead recognised by Southern blot screening of the paired isolate collection using clone 18 as the probe, itself isolated by PCR-subtractive hybridisation of isolate pair MI 506/MI 520. Rearrangement in the genes upstream of these loci (HP1342/jhp1261) was on the other hand detected in isolates MI 458/MI 459 by applying subtractive hybridisation and thus, both adjacent genes are implicated in the rearrangement of the genome that has taken place at this extended chromosomal site.

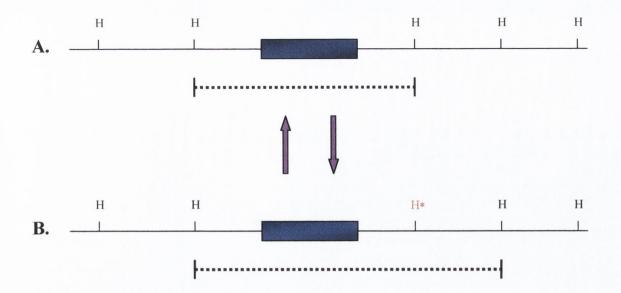


Fig. 4.8. Schematic representation of possible cause of the hybridisation pattern observed for clones 3 and 9 isolated by PCR-subtractive hybridisation of isolates MI 355 and MI 356.

The probe is represented by the shaded rectangle. 'H' refers to the position of *Hin*dIII restriction sites and the dotted line represents the size of the restriction fragment to which the probe hybridises. In (A), hybridisation of the probe to a smaller restriction fragment takes place. Removal or creation of a restriction site (H*) by a spontaneous point mutation would cause the same probe to hybridise to a longer restriction fragment (as seen between isolates MI 355/MI 356 for clone 3 and clone 9) as depicted in (B). It follows, therefore, that no physical rearrangement of the DNA sequence to which the probe hybridises is required to bring about the hybridisation pattern observed by Southern blotting (Fig. 4.5. B).

4.3.3.7. MI 571/MI 575 - Clone 41.

PCR-subtractive hybridisation of strain pair MI 571/MI 575 resulted in the isolation of two ORFs that encoded Omps. The DNA fragment designated clone 41 from isolate pair MI 571/MI 575 contained sequence with high identity to the *homC* gene encoded at loci HP0373/jhp1008. Rearrangement involving this DNA sequence was also detected among other isolate sets, namely, the strain pair MI 545A/MI 545C and between the isolates MI 561/MI 562/MI 563 (Fig. 4.6). Interestingly, the observation that probes to clones 3 and 9 from strains MI 355/MI 356 hybridised to the same *HindIII* fragment from strain MI 355 had parallels in isolates MI 571/MI 575 when considering clones 41 and 47. The *HindIII* fragment of between 1.65 and 2 kb in isolate MI 571 from which the most intense signal was obtained for clone 47 (Fig. 4.5. A) is the same fragment to which the clone 41 probe partially hybridises. Given that clone 47 contains DNA sequence of the *hopQ* gene that specifies the putative HopQ porin and that strain MI 575 does not contain this 1.65–2-kb hybridising fragment with either probe suggests that this may be evidence to indicate that the genes that encode both these Omps (*homC* and *hopQ*) had undergone rearrangement with each other.

4.3.3.8. MI 571/MI 575 - Clone 47.

As referred to in the discussion of rearrangement of the *homC* gene (clone 41) from this isolate pair, a second clone (clone 47) was isolated by subtractive hybridisation of isolates MI 571/MI 575 that consisted of sequence for the *hopQ* gene. Rearrangement of *hopQ* DNA sequences was also detected between the paired isolates MI 545A/MI 545C and MI 862A/MI 862C by Southern blotting (Fig. 4.6).

4.3.3.9. MI 458/MI 459 - Clone 2.

Within strain pair MI 458/MI 459 rearrangements were found in two Omp-encoding loci. Clone 2 from isolate pair MI 458/MI 459 consists of sequence of an ORF that exists within the genomes of both strains 26695 and J99 as a duplicated gene. These genes encode the Omps designated HopM (HP0227/jhp0212) and HopN (HP1342/jhp1261). In strains 26695 and J99 the chromosomal location of the two alleles are identical and are flanked in both strains by ORFs which encode the same putative proteins. It is likely that the occurrence of this gene duplication event took place early in *Helicobacter* evolution, given its initial finding in two strains from different geographic locations and from patients who had gastric pathology of differing severity.

Investigations carried out by Alm et al. (2000) in which 19 H. pylori strains were analysed for duplicate copies of the hopM/N genes revealed that in four isolates the N-terminal region of locus jhp1261 had been deleted. However, in all strains examined, duplicated alleles were found and at the same chromosomal locations as existed in strains 26695 and J99. It seems unlikely, therefore, that the occurrence of rearrangement in isolate pair MI 458/MI459 involving DNA homologous to alleles HP0227/jhp0212 and HP1342/jhp1261 might be due to a gene duplication event. However, the pattern of hybridisation of probe DNA for clone 2 to genomic digests in Southern blotting suggests that this could be the case, namely, one hybridising band in isolate MI 459 indicating the presence of one allele perhaps, while two hybridising bands in isolate MI 458 might be evidence for allele duplication. It is also possible and probably more plausible given that duplicate alleles were found in all isolates examined, that what has been observed in isolate pair MI 458/MI 459 is the occurrence of rearrangement within existing duplicated genes, e.g., homologous recombination between repeated chromosomal sequences. These possibilities do not to rule out other mechanisms of site-specific DNA rearrangement that could also account for the rearrangement observed. These could include the involvement of insertion sequence elements in the excision and relocation of DNA or indeed, the acquisition of exogenous DNA by horizontal transfer followed by recombination with the chromosome at or adjacent to the alleles in question.

4.3.3.10. MI 458/MI 459 - Clone 6.

Clone 6 from isolate pair MI 458/MI 459, the second Omp-encoding locus, contains sequence of the *hopL* gene at loci designated HP1157/jhp1084. This gene was found to be involved in rearrangement in an additional seven sets of paired isolates making it the gene which appears to undergo rearrangement between paired isolates with the greatest prevalence of all the loci identified by subtractive hybridisation (Fig. 4.6.A and Table 4.2). Table 4.3 details the comparison between strains 26695 and J99 of the Omp-encoding proteins that were isolated in the present study by PCR-subtractive hybridisation.

The finding that rearrangement occurring between paired isolates involved molecules located on the bacterial cell surface was not a surprising one. As previously mentioned, if the occurrence of microevolution, even sometime, is evidence for the adaptation of *H. pylori* to a changing microenvironment, it would seem most sensible and in the best interests of the bacterium to alter surface-exposed molecules, perhaps to exacerbate the chronicity of

infection by evading an immune response that may have been mounted to antigenic epitopes displayed on the pre-existing protein.

Several examples exist among bacterial genera of organisms that rearrange their chromosomal DNA in order to achieve antigenic variation of surface molecules. The *vsa* locus of *Mycoplasma pulmonis* can rearrange silent copies of the *vsa* allele such that they are relocated at a transcriptionally active 5'-end region that contains a promoter and ribosome-binding site. Antigenic variation is achieved by reassortment of 3'-regions from transcriptionally silent alleles with the 5'-end region from the expressed gene (Bhugra *et al.*, 1995). Rearrangement of flagellin genes in *Proteus mirabilis* causes the expression of a hybrid flagellin protein that is antigenically distinct from the wild-type FlaA and FlaB proteins. Recombination occurring between homologous regions of the *flaA* and *flaB* genes places the promoter and 5'-region of the *flaA* gene adjacent to previously silent regions of the *flaB* gene (Murphy and Belas, 1999).

Both of these mechanisms of DNA rearrangement and antigenic variation are reminiscent of the classical example of such a mechanism of surface molecule variation described for the type 4 pili of *Neisseria gonorrhoeae*. Intragenic recombination occurs to juxtapose silent copies of the *pilS* gene with a so-called expression locus (*pilE*). Borrelia hermsii also undergoes antigenic variation by a similar mechanism (Robertson and Meyer, 1992). Further, the Opa outer membrane proteins of *N. meningitidis* undergo antigenic variation by varying expression from multiple *opa* loci. These Opa proteins have different antigenic characteristics by virtue of the presence in *opa* alleles of two hypervariable regions and a semivariable region (Hobbs *et al.*, 1994).

4.3.4. General remarks.

PCR-subtractive hybridisation has allowed the definition herein of the genetic loci that have undergone rearrangement between paired *H. pylori* isolates and the gaining of a better understanding of the genetic events involved in the observed differences in oligofingerprint and RAPD-PCR profiles indicating evidence of microevolution. It is only through direct analysis of isolates such as these and definition of the genes involved, that the development of quasispecies can be understood at a molecular level. This is necessary in order to establish whether the occurrence of microevolution solely concerns spontaneous mutation for which no adaptive or beneficial phenotype can be deduced, or whether microevolution in *H. pylori* can

be inferred to have true adaptive significance and an important role to play in maintaining chronic colonisation of the stomach and in determining the outcome of infection. DNA sequence analysis of the genomic fragments believed to be rearranged between paired *H. pylori* isolates implicates mechanistically the occurrence of horizontal gene transfer, transposition of IS sequences or recombination between them, and spontaneous mutations within the chromosome of *H. pylori* during infection. Instances where rearrangements in the *H. pylori* genome might have become amplified in the population because they are advantageous to the organism in its adaptation to the stomach can also be inferred because of the frequency with which Omp-encoding loci were detected. In the restricted stomach niche over prolonged colonisation it is likely that the genetic variants that arise broaden the dynamics of the now variable colonising population, allowing survival of certain adapted clones in the changing stomach environment or permitting dissemination of infection to other regions of the stomach with different selective pressures. Successful colonisation of new hosts by newly adapted strains might also be possible.

Furthermore, the recently published study by Israel et al. (2001b) described the use of a whole genome microarray in an attempt to address similar questions as those that have been described in the present studies. It was found that among isolates of H. pylori strain J99 taken from the same infected individual 6 years after the sample from which the sequenced strain was obtained, evidence existed that the loss and acquisition of exogenous DNA from the genome of this strain had occurred. Genes of the J99 plasticity zone as well as, for example, genes that encode the putative partitioning protein (ParA) and a putative conjugal transfer protein (TraG) were among those identified to be involved in this genetic drift. While only the second report to examine the nature of diversity present between isolates taken from the same niche and to define the loci involved in in vivo DNA exchange (Kersulyte et al., 1999), microarray analysis as carried out by Israel and co-workers (2001b), while powerful and informative, does not allow as much novel data to be gleaned as does PCR-subtractive hybridisation. One of the most advantageous aspects of the subtractive hybridisation technique for analyses such as these herein is that it provides for the isolation of partial differences between the experimental genomes. Thus, using PCR-subtractive hybridisation, DNA recombinations can be detected while, in contrast, microarray technology gives information only about the presence or lack thereof of a particular allele in a chromosome. It follows that investigations of the continuous microevolution of H. pylori within a host are more far-reaching if they include the opportunity to identify the occurrence of intra-genomic

recombination or horizontal transfer of alleles that may become chromosomally inserted at a site removed from the endogenous location of the allele, as well the loss or acquisition of novel genetic sequences.

Chapter 5

Characterisation of genomic rearrangements at loci HP0227/jhp0212 (hopM) and HP1342/jhp1261 (hopN) in H. pylori strains and phenotypic analysis of paired isolates.

5.1.	Introduc	ction.	122
5.2.	Results.		125
	5.2.1.	Further analysis of the HP0227/jhp0212 and HP1342/jhp1261 loci	
		in isolate pair MI 458/MI 459 by Southern hybridisation	125
	5.2.2.	PCR amplification using the primers 1260/1262 and 211/213 in	
		H. pylori paired isolates where rearrangement at the loci HP0227/jhp126	1
		(hopM) and HP1342/jhp1261 (hopN) was not apparent by Southern	
		hybridisation.	126
	5.2.3.	Southern blot analysis of PCR products generated in Section 5.2.2 with	
		a probe for clone 2.	127
	5.2.4.	Amplification across the alleles HP0227/jhp0212 and HP1342/jhp1261	
		by PCR in H. pylori strains MI 458/MI 459 and MI 545A/MI 545C	127
	5.2.5.	Experimental approaches to characterising the nature of the	
		rearrangement that had taken place at alleles HP0227/jhp1261 (hopM)	
		and HP1342/jhp1261 (hopN) in isolate pairs MI 458/MI 459 and	
		MI 545A/MI 545C.	128
	5.2.6.	Alternative PCR approach to analysing the rearrangement at the	
		HopM- and HopN-encoding alleles in isolate pairs MI 458/MI 459	
		and MI 545A/MI 545C	129
	5.2.7.	Confirmation by DNA sequencing that specific amplification of the hopM	
		and hopN alleles had been achieved and determination of the DNA	
		sequences responsible for the diversity apparent by PCR	130
	5.2.8.	PCR product sequencing reveals distinct mechanisms of rearrangement	
		at the loci of interest in paired isolates MI 458/MI 459 and	
		MI 545A/MI 545C.	131
	5.2.9.	Individual H. pylori isolates from paired sets exhibit differences in growth	2
		rate in liquid medium	132
	5.2.10	. Isolation of Omps from H. pylori paired isolate sets MI 458/MI 459 and	
		MI 571/MI 575	133
5.3.	Discussi	on	135
	5.3.1.	Microevolution.	135
	5.3.2.	PCR-subtractive hybridisation findings.	136
	5.3.3.	Insights about the diversity present between MI 458 and MI 459	
		at hopM/hopN loci deduced from Southern blot analysis	137

5.3.4.	Polymorphisms between MI 458 and MI 459 at Omp-encoding loci	
	defined	138
	5.3.4.1. PCR from flanking ORFs.	138
	5.3.4.2. Localisation of heterogeneity.	139
	5.3.4.3. Diversity upstream of the <i>hopN</i> gene in strain MI 459	140
	5.3.4.4. Adaptive significance of horizontal gene transfer between	
	isolates MI 458 and MI 459.	141
	5.3.4.5. Summary of evidence in relation to isolates MI 458 and	
	MI 459 for rearrangement at the HopM/HopN-encoding genes.	143
5.3.5.	Characterisation of microevolution at the hopM/hopN loci between	
	MI 545A and MI 545C.	143
	5.3.5.1. Insertion of the transposable element ISHp608 downstream of the	e
	hopN gene in isolate MI 545A.	143
	5.3.5.2. IS elements in <i>H. pylori</i> .	145
	5.3.5.3. IS element transposition and host-specific adaptation	145
	5.3.5.4. Apparent anomalies in PCR evidence.	146
	5.3.5.5. Future perspective with regard to <i>H. pylori</i> Omp-encoding gene	
	rearrangements.	146
5.3.6.	Culture of paired isolates in liquid medium.	147
5.3.7.	Examination of the Omp profile of MI 458/MI459 and MI 571/MI 575	
	hy SDS-P4GF	149

5.1. Introduction.

The analysis of microevolution in *H. pylori* was first approached by identifying which loci had undergone rearrangement between epidemiologically related *H. pylori* isolates as described in the preceding chapter. The next phase of the present studies was to characterise to some extent the precise nature of the chromosomal changes that had taken place in order to define mechanisms by which intra-strain diversity had arisen. Phenotypic studies were also undertaken to ascertain if the heterogeneity that existed at the loci identified by PCR-subtractive hybridisation between paired isolates (which otherwise had very high overall chromosomal fingerprint identity) was evident somehow at a level that could be detected by simple phenotypic analyses, or to determine if genotypic information could be correlated with observed phenotypic traits.

As detected by oligofingerprinting in the studies described previously (Marshall, 1996) and herein, the emergence of *H. pylori* subtypic variants in the gastric niche of a single individual represents an opportunity to investigate the in vivo evolution of this organism. The publication and comparison of two complete genome sequences of H. pylori (Tomb et al., 1997; Alm et al., 1999; Alm and Trust, 1999) gave valuable insights about the diversity of the species, including the surprising yet highly significant discovery that the level of diversity displayed by H. pylori had apparently been over-estimated. The occurrence of a high level of nucleotide mutation throughout the genome of H. pylori accounts for much of the diversity detected with nucleotide-based typing techniques. Further, non-synonymous variation was found to be rare and implies that much of the mutation that was detected does not alter the proteome (Suerbaum et al., 1998; Alm and Trust, 1999). These insights notwithstanding, the diversity exhibited by H. pylori still exists at an unprecedented level and is accepted to arise from both a high level of recombination combined with a high frequency of nucleotide mutation. Thus, horizontal gene transfer along with other mechanisms of intragenomic recombination, e.g., chromosomal deletions, tandem duplications, translocations, are heavily implicated in the genesis of this diverse bacterium (Marshall et al., 1998; Suerbaum et al., 1998; Wang et al., 1999b).

Although these mechanisms of recombination have been put forward as the origins of the diversity observed in *H. pylori* and much as this theorising arose from informed, carefully formed opinions based on the evidence at hand, there is still much to discover. That subtypic variants or quasispecies arise is thought to be a reflection of the persistent accumulation of

mutations within individual strains and evidence that *H. pylori* undergoes microevolution in isolation in each individual host (Kuipers *et al.*, 2000). It follows therefore, that the analysis of the genomic changes that occur within a single chromosome during long-term host colonisation has much highly meaningful information to contribute. Towards illuminating further the subject of the diversity of *H. pylori*, it was of considerable interest to attempt to characterise mechanisms by which DNA rearrangements occur *in vivo*, having first identified the loci involved using PCR-subtractive hybridisation.

Although *H. pylori* isolates recovered from the same stomach were observed to have 'minor differences' in fingerprint profile in the early typing studies that were carried out (Oudbier *et al.*, 1990), few investigations in the intervening time have reported attempts to identify, let alone characterise, the events involved in eliciting fingerprint profile heterogeneity. The studies by Kersulyte (1999) and Israel (2001b) represent valuable additions to the study of *H. pylori* evolution, the former in particular demonstrating loss of the *cag*PAI by horizontal transfer of DNA containing the *cag* 'empty site' in isolates from a clonal population as well as other defined genomic recombinations. Employing a mouse model, Sozzi *et al.* (2001) also detected instability in the *cag*PAI that they believe to have taken place *in vivo*. Similarly, Björkholm *et al.* (2001a) reported the finding of subclones of *H. pylori* strains taken from the same individual that differed in the presence of the *cag*PAI. The application of microarray technology to isolates of strain J99 by Israel *et al.* (2001b) also revealed substantial diversity within what had been perceived as a homogeneous population and allowed the identification of the loci involved using this macroscale approach.

It is only possible to speculate about the significance the rearrangements identified in clonal isolates of *H. pylori* using PCR-subtractive hybridisation might have for the adaptation of the organism *in vivo*. Without knowledge of the mechanisms by which such rearrangements are brought about, and the precise chromosomal changes that result from these events, the potential, if any, that the occurrence of microevolution has to modify the antigenic characteristics of the organism is mere conjecture. In order, therefore, to remove from the realm of supposition the notion that *H. pylori* undergoes rearrangement at Omp-encoding loci as a mechanism of adaptation, characterisation of the rearrangement identified at the HP0227/jhp0212 (*hopM*) and/or HP1342/jhp1261 (*hopN*) loci was undertaken in isolate pairs MI 458/MI 459 and MI 545A/MI 545C. The growth rate of these strains was also analysed in

liquid medium. In addition, a pair-wise comparison of the profile of Omps produced by each isolate was undertaken using SDS-PAGE.

5.2. Results.

5.2.1. Further analysis of the HP0227/jhp0212 (hopM) and HP1342/jhp1261 (hopN) loci in isolate pair MI 458/MI 459 by Southern hybridisation.

In order to investigate further the nature of the rearrangement believed to have taken place at loci HP0227/jhp0212 and HP1342/jhp1261 based on Southern blot analysis from isolate pair MI 458/MI 459, further Southern blot analysis was carried out. As outlined in Section 4.2.12, these loci encode proteins from the Hop family of Omps, HopM and HopN, and are found in strains J99 and 26695 as duplicated alleles (jhp1261/jhp0212 and HP1342/HP0227, respectively). Based on the Southern blot hybridisation profile of the labelled clone 2 probe to strains MI 458 and MI 459 (Fig. 4.5.A), it appeared as if a gene duplication event could have taken place in this particular strain pair. Although, as discussed in Section 4.3, this possibility was realistically deemed remote based on several lines of evidence, the hybridisation profile made the argument for the occurrence of duplication of the hopM/hopN alleles in strain MI 458 a compelling one. The possibility that a mutational as opposed to recombinational mechanism had generated the diversity observed between isolates MI 458 and MI 459 using probe 'clone 2' was also suggested by the hybridisation profile observed with Southern blotting.

Towards further addressing both these possibilities, it was decided to assess the hybridisation profiles obtained using probes made from the insert in clone 2 cut into two. These would then be used to probe *Hin*dIII-restricted genomic DNA from isolates MI 458 and MI 459. If gene duplication had taken place, it was anticipated that all hybridising fragments (two in strain MI 458 and one in strain MI 459) should hybridise with the same intensity to the truncated probes as was detected with the full-length probe.

The insert in clone 2 was removed from the vector by digestion with the enzymes *Eco*RV and *Hin*dIII yielding a restriction fragment of approximately 600 bp in length that consisted of the insert sequence flanked on either side by short regions of vector sequence. This full-length insert was then restricted with the endonuclease *Mse*I which yielded restriction fragments of 300 bp and ~260 bp. These two fragments were extracted from the agarose gel and labelled with DIG by random priming. The full-length probe and each of the two new shorter probes were then used in hybridisation with a Southern blot of *Hin*dIII-restricted total genomic DNA of isolates MI 458 and MI 459 by sequential stripping and re-probing of the filter. After detection, the blots shown in Fig. 5.1 were obtained.

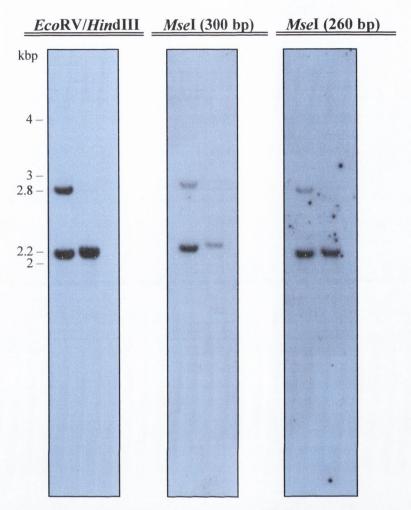


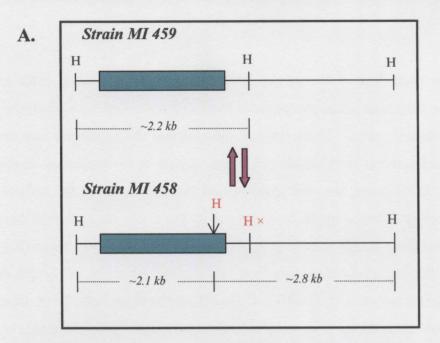
Fig. 5.1. Southern hybridisation analysis with full-length and cleaved clone 2 probe from PCR-subtractive hybridisation of isolate pair MI 458/MI 459 with *Hin*dIII digests of these strains.

As described in Section 5.2.1, the insert in clone 2 was extracted by digestion with the enzymes *Eco*RV/*Hin*dIII and was subsequently cleaved into two fragments using the enzyme *Mse*I. These DNA fragments were labelled with DIG and used to probe a Southern blot of *Hin*dIII-restricted genomic DNA of isolates MI 458 and MI 459 by sequential stripping and re-probing of the filter. In each of the panels above, isolate MI 458 DNA was electrophoresed in the left lane while isolate MI 459 DNA was electrophoresed in the right lane. The probes used are as indicated. The positions of DNA molecular size standards (kbp) are indicated to the left of the left-most panel and apply to all panels.

There are clear differences in the intensity of the signal from each of the hybridising restriction fragments when the probe is truncated compared with what was observed for the original clone 2 probe. These findings indicate that the hypothesis that the two hybridising restriction fragments in strain MI 458 represent each of the duplicated alleles is unlikely to be true and, furthermore, that the hybridisation profiles observed for isolates MI 458/MI 459 with the clone 2 probe might be explained by the occurrence of point mutations within the genomes of these strains as well as implicating a recombinational event (Fig. 5.2). As further evidence that point mutations may have taken place, when the full-length clone 2 probe was used in hybridisation with *Eco*RI- and *Hae*III-restricted genomic DNA from isolates MI 458 and MI 459 this resulted in identical hybridisation profiles between strains (data not shown).

5.2.2. PCR amplification using the primers 1260/1262 and 211/213 in H. pylori paired isolates where rearrangement at the loci HP0227/jhp0212 (hopM) and HP1342/jhp1261 (hopN) was not apparent by Southern hybridisation.

It was decided to ascertain using PCR if the HopM- or HopN-encoding alleles were both present in other H. pylori strains from paired isolate sets (MI 517/MI 566, MI 577A/MI 577C and MI 9613A/MI 9613C) in which rearrangement was not apparent by Southern blotting and all of which differed in the profile of restriction fragments which hybridised to the clone 2 probe (Fig. 5.3.A). In *H. pylori* strains 26695 and J99, the *hopM* and *hopN* genes exist in both strains flanked by the same ORFs. Thus, primers were designed that would amplify across HP0227/jhp0212 (hopM) and HP1342/jhp1261 (hopN) by binding to the DNA in flanking genes and initiating elongation from there. These primers were designated 1260 and 1262 (amplify across HP1342/jhp1261) and 211 and 213 (amplify across HP0227/jhp0212) corresponding to the loci from strain J99 to which they bound. These primers also bound to loci HP1341, HP1343, HP0226 and HP0228 from strain 26695. The primers 1260/1262 and 211/213 were utilised in the PCR using as templates DNA from strains MI 517, MI 577A and MI 9613A which, by reference to Fig. 5.3.A, all displayed different patterns of hybridisation to this probe. Gel electrophoresis of the PCR reactions showed that amplification products were obtained with primers for the loci corresponding to both HP1342/jhp1261 and HP0227/jhp0212 in each of these strains (Fig. 5.3.B).



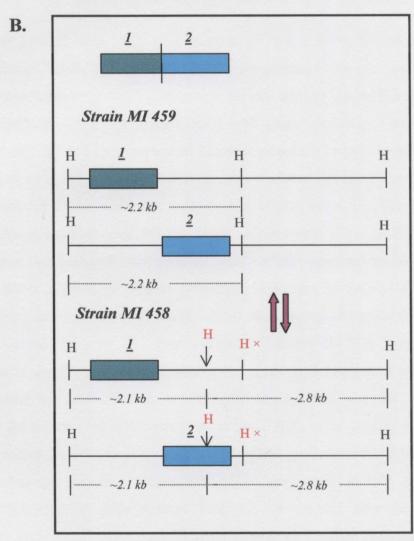
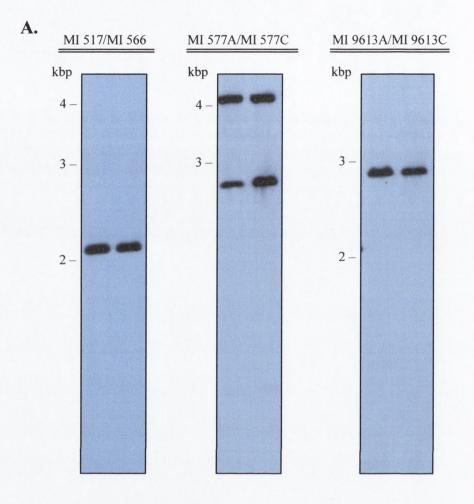
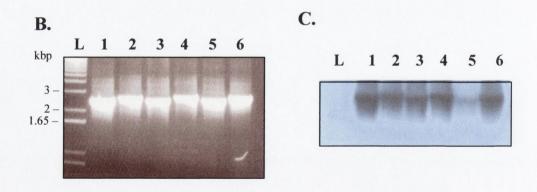


Fig. 5.2. Schematic diagrams showing (A) how the hybridisation pattern observed with probe 'clone 2' in isolate pair MI 458 and MI 459 might be accounted for by the occurrence of nucleotide mutation which changes the presence and location of restriction sites and hence explains the hybridisation pattern observed between these two strains, and (B) an explanation of how both mutation and recombination could be considered likely to have taken place based on inspection of the results of Fig. 5.1.

- (A). Possible explanation for Southern blot hybridisation profile observed with isolate pair MI 458 and MI 459 with the probe 'clone 2' by the occurrence of point mutation. Nucleotide mutation within the region of the probe which creates a *Hin*dIII restriction site (symbolised as H) in isolate MI 458 and a second mutation outside this which removes an existing site (symbolised as H×) could cause there to be two hybridising fragments in strain MI 458 as was observed. The original fragment would now be slightly smaller than that which hybridised in isolate MI 459 (depicted in strain MI 458 as ~2.1 kb) with the new larger fragment expected to give a much weaker probe signal.
- (B). With reference to Fig. 5.1. and the schematic shown to the left, it is considered likely that the pattern of hybridisation observed with probe 'clone 2' in isolates MI 458 and MI 459 can be accounted for both by mutational and recombinational events. By splitting the probe in two and examining the intensity of the hybridising fragments obtained when each truncated probe was used in hybridisation with the same *HindIII*-restricted genomic DNA, it follows first that, if nucleotide mutation had occurred, that at least half of the probe (labelled <u>1</u>) would give the same intensity of hybridisation between MI 458 and MI 459 (as was observed for the 260 bp *MseI* fragment of the clone 2 probe in Fig. 5.1), and secondly, that the segment of probe labelled <u>1</u> would not hybridise to the ~2.8 kb fragment in isolate MI 458. Thus, for one half of the truncated probe, if point mutation had occurred, only one probereactive *HindIII* fragment would be visible in strain MI 458. Only one of these two expected observations was made in Fig. 5.1 with the truncated probes.





- Fig. 5.3. PCR amplification of the chromosomal loci corresponding to alleles HP0227/jhp0212 (hopM) and HP1342/jhp1261 (hopN) in H. pylori isolates where rearrangement at these ORFs was not apparent and analysis by Southern hybridisation of the PCR products obtained.
- (A.) Southern hybridisation of *Hin*dIII-cleaved genomic DNA of three *H. pylori* paired isolate sets (indicated at the top of each panel) with a DIG-labelled clone 2 probe. The sizes of the molecular weight markers are indicated (kbp).
- (B.) Agarose (1.4% w/v) gel electrophoresis of the PCR products generated with the primers 1260/1262 and 211/213 which amplify across the jhp1261/HP1342 and jhp0212/HP0227 loci, respectively. The lane numbers correspond to the following *H. pylori* isolates and (primer combinations); lane 1, MI 517 (1260/1262); lane 2, MI 517 (211/213); lane 3, 577A (1260/1262); lane 4, 577A (211/213); lane 5, 9613A (1260/1262); lane 6, 9613A (211/213). The sizes (kbp) of the relevant bands in the molecular size ladder are indicated to the left of the panel.
- (C.) Southern hybridisation of the PCR products shown in (B) with a DIG-labelled clone 2 probe. Lanes are as indicated in (B). In both (B) and (C) the lane labelled L corresponds to where the DNA molecular size ladder was electrophoresed.

5.2.3. Southern blot analysis of PCR products generated in Section 5.2.2 with a probe for clone 2.

In order to determine if the amplification products generated using primers 1260/1262 and 211/213 for isolates MI 517, MI 577A and MI 9613A as described in Section 5.2.2 did correspond to the *hopM* and *hopN* alleles, Southern blot hybridisation of these amplimers was carried out using a DIG-labelled clone 2 probe. As shown in Fig. 5.3.C, all of these PCR products hybridised with the probe indicating that in the cases of strains MI 517, MI 577A and MI 9613A and therefore, by extension, also strains MI 566, MI 577C and MI 9613C, these isolates contain the *hopM* (HP0227/jhp0212) and *hopN* (HP1342/jhp1261) alleles flanked by the same ORFs as were found in strains 26695 and J99. Interestingly, although one might deduce from the Southern blot hybridisation pattern of strain pairs MI 517/MI 566 and MI 9613A/MI 9613C with the clone 2 probe that each isolate has only one or other of these alleles by virtue of there being a single probe-reactive *Hin*dIII fragment in both isolates, the PCR and Southern blot evidence indicates the contrary, i.e., that duplicate alleles exist in this strain. This, therefore, further points to the erroneous surmise in relation to isolates MI 458/MI 459 and MI 545A/MI 545C that what had been detected by Southern blotting was evidence of allele duplication as outlined earlier.

5.2.4. Amplification across the alleles HP1342/jhp1261 and HP0227/jhp0212 by PCR in H. pylori strains MI 458/MI 459 and MI 545A/MI 545C.

The Southern blot evidence presented in Chapter 4 and Fig. 5.1 demonstrates a definite difference between isolates MI 458 and MI 459 in relation to the *hopM* and *hopN* alleles and, in particular, reinforces the belief that rearrangement occurred at these loci. However, based on what was deduced from the results presented in Fig. 5.1 and Fig. 5.3, it was not thought that the Southern blot data were evidence of a gene duplication or indeed that loss of one copy of this allele had taken place between the paired isolates in question. To ascertain if both of the *hopM* and *hopN* alleles were present in paired strains MI 458/MI 459 and MI 545A/MI 545C, the primers 1260/1262 and 211/213 were used to amplify from genomic template preparations of DNA from these strains. Aliquots of the completed PCR reactions carried out on strains MI 458, MI 459, MI 545A and MI 545C were electrophoresed. The results obtained are shown in Fig. 5.4.A and B.

Evidently, because PCR products of between 2–3 kb are generated with both primer combinations, no vacant sites exist in the genomes of strains MI 458/MI 459 and MI 545A/MI

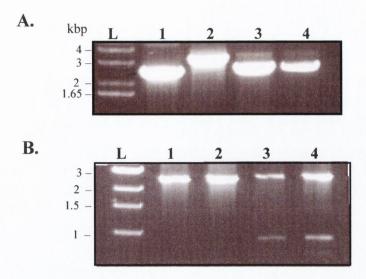


Fig. 5.4. Agarose gel electrophoresis of PCR products obtained from amplification of loci HP0227/jhp0212 (hopM) and HP1342/jhp1261(hopN) from H. pylori strains MI 458 and MI 459 and MI 545A and MI 545C.

- (A.) Agarose (1.3% w/v) gel electrophoresis of the PCR products generated with following strains and (primer combinations). Aliquots of the following PCR reactions were electrophoresed in the lanes labelled 1–4: lane 1, MI 458 1260/1262; lane 2, MI 459 (1260/1262); lane 3, MI 458 (211/213); lane 4, MI 459 (211/213).
- **(B.)** Agarose (1.3% w/v) gel electrophoresis of PCR products generated with the primers 1260/1262 and 211/213 with strains MI 545A and MI 545C. The lanes labelled 1–4 in (B) correspond to the same primer combinations as are detailed for (A) above with the following strains used as template in each lane: lane 1, MI 545A; lane 2, MI 545C; lane 3, MI 545A; lane 4 MI 545C. Molecular size standards were electrophoresed in the lane labelled L in both (A) and (B) and their sizes in kbp are indicated to the left of both panels.

545C between the flanking ORFs, strongly suggesting that duplicated alleles exist in both of these sets of paired strains as anticipated by comparison with what was observed for strains MI 517/MI 566, MI 577A/MI 577C and MI 9613A/MI 9613C. This, of course, assumes that the alleles located between alleles HP1341/jhp1260 and HP1343/jhp1262 correspond in strains MI 458, MI 459, MI 545A and MI 545C to allele HP1342 /jhp1261 and similarly, that the allele that is present between allele HP0226/jhp0211 and HP0228/jhp0213 is allele HP0227/jhp0212 in these pairs of clinical isolates. The difference in the sizes of the PCR products amplified using the primers 1260 and 1262 between MI 458 and MI 459 is noteworthy (Fig. 5.4.A, lanes 1 and 2).

To demonstrate that *H. pylori* strains MI 545A and MI 545C were clonally related, RAPD-PCR fingerprinting was carried out using three pair-wise combinations of RAPD primers. The amplimer profiles obtained upon electrophoresis of these reaction products are shown in Fig. 5.4.C, confirming the genotypic identity of the strains.

5.2.5. Experimental approaches to characterising the nature of the rearrangement that had taken place at alleles HP0227/jhp0212 (hopM) and HP1342/jhp1261 (hopN) in isolate pairs MI 458/MI 459 and MI 545A/MI 545C.

From what has been described thus far in relation to characterising the rearrangement that had taken place at alleles HP1342/jhp1261 and HP0227/jhp0212 in isolate pair MI 458/MI 459, there was still appropriate data to gather. To this end, it was of interest to clone the HindIII fragments to which the clone 2 probe hybridised in these strains in order to determine their DNA sequences and, thus, define the mechanism by which rearrangement had taken place. The construction of a sub-genomic library was attempted using HindIII-restricted genomic DNA of strains MI 458, MI 459, MI 545A and MI 545C that had been fractionated on a sucrose gradient. The fractions containing the fragments of interest were identified by Southern blotting. Attempts were then made to insert the restriction fragments in the subgenomic fraction into one of several popular cloning vectors, e.g., pUC18, pBR322 and pACYC184 using commercially available E. coli host strains. Attempts to isolate the desired restriction fragments by electroelution from low-melting-point agarose were also carried out. These approaches were consistently problematic and attempts to construct a H. pylori subgenomic library to clone these alleles (and others) were unsuccessful. Similar difficulties have been encountered by other investigators as reported previously (Covacci et al., 1993; Phadnis et al., 1993).

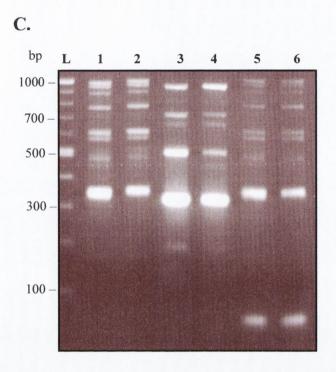


Fig. 5.4.C. Agarose gel electrophoresis of RAPD-PCR amplimers obtained using three combinations of primers for *H. pylori* strains MI 545A and MI 545C.

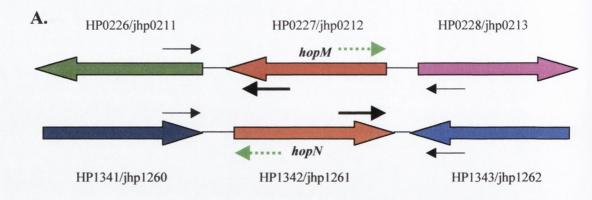
Agarose gel (1.8% w/v) electrophoresis of RAPD-PCR amplimers obtained using genomic DNA from isolates MI 545A and MI 545C as template and three different pairwise combinations of primers. Lanes correspond to the following strains and primers used: lanes 1 & 2, isolates MI 545A and MI 545C, respectively, primers 1254 and D8635; lanes 3 & 4, isolates MI 545A and MI 545C, respectively, primers D11344 and 1281; lanes 5 & 6, isolates MI 545A and MI 545C, respectively, primers 1281 and D8635. The molecular size markers (bp) are indicated to the left of the panel.

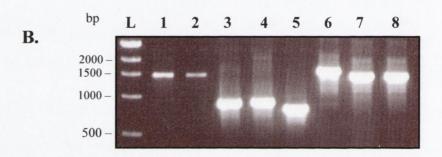
Towards circumventing these obstacles, various PCR-based approaches were adopted, including inverse PCR in which the target restriction fragment is amplified from a region of known sequence by outward-reading primers, the fragment having first been circularised by self-ligation of its cohesive ends (Collins and Weissman, 1984). Further, a recently described technique was used which was based on the inclusion in a PCR reaction that is carried out at low primer annealing temperatures of one primer specific for the region of known DNA sequence and one primer of arbitrary sequence. In a second round of amplification, another arbitrary primer is used (binds specifically to DNA sequence synthesised from the first arbitrary primer) at a slightly higher annealing temperature, thereby supposedly amplifying and enriching for the DNA in the region which flanks the area of known sequence (O'Toole and Kolter, 1998). Although these techniques held promise in light of the difficulties encountered in cloning exercises, they proved not to be viable for the purposes required in this study because spurious DNA fragments were amplified using arbitrary PCR and, the inverse PCR relied on the success of too many variables to be widely applicable.

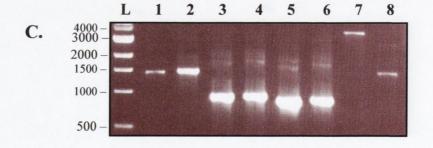
5.2.6. Alternative PCR approach to analysing the rearrangement at the HopM- and HopN-encoding alleles in isolate pairs MI 458/MI 459 and MI 545A/MI 545C.

As attempts to directly clone or indeed amplify the target *Hin*dIII restriction fragments failed, another approach was required to define the genomic events responsible for the rearrangement observed. For the purposes of inverse PCR (alluded to in Section 5.2.5) primers had been designed that bound to the DNA sequence of clone 2 but which read in opposite directions to one another. In order to investigate the possibility that the thus far undefined rearrangement involving the loci HP0227/jhp0212 (*hopM*) and HP1342/jhp1261 (*hopN*) had not been appraised fully by direct amplification across these loci from primers that bound at flanking DNA, not discounting the size discrepancy in the PCR products generated using primers 1260 and 1262 for strain pair MI 458/MI 459 as mentioned in Section 5.2.4, it was of interest to use these outward-reading primers specific for clone 2 (Cln2#1 and Cln2#2) in combination with both sets of primers to flanking DNA, i.e., Cln2#1 with both primers 1260 and 213; and Cln2#2 with both primers 1262 and 211. A schematic figure of these ORFs and location of primer-binding sites is shown in Fig. 5.5.A.

Aliquots of the PCR reactions were electrophoresed on a 1.3% agarose gel as shown in Fig. 5.5.B and C. Evidently, these primer combinations were considerably more useful at detecting and localising heterogeneity that exists at the *hopM* and *hopN* alleles between the







- Fig. 5.5. Schematic figure showing the *hopM* and *hopN*-encoding alleles, flanking genes and location of primer binding sites with agarose gel electrophoresis of PCR products generated with primers specific for clone 2 and for the ORFs flanking the HP0227/jhp0212 (*hopM*) and HP1342/jhp1261(*hopN*) loci.
- (A.) The duplicate alleles encoding the HopM and HopN proteins are represented by the red arrows. Flanking genes are as labelled. The green dashed arrows represent primer Cln2#1, the large black arrows represent primer Cln2#2 and the thin black arrows represent primers which bind to the various flanking genes.
- (B.) The PCR products shown were electrophoresed in a 1.3% (w/v) gel and were amplified from isolates MI 458 and MI 459 (with the primer combinations) which correspond to the following lane numbers: lanes 1 & 2, MI 458 (Cln2#2/211) and MI 459 (Cln2#2/211), respectively; lanes 3 & 4, MI 458 (Cln2#1/213) and MI 459 (Cln2#1/213), respectively; lanes 5 & 6, MI 458 (Cln2#1/1260) and MI 459 (Cln2#1/1260), respectively; and lanes 7 & 8, MI 458 (Cln2#2/1262) and MI 459 (Cln2#2/1262), respectively.
- (C.) The PCR products shown were amplified from *H. pylori* isolates MI 545A and MI 545C. The specific isolate (and primer combination) used to generate each product were as follows: lanes 1 & 2, MI 545A (Cln2#2/211) and MI 545C (Cln2#2/211), respectively; lanes 3 & 4, MI 545A (Cln2#1/213) and MI 545C (Cln2#1/213), respectively; lanes 5 & 6, MI 545A (Cln2#1/1260) and MI 545C (Cln2#1/1260), respectively; and lanes 7 & 8, MI 545A (Cln2#2/1262) and MI 545C (Cln2#2/1262), respectively. In both (B) and (C) the sizes of the indicated bands in the DNA size ladder (lanes labelled L) are shown in bp.

isolates in the two paired sets examined. As can be seen in Fig. 5.5.B, the primers Cln2#1 and 1260 yielded amplification products of different sizes between strains MI 458 and MI 459. A product of ~700 bp was amplified from strain MI 458 and one of 1.5–1.6 kb from strain MI 459. Additionally, as can be seen in Fig. 5.5.C., different sized PCR products were amplified from strains MI 545A and MI 545C using the primers Cln2#2 and 1262. A product of greater than 3 kb was obtained when MI 545A genomic DNA was used as template and a shorter product of approximately 1.5 kb when MI 545C DNA was analysed. For both sets of paired strains, the other primer combinations generated products of the same sizes from both strains.

5.2.7. Confirmation by DNA sequencing that specific amplification of the hopM and hopN alleles had been achieved and determination of the DNA sequences responsible for the diversity apparent by PCR.

The DNA sequences of selected PCR products shown in Fig. 5.5.B were determined in order to confirm that amplification of the HP0227/jhp0212 (hopM) and HP1342/jhp1261 (hopN) loci had been achieved using primers specific for flanking ORFs in combination with the clone 2-specific primers Cln2#1 and Cln2#2. The PCR products were cloned into a T-vector and sequenced from standard M13 primers.

Four PCR products were chosen for sequencing. These products were as follows: from strain MI 458, those amplified with primers Cln2#1 and 1260 (~700 bp) and with primers Cln2#2 and 1262 (~1.3 kb); and from strain MI 459, those amplified with primers Cln2#1 and 1260 (1.5 kb) and with primers Cln2#1 and 213 (~600 bp). BLAST analysis of these sequences with the *H. pylori* genome sequence databases confirmed that these primers (Cln2#1 and Cln2#2) specifically amplified the HP1342/jhp1261 and HP0227/jhp0212 loci from flanking genes and thus allowed the heterogeneity that existed at this extended chromosomal region, as evidenced by the PCR studies described above, to be localised better. As with isolates MI 458 and MI 459, the sequences of the PCR products amplified from both strains MI 545A and MI 545C with primers Cln2#2 and 1262 were determined following T-vector cloning of these PCR products. A schematic representation of the chromosomal regions where the *hopM*, *hopN* and flanking genes are located is shown in Fig. 5.6.B and C for both paired isolate sets based on PCR analysis and DNA sequence determination of selected products.

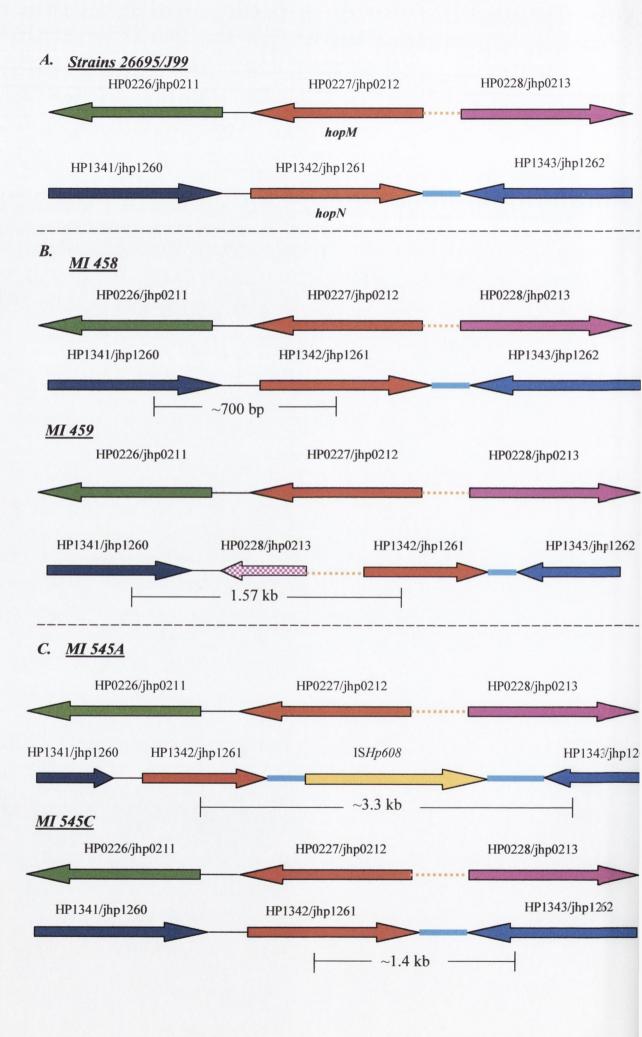


Fig. 5.6. Schematic representation of the structure of the chromosomal regions surrounding the HP0227/jhp0212 (hopM) and HP1342/jhp1261 (hopN) loci in isolate pairs MI 458/MI 459 and MI 545A/MI 545C deduced from analysis of PCR products and DNA sequencing.

- (A.) As shown in Fig. 5.5.A, the 'wild-type' configuration of the chromosomal regions that flank the HopM- and HopN-encoding alleles in the sequenced *H. pylori* strains 26695 and J99. The duplicate alleles (HP0227/jhp0212–hopM and HP1342/jhp1261–hopN) are represented by the red arrows with flanking genes depicted in different colours as labelled. The intergenic region between loci HP0227/jhp0212 and HP0228/jhp0213 is shown as an orange dotted line while the non-coding region between loci HP1342/jhp1261 and HP1343/jhp1262 is represented by a pale blue rectangle.
- (B.) The structure of the same regions deduced for isolates MI 458 and MI 459 with the insertion upstream of the *hopN* allele in isolate MI 459 of sequence from the HP0228/jhp0213 locus and the upstream sequence thereof. The partial HP0228/jhp0213 sequence (approximately half of the coding region of this gene) that exists in strain MI 459 upstream of the *hopN* allele is shown as a shaded pink arrow.
- (C.) The insertion downstream of the *hopN* allele in strain MI 545A of the mobile element IS*Hp608* is shown. Isolate MI 545C is essentially wild-type with respect to the structure of the chromosomal regions surrounding the Omp-encoding genes.

5.2.8. PCR product sequencing reveals distinct mechanisms of rearrangement at the loci of interest in paired isolates MI 458/MI 459 and MI 545A/MI 545C.

As depicted in the schematic diagram (Fig. 5.6), PCR product sequencing revealed that distinct mechanisms of rearrangement had taken place at the Omp-encoding genes in isolate pairs MI 458/MI 459 and MI 545A/MI 545C. Specifically, both isolates in the former set appeared to have what can be considered a wild-type structure at the chromosomal region of the hopM gene, while upstream of the hopN allele an insertion of approximately 900 bp of DNA was detected. Inserted between loci HP1341/jhp1260 and HP1342/jhp1261 (hopN) in isolate MI 459 was 605 bp of sequence from the HP0228/jhp0213 locus which represented 52% of the coding sequence of this gene. The start codon (ATG) is present and a little over half of the 5'-end of the gene. This inserted fragment of the HP0228/jhp0213 locus encodes this gene on the negative strand of the chromosome. Following immediately downstream of this is 270 bp of sequence from the intergenic region between locus HP0227/jhp0212 (hopM) and locus HP0228/jhp0213 which is located upstream of the gene HP0228/jhp0213 fragment in relation to the direction of transcription. The sequence of the fragment then reads into DNA of the HP1342/jhp1261 (hopN) allele. In strain MI 459 at the chromosomal region where HopM is encoded, evidence from PCR and sequencing analysis indicates that as is found in strains 26695 and J99, HP0227/jhp0212 is located on the chromosome flanked by HP0226/jhp0211 to the left and HP0228/jhp0213 to the right (Fig. 5.6. B). It is proposed that the ~900-bp fragment was acquired horizontally and became inserted into the chromosome of strain MI 459 by homologous recombination in a manner that is shown schematically in Fig. Double cross-over and fragment insertion is achieved putatively by recombination between identical sequences at the 5'-ends of HP0227/jhp0212 (hopM) and HP1342/jhp1261 (hopN) and between a 19-bp sequence that is located in the intergenic non-coding region between HP1341/jhp1260 and HP1342/jhp1261 that is 84% identical to a stretch of sequence located approximately half way through the coding region of HP0228/jhp0213.

In the case of isolate pair MI 545A and MI 545C, both of these strains again appear to have an unaltered *hopM* chromosomal region by comparison with strains 26695 and J99. In this instance, however, the polymorphism that is present between these strains is located downstream of the *hopN* gene and is caused by the insertion of ~1.9 kb of DNA, the majority of which consists of two ORFs that encode the putative transposase genes of a novel *H. pylori* insertion sequence, IS*Hp608* (Kersulyte *et al.*, 2002). Fig. 5.8.A depicts schematically the

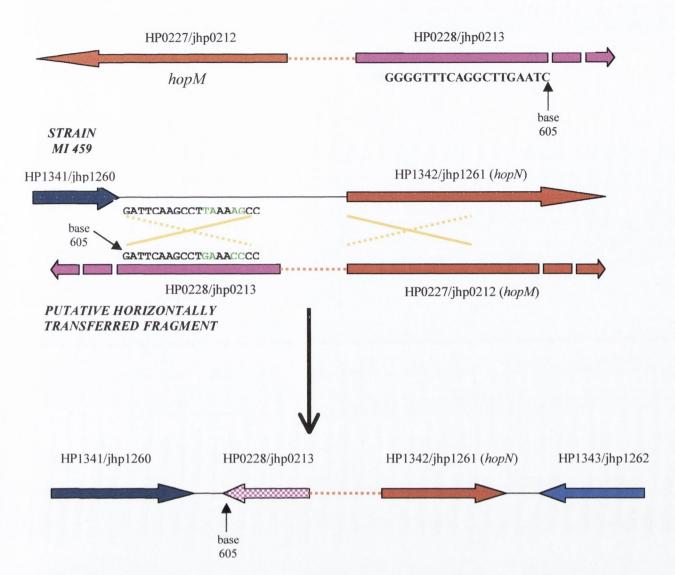


Fig. 5.7. Schematic representation of the putative mechanism of rearrangement that took place at the *hopN* locus between isolate MI 459 and a horizontally transferred fragment of DNA.

The coloured arrows represent the ORFs as labelled and correspond to the same genes as were depicted in Fig. 5.5. and Fig. 5.6. A mechanism is proposed for the chromosomal insertion of a DNA fragment likely to have been taken up horizontally by isolate MI 459. As explained in text, the exogenous fragment consisting of sequence of the genes HP0227/jhp0212 and HP0228/jhp0213 as well as DNA from the intergenic sequence between these two genes (orange dotted line) might become inserted into the chromosome of strain MI 459 by homologous recombination occurring between the 19-bp region of significant identity that exists approximately half way through the coding sequence of locus HP0228/jhp0213 and the sequence in the intergenic region between the HP1341/jhp1260 and HP1342/jhp1261 loci (the nucleotide mismatches between both these sequences are highlighted in green above). A second cross-over event occurring between the sequences of the identical alleles HP0227/jhp0212 and HP1342/jhp1261 would result in the insertion of the horizontally transferred fragment upstream of the hopN gene. The location of where homologous recombinations most likely occurred are depicted by yellow crosses, arrows are not to scale and the chequered pattern of the HP0228/jhp0213 gene when inserted in strain MI 459 signifies that only the 5'-half the gene is present.

insertion in strain MI 545A of this IS element and the location of its insertion point in the DNA sequence in the intergenic region between HP1342/jhp1261 (hopN) and HP1343/jhp1262. A multiple sequence alignment of the entire intergenic region between loci HP1342/jhp1261 and HP1343/jhp1262 from *H. pylori* strains J99, MI 545A and MI 545C is shown in Fig. 5.8.B. The DNA sequence of the left and right ends of IS*Hp608* as well as those sequences immediately flanking the ends of the element are shown in Fig. 5.8.C.

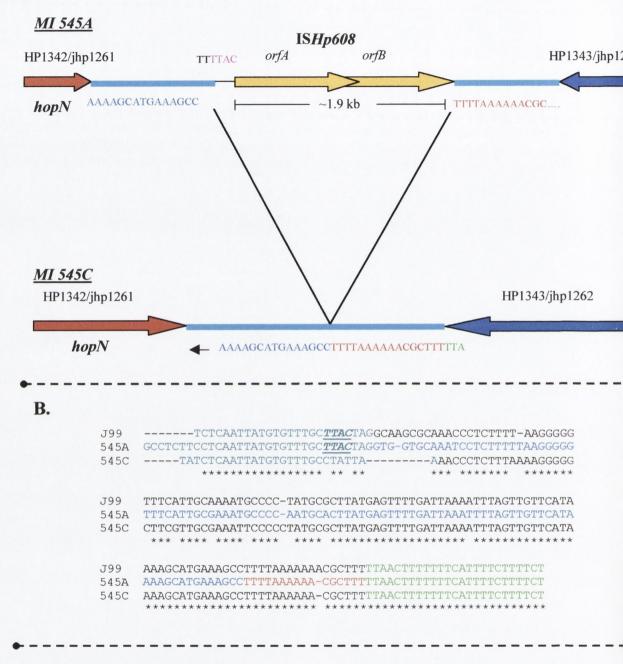
5.2.9. Individual H. pylori isolates from paired sets exhibit differences in growth rate in liquid medium.

Genotypic diversity occurring between two clonally related isolates is likely also to result in the observation of phenotypic diversity between these strains. Some DNA rearrangements might elicit phenotypic changes that would be impossible to detect because they occur at a level that cannot be analysed experimentally, e.g., in the genes for metabolic or biosynthetic enzymes for which no assays exist or in proteins whose presence cannot be detected because their function cannot be assayed for or to which no anti-serum has been raised.

One of the easiest phenotypes to observe in the bacterial kingdom is that of the growth rate of a particular organism. *H. pylori* colonies on agar plates can often be observed to have marked size differences, even among what is perceived to be a clonal population. This was found to be the case for several of the *H. pylori* paired isolates used throughout the studies described in this work (personal observation) where one isolate consistently grew with a colony size that had a greater diameter than that of the second isolate despite the high level of genotypic identity existing between them. Colony size variation is considered to reflect differences in growth rate exhibited by each isolate. In light of the findings from PCR-subtractive hybridisation analysis that several paired isolates were identified to have undergone genomic rearrangement which involved Omp-encoding genes, it was tempting to speculate that such rearrangements might somehow cause an alteration in expression of certain gene products.

Towards assessing the differences in growth rate that might exist between the paired isolates MI 458/MI 459, MI 545A/MI 545C, each of these individual strains was cultured in broth during which time growth was monitored by measuring the OD of the culture at 600nm. What had been observed as variable colony sizes on agar plates was borne out in liquid culture. Different doubling times were observed for isolate pair MI 458 and MI 459 and for isolate pair MI 545A and MI 545C. Four cultures of each of the isolates in the MI 458 and





C.

Fig. 5.8. Schematic diagram showing the DNA sequence in the intergenic region between HP1342/jhp1261 (hopN) and HP1343/jhp1262 where the ISHp608 element was found to be inserted in H. pylori strain MI 545A and multiple sequence alignment of this intergenic region from strains J99, MI 545A and MI 545C.

(A) The insertion sequence element ISHp608 was found to have become inserted in strain MI 545A downstream of the hopN allele in the non-coding region between this allele and the HP1343/jhp1262 locus. This intergenic region is represented above by the pale blue rectangle (exaggerated in length) between the coloured arrows which represent the indicated ORFs. The motif TTAC was found 5' to where the element inserted, although this is not believed to be part of the native sequence of the intergenic region of these isolates at the deduced insertion point. The nucleotide sequence shown is merely the DNA sequence immediately up and downstream of the insertion point. The black outwardpointing arrow symbolises that the intergenic sequence continues upstream of that shown. The sequence depicted in green at the end of the intergenic sequence represents the complement of the stop codon of the HP1343/jhp1262 locus. (B) ClustalW multiple sequence alignment of this intergenic region from isolates J99, MI 545A and MI 545C. A TTAC motif is highlighted in strains J99 and MI 545A with its notable absence from strain MI 545C. The insertion site of ISHp608 in the intergenic sequence in strain MI 545A is at the point where the red and blue font meet. The green font represents the 3'-end of the HP1343/jhp1262 gene and the pale blue font represents the 3'-end of the HP1342/jhp1261 (hopN) gene. (C) DNA sequence of the left and right ends of the ISHp608 element is shown in upper-case letters. The lower-case letters represent the sequences which flank the element with the TTAC motif underlined. Chromosomal sequences of strain MI 545A are depicted in blue, red and green font and correspond to the same sequences as are shown in (B).

MI 459 strain pair were analysed, each of these being carried out as independent experiments with different broth preparations inoculated with *H. pylori* cells grown separately for each experiment (Fig. 5.9). Strain MI 459 consistently grew with a doubling time of approximately 3 hours with little deviation from this (mean doubling time, 3 h±0.082, range 2 h 54 min–3 h 6 min). Strain MI 458 was found to grow more slowly than MI 459, however, and also exhibited considerable variability in doubling time (mean doubling time 4 h 49 min±0.89, range 4 h 12 min–6h 6 min). Examination of the growth rate of isolates MI 545A and MI 545C was carried out once only. Isolate MI 545A grew with a doubling time of 5 h 21 mins compared to a doubling time of 8 h 36 mins calculated for isolate MI 545C.

5.2.10. Isolation of Omps from H. pylori paired isolate sets MI 458MI 459 and MI 571/MI 575.

Given the observed genotypic differences at Omp-encoding loci in *H. pylori* paired isolate sets combined with the differences in growth rates exhibited by some of these same strains, SDS-PAGE analysis of outer-membrane protein fractions of these bacteria was performed to determine if differences could be detected. It is plausible that marked growth rate differences between two clonally related isolates might be founded in inadequate nutrient uptake by the bacteria because of recombinations that occur at loci which specify Omps responsible for nutrient transport. Few of the Omp family of proteins have been the subject of functional characterisation in *H. pylori* and their putative functions have been assigned based on amino acid sequence similarity with other proteins of known function. Thus, among the uncharacterised members of the *H. pylori* Hop family of Omps in particular, there may exist other porins that are highly selective for specific substrates, e.g., essential metabolic precursors whose access intracellularly may be vitiated by aberrant functioning or expression of a specific transport channel arising from recombination at the locus where it is encoded.

H. pylori isolates MI 458, MI 459, MI 571 and MI 575 were cultured in 25 ml of Brucella broth containing foetal calf serum and antibiotics for 48–72 h after which time the outer membranes of these strains were obtained followed by the isolation of a Sarkosyl-insoluble fraction of outer membrane proteins by differential centrifugation. The proteins present in the outer-membrane of each strain were resolved by SDS-PAGE and compared (Fig. 5.10). What is most striking about the outer membrane profiles of H. pylori strains is the lack of paralogues of the highly abundant non-selective OmpF and OmpC proteins that are found in E. coli and which stain very prominently in profiles of this nature from that organism. The H.

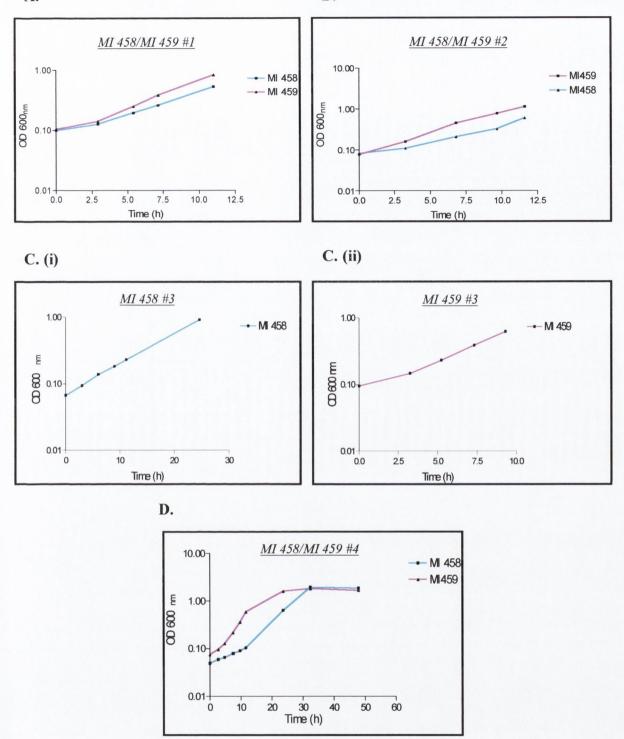


Fig. 5.9. Plots of growth of H. pylori strains MI 458 and MI 459 over time as measured by the OD of the bacterial culture at 600_{nm} .

(A)–(D) represent the growth of strains MI 458 and MI 459 as determined in four separate experiments by measuring the absorbance (optical density) of the bacterial culture at 600_{nm}. Isolate MI 459 consistently grows more quickly than isolate MI 459. Note, in the third experiment (C) above, the data presented are from samples of strains MI 458 and MI 459 cultured on different days.

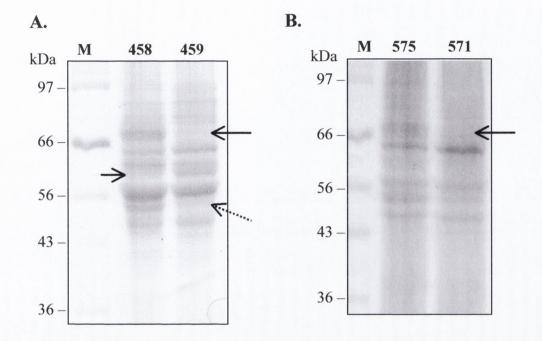


Fig. 5.10. SDS-PAGE analysis of outer membrane proteins isolated from *H. pylori* strains MI 458, MI 459, MI 571 and MI 575.

Aliquots of Omps isolated from the paired *H. pylori* strains MI 458/MI 459 (panel A) and MI 571/MI 575 (panel B) were resolved on a 12% (v/v) acrylamide gel. The lanes were loaded as labelled. Protein markers were electrophoresed in the lanes labelled M and the molecular weight (kDa) of the indicated bands are shown to the left of each panel. Discernible band differences in the Omp-profile between these paired strains are indicated with arrows with the dotted arrow pointing to a protein band that appears to be downregulated in expression in strain MI 459 when compared with the band at the same position in strain MI 458. The quality of the image in panel B is lacking although the arrow points to a position of the gel where a distinct band was present in isolate MI 575 but absent in isolate MI 571 when viewed with maximum illumination. The same profiles were obtained when the isolated Omps were electrophoresed in a urea-containing gel (data not shown).

pylori outer membrane profile contains several less abundant proteins as is visible in Fig. 5.10. The arrows in both panel A and panel B indicate where differences are discernible in protein bands between each respective isolate pair. Isolate MI 459 appears to lack at least two prominent protein species that are present in MI 458; one which resolves at just >66 kDa and a second which is slightly smaller than 56 kDa. Close inspection of the profile also reveals the presence of a protein band whose expression might be downregulated in strain MI 459 (indicated with a dotted arrow). The image of the outer membrane profile for isolates MI 571 and MI 575 is not as clear as that for strains MI 458 and MI 459. However, it is apparent that the same abundant band that resolves at a position immediately greater than the 66-kDa marker protein and that was absent in isolate M 459 is also absent in strain MI 571 and present in isolate MI 575. It is possible that this protein species of >66 kDa is a member of the Hop family of proteins (specifically the Y-Hop subgroup where Y signifies that the terminal amino acid residue of the protein is a tyrosine residue), as all but two of the members of this group are 70-80 kDa in size (Alm et al., 2000). Both HopM and HopN are members of the Y-Hop subgroup as is HopQ, the protein which is encoded by the DNA sequence of clone 47 and was isolated by PCR-subtractive hybridisation of isolates MI 571 and MI 575 (see Table 4.3). Similarly, the protein species at <56 kDa could be HopA, B or C as these three proteins have molecular sizes of 48, 49 and 50 kDa, respectively (Exner et al., 1995).

5.3. Discussion.

5.3.1. Microevolution.

Descriptions of the occurrence of microevolution in pathogenic microorganisms, including bacterial and fungal pathogens as well as *H. pylori*, have been reported several times in recent years (Achtman, 1994; Hobbs *et al.*, 1994; Franzot *et al.*, 1998; Ziebuhr *et al.*, 1999; Bavoil *et al.*, 2000; Israel *et al.*, 2001; Gee *et al.*, 2002). The emergence *in vivo* of variants of these organisms with new genotypic and phenotypic properties has been postulated to be a mechanism whereby survival in the host is enhanced, genetic instability providing a means to evade host defences or to adapt to changing conditions within the host microenvironment inhabited by the organism. Thus, this *in vivo*- or within-host variation of the species might be an essential component of the pathogenic process (Levin and Bull, 1994). However, much as the occurrence of microevolution lends itself to the theory that *in vivo*-occurring genetic variations result in advantageous phenotypes and subpopulations of the organism that are better adapted, it is only through the molecular characterisation of how the genetic diversity is created that one can fully appreciate the extent to which pathogenic microevolution might alter the antigenic characteristics of an organism in an advantageous way and hence can be considered to have genuine adaptive significance and explain the reasons for its selection.

Microevolutionary changes in clonal populations of several important pathogens are now understood at an in-depth level with extensive knowledge available about the molecular basis by which the genetic changes arise and their functional consequences. The well-documented examples of recombination and genetic exchange occurring within *N. meningitidis* at Opa protein-encoding genes and within *N. gonorrhoeae* at the loci where type 4 pili are specified represent perhaps the best understood of these systems of microevolutionary antigenic variation, together with what is known about antigenic variation of the surface-expressed variable major proteins of *Borrelia hermsii* (Hobbs *et al.*, 1994; Robertson and Meyer, 1992; Ziebuhr *et al.*, 1999).

Although genetic variants of *H. pylori* in clonal populations were observed in the relatively early work that was carried out on this organism (Oudbier *et al.*, 1990) followed by the observation of the adaptation of *H. pylori* strains *in vivo* in an experimental model of infection (Akopyants *et al.*, 1995), the degree to which the occurrence of microevolution in *H. pylori* is understood lags considerably behind investigations of the same subject in other microorganisms. This is probably founded in the fairly recent advent of microarray

technology as well as PCR-subtractive hybridisation as described in the preceding chapter as means to analyse microevolution directly. Where *H. pylori* is concerned, the application of these methodologies was necessary for an accurate appraisal of microevolution because in the time between its first recognition in this species and the development of these analytical approaches, phenotypic variability in specific antigens (known virulence factors or immunogenic proteins) had not been observed, albeit that variability in the 3'-end of the *cagA* gene and point mutations in the genes involved in LPS biosynthesis had been described. Hence, a method of assessing microevolution on a genome-wide scale was required.

5.3.2. PCR-subtractive hybridisation findings.

Employed as a means to isolate the differences between two highly similar H. pylori chromosomes, PCR-subtractive hybridisation effectively removed strain-specific differences in genetic content without the need to carry out entire genome sequencing. The DNA fragments that were isolated represented diverse loci which, based on Southern blot evidence, for the majority of cases were each rearranged between paired strains as opposed to being present in one and absent in the other. Further, as the analysis was chromosome-wide and not targeted to loci of any specific coding capacity, the full nature of what was isolated could not have been predicted or assumed prior to application of the technique. However, as discussed already (Section 4.3), it was not altogether surprising that Omp-encoding genes were identified as being involved in the generation of microevolution by comparison with what is known about microevolution in other bacteria. Indeed, the findings discussed in Chapter 4 opened up the exciting possibility that H. pylori Omp-encoding genes undergo recombination as a method of generating antigenic variation in a manner similar to what is observed in N. meningitidis or B. hermsii amongst others. It was of considerable interest, therefore, to attempt the characterisation of the rearrangement event that had taken place at an Ompencoding locus in H. pylori in order to advance the information gleaned from PCR-subtractive hybridisation analysis and to delve into the mechanisms of chromosomal recombination exhibited by H. pylori. The opportunity to perhaps define a mechanism whereby H. pylori undergoes adaptive mutation in vivo was potentially forthcoming.

5.3.3. Insights about the diversity present between MI 458 and MI 459 at hopM/hopN loci deduced from Southern blot analysis.

Clone 2 isolated by PCR-subtractive hybridisation of isolate pair MI 458/MI 459 consisted of DNA sequence from the loci HP1342/jhp1261 (hopM) and HP0227/jhp0212 (hopN) which are duplicate alleles in both strains 26695 and J99 and encode Omps from the Hop family which have been designated HopM and HopN (Alm et al., 2000). As explained in Section 4.2.1, it was necessary to establish whether what had taken place in isolates MI 458 and MI 459 was due to the occurrence of a gene duplication event because analysis of Southern blot evidence indicated this might be a plausible possibility or whether a mutational versus a recombinational event was responsible for the Southern blot hybridisation pattern observed.

When HindIII-restricted DNA from both strains MI 458 and MI 459 was probed with a DIGlabelled clone 2 probe that was cleaved in two, the signal intensities of bands in the profile differed with each of these truncated probes by comparison with what was detected when the full-length probe was utilised (Fig. 5.1). It had been suspected that if, for example, in isolate MI 458 each probe-reactive HindIII fragment represented individual fragments where each of the duplicated alleles was located, hybridisation signals of the same intensity would be observed with each truncated probe as was observed with the original probe and that these would be comparable between strains MI 458 and MI 459. Clearly, the DNA located on the ~2.2-kb HindIII-fragment in strain MI 459 is not identical to that located on the fragment of the same size in strain MI 458 based on signal intensity differences. Furthermore, the slight difference in signal intensity apparent between the ~2.2-kb HindIII fragment and the ~2.8-kb HindIII fragment in strain MI 458 (Fig. 4.5.A and Fig. 5.1) is borne out when each of the truncated probes is used, indicating unequivocally that these fragments do not represent exact duplicates of the same DNA sequence. Thus, the surmise that isolate MI 459 contained a single allele which became duplicated in isolate MI 458 was incorrect. In addition, as explained in Fig. 5.2 using a schematic diagram, it was also deemed unlikely that the occurrence of nucleotide mutations in the HP1342/jhp1260 and HP0227/jhp0212 loci could account entirely for what was observed by Southern blotting in strains MI 458 and MI 459. Evidently, something more intricate than could be inferred from Southern hybridisation analysis had taken place involving the HopM- and HopN- encoding alleles in strains MI 458 and MI 459.

5.3.4. Polymorphism between MI 458 and MI 459 at Omp-encoding loci defined.

5.3.4.1. PCR from flanking ORFs.

Genotypic comparison using PCR of *H. pylori* strains, each of whose Southern blot hybridisation profiles differed with the clone 2 probe, for the presence of both HP0227/jhp0212 (*hopM*) and HP1342/jhp1261 (*hopN*) revealed that Southern blot evidence (i.e., one hybridising fragment as a possible indicator of the presence of one allele) cannot be directly equated with the real chromosomal scenario (Fig. 5.3).

Are duplicate alleles present both in isolates MI 458 and MI 459 and in isolates MI 545A and MI 545C? Amplification across the HP0227/jhp0212 and HP1342/jhp1261 loci by PCR using the primers 1260/1262 and 211/213 revealed that the *hopM* and *hopN* genes appeared to be present in both strains MI 458 and MI 459 (Fig. 5.4.A) while a different sized PCR product was amplified from each of strains MI 458 and MI 459 using the *hopN*-specific primers (primers 1260 and 1262). Taken together, the Southern blot evidence discussed above for strains MI 458/MI 459 and this PCR information, which testifies to the presence in both strains of each of these *hop* alleles, was a further indication that erroneous assumptions had been made about the possible genotype of these paired *H. pylori* strains in relation to these alleles based on initial hybridisation evidence. Amplification of the *hopM* and *hopN* alleles from flanking ORFs in isolate pair MI 545A and MI 545C which also displayed evidence of rearrangement involving these loci by Southern blotting (Fig. 4.6.D) revealed that, similarly to isolate pair MI 458/MI 459, isolate pair MI 545A and MI 545C also appeared to have both alleles (Fig. 5.4. B).

Furthermore, the indications from the PCR results were that there was a size discrepancy between isolates MI 458 and MI 459 at the *hopN* locus. The presence of *hopM* and *hopN* orthologues in strains MI 458/MI 459 and MI 545A/MI 545C is in keeping with the findings of the study by Alm *et al.* (2000) who found duplicate orthologues in all but one of the nineteen strains examined. Significantly, in their study what the authors described as a deletion of the N-terminal region of HP1342/jhp1261 was detected in four strains and, thus far, what had been observed in strains MI 458 and MI 459 at locus HP1342/jhp1261 as a PCR product-size difference seemed reminiscent of what had been found by Alm and co-workers (2000).

5.3.4.2. Localisation of heterogeneity.

A more in-depth investigation of the PCR product size discrepancy present between MI 458 and MI 459 when the primers 1260 and 1262 were used was merited given that the objective of the studies was to characterise the rearrangement at the Omp-encoding loci that had been identified by PCR-subtractive hybridisation in tandem with Southern hybridisation. As outlined in Section 5.2.5, attempts to construct a sub-genomic library in order to clone the restriction fragments of strains MI 458 and MI 459 which hybridised to the clone 2 probe met with little success as did efforts to directly amplify these restriction fragments. Cloning and amplification attempts were made in order that the DNA sequence of the restriction fragments could be determined and the mode of rearrangement defined in this way. Indeed, the realisation that the rearrangement event being investigated might be deciphered from simple PCR investigations was somewhat fortuitous, and also meant that shorter tracts of DNA would have to be sequenced than would have been required with alternative approaches. Thus, as it transpired, isolation of DNA polymorphic between isolates MI 458 and MI 459 at the Omp-encoding loci in question appeared set to be markedly easier than employing any previously discussed method of restriction fragment isolation.

In order to localise the diversity to a smaller chromosomal region, the PCR primers Cln2#1 and Cln2#2 which bound to the clone 2 or HP0227/jhp0212- and HP1342/jhp1261-specific sequences and read outwards from each other were used in combination with the primers specific for ORFs flanking the HP0227/jhp0212 and HP1342/jhp1261 loci. In this way it was possible to ascertain the precise location of the diversity at these loci and whether or not it involved one or both of the hopM and hopN alleles. The PCR reactions described in Section 5.2.6 were also carried out for isolates MI 545A and MI 545C with the amplimers obtained for both sets of strains visible by gel electrophoresis in Fig. 5.5.B and C. electrophoresis of the PCR products it was evident that between strains MI 458 and MI 459, polymorphism existed at the chromosomal region upstream of the hopN gene as evidenced by the size difference in PCR products amplified using the primers 1260 and Cln2#1 (Fig. 5.5.B; lanes 5 and 6). The uniformity in size of the remaining PCR products between strains was evident by gel electrophoresis and indicated that, in as much as can be defined using this approach to analyse the diversity that exists, a region of polymorphism had been successfully localised. Similarly, between isolates MI 545A and MI 545C all of the PCR products were comparable between both strains save for those amplified using the primers Cln2#2 and 1262

(Fig. 5.5.C; lanes 7 and 8), again localising the chromosomal diversity to the region downstream of the *hopN* gene in this instance.

The DNA sequences of the PCR products amplified with the primers Cln2#1 and 1260 from strains MI 458 and MI 459 and of those amplified from strains MI 545A and MI 545C with primers Cln2#2 and 1262 were likely to be highly informative in terms of assessing the mode of rearrangement that had taken place.

5.3.4.3. Diversity upstream of the *hopN* gene in strain MI 459.

The sequences determined for each of the isolate pairs MI 458/MI 459 and MI 545A/MI 545C with the primers Cln2#1/1260 and Cln2#2/1262, respectively, are depicted schematically in Fig. 5.6. As is shown, in isolate MI 458 what exists is essentially the wild-type scenario where strains 26695 and J99 are considered wild-type strains, i.e., one finds each of the alleles which encode the HopM and HopN proteins flanked on either side by the same flanking genes that are found adjacent to the *hopM* and *hopN* genes on both of the sequenced genomes. In isolate MI 459, however, the primers Cln2#1 and 1260 amplified a PCR product some 900 bp longer than was amplified from strain MI 458 using the same primers. The sequence of the additional DNA present in isolate MI 459 was found largely to be that of the gene HP0228/jhp0213 (Fig. 5.6.B).

These data suggest the possible insertion of a horizontally transferred fragment of DNA into strain MI 459 that consisted of half of the coding sequence of the HP0228/jhp0213 gene with upstream non-coding DNA and most likely also some of the 5'-sequence from the *hopM* (HP0227/jhp0212) gene, which (as shown in Fig. 5.6) is transcribed in the opposite direction to the HP0228/jhp0213 gene. Upon arrival into strain MI 459, the putatively horizontally transferred piece of DNA (which could have been considerably longer than the 900-bp fragment that became inserted) might become chromosomally inserted by homologous recombination as described in Section 5.2.8 and depicted in Fig. 5.7. The presence of a seemingly intact HP0228/jhp0213 locus downstream of the *hopM* gene in both MI 458 and MI 459 based on PCR results is a further indication that the DNA inserted upstream of the *hopN* gene in strain MI 459 is likely to have been acquired from an exogenous source. If rearrangement between the two Omp-encoding loci had taken place to elicit the chromosomal

structure found in strain MI 459 one would expect polymorphism to exist between strains MI 458 and MI 459 at the chromosomal region upstream of the *hopM* gene.

It is intriguing to detect the presence and chromosomal insertion of a horizontally transferred DNA fragment between the chromosomes of two H. pylori strains of proven clonality which involves the HP0227/jhp0212 and HP1342/jhp1261 genes. Kersulyte et al. (1999), in a study based on a similar rationale to that described in the present and preceding chapter, reported evidence of recombinational scrambling between these two same alleles which was detected initially by sequencing of RAPD-PCR fragments polymorphic between isolates in what were considered two distinct yet clonal lineages. A deletion of a large region (1.6 kb) of the 5'-end of the HP1342/jhp1261 allele was also found in subtypic recombinants in this group of strains (Kersulyte et al., 1999). Thus, between isolates MI 458 and MI 459, based on the isolation by PCR-subtractive hybridisation of a DNA fragment which consisted of sequence from the HP0227/jhp0212 and HP1342/jhp1261 alleles, the possibility existed that a similar recombination event to what had been detected in the study by Kersulyte and co-workers (1999) might also have occurred in these Irish strains. However, as described in Section 5.2.8 and Fig. 5.7, a distinct recombinant was revealed arising from horizontal transfer of DNA. However, the recombination that would have been required between the HP0227/jhp0212 (hopM) sequence on the transferred fragment and the chromosomal copy of the HP1342/jhp1261 (hopN) locus serving as one of the cross-over events required to achieve chromosomal insertion of the fragment by homologous recombination may have contributed to the recombinational scrambling of these alleles that was reported extensively by Kersulyte et al. (1999). Although the recombination involving the alleles of the duplicate hopM and hopN genes was evidenced purely on the basis of DNA sequence information by Kersulyte et al. (1999), the physical observation of the requirement for recombination between the hopM and hopN genes as found in the present analysis advances the findings of the previously published study further by describing an in vivo scenario in which recombination has taken place.

5.3.4.4. Adaptive significance of horizontal gene transfer between isolates MI 458 and MI 459.

The use of PCR and DNA sequencing has thus allowed the characterisation of a 'microevolutionary event' between isolates MI 458 and MI 459. Although recombination

between Omp-encoding genes was involved in chromosomal insertion of the putatively horizontally acquired fragment in strain MI 459, what was detected was not rearrangement between Omp-encoding genes *per se* in the sense that two distinct Omp genes (as opposed to duplicate genes of identical coding sequence) sharing a high level of DNA sequence similarity were not found to have recombined their DNA sequences. With this knowledge at hand, it is possible to make a more informed judgement about the relevance of the described event as a possible mechanism of host-specific adaptation of this *H. pylori* strain. Kersulyte *et al.* (1999) assert that the recombinations in the regions identified, including the sequence scrambling and the 1.6-kb deletion in the HP0227/jhp0212 and HP1342/jhp1261 loci, were likely to have been specifically amplified by selection for good growth or persistence given their high detection level in single colony isolates from this patient.

In relation to the mechanism of recombination defined for isolates MI 458 and MI 459 (Fig. 5.7), the recombinant chromosomal structure present upstream of the hopN gene in isolate MI 459 does not immediately strike one as likely to be highly advantageous (the inserted DNA encodes half of the HP0228/jhp0213 gene and also contains non-coding DNA) and therefore, is difficult to rationalise as being significant in terms of strain adaptation. Indeed, it must be borne in mind that this specific recombinant might have been fortuitously selected for and amplified among other H. pylori subtypes in this individual's stomach by virtue of the selection of other useful phenotypes generated through recombination at other loci in this same strain pair (see Tables 4.1. and 4.2). As a counterpoint, however, the mixing of the gene sequences for the HP0227/jhp0212 and HP1342/jhp1261 alleles alone through homologous recombination, the concomitant insertion of HP0228/jhp0213 sequence and of upstream noncoding DNA aside, might have generated a genotype for the hopN allele that was best suited in this particular patient. Additionally, the HP1342/jhp1261 allele is likely in strain MI 459 to be operating under the control of the promoter for the HP0227/jhp0212 gene, assuming the functionality of this promoter region at its displaced chromosomal location. A hopN gene which is functioning better could result. However, the possibility also exists that the presence of this essentially foreign DNA fragment might have a detrimental effect on the expression of HP1342/jhp1261 in isolate MI 459.

5.3.4.5. Summary of evidence in relation to isolates MI 458 and MI 459 for rearrangement at the HopM/HopN-encoding genes.

It is pertinent to attempt to marry the entire collection of evidence in relation to isolate pair MI 458 and MI 459 and the perceived rearrangement of the HopM/HopN-encoding genes.

- PCR-subtractive hybridisation of isolate pair MI 458/MI 459 resulted in the isolation of the 'clone 2' fragment which appeared to differ between these two strains.
- Southern blot analysis (Fig. 4.5.A and Fig. 5.1) indicated that the pattern of hybridisation
 of the clone 2 fragment to strains MI 458 and MI 459 could not entirely be accounted for
 by a mutational basis of diversity, and that hopM/hopN allele duplication had not either
 taken place between these strains.
- PCR and DNA sequencing revealed an insertion upstream of the hopN gene in isolate MI
 459 of a ~900-bp DNA fragment that was likely to have been acquired by horizontal
 transfer which seemingly left the hopM and hopN gene sequences intact in both strains MI
 458 and MI 459.
- With reference to Fig. 4.6.D, isolates MI 458 and MI 459 also differed at the HP1343/jhp1262 gene, the locus located downstream of *hopN*.

The conclusions, therefore, in relation to this entire collection of evidence are twofold: (i) that the chromosomal insertion upstream of the *hopN* gene as a defined microevolutionary event seems to have been serendipitously discovered; and (ii) that definitive answers as to why the clone 2 fragment was isolated by PCR-subtractive hybridisation and at what level the apparent rearrangement of the HP1343/jhp1262 gene is connected with that in the Omp-encoding genes can most appropriately be addressed by determining the sequences of the restriction fragments to which the probes bind. Arguably, PCR analysis has not unfolded the full story in relation to rearrangement of the *hopM/hopN* alleles in isolate pair MI 458/MI 459.

- 5.3.5. Characterisation of microevolution at the hopM/hopN loci between MI 545A and MI 545C.
- 5.3.5.1. Insertion of the transposable element IS*Hp608* downstream of the *hopN* gene in isolate MI 545A.

Microevolution in isolate pair MI 545A/MI 545C at the same Omp-encoding genes was defined mechanistically in a different way to that described for strains MI 458 and MI 459. In this instance, the insertion downstream of the *hopN* gene of a novel *H. pylori* insertion sequence element (IS*Hp608*) was detected (Fig. 5.6 and Fig. 5.8). IS*Hp608* is a new member

of the IS605 transposable element family and contains two overlapping ORFs unlike simpler one-gene transposable elements. The gene designated orfA encodes a transposase which is required for transposition of the element. The second gene, orfB, is not essential for transposition and also encodes a putative transposase (Kersulyte et al., 2002). ISHp608 became inserted in the intergenic region between the hopN and HP1343/jhp1262 genes in isolate MI 545A (Fig. 5.8). Interestingly, BLAST analysis of the sequence of the PCR product obtained for the Irish isolate MI 545A with the primers Cln2#2/1262 revealed that the ISHp608-encoding portion of this sequence contained highest sequence identity to the sequence of the element found only in Peruvian and Alaskan strains by Kersulyte et al. (2002). The sequence found in strains from these populations represents one of the two types of the element detected, the second being widely geographically distributed.

According to Kersulyte et al. (2002), the element generates simple insertions upon transposition and not cointegrates. It inserts downstream of the motif TTAC and does not duplicate target sequences during transposition. ISHp608 does not either have terminal inverted repeats that are characteristic of many mobile elements (Mahillon and Chandler, 1998). By reference to Fig. 5.8, the precise DNA sequence in isolate MI 545A between which the ISHp608 element became inserted is shown. The sequence of the intergenic region between the HP1342/jhp1261 and HP1343/jhp1262 genes in isolates MI 545A and MI 545C has higher identity to that from strain J99 than strain 26695. This is most likely an indication of the progressive divergence of this region of sequence in different H. pylori populations over time. The TTAC motif, downstream of which the transposable element has a propensity to insert (Kersulyte et al., 2002), is found in strain J99 and MI 545A at the 3'-end of the hopN gene. Significantly, however, the TTAC sequence is not present in strain MI 545C, which is considered wild-type with respect to this chromosomal region (Fig. 5.6). Therefore, the TTAC sequence found almost immediately downstream of the insertion point of ISHp608 in strain MI 545A (Fig. 5.8.C) is not considered part of the native sequence of this region at the insertion point of the element. Alternatively, the sequence of this region determined for isolate MI 545C might reflect the scenario where the mobile element has been lost from this site with the movement of this insertion-specific sequence.

5.3.5.2. IS elements in *H. pylori*.

Mechanisms of insertion aside, the transposase proteins encoded on the IS elements are responsible for the movement of the mobile DNA segment from one position of the genome to another and, hence, contribute to genome evolution by changing its structure. Indeed, the first such element identified in H. pylori (IS605) was found in the cagPAI and was found to be capable of disrupting the structure of this region, thereby contributing to its evolution (Censini et al., 1996). Furthermore, IS605 was identified as being an ancient component of the H. pylori gene pool and disseminated in the species by horizontal transfer and homologous recombination (Höök-Nikanne et al., 1998). Another member of this family of IS elements, IS606, was also identified in the genomes of the H. pylori sequenced strains and, as was found for IS605, can exist in multiple copies and as full-length or partial elements (Marshall et al., 1998). As well as being mobile in nature and capable of movement between different chromosomal sites, IS elements additionally, by acting as repetitive DNA owing to their multiplicity throughout the *H. pylori* chromosome, may lead to recombinational events resulting in chromosomal rearrangement. It was particularly fortuitous to locate the transposable element ISHp608 at a chromosomal site in one isolate from a clonally related pair and its absence from the same site in the second isolate. Without the availability of information as to the copy number of this element in these strains, one can only say at this juncture that the element has either been acquired horizontally from an exogenous source and become inserted downstream of the hopN gene in isolate MI 545A, or, that it has become transposed from an alternative chromosomal site in these isolates to the location described.

5.3.5.3. IS element transposition and host-specific adaptation.

The horizontal transfer or intra-chromosomal transposition of ISHp608 undoubtedly contributed to the microevolution of the genomes of isolates MI 545A and MI 545C. However, as was encountered with isolate pair MI 458/MI 459 and the chromosomal insertion upstream of the hopN gene described earlier, the possible usefulness in terms of host-specific adaptation this event has is not immediately obvious, although mutation caused by movement of the element could alter expression of genes nearby the site from which the element became transposed and genes adjacent to its new chromosomal location. Some of the impact of the presence in a genome of transposable elements is likely to be felt in relation to this latter property. However, the exciting possibility exists that ISHp608 might effect other chromosomal rearrangements by recombining with additional copies of the element that might exist elsewhere on the chromosome by an extension of the fact that IS605 and IS606 are

consistently found as multiple copies (Tomb *et al.*, 1997; Alm *et al.*, 1999). Analysis by Kersulyte *et al.* (2002) found the presence of IS*Hp608* in duplicate in three isolates examined. As detailed in Table 4.2, rearrangement at two other loci was also detected in isolate pair MI 545A/MI 545C, albeit two Omp-encoding loci.

5.3.5.4. Apparent anomalies in PCR evidence.

When isolates MI 545A and MI 545C were analysed for the presence of both the hopM and hopN genes by PCR using primers that bound at flanking genes, what was observed is shown in Fig. 5.4.B. The amplification products obtained indicate that both of these isolates have genes at the chromosomal location which corresponds to the sites of the two Omp-encoding genes. However, in light of what was detected by PCR as shown in Fig. 5.5.C., what is shown in Fig. 5.4 is anomalous. In hindsight, the presence between the HP1342/jhp1261 and HP1343/jhp1262 genes of the element ISHp608 in strain MI 545A should interfere with amplification across the hopN allele using primers 1260 and 1262. However, this was not encountered. By way of reconciling these anomalous observations, is it possible that in vitro culturing of these H. pylori strains leads to movement of the IS element between chromosomal sites, thereby allowing for an apparently wild-type chromosomal configuration to exist in both isolates MI 545A and MI 545C downstream of the hopN gene despite the onetime presence there in isolate MI 545A of ISHp608? Furthermore, this anomalous observation may arise from the manner in which isolate MI 545C was stored after primary As detailed in Section 2.1.4, isolate MI 545C was stored as pooled colonies because it consistently grew poorly during culture. In effect therefore, the presence of multiple clones within this strain cannot be ruled out and may account for the observed anomalies or furthermore, co-culture of multiple clones in vitro may allow for the acquisition of DNA through natural transformation and homologous recombination and thus, the transposition of the ISHp608 element. This therefore, potentially casts doubt over whether this characterised microevolutionary event is a true representation of in vivo diversity in H. pylori.

5.3.5.5. Future perspective with regard to *H. pylori* Omp-encoding gene rearrangements.

The genetic events characterised for isolate pairs MI 458/MI 459 and MI 545A/MI 545C as detailed in the present chapter do not directly alter the integrity of the coding sequences of either the HP0227/jhp0212 or HP1342/jhp1261 alleles or those of the respective flanking genes in as much as can be determined from the PCR and sequencing analysis that has been

carried out. The effect of the chromosomal insertions upstream and downstream of the *hopN* gene in isolates MI 459 and MI 545A, respectively, could, as alluded to previously, change the expression of nearby genes. This, however, is purely speculative and would require extensive experimentation to prove. The objectives of the studies carried out in this chapter were to characterise mechanistically how microevolution arises in clonal isolates of *H. pylori* and to ascertain the extent to which these recombinant clonal subtypes might have been amplified in the resident population of this organism through selection for useful recombinant genotypes.

As described in Chapter 4, the widespread finding of what was perceived as rearrangement of *H. pylori* Omp-encoding genes between paired isolates strongly indicated that this organism might undergo adaptation *in vivo* by the recombination of Omp-encoding DNA sequences, the mixing of sequences between different genes generating proteins with altered antigenic properties. This, however, failed to be proven in the studies described in this chapter. Yet, it is believed that had more of the individual rearrangement events at Omp-encoding genes that were identified in the preceding chapter been characterised fully, evidence of recombination between distinct Omp genes would be forthcoming. The equivocal issues in relation to recombination of *H. pylori* Omp-encoding genes notwithstanding, phenotypic evidence accumulated for paired *H. pylori* isolates indicates that the situation in relation to rearrangement of *H. pylori* Omp genes requires that much new information be uncovered and definitely merits complete investigation.

5.3.6. Culture of paired isolates in liquid medium.

When isolates MI 458 and MI 459 (as well as other paired strains used throughout these studies) were cultured on solid medium, colony size differences between the two strains were apparent (personal observation) and suggested that each individual isolate had a different growth rate. To examine this more closely, these strains were cultured in liquid medium and it was found that within the two pairs analysed, each isolate had a different growth rate.

These findings are reminiscent of what was observed on agar culture by Kersulyte *et al.* (1999) as fast- and slow-growing phenotypes of *H. pylori* clonal subtypes. Growth rate differences in *H. pylori* strains MI 458/MI 459 and MI 545A/MI 545C, which display high overall fingerprint identity between pairs, are thought most likely to be founded in genotypic differences arising from the occurrence of chromosomal microevolution, and not, as asserted

by Kersulyte *et al.* (1999) from physiological adaptation. As alluded to in Section 5.2.9, with information at hand about the involvement of Omp-encoding ORFs in the generation of subtypic variants, it makes for compelling speculation that growth-rate differences might be correlated with aberrant functioning or expression of certain Omps because of chromosomal recombination of genes which encode these proteins. This in turn could hamper the uptake of nutrients from the medium *in vitro* or surrounding gastric tissue *in vivo*, causing the growth rate of the isolate to be noticeably altered. Further investigation of this question, working in essence from reverse by starting with a known phenotype, could lead to the functional characterisation of a *H. pylori* porin with a narrow range of nutrient specificity perhaps. Alternatively, undefined polymorphism between paired isolates in genes which encode enzymes responsible for nutrient turnover might also cause growth-rate differences to be observed between strains if the metabolic precursors cannot be converted.

It is important to bear in mind the consequences that differences in the vigour of bacterial growth between clonal isolates might have in the physiological situation. It follows that in the in vivo environment, strain MI 459 can achieve a greater density of culture than does isolate MI 458. This sole phenotypic trait might indicate that strain MI 459 has through microevolution become adapted for more vigorous growth and hence become a more virulent recombinant subtype. Indeed, a study by Atherton et al. (1996) assessed the density of H. pylori in vivo by quantitative culture and histological examination of biopsy specimens. It was found that bacterial density (i.e., vigour of growth) was positively correlated with the presence of the virulence marker genes cagA and vacA type s1, and that mean densities were higher in strains of this genotype than in cagA-/vacA s2 strains. More significantly, where higher culture densities existed, associations were found with mucosal neutrophil and lymphocyte infiltration and with epithelial injury. Duodenal ulcer patients were also found to have a 5-fold higher mean density of bacteria in the antrum of their stomachs than patients with other disease outcomes. Thus, it would appear that the density of H. pylori in vivo plays a significant role in pathogenesis and thus heightens the importance of the complete set of genotypic and phenotypic information which has been garnered for isolate pairs MI 458/MI 459 and MI 545A/MI 545C in terms of physiological relevance.

5.3.7. Examination of the Omp profile of MI 458/MI 459 and MI 571/MI 575 by SDS-PAGE.

Considerable information had been accumulated which suggested that microevolution had elicited changes in the outer membranes of paired isolates. With a view to observing this directly, Omps were isolated from each isolate in the paired strains MI 458/MI 459 and MI 571/MI 575. The profile of Sarkosyl-insoluble outer membrane proteins produced by each isolate were visualised by staining the protein bands resolved by SDS-PAGE as shown in Fig. 5.10. Panel A of this figure shows the Omp profiles of strains MI 458 and MI 459 and panel B those of strains MI 571 and MI 575. Differences in the profiles of the Omps present between the isolates in each pair suggest at least a link with the findings of the PCR-subtractive hybridisation investigation and the differences in growth rate observed between paired strains (Section 5.3.5).

In the findings of a novel study, Keenan et al. (2000a) report the shedding from H. pylori cells in vivo of outer membrane vesicles in which the presence of porins and biologically active VacA cytotoxin were found. The vesicles are postulated to play what is described as an offensive role in *H. pylori* pathogenesis by acting as a vehicle whereby damaging molecules might be delivered to the mucosa, thereby contributing to chronic inflammation. This is a further indication that H. pylori Omps might function significantly in the pathogenesis of infection, their role as semi-selective channels and adhesins aside. Indeed, a 34-kDa proinflammatory Omp (HP0638) was identified by Yamaoka et al. (2000) and put forward as likely to be an important factor in the risk of clinically significant pathology. The locus HP0638 was designated as the oipA gene. An outer membrane lipoprotein (Lpp20) identified in H. pylori outer membrane vesicles by Keenan et al. (2000b) was also found to be immunogenic and, moreover, to elicit protective immunisation. It would seem, therefore, that the roles of *H. pylori* porins in pathogenesis, reaching beyond their transport functions, further reinforces the likelihood that these molecules could be subject to selection pressure from which the necessity to adapt arises, and that what is observed in Fig. 5.10 as the alteration in outer membrane protein profiles between clonally related isolates is direct evidence of this fact as was also evidenced in the results obtained in PCR-subtractive hybridisation analysis.

Chapter 6

Genotypic analysis of familial H. pylori isolates and an appraisal of the occurrence of clonal variation in isolates from different individuals

6.1.	Introd	luction.	152
6.2.	s	154	
	6.2.1.	Oligofingerprint analysis of familial isolates.	154
	6.2.2.	RAPD-PCR analysis of familial isolates.	154
	6.2.3.	Estimation of similarity coefficients present between isolates	
		intrafamilially	155
	6.2.4.	PCR-subtractive hybridisation analysis of two clonally related	
		isolates from each family group.	155
	6.2.5.	DNA sequence analysis of clones from putatively rearranged genes	
		obtained from PCR-subtractive hybridisation of MI 431 and MI 577A	156
	6.2.6.	Examination of clonally related isolates from Family 1 and	
		paired familial isolates for diversity within the cagA gene	157
6.3.	Discu	ssion.	158

6.1. Introduction.

A study of the pathogenesis of infection of a micro-organism will always seek to identify the natural reservoir of infection of the organism and to establish its likely mode of transmission between individuals. Both of these are still controversial topics when considering *H. pylori*, with no consensus viewpoint existing on the epidemiology of the organism. No environmental reservoir of *H. pylori* has been found, but a zoonotic reservoir and food as a vehicle of transmission are thought unlikely to be involved in acquisition of the organism (Feldman *et al.*, 1998). Water has been suggested as a route of transmission based on studies carried out in South America in which PCR products of *H. pylori* were amplified from water sources (Hulten *et al.*, 1996), but this has not been a universal finding (Malatay *et al.*, 1996).

Evidence exists for the likely acquisition of *H. pylori* infection by direct contact between individuals. *H. pylori* has been detected in the oral cavity and in faeces by both culture and PCR (Namavar *et al.*, 1995; Dowsett *et al.*, 1999). In particular, there are many proponents of the notion that *H. pylori* resides in dental plaque (Pytko-Polonczyk *et al.*, 1996; Song *et al.*, 2000). Thus, transmission by oral-oral, gastro-oral and faecal-oral routes are all considered possibilities, although no overwhelming evidence exists to support unequivocally one of these modes of transmission. Nonetheless, oral-oral spread between children is thought to be the likely mode of transmission in the developed world with faecal-oral transmission playing a more significant part in *H. pylori* acquisition in developing countries (Feldman *et al.*, 1998).

Transmission of *H. pylori* from person-to-person is also strongly supported by the demonstration of the clustering of infection within families (Wang *et al.*, 1993a; Drumm *et al.*, 1990; Nwokolo *et al.*, 1992). Indeed, molecular analysis of *H. pylori* isolates from multiply colonised family members has provided evidence for the transmission intrafamilially of genotypically identical or highly similar *H. pylori* strains (Bamford *et al.*, 1993; Miehlke *et al.*, 1999; Han *et al.*, 2000). Such intrafamilial transmission of *H. pylori* isolates is perhaps the most prevalent example of the isolation from different individuals of clonally related strains of this organism. The population structure of *H. pylori* has been described as being panmictic or freely recombining (Go *et al.*, 1996; Suerbaum *et al.*, 1998) and the isolation of identical clones from individuals in the general population happens rarely, if at all.

The genotypic diversity present between isolates of *H. pylori* consists largely of nucleotide substitutions at the third base of codons which do not modify the encoded protein sequence

(Alm et al., 1999). Therefore, the proteome of H. pylori is relatively conserved between isolates despite the significant diversity at the level of the genome that molecular fingerprinting analysis testifies to. Yet, the accumulation of genetic variability within H. pylori by various mechanisms in the stomachs of individuals results in a diverse H. pylori population generally and also with respect to certain phenotypic traits (Marshall et al., 1998; Suerbaum and Achtman, 1999). If the transmission of H. pylori occurs by direct contact between individuals as suggested by the evidence from familial data, it follows that when transmitted, identical clones of H. pylori undergo continuous evolution in isolation in the environment of the newly infected host. Despite clonality of H. pylori strains over short periods of time after natural transmission (again, from familial evidence) (Suerbaum, 2000), years or decades of amassed genetic change would eventually obliterate evidence of clonal relationships that once existed between strains.

In the present study the possible transmission of clonal isolates of *H. pylori* between family members was investigated using the highly discriminatory oligofingerprinting genotyping method as well as RAPD-PCR. The strains were also analysed by Southern blotting using a *cagA* gene probe (see Section 3.2.1) to determine if diversity existed between clonally related isolates from different family members within this allele, i.e., upon transmission to a new host did the *cagA* allele undergo alteration in copy number of the 3'-end associated repeated sequences. PCR-subtractive hybridisation was also carried out using genotypically identical or highly similar isolates from each family. PCR-subtractive hybridisation carried out on identical clones from different individuals potentially allows direct analysis of the microevolutionary changes that may have taken place in a single clone upon transmission to a new host. A different host environment and a set of environmental conditions that are not likely to be completely uniform between two individuals are factors that might result in rearrangement taking place in certain genes in order to exert maximal chances of establishing successful colonisation in the new host and to potentiate virulence through genetic adaptation to a clone that is best suited to a new host.

6.2. Results.

6.2.1. Oligofingerprint analysis of familial isolates.

Familial contacts of index patients were recruited following successful culture of an *H. pylori* isolate from these symptomatic individuals. The largest family (Family 1) consisted of six individuals from whom a total of 12 isolates were obtained and the second largest, Family 2, of four members from whom a total of 7 isolates were obtained. Genomic DNA was isolated from each of these strains, restricted with the endonuclease *Hin*dIII and electrophoresed. Southern blots of these restricted DNA samples were probed with the oligonucleotide (GTG)₅ (Fig. 6.1.A and 6.2.A).

Examination of the fingerprint profiles of isolates obtained from each member of Family 1 (Fig. 6.1) revealed that each of the four children from this family colonised by *H. pylori* carried isolates which were almost indistinguishable from one another (Fig. 6.1, Lanes 1–7). However, isolates MI 431, MI 9613A, MI 9613C, MI 018 and MI 613 (lanes 1, 4, 5, 6 and 7, respectively) had greater identity to each other than to isolates MI 577A or MI 577C (lanes 2 and 3, respectively). The area of the fingerprint profiles where band differences exist between isolates MI 577A/MI 577C and the isolates from the remaining children are indicated with arrows in Fig. 6.1. The isolates carried by the mother and father of this family were different from each other and from the strains that colonised the children.

Analysis of the fingerprint profiles obtained with the same oligonucleotide probe for the isolates of Family 2 (Fig 6.2.A) indicated that mother and daughter were colonised with *H. pylori* isolates of highly similar oligofingerprint profile. The remaining individuals from whom isolates were obtained in this family group all had dissimilar fingerprint profiles to each other.

6.2.2. RAPD-PCR analysis of familial isolates.

Pair-wise combinations of arbitrary primers were used to generate RAPD-PCR profiles for each isolate from the two familial groups. Fig. 6.1.B and Fig. 6.2.B show the fingerprints obtained for Family 1 and Family 2, respectively, by RAPD-PCR. For both families, the profiles generated by RAPD-PCR and by oligofingerprinting are in agreement with each other with the children from Family 1 (Fig. 6.1.B, lanes 1–7) having fingerprints of high similarity, with minor band differences existing between isolates MI 577A and MI 577C and the remaining sibling strains. Similarly for Family 2, isolates MI 218A/MI 218C (mother) and

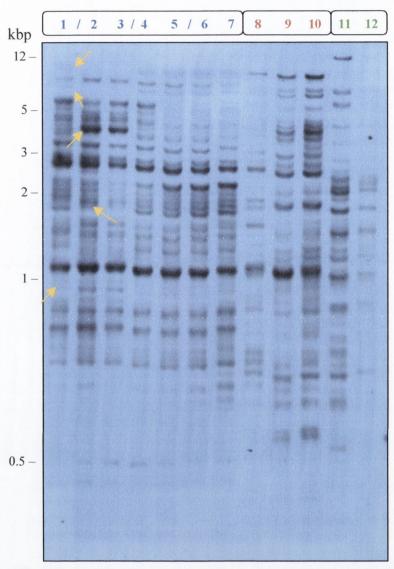


Fig. 6.1 A. Oligofingerprinting analysis of 12 *H. pylori* isolates from six members of Family 1.

The oligonucleotide (GTG)₅ was labelled with DIG and used to probe a Southern blot of *Hin*dIII-digested genomic DNA of 12 *H. pylori* isolates obtained from six members of a family (Family 1). The lanes labelled 1–7 correspond to the isolates obtained from the children of the family (7 isolates from 4 individuals) and are designated as follows: lane 1, MI 431 (index patient); lanes 2/3, MI 577A and MI 577C, respectively; lanes 4/5, MI 9613A and MI 9613C, respectively; lanes 6/7, MI 018 and MI 613, respectively. Lanes 8, 9 and 10 correspond to the isolates obtained from the mother of the family: lane 8, MI 561; lane 9, MI 562; lane 10, MI 563. The isolates in lanes 11 and 12 were taken from the father of the family: lane 11, MI 610; lane 12, MI 611. The location of band polymorphisms in the fingerprint profile between isolates MI 431 and MI 577A are indicated with yellow arrows. The sizes (kbp) of the molecular weight markers are indicated to the left of the panel.

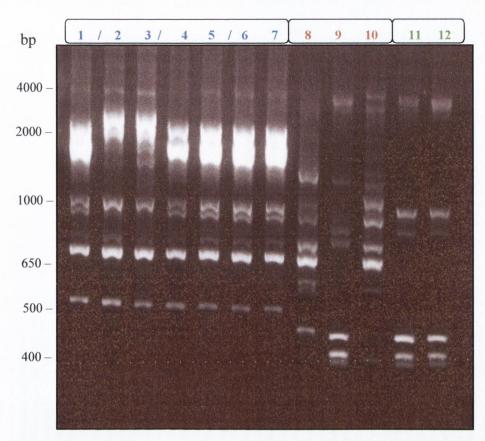


Fig. 6.1 B. RAPD-PCR analysis of 12 *H. pylori* isolates obtained from six members of Family 1.

Agarose gel (1.8% w/v) electrophoresis of RAPD-PCR amplimers of the 12 *H. pylori* isolates taken from individuals in Family 1. The lanes labelled 1–12 correspond to the following strains: 1, MI 431; 2, MI 577A; 3, MI 577C; 4, MI 9613A; 5, MI 9613C; 6, MI 018; 7, MI 613; 8, MI 561; 9, MI 562; 10, MI 563; 11, MI 610; 12, MI 611 Random primers 1254 and D8635 were used in the PCR reactions. Isolates from family members are colour-coded to correspond to panel A. The sizes of the molecular weight markers are indicated (bp).

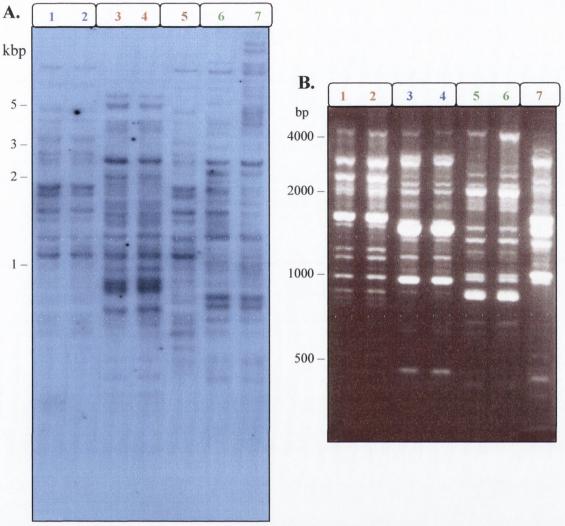


Fig. 6.2. (A.) Oligofingerprinting and (B.) RAPD-PCR analysis of seven *H. pylori* isolates from four family members of Family 2.

- (A.) A Southern blot of *Hind*III-cleaved genomic DNA was probed with DIG-labelled oligonucleotide (GTG)₅. Lanes 1–7 correspond to the following *H. pylori* isolates: 1, MI 218A and 2, MI 218C (mother); 3, MI 605A and 4, MI 605C (son); 5, MI 221C (daughter); 6, MI 222A and 7, MI 222C (daughter's spouse). The sizes (kbp) of the molecular weight markers are indicated.
- **(B.)** Agarose gel (1.8% w/v) electrophoresis of RAPD-PCR amplimers obtained from the isolates from Family 2. Lanes correspond to the following strains: 1, MI 605A and 2, MI 605C (son); 3, MI 218A and 4, MI 218C (mother); 5, MI 222A and 6, MI 222C (daughter's spouse); 7, MI 221C (daughter). Random primers D11344 and 1281 were used in the PCR reaction. Amplimer bands range in size from approximately 4 kbp to 500 bp as indicated to the left of the panel. Lanes in A and B where the same isolate was electrophoresed are depicted by lane numbers of the same colour.

MI 221C (daughter) have RAPD-PCR fingerprint profiles which approach identity (Fig. 6.2.B, lanes 3, 4 and 7, respectively). The use of unequal quantities of MI 218A/MI 218C and MI 221C template DNA in the PCR reaction may account for the differences in band intensity observed between the profiles of these strains.

6.2.3. Estimation of similarity coefficients present between isolates intrafamilially.

The percentage similarity of the strains in the two familial groups to one another was estimated by applying the similarity coefficient formula (Section 2.12.3) to a pair-wise analysis of strains in order to determine numerically the degree of genetic relatedness that existed between isolates intrafamilially. Despite the obvious clonality that was present between isolates in each family group by visual examination of the oligofingerprint and RAPD-PCR profiles as discussed previously, a more accurate determination of strain relatedness is allowed by an estimation of percentage similarity coefficients of one strain to another in a method based on band by band analysis. As described in Section 2.12.3, the number of identical bands between the two profiles is multiplied by two with the resulting figure then divided by the total number of bands in both profiles. The percentage similarity (%S_D) of one strain to another is thus determined. The data for each family are presented in matrix format in Fig. 6.3 where it can be seen that among the members of Family 1, as was perceived from visual examination, each child is colonised by a clonally related strain, whereas both parents have different strains to those carried by the children and to each other. The mother and daughter in Family 2 were determined by similarity coefficient analysis to be colonised by a clonally related strain which differed from those colonising the other two family members analysed.

6.2.4. PCR-subtractive hybridisation analysis of two clonally related isolates from each family group.

The isolates MI 431 and MI 577A were chosen for analysis from Family 1 because of the high genotypic similarity that existed between them while also displaying minor differences in fingerprint profile, akin to what was termed previously as 'microevolution' between epidemiologically related isolates taken from a single individual. Isolates MI 218A and MI 221C from Family 2 were analysed similarly and were the only isolates from this group of strains that had fingerprint profile identity or high similarity.



	431	577A	577C	9613A	9613C	018	613	561	562	563	610	611
431												
577A	90											
577C	90	100										
9613A	98	90	90									
9613C	98	90	90	100								
018	98	98	98	98	98							
613	98	98	98	98	98	100						
561	71	71	71	71	71	71	71					
562	71	71	71	71	71	71	71	94				
563	71	71	71	71	71	71	71	100	94			
610	79	79	79	79	79	79	79	76	70	76		
611	79	79	79	79	79	79	79	76	70	76	100	

B.

	218A	218C	605A	605C	222A	222C	221C
218A 218C 605A 605C 222A 222C 221C							
218C	100						
605A	79	79					
605C	79	79	100				
222A	58	58	73	73			
222C	58	58	73	73	100		
221C	90	90	66	66	48	48	

Fig. 6.3. Similarity coefficient matrices of isolates from (A) Family 1 and (B) Family 2.

Numbers refer to percentage values obtained after applying the similarity coefficient formula to pair-wise combinations of isolates. Brackets join isolates from an individual.

Screening of the amplimers generated upon secondary PCR of the mixture of DNA fragments obtained after subtraction and enrichment of isolate MI 431 with isolate MI 577A yielded several putative clones that, based on probing with labelled genomic DNA from both of these strains, appeared as if they might differ between the two. Using these labelled PCR products to probe Southern blots of *Hin*dIII-cleaved genomic DNA from each of MI 431 and MI 577C, two DNA fragments were isolated that displayed a different hybridisation pattern to *Hin*dIII restriction fragments in each strain (Fig. 6.4.A).

Despite the presence of amplimers after secondary PCR of the subtracted and enriched pool of DNA fragments generated from subtraction of MI 221C with MI 218A, none of the cloned amplimers contained a DNA fragment that was unique to isolate MI 221C (used as the tester strain) or differed between this strain and MI 218A. As with MI 431 and MI 577A, this latter point was determined by using DIG-labelled secondary PCR products to probe a Southern blot of *Hin*dIII-restricted genomic DNA of MI 218A and MI 221C. A representation of some of the clones analysed in this manner but which gave identical hybridisation patterns between both of these strains are shown in Fig. 6.4.B. Oligofingerprint blots of the two sets of paired familial strains are shown in Fig. 6.4.C.

6.2.5. DNA sequence analysis of clones from putatively rearranged genes obtained from PCR-subtractive hybridisation of MI 431 and MI 577A.

The DNA sequence of inserts in the clones designated clone 6 and clone 19, determined by Southern blotting to have a different hybridisation pattern between isolates MI 431 and MI 577A, were analysed. The secondary PCR product of clone 6 was approximately 650 bp in length. BLAST analysis of the DNA sequence obtained for this clone against the two *H. pylori* genome sequence databases showed sequence identity of 98% in a region of overlap with chromosomal sequence from both strains J99 and 26695 of 538 bp. This region contains coding sequence from two contiguous ORFs. One ORF (jhp1151/HP1230) encodes a protein annotated as putative and has no DNA sequence identity or similarity to any previously characterised gene sequence. The second ORF (jhp1152/HP1231) is the gene designated *holB* that putatively encodes the delta subunit of DNA polymerase III.

The insert in clone 19 was almost 1 kb longer than that present in clone 6 at approximately 1650 bp. As with clone 6, the nucleotide sequence of clone 19 consisted of coding sequence of two genes located adjacent to each other in strains 26695 and J99. In an overlap of 794

B.

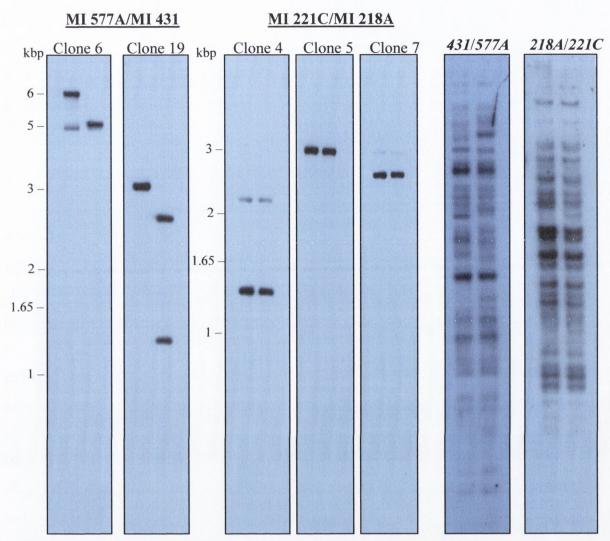


Fig. 6.4. Southern blot analysis of the clones isolated by PCR-subtractive hybridisation of clonally related familial *H. pylori* isolates MI 431/MI 577A and MI 218A/MI 221C.

(A.) HindIII digests of isolates MI 577A (left hand lanes) and MI 431 (right hand lanes) were probed with DIG-labelled probes of the clones indicated at the top of each panel. Similarly (B.) shows the hybridisation of three clones isolated by PCR-subtractive hybridisation of MI 221C (left hand lanes) with MI 218A (right hand lanes) to Southern blots of HindIII digests of these strains. No genetic differences between these latter two strains were detected using the subtractive hybridisation technique. (C.) A DIG-labelled (GTG)₅ oligonucleotide probe was used in hybridisation with HindIII-restricted genomic DNA from each of the indicated strains. Profile bands range in size from \sim 12 kb to \sim 500 bp.

nucleotides, clone 19 shared 94.7% sequence identity with the ORF designated jhp1275/HP1357. This encodes a phosphatidylserine decarboxylase proenzyme. The remaining portion of the sequenced part of clone 19 showed nucleotide sequence identity of 87.9% in a region of overlap 141-bp long with the locus designated jhp1274/HP1356. The gene present at these homologous loci is called *nadA* and encodes a quinolinate synthetase enzyme. These data are summarised in Table 6.1.

6.2.6. Examination of clonally related isolates from Family 1 and paired familial isolates for diversity within the cagA gene.

A Southern blot of *Him*dIII-restricted genomic DNA of ten isolates from Family 1 was probed with the clone-1 *cagA* gene probe in order to ascertain if diversity existed between the clonally related isolates from this family in the *cagA* gene and, by comparison to the findings detailed in Chapter 3, at the 3'-end of this gene. The hybridisation profiles obtained for each of these isolates with this *cagA* probe are shown in Fig. 6.5. Isolates from each of the children from the family display an almost identical pattern of hybridisation with this probe, save for the RFLP present in isolates MI 577A and MI 577C when compared with the remainder of the sibling strains. No major band differences exist between isolates as was seen between paired strains in Chapter 3. Thus, within clonally related isolates of *H. pylori* obtained from different individuals, the *cagA* gene does not appear to undergo 3'-end repeat copy number variation. This was further borne out by analysis of these strains using primers specific for the 3'-end of *cagA* to amplify this region of the gene as described in Chapter 3. PCR products of identical size (~650 bp) were obtained for all of these sibling strains (data not shown).

Table 6.1. PCR-Subtractive hybridisation data obtained from the analysis of familial H. pylori isolates MI 431 and MI 577A.

Clone name	Approx. nested	No. of nucleotides	Blas	t analysis	% Identity	Number of	Putative function of ORFs
	PCR product size	sequenced	J99 Locus no.	26695 Locus no.		nucleotides overlap	
Clone 6	650 bp	567 bp	jhp1151	HP1230	98	538	Hypothetical protein
			jhp1152	HP1231			(holB) DNA polymerase III delta prime subunit
Clone 19	1650 bp	955 bp	jhp1274	HP1356	87.9	141	(nadA) Quinolinate synthetase
			jhp1275	HP1357	94.7	794	(psd) Phosphatidylserine decarboxylase proenzyme

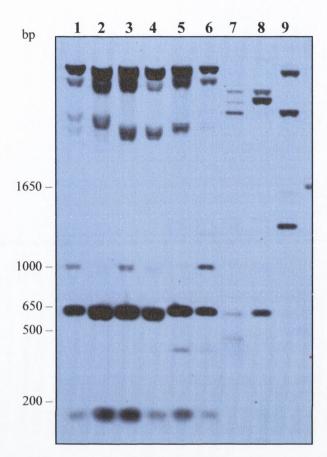


Fig. 6.5. Southern blot analysis of nine *H. pylori* isolates (representing each family member) from Family 1 with a *cagA* gene probe.

The following *H. pylori* isolates from Family 1 were electrophoresed in each lane: 1, MI 577A; 2, MI 577C; 3, MI 9613A; 4, MI 9613C; 5, MI 431; 6, MI 613 (lanes 1–6 correspond to the isolates from the children of the family); 7, MI 561 and 8, MI 562 (mother); 9, MI 610 (father). *HindIII* was used to digest total genomic DNA from each strain and the plasmid clone 1 harbouring a cloned *cagA* gene was DIG-labelled and used as a probe. The position of DNA molecular size standards (bp) are indicated to the left of the blot.

6.3. Discussion.

Efforts to define precisely the epidemiology of *H. pylori* have not kept pace with the advances that have been made in understanding many aspects of the biology of this organism. While not the most crucial factor in elucidating the molecular mechanisms of disease pathology, the possession of epidemiological information for any organism is necessary in order to either prevent initial infection or to establish the risk of reinfection following eradication treatment. By the same token, the route and vehicle of transmission might be significant factors in modulating the consequences of infection (Feldman *et al.*, 1997). Thus, the identification of a natural reservoir of infection and the route of transmission of *H. pylori* between individuals remain significant goals in adding to what is now a very broad picture of *H. pylori* disease aetiology. Moreover, molecular epidemiological studies of *H. pylori* can also be exploited to gain insights into the genomic events likely to be involved in the generation of some of the genotypic diversity observed in the species, keeping in mind that *H. pylori* isolates are believed to undergo evolution in isolation in individual stomachs upon their transmission and establishment of infection there.

The transmission of an organism between individuals or transmission to a host from a natural reservoir implies that the organism can exist in several habitats and can encounter more than one host environment. Indeed, the notion that an organism inhabits a niche means that it would be incorrect to consider bacterial cells as isolated, autonomous entities. Most bacteria exist in populations of many thousands to millions of cells which themselves live on environmental substrata or within animal hosts. It is, therefore, more accurate to think of bacteria as being members of a multicellular community capable of interacting with other cells in the community by molecular cross-talk and of being influenced by factors associated with the host environment (Joyce *et al.*, 2000; Freeman *et al.*, 2000; Sperandio *et al.*, 1999).

By this logic, in the present study clonally related isolates of *H. pylori* taken from the stomachs of different individuals were identified by discriminatory molecular fingerprinting methods. PCR-subtractive hybridisation was then used to detect the genetic differences that existed between these clonal variants taken from different host environments. It was thought likely that upon transmission to a new host by an as yet uncertain route, the influence of this new environment might be brought to bear on the newly infecting strain and effect some type of adaptive phenotypic change that may be detected at a genotypic level. Concomitantly, information about the evolution of clonal lineages in different individuals may also be

gleaned. These represent the likely initial stages of the break-down of these clonal relationships en route to marked diversity. The availability of these clonal strains and the use of PCR-subtractive hybridisation methodology allowed direct probing of this question. Further data to support the transmission of *H. pylori* between family members was also obtained.

H. pylori isolates from two families were analysed. These family groups consisted of mother, father and four children in the case of the family designated Family 1 with 12 H. pylori isolates analysed from these 6 individuals. The second family (Family 2) consisted of four members, namely mother, son, daughter and her spouse. Seven H. pylori isolates were analysed from these four individuals. From oligofingerprinting (Fig. 6.1.A and 6.2.A) and RAPD-PCR (Fig. 6.1.B and 6.2.B) analysis of the isolates from each of these family groups, combined with the coefficients of similarity calculated for these strains (Fig. 6.3), it was clear that the four children in Family 1 were colonised by H. pylori isolates which displayed a genetic relatedness approximating identity.

As previously mentioned, comparison of isolates MI 577A and MI 577C from one child with the strains from the other three siblings revealed that profile band differences were present between these strains. This is perceived as evidence of the occurrence of rearrangement in the genome of this common clone upon transmission of the strain to this family member. Indeed, the existence of clonal variants of *H. pylori* isolates infecting different individuals is reminiscent of the isolation from a single stomach of subtypic variants of the same strain. The isolates MI 431, MI 9613A, MI 9613C, MI 018 and MI 613 had similarity coefficients of 98% while the corresponding coefficient was 90% when these isolates were analysed against isolates MI 577A or MI 577C (the latter being identical to each other) (Fig. 6.3).

Generally, where cases of intrafamilial clustering of *H. pylori* infection have been reported, it has usually been found that parent(s) and child or children were colonised by the same strain (Nwokolo *et al.*, 1992; Bamford *et al.*, 1993; Wang *et al.*, 1993a; Miehlke *et al.*, 1999; Han *et al.*, 2000). Transmission was usually considered to have occurred from parent to offspring or between siblings (Goodman, 2000; Han *et al.*, 2000; Miehlke *et al.*, 1999). However, there is also the possibility that family members become colonised with a particular isolate from a common source. Both visual analysis and examination of the percentage identity of the parental isolates in Family 1 with those of the isolates from the four children indicated that in

this family group neither parent was colonised by isolates which were sufficiently closely related genetically to be considered as even clonal variants of the isolates carried by each child. *H. pylori* isolates from mother and father had percentage identities of 70–76%. Thus, it seems likely in this family that *H. pylori* isolates have been transmitted from person-to-person among siblings but that the children had not acquired the strain from either parent.

Three isolates were obtained from the mother of the family, namely MI 561, MI 562 and MI 563. Isolates MI 561 and MI 562 were discussed in detail in Chapter 3 having undergone analysis of the cagA gene from these strains. As discussed in relation to this aspect of the present studies, isolates MI 561 and MI 562 were deemed to be clonally related strains whose oligofingerprints showed evidence of the occurrence of microevolution with minor differences in band profile present in fingerprints that were otherwise highly similar. The similarity coefficient calculated for these strains was 94% and was 100% between MI 561 and MI 563 (Fig. 6.3). RAPD-PCR of MI 561 and MI 562 (Fig. 6.1.B, lanes 8 and 9) also testifies to the clonal variation that exists between these two isolates and indeed, considering solely RAPD-PCR profiles without oligofingerprint data, these isolates might be perceived as being more distantly related. However, the nature of RAPD-PCR analysis means that a fingerprint is generated for an isolate based on where the arbitrary primers bind to the genome under the conditions used (Akopyanz et al., 1992a). It does not, therefore, examine the entire genome as is the case with oligofingerprint analysis. For this reason, the oligofingerprint evidence of high genetic similarity and the corroborating similarity coefficient of 94% calculated for oligofingerprint profiles testify that isolates MI 561 and MI 562 are clonal variants.

Molecular fingerprinting analysis of the seven isolates obtained from the members of Family 2 showed that mother and daughter were colonised by almost identical strains, MI 218A/MI 218C and MI 221C, respectively. This is further evidence of the likely parent-to-daughter transmission of *H. pylori*. Son and daughter were colonised by strains which only displayed 66% identity indicating that *H. pylori* was not transmitted between these two individuals. The dissimilarity of the *H. pylori* strains isolated from the daughter of Family 2 and her spouse shows that *H. pylori* was not transmitted between spouses in this instance as has been reported in other investigations (Georgopoulos *et al.*, 1996).

The availability of clonally related isolates obtained from different hosts, and the application of PCR-subtractive hybridisation methodology to the analysis of such strains allowed a direct

assessment of the evolution that has taken place in the genome of this clone upon its transmission to the stomach of a new individual. Extensive evidence from the analysis of familial strains indicates that *H. pylori* is clonal over short time periods (possibly up to five years or more) and more importantly, by extension, implies that the diversity that characterises this *Helicobacter* species takes place through the evolution in isolation in individual hosts of what were once clonal lineages (Suerbaum, 2000). Indeed, within the limits of isolate availability, the clonal variation that exists between isolates MI 577A/MI 577C and the isolates from the remaining siblings of Family 1 took place after the isolate was transmitted. However, it must be borne in mind that, if a greater number of colonies from primary isolation plates had been examined, each individual child may have been shown to be colonised with both types of subtypic variant strain. Although, the belief that evolution took place post-transmission might be strengthened by the fact that three children from this family were colonised by *H. pylori* isolates whose fingerprint profiles approached identity. It seems reasonable to imagine that if clonal variants of this strain had been present in another individual simultaneously that they would have been detected in at least one other child.

PCR-subtractive hybridisation carried out using isolates MI 431 and MI 577A from Family 1 resulted in the isolation of two genomic fragments that correspond to genes that, based on differences in hybridisation pattern on Southern blots (Fig. 6.4), appear to have undergone rearrangement between isolates MI 431 and MI 577A. DNA sequence analysis of these cloned fragments revealed that two adjacent ORFs from the sequenced *H. pylori* strains were spanned by the sequences of both fragments. As detailed in Table 6.1, the genes encoded at these loci were involved in housekeeping functions in the *H. pylori* cell. The *nadA* gene belongs to a group of loci involved in the biosynthesis of cofactors, prosthetic groups and carriers, while the *psd* gene is a member of a group of genes that function in fatty acid and phospholipid metabolism (Alm *et al.*, 1999). As the *holB*, *nadA* and *psd* genes all encode intracellular enzymes it is hard to rationalise an adaptive basis for the occurrence of rearrangement at these loci upon transmission of an *H. pylori* isolate from one host to another or, indeed, what possible selective impetus would be present in a new host to effect such possible genotypic alteration.

The corollary cannot be ignored either. The genetic events that brought about clonal variation between strains MI 577A/MI 577C and the other siblings' isolates may also have been due to rearrangement at a chromosomal hot spot. There are four copies of a short, 10-bp DNA

repeat of the consensus sequence AACGCTCTTT located within the psd coding sequence of clone 19. Furthermore, the sequences $A/G/C(T)^{3-8}$ occur at over 30 locations in total in the genes nadA and psd. It is not beyond the realm of possibility that rearrangement between these repeats at this local chromosomal site, or between these repeats and more widely dispersed copies of the sequences might be responsible for the differences in Southern blot hybridisation patterns observed with this probe. Interaction between small dispersed repeats leading to chromosomal duplications and deletions have been documented in $E.\ coli$ (Petes and Hill, 1988).

Although only two DNA fragments representing genetic differences between isolates MI 431 and MI 577A were isolated by subtractively hybridising these strains, it is significant in terms of the findings detailed in Chapters 3 and 4 that both these fragments should consist of DNA that encodes housekeeping functions. The involvement of Omp-encoding genes in rearrangement events occurring between paired isolates as found for isolates MI 458/MI 459 and MI 571/MI 575 and the occurrence of 3'-end repeat copy number alteration in the cagA gene between paired isolates are all rearrangement events highly likely to be selected for by virtue of the mounting of an immune response by the host to these surface-located molecules. In a stomach devoid of H. pylori infection until the arrival of a newly transmitted founder isolate, no antibodies, either circulating or secreted, to H. pylori proteins would be found. There would not, therefore, be any selective basis for the occurrence of rearrangement at alleles that encode molecules displayed on the bacterial surface in a host environment newly infected with H. pylori. Based on this minimal information about rearrangements in H. pylori clonal variants from different hosts, the likelihood that some of the rearrangements found between paired isolates from single individuals (as detailed in Chapter 4) became amplified by selection arising from the need to adapt phenotypically is strengthened further. This is further borne out by the finding that all of the isolates from the children of Family 1 have cagA 3'-ends of the same structure based on the identity of PCR product sizes (data not shown) and the identity of hybridisation profile when Southern blots of these strains were probed with a *cagA* gene probe.

There is much to be contributed to our knowledge of the evolution of *H. pylori* by further and more extensive analysis of the type described in the present study on the analysis of *H. pylori* clonal variants from different individuals. Although *H. pylori* is clonal over short time

periods relative to the usual life-long nature of the infection caused by this organism, evolution of the bacterium in isolation in separate stomachs with few, if any, bacterial competitors of other genera, leads to the *H. pylori* population structure that exists today, that of a freely recombining organism of marked genetic diversity with the presence of weak clonal lineages in specific geographical populations (Achtman *et al.*, 1999).

Comparative analysis of the sequenced genomes of *H. pylori* strains J99 and 26695 (Alm *et al.*, 1999; Alm and Trust, 2000) revealed that this species was not as diverse at the level of the proteome as DNA fingerprinting evidence would have predicted. A high level of synonymous substitution at the third base position of codons was detected that change the DNA sequence but not the sequence of the encoded protein. However, this does not account for evolution in *H. pylori* in its entirety and the almost complete disappearance over time of clonal relationships. A more complex scenario exists with chromosomal rearrangements, horizontal transfer of DNA between isolates and the involvement of IS elements all contributing to the phenomenon. The present studies provide preliminary information on the initial DNA rearrangements that have taken place between two *H. pylori* isolates, although the full nature of the rearrangement events have not yet been characterised.

Chapter 7

Analysis of the putative molecular mechanism for H. pylori LPS Lewis antigen phenotypic variation

7.1.	Introd	luction.	166				
7.2.	Results.						
	7.2.1.	Phenotypic analysis of H. pylori isolates for the expression of Lex					
		and Le ^Y by Western immunoblotting	170				
	7.2.2.	Oligofingerprinting analysis of paired isolates MI 561/MI 562					
		and MI 222A/MI 222C to confirm that the phenotypic diversity					
		observed has taken place between genetically related isolates	171				
	7.2. 3 .	DNA sequence analysis of a region of the H. pylori fucT2 gene	171				
	7.2.4.	Second round of sequence analysis in order to confirm the data					
		obtained in the primary analysis.	173				
	7.2.5.	Additional round of PCR amplification and sequencing of a					
		region of the fucT2 gene from isolate MI 562.	174				
7.3.	Discu	ssion.	175				

7.1. Introduction.

When it was recognised that sera from H. pylori-infected patients or from mice immunised with the bacterium contained autoantibodies that reacted with the gastric mucosa of humans and mice, it was suspected that H. pylori induced these autoantibodies and that they played an important role in the pathogenesis of gastritis and gastric atrophy (Negrini et al., 1991; Negrini et al., 1993). The bacterial antigens that induced host autoimmunity were unknown at that time but candidate antigens did present themselves. Structural analysis of H. pylori lipopolysaccharide (LPS) molecules revealed that the O-antigen region of the molecule has unusual properties. The LPS present in the outer membrane of H. pylori strain NCTC11637 was determined structurally to consist of extended N-acetyl-lactosamine (LacNAc) units in its O-antigen region that terminated in di- or trimeric Lewis^x (Le^x) determinants (Aspinall et al., 1996). Both LacNAc and Le^x mimic human cell-surface glycoconjugates present on various human cell types. The chemical structures of the LPS O-antigen region of other H. pylori strains (P466 and MO19) have also been determined (Aspinall and Monteiro, 1996) and were found to have internal Le^x units but terminated in Lewis (Le^x) (strain P466) and, in the case of strain MO19 to consist of a single Le^Y epitope as the terminal sugar residue in the Oantigen chain, linked to an inner core region that was identical to that present in strains NCTC11637 and P466. More recently, strains 26695 and J99 as well as the *H. pylori* Sydney strain and strain UA915 have had the molecular makeup of their LPS molecules defined (Monteiro et al., 2000) and were found to express sially Le^x and also Lewis^b (Le^b) in the case of strain UA915. The latter antigen is widely expressed by cells of the human gastric mucosa (Borén et al., 1993). H. pylori LPS has also been found to express H-type I blood group structures (Appelmelk et al., 1996).

The available structural evidence for *H. pylori* LPS molecules which delineates their expression of human blood group antigens (Le^x, Le^y, Le^b, sialyl Le^x and H type I) strongly implicated these molecules in *H. pylori*-linked autoimmunity. Appelmelk and coworkers (1996) demonstrated the potential part these antigens played in autoimmune responses by showing that anti-Le^x, anti-Le^y and anti-H type I antibodies bound to a panel of human and murine cell types including gastric glandular tissue, adenocarcinoma cells and MALT cells and further, that growth in mice of a hybridoma that secretes anti-Le^y mAbs resulted in histopathological evidence of gastritis, thereby indicating a direct role in pathogenesis for these antibodies. Indeed, in a study carried out by Wirth *et al.* (1996), it was determined that 89% of the *H. pylori* isolates examined expressed Le determinants on their LPS molecules

and some isolates were found to express both Le^x and Le^y on their LPS O-antigen side-chains. Taken together, the data suggest that the apparent restricted diversity in O-antigen expression of *H. pylori* points to a specific role for Le epitopes in pathogenesis (Appelmelk *et al.*, 2000).

Studies concerning the expression by H. pylori LPS of Le antigens have revealed that the Le determinants might have myriad functions in the pathogenesis of H. pylori disease (Appelmelk et al., 2000). Microbial expression of host antigens, as discussed above in relation to H. pylori, is speculated to be a mechanism whereby the bacterium escapes elimination by antigen-specific host defences. Further, it was believed by some that correlations that were found between host Lewis phenotypes and that expressed by colonising H. pylori strains represented the basis for the escape from the host response by bacterial Lewis antigen mimicry. However, although compelling theoretically, the argument for the evasion of host responses by the expression on H. pylori LPS of Le antigens has little substantive evidence. Similarly, the likelihood that H. pylori Le antigens are 'pathogenic', because they induce autoantibody production and contribute to gastric pathology through Tcell induction and concomitant inflammatory mediator release resulting in visible histopathology, is now also considered tenuous given that no proof exists to show that the antibodies bind to the gastric mucosa. Furthermore, these autoantibodies directed to gastric parietal cells cannot be absorbed with H. pylori whole cells or whole cell lysates (Appelmelk et al., 2000). Significantly, a role for Le^x in mediating adhesion of H. pylori to the gastric epithelium has been demonstrated and suggests that the molecule might function in the colonisation of stomach tissue by the bacterium (Edwards et al., 2000).

Similarly to the carbohydrate surface structures of *Haemophilus influenzae* and *Neisseria* spp. (High *et al.*, 1993; Roche and Moxon, 1995; Saunders *et al.*, 2000), phenotypic variation of *H. pylori* LPS molecules has also been observed (Appelmelk *et al.*, 1998; Gibson *et al.*, 1998; Appelmelk *et al.*, 1999; Wang *et al.*, 1999a; Wirth *et al.*, 1999). Initial observations of spontaneous high-frequency phase variability of *H. pylori* LPS were made following *in vitro* UV irradiation of cells, with spontaneous variation observed at lower frequency for strain NCTC11637 and two clinical isolates which had not been irradiated (Appelmelk *et al.*, 1998). These observations were later extended, with demonstrations of Le expression variation among freshly isolated *H. pylori* populations from single biopsy samples or single hosts (Wirth *et al.*, 1999). Further, RAPD-PCR fingerprinting of these isolates showed that they were indistinguishable at the genetic level and that phenotypic or phase variation in Le

antigen expression suggested microdiversity within a single population as opposed to mixed strain *H. pylori* infection.

Phase variability of surface antigens in bacteria has important implications for bacterium-host interactions, although in many cases the precise significance of events that generate phenotypic diversity remain unclear. The possibilities are raised that phase variation provides for the adaptation of the bacterium to more than one environment and, indeed, that reversible phase-variable switching could generate mixed populations in which a subset of cells exist, already poised to survive in a new environment because they express a particular phenotype (Henderson *et al.*, 1999). Additionally, immune evasion and transmission of bacteria to other hosts are thought likely to be facilitated by phase variation of surface molecules, the virulence of the organism being modulated in this manner (Henderson *et al.*, 1999). In relation to *H. pylori*, despite some disparate opinion about why this organism expresses Lewis antigens (Appelmelk *et al.*, 2000; Falk, 2001), the fact that they can be phase variable is thought likely to be useful to the organism by conferring on it a selective advantage to survive and persist during the changes that occur in the host microenvironment during infection (Wang *et al.*, 2000).

The molecular mechanisms by which phenotypic diversity in Le antigens is generated in H. pylori LPS have been investigated closely (Appelmelk et al., 1999; Wang et al., 1999, 2000). Indeed, prior to the availability of experimental evidence, it was speculated that H. pylori LPS might be phase variable by virtue of the presence of homopolymeric tracts in genes which encoded glycosyltransferases involved in the biosynthesis of LPS. Based on genome sequence scrutiny, the phase-variable elements of H. pylori were listed as being the proteins involved in LPS biosynthesis, DNA R-M systems and cell-surface-associated proteins (Tomb et al., 1997). The DNA sequences of the diverse group of genes that encode these proteins all contain simple oligonucleotide repeat regions which by their nature are hypermutable. The nucleotide repeats which comprise homopolymeric tracts are believed to be involved in slipped-strand mispairing as a mechanism of phase variation leading to translational frameshifting (Wang et al., 2000). Glycosyltransferases function in LPS biosynthesis by transferring polysaccharide residues to core sugar residues in the backbone of the molecule (Martin et al., 1997; Ge et al., 1997; Wang et al., 1999a). Aberrant functioning of these transferases arising from replication or translation-directed mutation at these hypermutable chromosomal hot spots would, thus, cause the phenotype expressed by the LPS to be altered.

The present study describes an investigation into the molecular mechanism for LPS phase variability detected in two sets of clonally related *H. pylori* isolates. Although a small study, the data described raise some interesting questions, largely arising out of the anomalous nature of the findings. Questions for future research are raised in light of the results reported here.

7.2. Results.

7.2.1. Phenotypic analysis of H. pylori isolates for the expression of Le^x and Le^y by Western immunoblotting.

Total cell lysates of paired and single *H. pylori* isolates were treated with proteinase K. The remaining polysaccharide material was electrophoresed by SDS-PAGE and transferred to nitrocellulose membrane followed by probing with a monoclonal antibody (mAb) to the Le^x antigen. After detection, bound Le^x mAb was removed by stripping the membrane which was then re-probed with anti-Le^x mAb. A representation of the Western blots obtained from these analyses are shown in Fig. 7.1.A and B.

In all, two sets of paired isolates (MI 222A/MI 222C and MI 561/MI 562) were identified where each isolate expressed different Le phenotypes. Interestingly, both antral isolates (MI 222A and MI 561) expressed Le^x, while Le^y was expressed on the isolates obtained from the corpus of the stomach (MI 222C and MI 562). In order to confirm these data, expanded single colony populations of each strain were cultured and analysed in a similar fashion by Western blotting (Fig. 7.1.C–F). These blots (Fig. 7.1.C–F) confirm the phenotypic diversity in Le antigen expression that exists between both these sets of paired isolates, although the numbers of colonies examined in the form of expanded populations represent only a very small sample of the population of each strain. Thus, the likelihood that these strains expressed the given Le phenotypes *in vivo* is suggested.

Furthermore, isolates MI 222A and MI 222C were analysed after being freshly isolated from biopsy tissue and had undergone a minimum of subculturing. Isolates MI 561 and MI 562 had been cultured on laboratory medium more frequently prior to analysis by Western blotting. There are, however, literature reports that testify to the stability of Le antigens expressed by *H. pylori* when the isolates are subject to *in vitro* passage (Wirth *et al.*, 1999). Taken together, the divergence in the Le phenotype expressed between the isolates in the paired sets MI 222A/MI 222C and MI 561/MI 562 is thought likely to represent a reasonably good reflection of the Le phenotype expressed by these strains in their *in vivo* environment in as much as can be determined using these analytical approaches. Nonetheless, this is highly speculative and the possibility cannot be ignored that the observed phenotypic diversity arose during *in vitro* culture and storage of these strains. Additionally, the biopsy tissue from which each of these strains was isolated could feasibly have contained *H. pylori* of both phenotypes

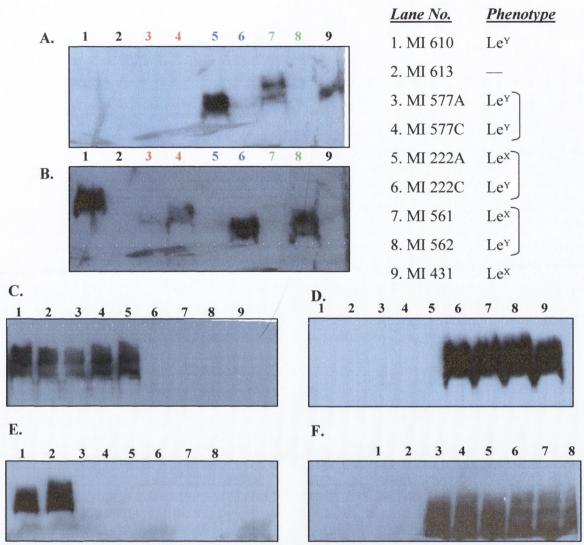


Fig. 7.1. Western immunoblot analysis of *H. pylori* isolates with anti-Le^x and anti-Le^y monoclonal antibodies.

(A) and (B), Western immunblot of the strains indicated to the right of this panel which correspond to the lane numbers shown above both blot A and blot B. Isolates from matched sets are indicated in red, blue and green coloured font. Blot A was probed with anti-Le^x mAbs. Blot B was probed with anti-Le^x mAbs. The deduced phenotypes of these strains are also shown to the right of panels A and B. (C) and (D), Anti-Le^x and anti-Le^x antibodies, respectively, were used to probe Western immunoblots of populations of *H. pylori* cells expanded from single colonies of strains MI 561 and MI 562. Cell lysates of five expanded colonies of strain MI 561 were electrophoresed in lanes 1–5 on blot C with four expanded colonies of strain MI 562 run in lanes 6–9 on this blot. The same sets of lysates were electrophoresed in blot D and probed with anti-Le^x mAbs. (E) and (F), As for blots C and D, lysates of expanded single colonies of strains MI 222A and MI 222C were probed with anti-Le^x (blot E) and anti-Le^x antibodies (blot F). Lanes in both E and F were loaded as follows: 1–2, two expanded single colonies of strain MI 222A; lanes 3–8, six expanded single colonies of strain MI 222C.

although in order to address this, the analysis of multiple isolates from the same specimen would be required.

7.2.2. Oligofingerprinting analysis of paired isolates MI 561/MI 562 and MI 222A/MI 222C to confirm that the phenotypic diversity observed has taken place between genetically related isolates.

Although the strains in each of isolate pairs MI 222A/MI 222C and MI 561/MI 562 were taken from the stomachs of the same individuals, it was necessary to confirm by discriminatory molecular fingerprinting analysis that they were genotypically identical or highly similar. The differential expression of Le antigens between proven clonally related *H. pylori* strains affords the opportunity to investigate the likely mechanism by which this phenotypic diversity arises *in vivo*. To this end, genomic DNA was prepared from *H. pylori* isolates MI 561/MI 562, MI 222A/MI 222C and restricted with *Hin*dIII. In an identical manner to the oligofingerprinting analysis described in Chapters 3, 4 and 6, the Le antigen phase-variable strains were probed with DIG-labelled oligonucleotide (GTG), for isolates MI 561 and MI 562, and (GACA)₄ for isolates MI 222A and MI 222C, and their genetic relatedness assessed. The fingerprint profiles obtained for these strains are shown in Fig 7.2 and reveal the high genotypic identity that exists between the paired strains, although band differences are apparent between them (indicated with yellow arrows).

7.2.3. DNA sequence analysis of a region of the H. pylori fucT2 gene.

The on-and-off or phase-variable switching of bacterial surface structures is usually accompanied by a parallel genetic switch in the genes that code for these molecules (Appelmelk *et al.*, 1998). The enzyme responsible for the biosynthesis of the Le^Y epitope from the Le^X antigen, by the addition of an $\alpha(1,2)$ -linked fucose residue to the existing Le^X structure, is the α 1,2-fucosyltransferase and is encoded by the gene designated *fucT2* (Wang *et al.*, 1999a) (see Fig. 7.3 over).

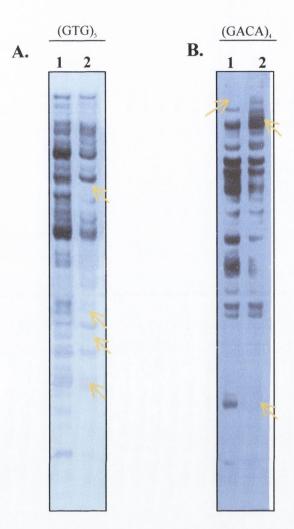


Fig. 7.2. Analysis of *H. pylori* isolate pairs MI 561/MI 562 and MI 222A/MI 222C by oligofingerprinting.

- (A) HindIII-restricted genomic DNA from isolates MI 561 and MI 562 (lanes 1 and 2, respectively) was probed with the DIG-labelled oligonucleotide (GTG)₅. The locations of fingerprint band differences between these two strains are indicated by yellow arrows.
- (B) HindIII-restricted genomic DNA from isolates MI 222A and MI 222C (lanes 1 and 2, respectively) was probed with the DIG-labelled oligonucleotide (GACA)₄ resulting in the fingerprint profile for these strains shown above. Positions where there are obvious differences in band pattern between the strains are highlighted with yellow arrows.

In both (A) and (B), the hybridising bands in the oligofingerprint profile range in size from 12 kbp to approximately 500 bp. Note, oligofingerprints of these same paired strains were presented in Chapter 6 in the course of oligofingerprint analysis of familial *H. pylori* isolates. Oligofingerprints of isolates MI 561 and MI 562 were also displayed in Chapter 3 in the analysis of the 3'-end of the *cagA* gene between these strains. The oligofingerprints presented here are, however, from repeat experiments.

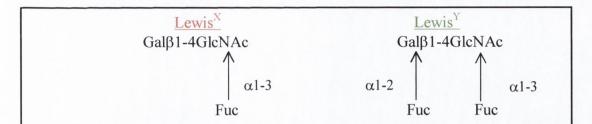


Fig. 7.3. Schematic structures of the Lewis X and Lewis Y antigens. Abbreviations Gal, D-galactose; Fuc, L-fucose; GlcNAc, N-acetyl-D-glucosamine.

Therefore, it seemed probable that phase-variable switching between expression of Le^x and Le^y could be mediated by an aberrant $\alpha 1,2$ -fucosyltransferase activity. The DNA sequence of the *fucT2* gene contains numerous simple oligonucleotide repeat regions which by their nature are hypermutable. Furthermore, a region of imperfect repeats of the nucleotides TAA designated the 'A-rich region' has also been identified and is postulated to be a hot spot for mutation (Wang *et al.*, 1999a). Thus, the sequence information strongly suggested that there was potential within the DNA sequence of the *fucT2* gene for the occurrence of a number of mutagenic events that might give rise to the aberrant functioning of this gene and hence affect the Lewis phenotype being expressed in a particular isolate.

In order to investigate if this possibility might be true in the paired *H. pylori* isolates identified by Western immunoblotting (Section 7.2.1), a region of the *fucT2* gene was amplified by PCR from each of isolates MI 561, MI 562, MI 222A and MI 222C. The primers (fucT7 and 1,2fucR) amplified a region of the *fucT2* gene of approximately 980 bp that encompassed the promoter region and almost the entire coding region of the gene. The resultant PCR products were cloned into a T/A vector and sequenced. Approximately 641 nucleotides of DNA sequence was determined for each of the PCR products. Pair-wise alignments of the sequences obtained for each set of paired isolates using these primers are shown in Fig. 7.4.A and Fig. 7.5.A.

For isolate pair MI 222A and MI 222C, the DNA sequence obtained for the *fucT2* gene is identical save for the divergence of one nucleotide in the polyC tract (Fig. 7.4.A). Isolate MI 222C has an additional C base in this region (13 C nucleotides in total) requiring that a gap be created in the corresponding sequence of MI 222A. The result of translating this DNA sequence information gives a protein sequence that is at variance with that expected based on

222A	GAACACTCACACGCGTCTTTT
222C	GAACACTCACACGCGTCTTTT
222A 222C	TCAAATAAAAATTCAAATGATTTGAAAGCGATGCTTGCCCCGCTTTTTGGGCTTTTATT TCAAATAAAAAATTCAAATGATTTGAAAGCGATGCTTGCCCCGCTTTTTTGGGCTTTTATT *******************************
222A 222C	GAAAAAGGGCTTTAAAATGAGCTAAAATGAGCGTTTCATTTGAAAAATAAAGGGATTGAA GAAAAAGGGCTTTAAAATGAGCTAAAATGAGCGTTTCATTTGAAAAATAAAGGGATTGAA ****************
222A 222C	TGGCTTTTAAGGTGGTGCAAATTTGCGGAGGGCTTGGGAATCAAATGTTTCAATACGCTT TGGCTTTTAAGGTGGTGCAAATTTGCGGAGGGCTTGGGAATCAAATGTTTCAATACGCTT **********************************
222A 222C	TCGCTAAAAGTTTGCAAAAACACTCTAATACGCCTGTGCTGTTAGATATCACTTCTTTTG TCGCTAAAAGTTTGCAAAAACACTCTAATACGCCTGTGCTGTTAGATATCACTTCTTTTG ******************************
222A 222C	ATTGGAGTAATAGGAAAATGCAATTAGAACTTTTCCCTATTGATTTGCCCTATGCGAGCG ATTGGAGTAATAGGAAAATGCAATTAGAACTTTTCCCTATTGATTTGCCCTATGCGAGCG ********************************
222A 222C	CTAAAGAAATCGCTATAGCTAAAATGCAACACCTCCCCAAGCTAGTAAGAGATGCACTCA CTAAAGAAATCGCTATAGCTAAAATGCAACACCTCCCCAAGCTAGTAAGAGATGCACTCA *********************************
222A 222C	AATGCATGGGATTTGATAGGGTGAGTCAAGAAATCGTTTTTGAATACGAGCCTAAATTGC AATGCATGGGATTTGATAGGGTGAGTCAAGAAATCGTTTTTGAATACGAGCCTAAATTGC ***********************************
222A 222C	TAAAGCCAAGCCGCTTGACTTATTTTTATGGCTATTTCCAAGATCCACGATACTTTGATG TAAAGCCAAGCC
222A 222C	CTATATCCCCTTTAATCAAGCAAACCTTCACTCTACCCCCCCC
222A 222C	AATAATAAT <mark>AAAAAAG</mark> AGGAAGAATACCACCGCAAGCTTTCTTTGATTTTAGCCGCTAAA AATAATAAT <mark>AAAAAAG</mark> AGGAAG <u>AATACCA</u> CCGCAAGCTTTCTT <u>TGATTTT</u> AGCCGCTAAA **********************************
222A 222C	AACAGCGTGTTTGTGCATATA

Fig. 7.4. A. Alignment of the DNA sequences determined for a region of the *fucT2* gene from *H. pylori* isolates MI 222A and MI 222C.

A region of the *fucT2* gene from isolates MI 222A and MI 222C was amplified using the primers fucT7 and 1,2fucR from which the above DNA sequences were determined. The sequences were aligned using the ClustalW alignment tool as shown above. The start codon (ATG) of the gene is highlighted in red and the region of the polyC tract is shown in blue. The additional C nucleotide present in strain MI 222C in this tract would cause a stop codon to be read in the translated sequence of this partial gene sequence immediately downstream of the polyC tract (depicted in pink, underlined font). The putative translation frameshift motifs (TFM) are highlighted in pale blue, with the Shine-Dalgarno-like sequence located upstream of the second TFM shown in bright green font and the sequences involved in the formation of the putative downstream stem-loop structure shown underlined and in green, with the nucleotides that would be read as a stop codon if this heptanucleotide sequence were in frame shown in bold face. As detailed in text, this DNA sequence information does not correlate with the phenotypes observed for these strains by Western immunoblotting.

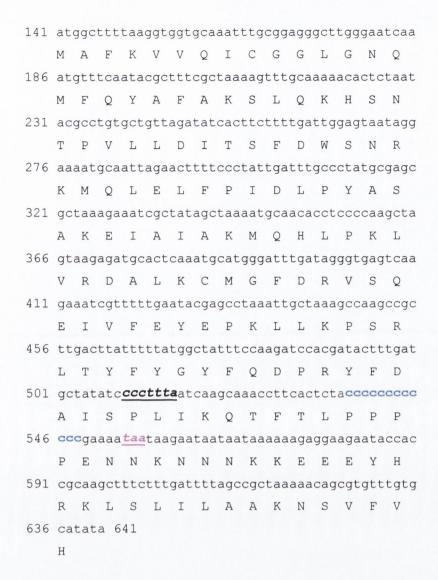


Fig. 7.4. B. Amino acid sequence predicted from the translation of the *fucT2* DNA sequence determined for isolate MI 222A.

The putative sequence of amino acids in the translated DNA sequence of the *fucT2* gene from isolate MI 222A is shown above in the one letter amino acid code (uppercase font) below the DNA sequence of the region which is shown in lower-case font. The region of the polyC tract is highlighted in blue and the putative translation frameshift motif (C CCT TTA) upstream of this that is in-frame with the initiation codon is shown bold, italicised and underlined. The sequence downstream of the polyC tract that would have been read as a stop codon (TAA) in isolate MI 222C because of the additional C nucleotide present in the homopolymeric region in this isolate is highlighted in pink. Translational read-through takes place, therefore, in isolate MI 222A as depicted above.

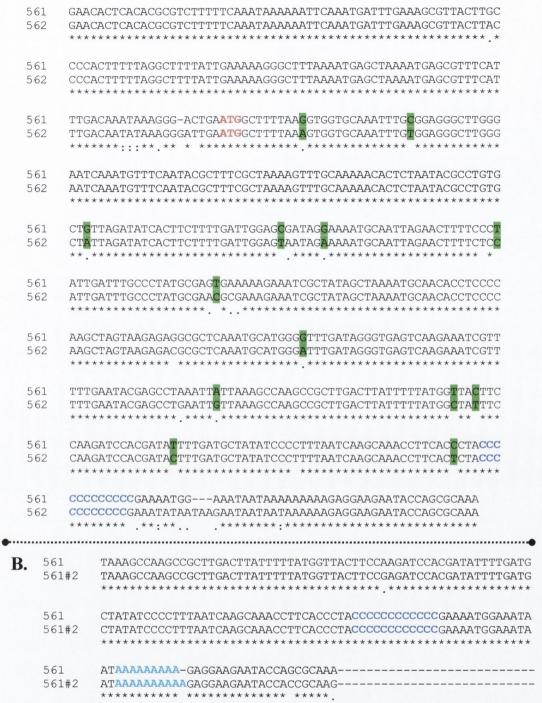


Fig. 7.5. (A) Alignment of the partial *fucT2* gene sequences determined for isolates MI 561 and MI 562 and (B) alignment of a short region of the *fucT2* gene of isolate MI 561 determined from two rounds of sequencing.

- (A) When the partial sequences of the *fucT2* gene from isolates MI 561 and MI 562 were aligned, the polymorphisms that were present between the strains were not confined to a distinct region of the locus but were present throughout the gene. The sites of synonymous nucleotide substitutions between the strains are highlighted in green. The start codon (ATG) is shown in red font and the region of the polyC tract is shown in blue font.
- **(B)** When a second round of sequencing was carried out for isolate MI 561, the sequences obtained were identical apart from the mismatches apparent in the above alignment and the presence of a heterogeneous number of A nucleotides in the 'Arich region' (highlighted in pale blue, bold font) downstream of the polyC tract.

knowledge of the Le phenotype of these strains (Table 7.1). As isolate MI 222C expresses Le^{Y} , it was anticipated that the *fucT2* gene in this strain would produce an active $\alpha 1,2$ -fucosyltransferase enzyme. However, the presence of the additional C nucleotide in the homopolymeric tract results in the premature reading of a stop codon (TAA) in the region immediately downstream of the polyC nucleotides. In isolate MI 222A where there are 12 C nucleotides in the polyC tract, the downstream region is in frame and it is assumed that a full length $\alpha 1,2$ -fucosyltransferase enzyme is translated (Fig. 7.4.B). As isolate MI 222A expresses the Le^{X} antigen, it was expected that a truncated enzyme might be translated in this strain, rendering it impossible for Le^{Y} to be expressed on LPS molecules.

As is evident from the alignment of the sequences obtained for strains MI 561 and MI 562 (Fig. 7.5.A), there is considerable divergence of sequence in the *fucT2* gene between these two isolates. This divergence exists throughout the sequenced region of the gene and is not only confined to the homopolymeric tract and downstream 'A-rich region', although both of these constitute areas of high divergence. While most of the nucleotide substitutions in the remainder of the gene occur at synonymous sites, non-synonymous substitutions are evident also. Furthermore, the promoter region of strain MI 561 features a nucleotide position where a gap was created in the alignment. As with strain MI 222C, translation of the *fucT2* DNA sequence obtained for strain MI 562 resulted in premature termination of translation some 55 nucleotides after the polyC tract (not shown in Fig. 7.5.A). As Western immunoblotting determined a Le^Y phenotype for this isolate, this again was surprising given that the phenotypic information predicted the translation of a functional enzyme. Isolate MI 561 which displayed an Le^X antigen also displayed DNA sequence information that could not be correlated with the deduced phenotype of the isolate.

7.2.4. Second round of sequence analysis in order to confirm the data obtained in the primary analysis.

As the findings of the sequencing analysis of the *fucT2* gene for isolate pairs MI 222A/MI 222C and MI 561/MI 562 were not what was anticipated, a second round of sequencing was carried out using freshly cloned PCR products in order to ascertain if the results of the primary investigation were reproducible. As before, PCR was employed to amplify a region of the gene. The same forward primer (fucT7) was used in this instance with a new reverse primer (fucT6) that binds to the *fucT2* gene downstream of the 'A-rich region'. The data

Table 7.1. Summary of the phenotypes of the two *H. pylori* paired isolate sets which showed phenotypic variation, also giving the hypothetical status of the *fucT2* gene and the status of the gene as determined by the initial DNA sequencing study.

H. pylori	Lewis	Hypothetical	Status of the fucT2 gene from primary
strain	phenotype	status of fucT2	DNA sequence and deduced amino
	expressed	gene	acid sequence analysis
MI 222A	Le ^x	OFF	ON
MI 222C	Le^{Y}	ON	OFF
MI 561	Le ^x	OFF	ON
MI 562	Le ^Y	ON	OFF

gleaned from this additional sequence information heightened the complexity of the issues raised by the phenotypic diversity observed between the two paired isolate sets and, as discussed in the preceding section, the genetic data that did not corroborate the phenotypes determined. Second round sequencing revealed slightly different sequences again for some of the isolates to those which had been determined in the first analysis. The sequences obtained for strains MI 222A and MI 222C were now identical, each strain having 12 C residues in the polyC tract (MI 222A has the same sequence as before, therefore). For strain MI 562 this second round of sequencing resulted in a divergent DNA sequence to that obtained when the gene was initially sequenced as was also found for isolate MI 561. Interestingly, between the two rounds of sequencing, strain MI 561 displayed a small number of nucleotide mismatches, had identical numbers of nucleotides in the polyC tract but had different numbers of A nucleotides in a homopolymeric stretch in the 'A-rich region' downstream of the polyC tract as highlighted in Fig. 7.5.B.

7.2.5. Additional round of PCR amplification and sequencing of a region of the fucT2 gene from isolate MI 562.

A third round of PCR and sequencing was carried out for the *fucT2* gene from isolate MI 562 in an attempt to reproduce either of the two existing sequences that had been determined for this gene, because both exhibited divergence that was greater than would have been expected. Only minor nucleotide changes were expected in the *fucT2* gene within a strain at different sequencing occasions as was found for isolates MI 222C and MI 561. When this third sequence obtained for strain MI 562 was aligned with the previous sequences that had been determined, it was clear that no two of these sequences were identical to each other (Fig. 7.5.C).

562	GAACACTCACACGCGTCTTT
562#2	GAACACTCACACGCGTCTTT
562#3	GAACACTCACACGCGTCTTT
002110	*********
562	TTCAAATAAAAATTCAAATGATTTGAAAGCGTTACTTACCCCACTTTTTAGGCTTTTAT
562#2	TTCAAATAAAAATTCAAATGATTTAAAG-CGATGCTTGCCCCGTTTTTTTGGGCTTTTAT
562#3	TTCAAATAAAAATTCAAATGATTTGAAAGCGTTACTTGCCCCACTTTTTAGGCTTTTAT

562	TGAAAAAGGGCTTTAAAATGAGCTAAAATGAGCGTTTCATTTGACAATATAAAGGGATTG
562#2	TGAAAAAGGGCTTTAAAATGGGCTAAAATTAGGGTTTTATTTGACAATATAAAGGGATTG
562#3	TGAAAAAGGGCTTTAAAATGAGCTAAAATGAGCGTTTCATTTGACAAATAAAGGG-ATTG

562	$\mathbf{A}^{\mathbf{ATG}}_{\mathbf{G}}\mathbf{GCTTTTAAAGTGGTGCAAATTTGTGGAGGGCTTGGGAATCAAATGTTTCAATACGC}$
562#2	AATGCTTTTAAGGTGGTGCAAATTTGCGGAGGGCTTGGGAATCAAATGTTTCAATACGC
562#3	AATGCTTT-AAGGTGGTGCAAATTTGCGGAGGGCTTGGGAATCAAATGTTTCAATACGC
	******* ** .******** **************
562	TTTCGCTAAAAGTTTG-CAAAAACACTCTAATACGCCTGTGCTATTAGATATCACTTCTT
562#2	TTTCGCTAAAAGTTTG-CAAAAACACTCTAATACGCCTGTGCTGTTAGATATCACTTCTT
562#3	TTTCGCTAAAAGTTTGCAAAAAACACTCTAATACGCCTGTGCTGTTAGATATCACTTCTT
302#3	************
	•
562	TTGATTGGAGTAATAGAAAATGCAATTAGAACTTTTCTCCATTGATTTGCCCTATGCGA
562#2	TTGATTGGAGCGATAGGAAAATGCAATTAGAACTTTTCCCTATTGATTTGCCCTATGCGA
562#3	TTGATTGGAGCGATAGGAAAATGCAATTAGAACTTTTCCCTATTGATTTGCCCTATGCGA

562	ACGCGAAAGAAATCGCTATAGCTAAAATGCAACACCTCCCCAAGCTAGTAAGAGACGCGC
562#2	GCGCGAAAGAAATCGCTATAGCTAAAATGCAACACCTCCCCAAGCTAGTAAGAGAGGCGC
562#3	GTGAAAAAGAAATCGCTATAGCTAAAATGCAACACCTCCCCAAGCTAGTAAGAGAGGCGC
302113	* . * * * * * * * * * * * * * * * * * *
562	TCAAATGCATGGGATTTGATAGGGTGAGTCAAGAAATCGTTTTTGAATACGAGCCTGAAT
562#2	TCAAATGCATGGGGTTTGATAGGGTGAGTCAAGAAATCGTTTTTGAATACGAGCCTAAAT
562#3	TCAAATGCATGGGGTTTGATAGGGTGAGTCAAGAAATCGTTTTTGAATACGAGCCTAAAT

562	TGTTAAAGCCAAGCCGCTTGACTTATTTTTATGG-CTATTTCCAAGATCCACGATACTTT
562#2	TGTTAAAGCCAAGCCGCTTGACTTAATTTTTATGGCTATTTTCAAGATCCACGATACTTT
562#3	TATTAAAGCCAAGCCGCTTGACTTATTTTTATGG-TTACTTCCAAGATCCACAATATTTT
	*.****************************
562	GATGCTATATCCCTTTTAATCAAGCAAACCTTCACTCTACCCCCCCC
562#2	GATGCTATATCCCTTTTAATCAAGCAAACCTTCACTCTACCCCCCCC
562#3	GATACTATATCCCCTTTAATCAAGCAAACCTTCACCCTACCCCCCCC
J 02#J	***.****** ************* ************ *:***
5.60	
562	AAGAATAATAAAAAAAGAGGAAGAATACCAGCGCAA
562#2	AAGAATAATAAAAAAAGAGGAAGAATACCGCT
562#3	GGAAATAATAAAAAAGAAGAGGAAGAATACCACCGCAA
	******* *** ********

562

Fig. 7.5.C. Alignment of the DNA sequences obtained for the *fucT2* gene from isolate MI 562 determined at three separate sequencing rounds after PCR on three distinct occasions.

A region of the *fucT2* gene was amplified from strain MI 562 on three distinct occasions and the sequence of the PCR products were determined. The aligned sequences are shown above with sites of nucleotide mismatches and positions where gaps were created in the alignment indicated by the absence of the asterix symbol beneath the sequences. The ATG start codon is highlighted in red bold font, the polyC tract is highlighted in blue font and the 'A-rich region' is shown in red.

7.3. Discussion.

The phenotypic variability of surface molecules is a theme that features widely among many genera of bacteria but which is exemplified in large part in Gram-negative organisms (Henderson et al., 1999). The characteristic spontaneous, high frequency, reversible ON-OFF switching exhibited by phase-variable systems serve myriad potential roles in mediating bacterium-host interactions. The heterogeneous population created by the processes of phenotypic variation are generally considered to function as a means of escaping the host immune response and as a mechanism of microenvironmental adaptation (Robertson and Meyer, 1992). Indeed, the success of certain pathogens relies heavily on their ability to undergo phenotypic variation and generally as mentioned above, in pathogenic prokaryotes such variation involves surface components. Although considered to be 'random' processes, the occurrence of phenotypic (phase) variations can also be viewed as programmed events because genome sequences exist that increase the likelihood that certain events will take place (Henderson et al., 1999). Complete genome sequencing of H. pylori brought to light that the chromosome of the organism contains many genes (thirty were identified in strain 26695) which contain homopolymeric tracts or dinucleotide repeats, whose presence predicts the phase variability of the ORFs they are associated with. This information led to descriptions of phase-variable systems in H. pylori as being associated with genes that encoded cell-surface associated proteins, DNA R-M systems and enzymes that are involved in LPS biosynthesis (Tomb et al., 1997; Saunders et al., 1998).

That *H. pylori* LPS is phase variable with respect to the expression of Le antigens is now a well-documented phenomenon and one which has also been elucidated mechanistically. The random shortening and lengthening of homopolymeric tracts located within the glycosyltransferase genes responsible for LPS biosynthesis cause reversible frameshifting within the coding region of the gene and can result in gene inactivation. Genetic and serological evidence to support this has been accumulated and reversible frameshifting is accepted as the mechanism whereby LPS phase variation is achieved (Appelmelk *et al.*, 1999; Wang *et al.*, 1999a). In the present study, two sets of clonally related *H. pylori* isolates (MI 222A/MI 222C and MI 561/MI 562) were identified which differed within each pair in the Le phenotype that was expressed by each isolate. Both antral isolates (MI 222A and MI 561) expressed LeX, whereas LeY was expressed on the isolates obtained from the corpus of the stomach (MI 222C and MI 562). This was initially perceived as an indication that different gastric regions may be colonised by phenotypically variant *H. pylori* populations with respect

to Lewis antigen expression. However, more in-depth investigation negated this possibility (data not shown). It was of interest to attempt to define the mechanism by which this phenotypic variation had arisen in these clonally related isolates. This seemed to offer a novel approach in illuminating how the phase variability of the *H. pylori* LPS molecule is brought about, i.e., the comparison of phenotypically variable yet genetically highly similar clinical isolates obtained from the stomachs of single individuals. The literature reports studies on phenotypic variants derived from laboratory strains (Appelmelk *et al.*, 1999; Wang *et al.*, 1999a) and also on fucosyltransferase genes from different clinical strains with distinct Le phenotypes but not, as was carried out in this instance, on two pairs of clonally related clinical isolates which displayed LPS phenotypic variation.

The cloning of the gene that specified an $\alpha(1,3)$ -fucosyltransferase activity had been reported independently by two groups (Ge *et al.*, 1997; Martin *et al.*, 1997). Indeed, the defining step of Le^x biosynthesis is catalysed by this enzyme by transferring GDP-fucose to core Gal β 1-4GlcNAc structures. In the genome sequence of strain 26695 two ORFs were identified that putatively specified an $\alpha(1,3)$ -fucosyltransferase enzyme, $\alpha(1,3)$ -fucT, but no $\alpha(1,2)$ -fucT gene was identified. The final step of Le^x biosynthesis requires the action of a fucosyltransferase enzyme with $\alpha(1,2)$ -linkage specificity that transfers a fucose sugar residue to the terminal galactose on the existing Le^x moiety (Fig. 7.3). Thus, $\alpha(1,3)$ - and $\alpha(1,2)$ -fucosyltransferases are required for Le^x biosynthesis. The ORFs designated HP0093 and HP0094 were predicted to be a putative $\alpha(1,2)$ -fucT gene by Berg *et al.* (1997), although they were not originally annotated as such (Tomb *et al.*, 1997). The gene, designated variously fucT2 or futT2 that encodes an $\alpha(1,2)$ -fucosyltransferase was cloned and subjected to mutational and biochemical analysis by Wang *et al.* (1999a) who confirmed a role for this enzyme in the biosynthesis of Le^x and the phase-variable expression of this antigen.

To ascertain if variability existed in the sequence of the *fucT2* gene between isolates MI 561 and MI 562, and between MI 222A and MI 222C, a region of the *fucT2* gene was amplified by PCR, the DNA sequence of which was then determined. Within each of these paired isolate sets, phenotypic variability between expression of Le^X and Le^Y had taken place, strongly implicating the aberrant functioning or expression of the *fucT2* gene product in mediating the change in Le antigen displayed by each isolate. However, the sequence information does not correlate with the phenotypes observed for these isolates (Fig. 7.4.A,

7.5.A and Table 7.1). For strains MI 222A and MI 222C, Western immunoblotting determined that they expressed the Le^x and Le^y antigens, respectively, which predicted that the $\alpha(1,2)$ fucosyltransferase enzyme would be expressed by isolate MI 222C but not expressed by isolate MI 222A. Between these strains, the *fucT2* gene has an identical sequence (Fig. 7.4.A). However, strain MI 222C has one additional C nucleotide present in the polyC tract giving 13 in total (highlighted in blue font). If these DNA sequences are translated, a stop codon (TAA) is read in strain MI 222C almost immediately downstream of the polyC tract because of the presence within this homopolymeric region of an extra C nucleotide. At variance to this, strain MI 222A, with one less C nucleotide present is not frameshifted and translational read-through is predicted (Fig. 7.4.B).

It must be borne in mind, however, that a sizeable portion of DNA sequence at the 3'-end of the fucT2 gene was not determined. Thus, the possibility cannot be ruled out that a mutagenic event further downstream in isolate MI 222A is responsible for the phenotype observed. Nonetheless, in as much as can be determined from the information at hand, although isolate MI 222A expresses the Le^x antigen, it appears likely to translate a full-length fucT2 gene product while isolate MI 222C has translation of the gene terminated prematurely, even although it expresses the Le^Y phenotype. In an attempt to confirm these conclusions, a second round of PCR and sequencing was carried out in order to reproduce the data. On this occasion, an identical sequence was determined for strain MI 222A (12 C nucleotides in the polyC tract as before). However, strain MI 222C had now lost a C base from the tract and now also had 12 C nucleotides in this region. Thus, as before, isolate MI 222A was deemed to express the $\alpha(1,2)$ fucosyltransferase enzyme, at odds with its Le^x phenotype while strain MI 222C now had a fucT2 gene sequence that could be correlated with the Le^Y phenotype it expressed. The findings are anomalous, therefore, but do suggest that slipped-strand mispairing took place in the polyC region during replication because of the sequence polymorphism apparent between the two isolates. However, an alternative mechanism is also implicated in bringing about the variable phenotypes observed because the data do not correlate.

There are several conclusions that can be drawn from the results of these investigations. First, the variable sequences obtained for isolate MI 222C at two distinct rounds of PCR and sequencing are considered evidence for the reversible nature of the phase-variable switching

phenomenon. As indicated by the DNA sequence information, strain MI 222C appears to have undergone lengthening and shortening of the polyC tract resulting, presumably, in reversible phenotypic switching of the LPS molecule. Although the H. pylori LPS molecule has been shown to stably express Le antigens in vitro (Wirth et al., 1999), the DNA sequence determined for strain MI 222C appears to suggest otherwise. However, only one phenotype (Le^Y) was detected for strain MI 222C by Western immunoblotting (Fig. 7.1) which might indicate that the antigen is stable in vitro and that the polymorphic PCR product sequences obtained reflect the manner in which the analysis was performed. It follows that if a heterogeneous population existed in vivo (and very likely also in vitro, Section 2.1.4) with respect to the Le phenotype expressed by the LPS of isolate MI 222C but where Le^Y was expressed by the majority of the cells, Western immunoblotting would detect a Le^Y phenotype but PCR would permit amplification of a translatable fucT2 PCR product from cells that were of the perceived phenotype, i.e., Le as in the second round of sequencing from strain MI 222C, while also able to amplify from a cell in which phase-variable switching had taken place resulting in the determination of a DNA sequence which likely reflects a cell that expressed Le^x phenotypically (as found in the first round of sequencing). By extrapolation, therefore, the same reversible switching can also take place in strain MI 222A. However, the DNA sequences determined in this investigation only bear witness to a putative Le^Y phenotype for this isolate.

Additionally, as discussed in Chapter 3 in relation to the DNA sequences of the *cagA* gene 3'end and apparent sequencing anomalies therein, the data described here for the *fucT2* gene
were similarly obtained using *Taq* DNA polymerase which has no proof-reading ability.
Furthermore, sequencing was carried from cloned PCR products on a single strand only.
These represent significant limitations of these studies given that the detection of single base
changes is central to the discussion of Le antigen phenotypic variability in *H. pylori*.

In their in-depth analysis of the *H. pylori fucT2* gene and the genetic basis for variable Le antigen expression, Wang *et al.* (1999a) also encountered a scenario where a strain (namely 26695) which had a hypothetical OFF status of the *fucT2* gene actually expressed the Le^Y antigen. To account for this, sequences were identified in the coding region of the *fucT2* gene which resembled translation frameshift cassettes from *E. coli* (Farabaugh, 1996). According to these authors (Wang *et al.*, 1999a) such cassettes could provide a reasonable mechanism

whereby the full-length protein could be produced by certain OFF status *fucT2* genes. Translation frameshifting is a process whereby a normal tRNA with attached amino acid reads an mRNA in a reading frame that is shifted in the -1 or +1 direction from the reading frame of the initiation codon (Engelberg-Kulka and Schoulaker-Schwarz, 1994). It is tempting to speculate that -1 frameshifting may also have taken place in isolate MI 222A to render the production of a truncated protein and hence an Le^x phenotype for this isolate, even although the predicted translation product of the partial gene sequence indicates that a full-length protein might be expressed.

The translation frameshift sequence of the consensus X XXY YYZ is highlighted in the sequence of strains MI 222A and MI 222C (Fig. 7.4.A) and exists in these strains as the sequence C CCT TTA upstream of the polyC tract. A second so-called 'slippery sequence' (AAAAAAG) is located downstream of the polyC region (also highlighted in Fig. 7.4.A) and is identical to the translation frameshift signal present in the mRNA of the E. coli dnaX gene. The almost homopolymeric heptanucleotide is followed by a region of secondary structure in the dnaX gene. A region of DNA that can form a secondary stem-loop structure is found downstream of the frameshift sequence in the fucT2 gene with a Shine-Dalgarno (SD)-like sequence found upstream, both serving as stimulators of frameshifting (Farabaugh, 1996). However, the heptanucleotide sequence is not in the correct reading frame. Only the C CCT TTA sequence exists in the reading frame of the initiation codon in the fucT2 gene of strain MI 222A, implicating programmed translation frameshifting at this sequence as a possible mechanism to account for the observed phenotype of this strain, despite DNA sequence evidence which suggests an alternative phenotype, most likely Le^Y. The same -1frameshifting in strain MI 222C would still cause a stop codon (TGA) to be read approximately 60 bases downstream of the polyC tract and would not, therefore, modify the Le phenotype expressed by this strain if the polyC tract contains 13 nucleotides.

The same analyses described for isolates MI 222A and MI 222C were also carried out for isolates MI 561 and MI 562 between which variability existed in the Le phenotype expressed by each isolate (Fig. 7.5.A). The diversity between isolates MI 222A and MI 222C in the *fucT2* gene was confined to the region of the polyC tract. However, it is immediately obvious upon examination of the alignment of the sequenced region of the *fucT2* gene from isolates MI 561 and MI 562 that a considerably more complex scenario exists in these strains. Polymorphisms between these isolates are present throughout the gene with several nucleotide

substitutions apparent. Those that are present at the position of the third base in the reading frame of the initiation codon are highlighted in green in Fig. 7.5.A. Close inspection of the sequence also indicates that several non-synonymous substitutions also took place and that very marked diversity exists between the strains in the region downstream of the polyC tract, which itself is heterogeneous in nucleotide number between MI 561 and MI 562. The so-called region of imperfect TAA repeats or 'A-rich region' that is present in the *fucT2* gene downstream of the homopolymeric tract was also identified by Wang *et al.* (1999a) as being polymorphic between independent isolates and a likely position at which slipped-strand mispairing might occur during replication causing phenotypic variation.

This DNA sequence information raises potentially intriguing questions. Does the extensive nature of the polymorphisms that exist between MI 561 and MI 562 in the fucT2 gene call into question the relatedness and hence clonality of these isolates? It is pertinent to doubt that these isolates are subtypic variants based on the DNA sequence information presented here. Indeed, amplification during PCR of contaminating DNA may be a valid explanation for the level of diversity observed in the three independent sequencing experiments for the fucT2 gene from isolate MI 562 (Fig. 7.5.C). However, it must also be borne in mind (assuming that no contamination took place during PCR) that the polymorphism that exists in the fucT2 gene might reflect the strong selective pressure that may be applied to the fucT2 gene product in vivo and that what is being observed in the polyC tract and downstream thereof is evidence of many rounds of mutation (i.e., slipped-strand mispairing) within the gene leading to ever-increasing divergence of the gene sequence between these strains. Furthermore, the oligofingerprinting method of discriminating between isolates testifies to a clonal relationship existing between these strains (Fig. 7.2.A). Additional rounds of PCR and sequencing of the same region of the fucT2 gene from strain MI 561 and, in particular, strain MI 562, gave still more compelling evidence of the likelihood that within the fucT2 locus mutation may be continuous, eliciting presumably many phenotypic variants over the prolonged period of colonisation (Fig. 7.5.B and C). As was encountered with isolates MI 222A and MI 222C, the sequence of the predicted translated fucT2 gene products determined for isolates MI 561 and MI 562 do not in all instances correlate with the Le phenotypes observed for these strains by Western immunoblotting. The occurrence of programmed translation frameshifting, as discussed in relation to MI 222A and MI 222C, in the fucT2 gene from MI 561 and MI 562 might, along with slipped-strand mispairing, mediate phase-variable switching of Le antigens.

It must be acknowledged that the data described in the present chapter are inconclusive and somewhat confusing. It is not possible to say if the sequence polymorphisms that were detected took place *in vivo* as a mechanism of adaptation by *H. pylori*, or if the sequences reflect the reversible, high-frequency nature of the phenomenon of phase variation and arose during *in vitro* culture. Merely being mentioned as an anecdote, very brief colony blot analysis of these phase-variable *H. pylori* isolates was carried out in order to determine if individual colonies that had undergone phase-variable switching could be detected in the population. When colonies were lifted from agar plates and the intact whole cells were probed with monoclonal antibodies to the Le^x and Le^y antigens it was found that the phenotype displayed by the colonies (numbering hundreds) was homogeneous. No colonies that were non-reactive, or indeed sectored, were found. Thus, this might be tentatively taken as evidence that the Le phenotype expressed by *H. pylori* LPS undergoes very low-frequency switching *in vitro* to the point where these antigens are considered to be stable during artificial culture and hence, that the sequencing data presented in this chapter might, as alluded to previously, reflect switching that had taken place *in vivo*.

It is clear that much additional work would be required to allow definitive conclusions to be drawn about the *bona fide* involvement of programmed translation frameshifting as a means by which Le antigen phase variation is brought about in these *H. pylori* isolates. This small study has only really addressed the possibility that slipped-strand mispairing was responsible mechanistically for the variable phenotypes observed between these two paired isolate sets, with the involvement of the alternative, programmed translation frameshifting mechanism inferred from the DNA sequence data that was obtained. As such, the study has many shortcomings. The analysis of mRNA transcription from the *fucT2* gene in isolates MI 222A, MI 222C, MI 561 and MI 562 by Northern blotting would be necessary for an accurate appraisal of the involvement of replication slippage and/or programmed translation frameshifting events in eliciting the phase variability detected. For example, an mRNA transcript from this gene should be detectable in all scenarios given that the effect of both slipped-strand mispairing and programmed translation frameshifting in terms of the functionality of the enzyme is brought to bear during translation of the protein.

Additionally, as also noted by Wang et al. (2000), the demonstration that these proposed mechanisms of phase variation actually function in vivo would represent significant data. The

investigation of this subject using cells cultured on laboratory medium can give at best, indications that the various sequence elements described might be involved in the process of phase variation *in vivo*, even although their involvement in LPS phase variation *in vitro* is well documented. Importantly, in light of the findings described in the present work, the analysis of fucosyltransferase gene sequences in tandem with serological determinations of the Le phenotypes of cells which had been cultured *in vivo*, e.g., in animal models of *H. pylori* infection, would be necessary to enable genotypic and phenotypic data to be correlated at a highly meaningful level. Indeed, as suggested by Wang *et al.* (2000), direct PCR from gastric biopsies would be an ideal approach towards an accurate appraisal of the *in vivo* situation.

Chapter 8
General Discussion

8.1.	A diverse species	185
8.2.	Microevolution	185
8.3.	Adaptive mutation?	186
8.4.	Insights into microevolution in <i>H. pylori</i> from studies described herein	187
8.5.	Adaptive mutation in other bacteria.	188
8.6.	Future perspectives.	189
8.7.	Concluding remarks	190

8.1. A diverse species.

For a complete understanding of a bacterium such as H. pylori and the disease it causes many research disciplines are necessary. There are those who have an epidemiological perspective, those who are interested in the biology of the bacterium and those who focus on the host response to infection. Generally, where these latter two schools intersect and evidence for dynamic, interactive relationships between host and pathogen are found, the most interesting and meaningful data in terms of disease mechanisms are obtained. Its clinical relevance aside, H. pylori is a most fascinating organism with many unusual physiological properties. From the outset almost, it was recognised that H. pylori exhibited a level of genotypic diversity unseen among other bacterial pathogens. The definition of a panmictic population structure for this organism combined with genome sequencing has shed much light on the origins of this diversity. It is now agreed that genetic variation arises from both a high level of de novo mutation and frequent recombination (Suerbaum, 2000). The realisation that H. pylori seemingly undergoes continuous microevolution during infection raised the possibility that the in vivo-occurring genetic variants might be evidence for host-specific adaptation in the genome of H. pylori and brought new insights to the relationship between H. pylori and host.

8.2. Microevolution.

The occurrence of microevolution in bacteria and in bacterial pathogens specifically is a commonly encountered theme. Microevolution is not surprising given that bacteria live and survive in diverse environments – different hosts, different niches within a single host and perhaps, on various environmental substrata. A degree of flexibility is required, thus, to mediate adaptation to continuous environmental change, e.g., nutrient exhaustion or the triggering of host defences (Claverys *et al.*, 2000). Adaptive responses can involve the activation of regulatory networks that act at the level of gene expression. However, bacteria have also evolved strategies that allow adaptation to be achieved through the generation of genetic diversity. Indeed, genetic variation provides the heterogeneity on which natural selection works and the continuous, spontaneous generation of novel genetic variants is a prerequisite for biological evolution (Arber, 2000).

Thus, in *H. pylori* as in other bacteria in which genetic plasticity has been investigated, variability can be provided by both endogenous mechanisms (point mutations, genetic rearrangements) and horizontal gene transfer (Alm *et al.*, 1999; Suerbaum *et al.*, 1998). The

genetic diversity of *H. pylori* reflects a considerable aptitude for ongoing adaptation. Indeed, Ferrero and Jenks (2001) cite several examples of *H. pylori* host-specific adaptation in animal model experiments. When non-human primates were administered pools of *H. pylori* strains, strain selection took place *in vivo* and the animals became colonised with a predominant strain. Furthermore, *in vivo* passage of *H. pylori* has been reported to result in increased infectivity of the strain in the piglet model (Akopyants *et al.*, 1995). *H. pylori* protein synthesis has also been observed to undergo alteration when strains were passaged *in vivo* (Janvier *et al.*, 1999). Thus, the plasticity of the *H. pylori* genome would seem to confer on the bacterium an inherent ability to improve fitness through *in vivo* selection of spontaneous mutations

8.3. Adaptive mutation?

The continual restatement in the literature, namely, that the rapid generation of new genotypes (afforded by genetic rearrangement and horizontal transfer in *H. pylori*) is a mechanism used by *H. pylori* to maintain fitness where environmental adaptation is required, actually has little corroborating *genotypic* evidence. The frequent generation of genetic variants would be expected in an organism that has a recombinational population structure and the occurrence of microevolution, therefore, might reflect both the panmixis of the species and evidence for host-specific adaptation. It is unlikely to be exclusively one or other of these scenarios. Furthermore, the occurrence of host-specific adaptation might be mediated through regulatory networks that have consequences for gene transcription. That the genome of *H. pylori* undergoes microevolution is now a recognised biological feature of this organism, but to what extent the occurrence of defined microevolutionary events can be deduced to have adaptive significance is an important question for which information is lacking. Adaptive mutation is defined as the production of mutations under non-lethal selections that relieve the selective pressure whether or not other non-selected mutations are also produced (Foster, 2000). Thus, what propensity to undergo adaptive mutation does *H. pylori* exhibit?

The importance of analysing microevolution between paired *H. pylori* isolates as took place throughout the studies described herein is underscored by the opportunity to address this question at several highly appropriate levels. For example, the paired isolates between which rearrangements were investigated and defined, are of proven clonality and allow the genesis of diversity in a single chromosome to be illuminated. Furthermore, the isolates represent clinical samples and the data, therefore, are not blurred by concerns as to the relevance of the

findings to human infection, as might have been the case had other *Helicobacter* spp. and animal models been employed for the selection of adapted derivative strains. By extension therefore, the contribution of host and environmental selective pressures (although these are undefined) to the occurrence of microevolution in the *H. pylori* genome is being addressed, albeit currently dependent on the deduction of adaptive significance based on the coding capacity of rearranged genes.

8.4. Insights into microevolution in *H. pylori* from studies described herein.

The insights about microevolution in *H. pylori* gained from the studies described herein allow greater appreciation of the subject while also raising many questions that merit further investigation. As was suspected, the findings indicate that rearrangement in the genomes of clonal isolates reflect the considerable capacity of *H. pylori* to generate genetic diversity and are summarised in Fig. 8.1. The involvement of repetitive DNA (at the 3'-end of the *cagA* gene), IS elements, horizontal gene transfer, slipped-strand mispairing and possibly also point mutation in the genesis of diversity between paired *H. pylori* isolates was detected. As such, the genetic drift that takes place in the *H. pylori* chromosome *in vivo* touches on all of the classical mechanisms of prokaryotic DNA rearrangement.

Furthermore, among the isolates examined, rearrangement involving Omp-encoding genes was found with greatest prevalence. This suggests likely adaptive mutation among the members of the *H. pylori hop* multigene family to possibly relieve selective pressure on outer-membrane-located proteins that might be involved in bacterium-host interactions. It is an intriguing possibility and one that had been suspected since the extent of the Omp-encoding gene family in *H. pylori* was recognised and such a large number of sequence-related genes were shown to exist in this organism (Tomb *et al.*, 1997). Characterisation of rearrangements within a single isolate between members of this multigene family would represent a significant contribution to what is known about the dynamic nature of the interaction between *H. pylori* and host. Definitive conclusions could also be drawn on the issue of Omp-encoding gene recombination and provide evidence to substantiate the widespread speculations on this issue. The recombinational events characterised upstream and downstream of the *hopN* gene (Chapter 5) provide physical evidence for horizontal transfer and IS element transposition, respectively, between paired isolates but do not implicate recombination between members of the Omp-encoding gene family.

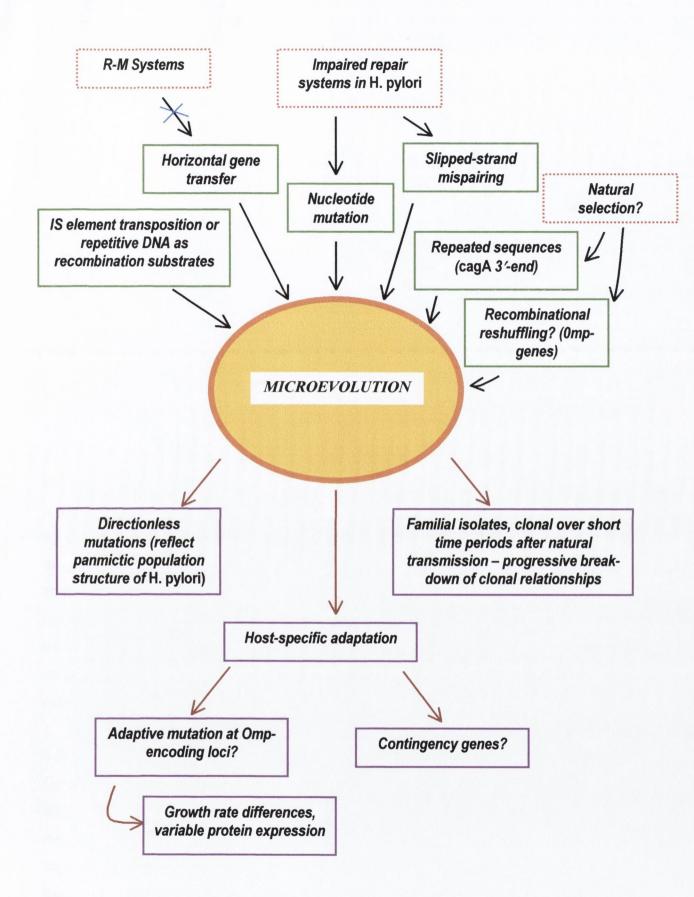


Fig. 8.1. Summary of findings from studies of microevolution in *H. pylori* paired isolates showing the sources of mutation, the barriers or stimulators thereof, and the various consequences for the *H. pylori* genome arising from *in vivo* diversity.

Nonetheless, differences in phenotype between paired *H. pylori* isolates in relation to growth rate and Omp expression strongly suggest a correlation with rearrangement involving Ompencoding genes. Indeed, herein lies a major question for future investigation. Differences in growth rate suggest mutation leading to a bacterial clone with improved fitness. With Ompencoding gene recombination as a possible mediator of this phenotypic difference, it would be interesting to use this as a basis for functional characterisation of the Omps involved. For example, N-terminal sequencing of the Omps that are variably expressed between paired isolates could allow correlation with an identified rearrangement at an Omp-encoding locus. Then, to create a scenario where genotypic and phenotypic information can be directly correlated, isogenic mutant strains could be constructed to analyse the infectivity of wild-type and mutant by quantitative culture in a colonisation model.

In terms of understanding *H. pylori* microevolution and how the *H. pylori* genome might undergo adaptive mutation, the ability to associate a particular genotype with a defined phenotype is the ultimate goal. Indeed, the extent of present knowledge about the *in vivo* adaptive significance of a spontaneous mutation event is probably best exemplified by spontaneous point mutations at positions 2143 and 2144 in the *H. pylori* 23S rRNA gene that confer a clarithromycin-resistant phenotype (Mégraud, 1998). Furthermore, slipped-strand mispairing in *H. pylori* fucosyltransferase genes mediates phase-variable expression of Lewis antigens expressed on the *H. pylori* LPS molecule, a phenomenon considered likely to be significant *in vivo* for host-specific adaptation, although it is not proven. Clearly, the data described herein suggest that much remains to be revealed and might form the foundation of a more in-depth study to ascertain the molecular mechanisms by which adaptive mutation at *H. pylori* Omp-encoding genes could take place.

8.5. Adaptive mutation in other bacteria.

Genetic plasticity in *S. pneumoniae* allows this organism to undergo environmental adaptation. Indeed, this bacterium is considered a paradigm among bacteria for recombination-mediated variability (Claverys *et al.*, 2000). It is comparable to *H. pylori* in its natural competence for transformation. It is this feature that allows, for example, switching of capsular types by intraspecies transformation, and the emergence of penicillin resistance by interspecies transformation, with the appearance of mosaic genes for penicillin-binding proteins that specify proteins with reduced affinity for the antibiotic. It is thought that natural transformation has evolved in *S. pneumoniae* to maximise genetic flexibility for

environmental adaptation (Claverys *et al.*, 2000). Indeed, the same might be true for H. pylori, although more is known about this subject in S. pneumoniae. Examples of defined genetic variability in S. pneumoniae resulting from recognised host and environmental selective pressures have been identified and include, as alluded to above, capsular serotype switching in response to host defences and altered penicillin-binding proteins arising from β -lactam treatment. There is some way to go yet before any aspect of host-specific adaptation can be precisely defined at a genetic level in H. pylori. However, the coming years might see H. pylori joining S. pneumoniae as both best exemplifying bacterial adaptation to its host and the interactive relationship that exists between them.

There are myriad other examples of bacteria that have evolved mechanisms for rapid and efficient phenotypic variation. As previously mentioned (Chapter 4), N. meningitidis, N. gonorrhoeae, Borrelia spp. and H. influenzae undergo antigenic variation of surface proteins or polysaccharide molecules that constitute major determinants of virulence (Deitsch et al., 1997). Furthermore, the existence of 'contingency' genes or hypermutable loci in bacteria has been proposed whereby the specific ability to generate variation at loci involved in hostpathogen interactions has evolved (Moxon et al., 1994). Such loci are characterised by the presence of certain nucleotide arrangements, e.g., homopolymeric tracts, tandem repeats or potential methylation sites, and allow high-frequency changes in the expression of genes where they are located. The term 'contingency' locus emphasises its role in facilitating rapid diversification by having a high mutation rate while coexisting with house-keeping genes that have low mutation rates. The fucosyltransferase genes of H. pylori might be candidate contingency genes that have evolved through natural selection. Indeed, the presence of a socalled 'mutator' phenotype (because of the absence of mismatch repair system components) in this organism increases the probability of mutations occurring in order to effect highfrequency phenotypic change of Lewis antigens. The illumination of the in vivo relevance of Lewis antigen phenotypic switching will clarify this possibility.

8.6. Future perspectives.

There is much scope to further the studies describe herein. With reference to the information obtained from PCR-subtractive hybridisation analysis of paired *H. pylori* strains, characterisation of the molecular basis for the genotypic difference between the paired isolates would be an important task, with priority given to Omp-encoding gene

rearrangements. Cloning and sequence analysis of the genomic fragments where evidence of rearrangement is apparent would probably be the best approach to address this question and would allow the full extent of PCR-subtractive hybridisation and Southern blotting data to be understood, although it was an approach fraught with difficulty in the present studies. Indeed, characterisation of putative Omp-encoding gene rearrangements or any of the rearrangement events identified is necessary in order to establish if *bona fide* rearrangement took place as opposed to nucleotide mutation, which in a minority of instances (Chapter 4) could have accounted for the hybridisation profile observed in Southern blotting. Simultaneously, it is possible to garner a more in-depth appreciation of the likelihood that a particular rearrangement is adaptive rather than directionless as previously discussed.

Clonally related familial *H. pylori* isolates constitute an important collection. Analysis of chromosomal recombinations in isolates derived from different hosts at a broader and more in-depth level than that which was described in Chapter 6 might enable specific host or environmental selective pressures to be pin-pointed more easily. It is an interesting possibility. Diversity in the *H. pylori* genome in the gastric niche of a single individual has the potential to allow identification of selective pressures and possible mechanisms of *H. pylori* adaptation. However, the opportunity to analyse the same isolate from a different host environment has also potentially meaningful information to contribute in this regard as the set of niche conditions encountered in the new host are entirely new.

Given the emergence of antibiotic-resistant isolates of *H. pylori*, the development of a prophylactic or therapeutic vaccine is receiving increased attention. The impact of microevolution on the efficacy of an *H. pylori* vaccine is difficult to assess. Nonetheless, because of the propensity of this organism for genetic variation, the identification of stable protective antigens would seem to be a priority in this field. Whether *H. pylori* can evolve *in vivo* to counteract vaccine-induced immune responses is unknown but the organism's ability to evolve to a distinct strain during long-term carriage *in vivo* would seem to be an omen for vaccine failures. Thus, microevolution of this organism is likely to remain an intriguing conundrum in future therapeutic approaches.

8.7. Concluding remarks.

The discovered association between *H. pylori* infection and gastroduodenal disease marked something of a watershed in medical science. The twenty years that have almost elapsed

since Warren and Marshall's first letter to The Lancet (1983) has been a prolific time in this arena not least of which has seen the sequencing and comparison of two *H. pylori* genomes set a precedent in bacteriology. There is a certain irony therefore in the fact that at a time when we are uniquely poised to discover more about the molecular mechanisms of *H. pylori* disease pathogenesis, social improvement has contributed to a decline in the global incidence of *H. pylori* infection. Nonetheless, the considerable clinical relevance of *H. pylori*-related disease is at the foundation of a field that will continue to be prodigious in output. To study this organism is to be immersed in a subject which itself is continually evolving and which has never been static. For a time, the microevolution of the *H. pylori* genome existed as a tangent to the story of the greater genomic diversity displayed by this species. As clarifications regarding this latter area were forthcoming, genomic microevolution is now regarded with a new patina of significance. Further dissection of the question holds great promise for an enhanced understanding of the dynamic relationship between *H. pylori* and host, and could yet provide pivotal insights into the chronic nature of the infection caused by *H. pylori* in humans.

Bibliography

- Achtman, M. (1994). Clonal spread of serogroup A meningococci: a paradigm for the analysis of microevolution in bacteria. *Molecular Microbiology* 11, 15-22.
- **Achtman, M. (2001).** Population structure of *Helicobacter pylori* and other pathogenic bacterial species. In Helicobacter pylori: *Molecular and Cellular Biology*, pp. 311-321. Edited by M. Achtman and S. Suerbaum. Wymondham, U.K.: Horizon Scientific Press.
- Achtman, M., Azuma, T., Berg, D.E., Ito, Y., Morelli, G., Pan, Z.-J., Suerbaum, S., Thompson, S.A., van der Ende, A. and van Doorn, L.-J. (1999). Recombination and clonal groupings within *Helicobacter pylori* from different geographical regions. *Molecular Microbiology* 32, 459-470.
- Akada, J.K., Shirai, M., Takeuchi, H., Tsuda, M. and Nakazawa, T. (2000). Identification of the urease operon in *Helicobacter pylori* and its control by mRNA decay in response to pH. *Molecular Microbiology* 36, 1071-1084.
- Akopyanz, N., Bukanov, N.O., Westblom, T.U., Kresovich, S. and Berg, D.E. (1992a). DNA diversity among clinical isolates of *Helicobacter pylori* detected by PCR-based RAPD fingerprinting. *Nucleic Acids Research* 20, 5137-5142.
- Akopyanz, N., Bukanov, N.O., Westblom, T.U. and Berg, D.E. (1992b). PCR-based RFLP analysis of DNA sequence diversity in the gastric pathogen *Helicobacter pylori*. *Nucleic Acids Research* 20, 6221-6225.
- Akopyants, N.S., Eaton, K.A. and Berg, D.E. (1995). Adaptive mutation and cocolonization during *Helicobacter pylori* infection of gnotobiotic piglets. *Infection and Immunity* 63, 116-121.
- Akopyants, N.S., Clifton, S.W., Kersulyte, D., Crabtree, J., Youree, B.E., Reece, C.A., Bukanov, N.O., Drazek, E.S., Roe, B.A. and Berg, D.E. (1998a). Analyses of the cag pathogenicity island of *Helicobacter pylori*. *Molecular Microbiology* 28, 37-53.
- Akopyants, N.S., Fradkov, A., Diatchenko, L., Hill J.E., Siebert, P.D., Lukyanov, S.A., Sverdlov, E.D. and Berg. D.E. (1998b). PCR-based subtractive hybridization and differences in gene content among strains of *Helicobacter pylori*. Proceedings of the National Academy of Sciences USA 95, 13108-13113.
- Allan, E., Dorrell, N., Foynes, S., Anyim, M. and Wren, B.W. (2000). Mutational analysis of the genes encoding the early flagellar components of *Helicobacter pylori*: evidence for transcriptional regulation of flagellin A biosynthesis. *Journal of Bacteriology* 182, 5274-5277.
- Alm, R.A. and Trust, T.J. (1999). Analysis of the genetic diversity of *Helicobacter pylori*: the tale of two genomes. *Journal of Molecular Medicine* 77, 834-846.
- Alm, R.A., Ling, L.-S.L., Moir, D.T. and 20 other authors. (1999). Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori*. *Nature* 397, 176-180.

- Alm, R.A., Bina, J., Andrews, B.M., Doig, P., Hancock, R.E.W. and Trust, T.J. (2000). Comparative genomics of *Helicobacter pylori*: analysis of the outer membrane protein families. *Infection and Immunity* **68**, 4155-4168.
- Ando, T., Israel, D.A., Kusugami, K. and Blaser, M.J. (1999). HP0333, a member of the dprA family, is involved in natural transformation in Helicobacter pylori. Journal of Bacteriology 181, 5572-5580.
- Appelmelk, B.J., Simoons-Smit, I., Negrini, R. and 17 other authors. (1996). Potential role of molecular mimicry between *Helicobacter pylori* lipopolysaccharide and host Lewis blood group antigens in autoimmunity. *Infection and Immunity* 64, 2031-2040.
- Appelmelk, B.J., Shiberu, B., Trinks, C. and 10 other authors. (1998). Phase variation in *Helicobacter pylori* lipopolysaccharide. *Infection and Immunity* 66, 70-76.
- Appelmelk, B.J., Martin, S.L., Monteiro, M.A. and 10 other authors. (1999). Phase variation in *Helicobacter pylori* lipopolysaccharide due to changes in the lengths of poly(C) tracts in α3-fucosyltransferase genes. *Infection and Immunity* 67, 5361-5366.
- Appelmelk, B.J., Monteiro, M.A., Martin, S.L., Moran, A.P. and Vandenbroucke-Grauls, C.M.J.E. (2000). Why *Helicobacter pylori* has Lewis antigens. *Trends in Microbiology* 8, 565-570.
- Arber, W. (2000). Genetic variation: molecular mechanisms and impact on microbial evolution. FEMS Microbiology Reviews 24, 1-7.
- **Aspinall, G.O. and Monteiro, M.A. (1996).** Lipopolysaccharides of *Helicobacter pylori* strains P466 and MO19: structures of the O antigen and core oligosaccharide regions. *Biochemistry* **35**, 2498-2504.
- Aspinall, G.O., Monteiro, M.A., Pang, H., Walsh, E.J. and Moran, A.P. (1996). Lipopolysaccharide of the *Helicobacter pylori* type strain NCTC11637 (ATCC 43504): structure of the O antigen chain and core oligosaccharide regions. *Biochemistry* 35, 2489-2497.
- Atherton, J.C. (1997). The clinical relevance of strain types of *Helicobacter pylori*. Gut 40, 701-703.
- Atherton, J.C. (1998). H. pylori virulence factors. British Medical Bulletin 54, 105-120.
- Atherton, J.C., Cao, P., Peek, R.M., Tummuru, M.K.R., Blaser, M.J. and Cover, T.L. (1995). Mosaicism in vacuolating cytotoxin alleles of *Helicobacter pylori*. The Journal of Biological Chemistry 270, 17771-17777.
- Atherton, J.C., Tham, K.T., Peek, Jr., R.M., Cover, T.L. and Blaser, M.J. (1996). Density of *Helicobacter pylori* infection *in vivo* as assessed by quantitative culture and histology. *The Journal of Infectious Diseases* 174, 552-556.

- Atherton, J., Peek, R.M., Jr, Tham, K.T., Cover, T.L. and Blaser, M. (1997). Clinical and pathological importance of heterogeneity in *vacA*, the vacuolating cytotoxin gene of *Helicobacter pylori*. *Gastroenterology* 112, 92-99.
- Atherton, J.C., Cover, T.L., Twells, R.J., Morales, M.R., Hawkey, C.J. and Blaser, M.J. (1999). Simple and accurate PCR-based system for typing vacuolating cytotoxin alleles of *Helicobacter pylori*. *Journal of Clinical Microbiology* 37, 2979-2982.
- Audibert, C., Burucoa, C., Janvier, B. and Fauchère, J.L. (2001). Implication of the structure of the *Helicobacter pylori cag* pathogenicity island in induction of interleukin-8 secretion. *Infection and Immunity* 69, 1625-1629.
- Ausubel, R.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. and Struhl, K. (1987). Current Protocols in Molecular Biology. Current Protocols, Greene Publishing Associates Inc., NY: John Wiley and Sons Inc.
- Backert, S., Ziska, E., Brinkmann, V., Zimny-Arndt, U., Fauconnier, A., Jungblut, P.R., Naumann, M. and Meyer, T.F. (2000). Translocation of the *Helicobacter pylori* CagA protein in gastric epithelial cells by a type IV secretion apparatus. *Cellular Microbiology* 2, 155-164.
- Bamford, K.B., Bickley, J., Collins, J.S., Johnston, B.T., Potts, S., Boston, V., Owen, R.J. and Sloan, J.M. (1993). *Helicobacter pylori*: comparison of DNA fingerprints provides evidence for intrafamilial infection. *Gut* 34, 1348-1350.
- Bavoil, P.M., Hsia, R.-C. and Ojcius, D.M. (2000). Closing in on *Chlamydia* and its intracellular bag of tricks. *Microbiology* 146, 2723-2731.
- Bereswill, S., Waidner, U., Odenbreit, S., Lichte, F., Fassbinder, F., Bode, G. and Kist, M. (1998). Structural, functional and mutational analysis of the *pfr* gene encoding a ferritin from *Helicobacter pylori*. *Microbiology* 144, 2505-2516.
- Berg, D.E., Hoffman, P., Appelmelk, B.J. and Kusters, J.G. (1997). The *Helicobacter pylori* genome sequence: genetic factors for long life in the gastric mucosa. *Trends in Microbiology* 12, 468-474.
- Bhugra, B., Voelker, L.L., Zou, N., Yu, H. and Dybvig, K. (1995). Mechanism of antigenic variation in *Mycoplasma pulmonis*: interwoven, site-specific DNA inversions. *Molecular Microbiology* 18, 703-714.
- Birnboim, H.C. and Doly, J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Research* 7, 1513-1523.
- Björkholm, B., Lundin, A., Sillén, A., Guillemin, K., Salama, N., Rubio, C., Gordon, J.I., Falk, P. and Engstrand, L. (2001a). Comparison of genetic divergence and fitness between two subclones of *Helicobacter pylori*. *Infection and Immunity* 69, 7832-7838.
- Björkholm, B.M., Oh, J.D., Falk, P.G., Engstrand, L.G. and Gordon, J.I. (2001b). Genomics and proteomics converge on *Helicobacter pylori*. Current Opinion in Microbiology 4, 237-245.

Blaser, M.J. (1990). *Helicobacter pylori* and the pathogenesis of gastroduodenal inflammation. *The Journal of Infectious Diseases* 161, 626-633.

Blaser, M.J. (1997a). Heterogeneity of Helicobacter pylori. European Journal of Gastroenterology & Hepatology 9 (Suppl. 1), S3-S7.

Blaser, M.J. (1997b). Not all *Helicobacter pylori* strains are created equal: should all be eliminated? *Lancet* 349, 1020-1022.

Blaser, M.J. (1993). *Helicobacter pylori*: microbiology of a 'slow' bacterial infection. *Trends in Microbiology* 1, 255-260.

Blaser, M.J. and Parsonnet, J. (1994). Parasitism by the 'slow' bacterium *Helicobacter pylori* leads to altered gastric homeostasis and neoplasia. *The Journal of Clinical Investigation* **94**, 4-8.

Bodger, K. and Crabtree, J.E. (1998). Helicobacter pylori and gastric inflammation. British Medical Bulletin 54, 139-150.

Boehringer Mannheim (1995). The DIG System User's Guide for Filter Hybridization. Germany: Boehringer Mannheim GmbH, Biochemica.

Borén, T., Falk, P., Roth, K.A., Larson, G. and Normark, S. (1993). Attachment of *Helicobacter pylori* to human gastric epithelium mediated by blood group antigens. *Science* 262, 1892-1895.

Bukanov, N.O. and Berg, D.E. (1994). Ordered cosmid library and high-resolution physical-genetic map of *Helicobacter pylori* strain NCTC11638. *Molecular Microbiology* 11, 509-523.

Calam, J. (1996). Clinicians' guide to Helicobacter pylori, pp. 48-52. London, UK: Chapman and Hall Medical.

Calam, J. (1998). Clinical science of *Helicobacter pylori* infection: ulcers and NSAIDs. *British Medical Bulletin* 54, 55-62.

Calia, K.E., Waldor, M.K. and Calderwood, S.B. (1998). Use of representational difference analysis to identify genomic differences between pathogenic strains of *Vibrio cholerae*. *Infection and Immunity* 66, 849-852.

Censini, S., Lange, C., Xiang, Z., Crabtree, J.E., Ghiara, P., Borodovsky, M., Rappuoli, R. and Covacci, A. (1996). cag, a pathogenicity island of Helicobacter pylori, encodes type I-specific and disease-associated virulence factors. Proceedings of the National Academy of Sciences USA 93, 14648-14653.

Chakravarti, D.N., Fiske, M.J., Fletcher, L.D. and Zagursky, R.J. (2001). Application of genomics and proteomics for identification of bacterial gene products as potential vaccine candidates. *Vaccine* 19, 601-612.

- Chen, M., Lee, A. and Hazell, S.L. (1992). Immunisation against *Helicobacter* infection in a mouse/*Helicobacter felis* model. *Lancet* 339, 1120-1121.
- Chevalier, C., Thiberge, J.-M., Ferrero, R. and Labigne, A. (1999). Essential role of *Helicobacter pylori* γ -glutamyltranspeptidase for the colonization of the gastric mucosa of mice. *Molecular Microbiology* 31, 13599-1372.
- Claverys, J-P., Prudhomme, M., Mortier-Barrière, I. and Martin, B. (2000). Adaptation to the environment: *Streptococcus pneumoniae*, a paradigm for recombination-mediated genetic plasticity? *Molecular Microbiology* 35, 251-259.
- Clontech, (1997). Clontech PCR-Select ™ Bacterial Genome Subtraction Kit User Manual. Palo Alto, CA, USA: Clontech Laboratories, Inc.
- Clyne, M., O'Croinin, T., Suerbaum, S., Josenhans, C. and Drumm, B. (2000). Adherence of isogenic flagellum-negative mutants of *Helicobacter pylori* and *Helicobacter mustelae* to human and ferret gastric epithelial cells. *Infection and Immunity* 68, 4335-4339.
- Clyne, M., Labigne, A. and Drumm, B. (1995). Helicobacter pylori requires an acidic environment to survive in the presence of urea. Infection and Immunity 63, 1669-1673.
- Collins, F.S. and Weissman, S.M. (1984). Directional cloning of DNA fragments at a large distance from an initial probe: a circularization method. *Proceedings of the National Academy of Sciences USA* 81, 6812-6816.
- Correa, P. (1992). Human gastric carcinogenesis: a multistep and multifactorial process. *Cancer Research* 52, 6735-6740.
- Covacci, A. and Rappuoli, R. (1998). Helicobacter pylori: molecular evolution of a bacterial quasi-species. Current Opinion in Microbiology 1, 96-102.
- Covacci, A., Censini, C., Bugnoli, M. and 8 other authors. (1993). Molecular characterization of the 128-kDa immunodominant antigen of *Helicobacter pylori* associated with cytotoxicity and duodenal ulcer. *Proceedings of the National Academy of Sciences USA* **90**, 5791-5795.
- Covacci, A., Telford, J.L., Del Giudice, G., Parsonnet, J. and Rappuoli R. (1999). *Helicobacter pylori* virulence and genetic geography. *Science* 284, 1328-1333.
- Cover, T.L., Dooley, C.P. and Blaser, M.J. (1990). Characterization of and human serologic response to proteins in *Helicobacter pylori* broth culture supernatants with vacuolizing cytotoxin activity. *Infection and Immunity* 58, 603-610.
- Cover, T.L., Tummuru, M.K.R., Cao, P., Thompson, S.A. and Blaser, M.J. (1994). Divergence of genetic sequences for the vacuolating cytotoxin among *Helicobactyer pylori* strains. *The Journal of Biological Chemistry* **269**, 10566-10573.
- Cover, T.L., Glupczynski, Y., Lage, A.P., Burette, A., Tummuru, M.K.R., Perez-Perez, G.I. and Blaser, M.J. (1995). Serologic detection of infection with $cagA^+$ Helicobacter pylori strains. Journal of Clinical Microbiology 33, 1496-1500.

- Crabtree, J.E., Farmery, S.M., Lindley, I.J., Figura, N., Peichl, P. and Tompkins, D.S. (1994). CagA/cytotoxic strains of *Helicobacter pylori* and interleukin-8 in gastric epithelial cell lines. *Journal of Clinical Pathology* 47, 945-950.
- Crabtree, J.E., Covacci, A., Farmery, S.M., Xiang, Z., Tompkins, D.S., Perry, S., Lindley, I.J. and Rappuoli, R. (1995a). *Helicobacter pylori* induced interleukin-8 expression in gastric epithelial cells is associated with CagA positive phenotype. *Journal of Clinical Pathology* 48, 41-45.
- Crabtree, J.E., Xiang, Z., Lindley, I.J., Tompkins, D.S., Rappuoli, R. and Covacci, A. (1995b). Induction of interleukin-8 secretion from gastric epithelial cells by a *cagA* negative isogenic mutant of *Helicobacter pylori*. *Journal of Clinical Pathology* 48, 967-969.
- Cuenca, R., Blanchard, T.G., Czinn, S.J., Nedrud, J.G., Monath, T.P., Lee, C.K. and Redline, R.W. (1996). Therapeutic immunization against *Helicobacter mustelae* in naturally infected ferrets. *Gastroenterology* 110, 1770-1775.
- Czajkowsky, D.M., Iwamoto, K., Cover, T.L. and Shao, Z. (1999). The vacuolating toxin from *Helicobacter pylori* forms hexameric pores in lipid bilayers at low pH. *Proceedings of the National Academy of Sciences USA* 96, 2001-2006.
- de Bernard, M., Papini, E., de Filippis, V., Gottardi, E., Telford, J.L., Manetti, R., Fontana, A., Rappuoli, R. and Montecucco, C. (1995). Low pH activates the vacuolating toxin of *Helicobacter pylori*, which becomes acid and pepsin resistant. *The Journal of Biological Chemistry* 270, 23937-23940.
- de Bernard, M., Burroni, D., Papini, E., Rappuloi, R., Telford, J. and Montecucco, C. (1998). Identification of the *Helicobacter pylori* VacA toxin domain active in the cell cytosol. *Infection and Immunity* 66, 6014-6016.
- **Deighan, P., Free, A. and Dorman, C.J. (2000).** A role for the *Escherichia coli* H-NS-like protein StpA in OmpF porin expression through modulation of *micF* RNA stability. *Molecular Microbiology* **38**, 126-139.
- **Deitsch**, K.W., Moxon, E.R. and Wellems, T.E. (1997). Shared themes of antigenic variation and virulence in bacterial, protozoal, and fungal pathogens. *Microbiology and Molecular Biology Reviews* **61**, 281-293.
- de la Cruz, F. and Davies, J. (2000). Horizontal gene transfer and the origin of species: lessons from bacteria. *Trends in Microbiology* 8, 128-133.
- Delaney, B.C. and O'Morain, C. (2001). Management of dyspepsia. Current Opinion in Gastroenterology 17 (Suppl. 1), S38-S42.
- Delany, I., Pacheco, A.B.F., Spohn, G., Rappuoli, R. and Scarlato, V. (2001). Iron-dependent transcription of the *frpB* gene of *Helicobacter pylori* is controlled by the Fur repressor protein. *Journal of Bacteriology* 183, 4932-4937.

- **Dice**, L.R. (1945). Measures of the amount of ecological association between species. *Ecology* 26, 297-302.
- **Doenges, J.L.** (1938). Spirochaetes in the gastric glands of *Macacus rhesus* and humans without definite history of related disease. *Proceedings of the Society for Experimental Biology and Medicine* 38, 536-538.
- Doidge, C., Gust, I., Lee, A., Buck, F., Hazell, S. and Manne, U. (1994). Therapeutic immunisation against *Helicobacter* infection. *Lancet* 343, 914-915.
- **Doig, P. and Trust, T.J. (1994).** Identification of surface-exposed outer membrane antigens of *Helicobacter pylori*. *Infection and Immunity* **62**, 4526-4533.
- Doig, P., Exner, M.M., Hancock, R.E.W. and Trust, T.J. (1995). Isolation and characterization of a conserved porin protein from *Helicobacter pylori*. *Journal of Bacteriology* 177, 5447-5452.
- Dowsett, S.A., Archila, L., Segreto, V.A., Gonzalez, C.R., Silva, A., Vastola, K.A., Bartizek, R.D. and Kowolik, M.J. (1999). Helicobacter pylori infection in indigenous families of Central America: serostatus and oral and fingernail carriage. Journal of Clinical Microbiology 37, 2456-2460.
- Drumm, B., Perez-Perez, G.I., Blaser, M.J. and Sherman, P.M. (1990). Intrafamilial clustering of *Helicobacter pylori* infection. *New England Journal of Medicine* 322, 359-363.
- Dundon, W.G., de Bernard, M. and Montecucco, C. (2001). Virulence factors of Helicobacter pylori. International Journal of Medical Microbiology 290, 647-658.
- Dunn, B.E., Cohen, H. and Blaser, M.J. (1997). Helicobacter pylori. Clinical Microbiology Reviews 10, 720-741.
- Eaton, K.A. (1999). Animal models of *Helicobacter* gastritis. Current Topics in Microbiology and Immunology, 241, 123-154.
- Eaton, K.A., Brooks, C.L., Morgan, D.R. and Krakowka, S. (1991). Essential role of urease in pathogenesis of gastritis induced by *Helicobacter pylori* in gnotobiotic piglets. *Infection and Immunity* 59, 2470-2475.
- Eaton, K.A.S., Suerbaum, S., Josenhans, C. and Krakowka, S. (1996). Colonization of gnotobiotic piglets by *Helicobacter pylori* deficient in two flagellin genes. *Infection and Immunity* 64, 2445-2448.
- Edwards, N.J., Monteiro, M.A., Faller, G., Walsh, E.J., Moran, A.P., Roberts, I.S. and High, N.J. (2000). Lewis X structures in the O antigen side-chain promote adhesion of *Helicobacter pylori* to the gastric epithelium. *Molecular Microbiology* 35, 1530-1539.
- El-Omar, E.M. and Malfertheiner, P. (2001). Helicobacter pylori infection and gastric cancer. Current Opinion in Gastroenterology 17 (Suppl. 1), S24-S27.

- Engelberg-Kulka, H. and Schoulaker-Schwarz, R. (1994). Regulatory implications of translational frameshifting in cellular gene expression. *Molecular Microbiology* 11, 3-8.
- Engstrand, L. and Hazell, S. (2001). Genomics and the evolution of *Helicobacter* spp. Current Opinion in Gastroenterology 17 (Suppl. 1), S6-S11.
- Enroth, H., Åkerlund, T., Sillén, A and Engstrand, L. (2000). Clustering of clinical strains of *Helicobacter pylori* analyzed by two-dimensional gel electrophoresis. *Clinical and Diagnostic Laboratory Immunology* 7, 301-306.
- Ernst, P.B. and Gold, B.D. (2000). The disease spectrum of *Helicobacter pylori*: the immunopathogenesis of gastroduodenal ulcer and gastric cancer. *Annual Review of Microbiology* **54**, 615-640.
- Evans, D.J. and Evans, D.G. (2001). *Helicobacter pylori* CagA: analysis of sequence diversity in relation to phosphorylation motifs and implications for the role of CagA as a virulence factor. *Helicobacter* 6, 187-198.
- Exner, M.E., Doig, P., Trust, T.J. and Hancock, R.E. (1995). Isolation and characterization of a family of porin proteins from *Helicobacter pylori*. *Infection and Immunity* 63, 1567-1572.
- Falk, P. (2001). Why does *Helicobacter pylori* actually have Lewis antigens. *Trends in Microbiology* 9, 61-62.
- Falk, P.G., Syder, A.J., Guruge, J.L., Kirschner, D., Blaser, M.J. and Gordon, J.I. (2000). Theoretical and experimental approaches for studying factors defining the *Helicobacter pylori*-host relationship. *Trends in Microbiology* 8, 321-329.
- Farabaugh, P.J. (1996). Programmed translational frameshifting. Annual Review of Genetics 30, 507-528.
- Fassbinder, F., van Vliet, A.H.M., Gimmel, V., Kusters, J.G., Kist, M. and Bereswill, S. (2000). Identification of iron-regulated genes of *Helicobacter pylori* by a modified Fur titration assay (FURTA-Hp). *FEMS Microbiology Letters* 184, 225-229.
- Feldman, R.A., Eccersley, A.J.P. and Hardie, J.M. (1997). Transmission of *Helicobacter pylori*. Current Opinion in Gastroenterology **8**, 8-12.
- **Feldman, R.A., Eccersley, A.J.P. and Hardie, J.M. (1998).** Epidemiology of *Helicobacter pylori*: acquisition, transmission, population prevalence and disease-to-infection ratio. *British Medical Bulletin* **54**, 39-53.
- **Ferrero, R.L. and Jenks, P.J. (2001).** *In vivo* adaptation to the host. In Helicobacter pylori: *Physiology and Genetics*, pp. 583-592. Edited by H.L.T. Mobley, G.L. Mendz and S.L. Hazell, S.L. Washington, D.C.: ASM Press.
- Figueiredo, C., van Doorn, L.J., Nogueira, C., Soares, J.M., Pinho, C., Figueira, P., Quint, W.G. and Carneiro, F. (2001). *Helicobacter pylori* genotypes are associated with

- clinical outcome in Portuguese patients and show a high prevalence of infections with multiple strains. Scandinavian Journal of Gastroenterology 36, 128-135.
- Forbes, K.J., Fang, Z. and Pennington, T.H. (1995). Allelic variation in the *Helicobacter pylori* flagellin genes *flaA* and *flaB*: its consequences for strain typing schemes and population structure. *Epidemiology and Infection* 114, 257-266.
- Foster, P.L. (2000). Adaptive mutation: implications for evolution. *BioEssays* 22, 1067-1074.
- Fox, J.G., Correa, P. and Taylor, N.S., Lee, A., Otto, G., Murphy, J.C. and Rose, R. (1990). *Helicobacter mustelae*-associated gastritis in ferrets. An animal model of *Helicobacter pylori* gastritis in humans. *Gastroenterology* 99, 352-361.
- Foxall, PA., Hu, L.-T. and Mobley, H.L.T. (1992). Use of polymerase chain reaction-amplified *Helicobacter pylori* urease structural genes for differentiation of isolates. *Journal of Clinical Microbiology* 30, 739-741.
- Franzot, S.P., Mukherjee, J., Cherniak, R., Chen, L.-C., Hamdan, J.S. and Casadevall, A. (1998). Microevolution of a standard strain of *Cryptococcus neoformans* resulting in differences in virulence and other phenotypes. *Infection and Immunity* 66, 89-97.
- Frazier, B.A., Pfeifer, J.D., Russell, D.G., Falk, P., Olsén, A.N., Hammar, M., Westblom, T.U. and Normark, S.J. (1993). Paracrystalline inclusions of a novel ferritin containing nonheme iron, produced by the human gastric pathogen *Helicobacter pylori*: evidence for a third class of ferritins. *Journal of Bacteriology* 175, 966-972.
- Freeman, J.A., Lilley, B.N. and Bassler, B.L. (2000). A genetic analysis of the function of LuxN: a two-component hybrid sensor kinase that regulates quorum sensing in *Vibrio harveyi*. *Molecular Microbiology* 35, 139-149.
- Fujimoto, S., Marshall, B. and Blaser, M.J. (1994). PCR-based restriction fragment length polymorphism typing of *Helicobacter pylori*. *Journal of Clinical Microbiology* 32, 331-334.
- Ge, Z., Chan, N.W.C., Palcic, M.M. and Taylor, D.E. (1997). Cloning and heterologous expression of an α1,3-fucosyltransferase gene from the gastric pathogen *Helicobacter pylori*. *The Journal of Biological Chemistry* 272, 21357-21363.
- Gee, S.F., Joly, S., Soll D.R., Meis, J.F., Verweij, P.E., Polacheck, I., Sullivan, D.J. and Coleman, D.C. (2002). Identification of four distinct genotypes of *Candida dubliniensis* and detection of microevolution *in vitro* and *in vivo*. *Journal of Clinical Microbiology* 40, 556-574.
- Georgopoulos, S.D., Mentis, A.F., Spiliadis, C.A., Tzouvelekis, L.S., Tzelepi, E., Moshopoulos, A. and Skandalis, N. (1996). *Helicobacter pylori* infection in spouses of patients with duodenal ulcers and comparison of ribosomal RNA gene patterns. *Gut* 39, 634-638.
- Gerhard, M., Lehn, N., Neumayer, N., Boren, T., Rad, R., Schepp, W., Miehlke, S., Classen, M. and Prinz, C. (1999). Clinical relevance of the *Helicobacter pylori* gene for

- blood-group antigen-binding adhesin. *Proceedings of the National Academy of Sciences USA* **96,** 12778-12783.
- Ghiara, P., Marchetti, M., Blaser, M.J., Tummuru, M.K.R., Cover, T.L., Segal, E.D., Tompkins, L.S. and Rappuoli R. (1995). Role of the *Helicobacter pylori* virulence factors vacuolating cytotoxin, CagA, and urease in a mouse model of disease. *Infection and Immunity* 63, 4154-4160.
- Ghiara, P., Rossi, M., Marchetti, M. and 9 other authors. (1997). Therapeutic intragastric vaccination against *Helicobacter pylori* in mice eradicates an otherwise chronic infection and confers protection against reinfection. *Infection and Immunity* 65, 4996-5002.
- Gibson, J.R., Chart, H. and Owen, R.J. (1998). Intra-strain variation in expression of lipopolysaccharide by *Helicobacter pylori*. Letters in Applied Microbiology 26, 399-403.
- Go, M.F., Kapur, V., Graham, D.Y. and Musser, J.M. (1996). Population genetic analysis of *Helicobacter pylori* by multilocus enzyme electrophoresis: extensive allelic diversity and recombinational population structure. *Journal of Bacteriology* 178, 3934-3938.
- Goldman, B. and Kranz, R. (1998). Evolution and horizontal transfer of an entire biosynthetic pathway for cytochrome c biogenesis: *Helicobacter*, *Deinococcus*, Archae and more. *Molecular Microbiology* 27, 871-874.
- Goodman, K.J. and Correa, P. (2000). Transmission of *Helicobacter pylori* among siblings. *Lancet* 355, 358-362.
- Goodwin, C.S. (1997). *Helicobacter pylori* gastritis, peptic ulcer, and gastric cancer: clinical and molecular aspects. *Clinical Infectious Diseases* 25, 1017-1019.
- Goodwin, C.S., Armstrong, J.A., Chilvers, T. (1989). Transfer of Campylobacter pylori and Campylobacter mustelae to Helicobacter gen. nov. as Helicobacter pylori comb. nov. and Helicobacter mustelae comb. nov., respectively. International Journal of Systematic Bacteriology 39, 397-405.
- Graham, D.Y. and Yamaoka, Y. (2000). Disease-specific Helicobacter pylori virulence factors: the unfulfilled promise. Helicobacter 5 (Suppl. 1), S3-S9.
- Gravekamp, C., Rosner, B. and Madoff, L.C. (1998). Deletion of repeats in the alpha C protein enhances the pathogenicity of group B streptococci in immune mice. *Infection and Immunity* 66, 4347-4354,
- Han, S.-R., Zschausch, H.-C.E., Meyer, H.-G.W., Schneider, T., Loos, M., Bhakdi, S. and Maeurer, M.J. (2000). *Helicobacter pylori*: clonal population structure and restricted transmission within families revealed by molecular typing. *Journal of Clinical Microbiology* 38, 3646-3651.
- Hazell, S.L., Andrews, R.H., Mitchell, H.M. and Daskalopoulous, G. (1997). Genetic relationship among isolates of *Helicobacter pylori*: evidence for the existence of a *Helicobacter pylori* species-complex. *FEMS Microbiology Letters* **150**, 27-32.

- **Henderson**, I.R., Owen, P. and Nataro, J.P. (1999). Molecular switches the ON and OFF of bacterial phase variation. *Molecular Microbiology* 33, 919-932.
- High, N.J., Deadman, M.E. and Moxon, E.R. (1993). The role of a repetitive DNA motif (5'-CAAT-3') in the variable expression of the *Haemophilus influenzae* lipopolysaccharide epitope $\alpha \text{Gal}(1-4)\beta \text{Gal}$. *Molecular Microbiology* 9, 1275-1282.
- Hobbs, M.M., Seiler, A., Achtman, M. and Cannon, J.G. (1994). Microevolution within a clonal population of pathogenic bacteria: recombination, gene duplication and horizontal genetic exchange in the *opa* gene family of *Neisseria meningitidis*. *Molecular Microbiology* 12, 171-180.
- Hoch, J.A. (2000). Two-component and phosphorelay signal transduction. *Current Opinion in Microbiology* 3, 165-170.
- Hofreuter, D., Odenbreit, S., Henke, G. and Haas, R. (1998). Natural competence for DNA transformation in *Helicobacter pylori*: identification and genetic characterization of the *comB* locus. *Molecular Microbiology* 28, 1027-1038.
- Hofreuter, D., Odenbreit, S. and Haas, R. (2001). Natural transformation competence in *Helicobacter pylori* is mediated by the basic components of a type IV secretion system. *Molecular Microbiology* 41, 379-391.
- Höök-Nikanne, J., Berg, D.E., Peek, Jr., R.M., Kersulyte, D., Tummuru, M.K.R. and Blaser, M.J. (1998). DNA sequence conservation and diversity in transposable element IS605 of Helicobacter pylori. Helicobacter 3, 79-85.
- Hulten, K., Han, S.W. and Enroth, H., Kelin, P.D., Opekun, A.R., Gilman, R.H., Evans, D.G., Engstrand, L., Graham, D.Y. and El-Zaatari, F.A. (1996). *Helicobacter pylori* in the drinking water in Peru. *Gastroenterology* 110, 1031-1035.
- Ikenoue, T., Maeda, S., Ogura, K., Akanuma, M., Mitsuno, Y., Imai, Y., Yoshida, H., Shiratori, Y. and Omata, M. (2001). Determination of *Helicobacter pylori* virulence by simple gene analysis of the *cag* pathogenicity island. *Clinical and Diagnostic Laboratory Immunology* 8, 181-186.
- Ilver, D., Arnqvist, A., Ögren, J., Frick, I., Kersulyte, D., Incecik, E.T., Berg, D.E., Covacci, A., Engstrand, L. and Borén, T. (1998). *Helicobacter pylori* adhesin binding fucosylated histo-blood group antigens revealed by retagging. *Science* 279, 373-377.
- Ilver, D., Rappuoli, R. and Telford, J.L. (2001). Protein export. In Helicobacter pylori: *Physiology and Genetics*, pp. 219-226. Edited by H.L.T. Mobley, G.L. Mendz and S.L. Hazell, S.L. Washington, D.C.: ASM Press.
- **Israel, D.A. (2001).** Genetic exchange. In Helicobacter pylori: *Physiology and Genetics*, pp. 313-319. Edited by H.L.T. Mobley, G.L. Mendz and S.L. Hazell, S.L. Washington D.C.: ASM Press.

- Israel, D.A., Salama, N., Arnold, C.N. and 8 other authors. (2001a). Helicobacter pylori strain-specific differences in genetic content, identified by microarray, influence host inflammatory responses. The Journal of Clinical Investigation 107, 611-620.
- Israel, D.A., Salama, N., Krishna, U., Rieger, U.M., Atherton, J.C., Falkow, S. and Peek, Jr., R.M. (2001b). *Helicobacter pylori* genetic diversity within the gastric niche of a single human host. *Proceedings of the National Academy of Sciences USA* 98, 14625-14630.
- Janvier, B., Grignon, B., Audibert, C., Pezennec, L. and Fauchère, J.L. (1999). Phenotypic changes of *Helicobacter pylori* components during experimental infection in mice. *FEMS Immunology and Medical Microbiology* **24**, 27-33.
- Jiang, Q., Hiratuka, K. and Taylor, D.E. (1996). Variability of gene order in different *Helicobacter pylori* strains contributes to genome diversity. *Molecular Microbiology* 20, 833-842.
- Josenhans, C., Eaton, K.A., Thevenot, T. and Suerbaum, S. (2000). Switching of flagellar motility in *Helicobacter pylori* by reversible length variation of a short homopolymeric sequence repeat in *fliP*, a gene encoding a basal body protein. *Infection and Immunity* 68, 4598-4603.
- Joyce, E.A., Bassler, B.L. and Wright, A. (2000). Evidence for a signalling system in *Helicobacter pylori*: detection of a *luxS*-encoded autoinducer. *Journal of Bacteriology* **182**, 3638-3643.
- Jungblut. P.R., Bumann, D., Haas, G., Zimny-Arndt, U., Holland, P., Lamer, S., Siejak, F., Aebischer, A. and Meyer, T.F. (2000). Comparative proteome analysis of *Helicobacter pylori*. *Molecular Microbiology* 36, 710-725.
- Kalantar, J., Eslick, G.D. and Talley, N.J. (1999). Chronic gastritis and nonulcer dyspepsia. *Current Topics in Microbiology and Immunology* **241**, 31-45.
- Kansau, I., Raymond, J., Bingen, E., Courcoux, P., Kalach, N., Bergeret, M., Braimi, N., Dupont, C. and Labigne, A. (1996). Genotyping of *Helicobacter pylori* isolates by sequencing of PCR products and comparison with the RAPD technique. *Research in Microbiology* 147, 661-669.
- Keenan, J., Day, T., Neal, S., Cook, B., Perez-Perez, G., Allardyce, R. and Bagshaw P. (2000a). A role for the bacterial outer membrane in the pathogenesis of *Helicobacter pylori* infection. *FEMS Microbiology Letters* 182, 259-264.
- Keenan, J., Oliaro, J., Domigan, N., Potter, H., Aitken, G., Allardyce, R. and Roake, J. (2000b). Immune response to an 18-kilodalton outer membrane antigen identifies lipoprotein 20 as a *Helicobacter pylori* vaccine candidate. *Infection and Immunity* 68, 3337-3343.
- Kersulyte, D., Chalkauskas, H. and Berg, D.E. (1999). Emergence of recombinant strains of *Helicobacter pylori* during human infection. *Molecular Microbiology* 31, 31-43.

- Kersulyte, D., Mukhopadhyay, A.K., Shirai, M., Nakazawa, T. and Berg, D.E. (2000). Functional organization and insertion specificity of IS607, a chimeric element of *Helicobacter pylori*. *Journal of Bacteriology* **182**, 5300-5308.
- Kersulyte, D., Velapatiño, B., Giedrius, D., Mukhopadhyay, A.K., Ito, Y., Cahuayme, L., Parkinson, A.J., Gilman, R.H. and Berg, D.E. (2002). Transposable element ISHp608 of Helicobacter pylori: nonrandom geographic distribution, functional organization, and insertion specificity. Journal of Bacteriology 184, 992-1002.
- Kim, J.-S., Chang, J.-H., Chung, S.-I. and Yum, J.-S. (2001). Importance of the host genetic background on immune responses to *Helicobacter pylori* infection and therapeutic vaccine efficacy. *FEMS Immunology and Medical Microbiology* 31, 41-46.
- Kotloff, K.L., Sztein, M.B., Wasserman, S.S., Losonsky, G.A., Di Lorenzo, S.C. and Walker, R.I. (2001). Safety and immunogenicity of oral inactivated whole-cell *Helicobacter pylori* vaccine with adjuvant among volunteers with or without subclinical infection. *Infection and Immunity* 69, 3581-3590.
- Krienitz, U. (1906). Ueber das Auftreten von Spirochaeten verschiedener Form im Mageninhalt bei Carcinoma ventriculi. Deutsche Medizinische Wochenschrift 32, 872.
- **Kuipers**, E.J. and Blaser, M.J. (1998). *Helicobacter pylori* and gastroduodenal disorders. In *Emerging Infections 1*, pp. 191-206. Edited by W.M. Scheld, D. Armstrong and J.M. Hughes. Washington, D.C.: ASM Press.
- Kuipers, E.J., Uyterlinde, A.M., Pena, A.S., Roosendaal, R., Pals, G., Nelis, G.F., Festen, H.P. and Meuwissen, S.G. (1995). Long-term sequelae of *Helicobacter pylori* gastritis. *Lancet* 345, 1525-1528.
- Kuipers, E.J., Israel, D.A., Kusters, J.G. and 8 other authors. (2000). Quasispecies development of *Helicobacter pylori* observed in paired isolates obtained years apart from the same host. *The Journal of Infectious Diseases* 181, 273-282.
- Labenz, J., Blum, A., Bayerdorffer, E., Meining, A., Stolte, M., Borsch, G. (1997). Curing *Helicobacter pylori* infection in patients with duodenal ulcer may provoke reflux oesophagitis. *Gastroenterology* 112, 1442-1447.
- Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of phage T4. *Nature* 227, 680-685.
- Lan, R. and Reeves, P.R. (2000). Intraspecies variation in bacterial genomes: the need for a species genome concept. *Trends in Microbiology* **8**, 396-401.
- Langenberg, W., Rauws, E.A.J., Widjojokusumo, A., Tytgat, G.N.J. and Zanen, H.C. (1986). Identification of *Campylobacter pyloridis* isolates by restriction endonuclease DNA analysis. *Journal of Clinical Microbiology* 24, 414-417.
- Lee, A. (1998). Animal models for host-pathogen interaction studies. *British Medical Bulletin* 54, 163-174.

Lee, A. (1999). Animal models of *Helicobacter* infection. *Molecular Medicine Today* 5, 500-501.

Le Gouill, C. and Dery, C.V. (1991). A rapid procedure for the screening of recombinant plasmids. *Nucleic Acids Research* 19, 6655.

Levin, B.R. and Bull, J.J. (1994). Short-sighted evolution and the virulence of pathogenic microorganisms. *Trends in Microbiology* 2, 76-81.

Levinson, G. and Gutman, G.A. (1987). Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Molecular Biology and Evolution* 4, 203-221.

Lin, L.-F., Posfai, J., Roberts, R.J. and Kong, H. (2001). Comparative genomics of the restriction-modification systems in *Helicobacter pylori*. *Proceedings of the National Academy of Sciences USA* 98, 2740-2745.

Lisitsyn, N., Lisitsyn, N. and Wigler, M. (1993). Cloning the differences between two complex genomes. *Science* 259, 946-951.

Logan, R.P.H. and Berg, D.E. (1996). Genetic diversity of *Helicobacter pylori*. Lancet 348, 1462-1463.

Lohman, T.M. (1992). Escherichia coli DNA helicases: mechanisms of DNA unwinding. Molecular Microbiology 6, 5-14.

Mahillon, J. and Chandler, M. (1998). Insertion sequences. *Microbiology and Molecular Biology Reviews* 62, 725-774.

Majewski, S.I.H. and Goodwin, C.S. (1988). Restriction endonuclease analysis of the genome of *Campylobacter pylori* with a rapid extraction method: evidence for considerable genomic variation. *The Journal of Infectious Diseases* 157, 465-471.

Malaty, H.M., Kim, J.G., Kim, S.D. and Graham, D.Y. (1996). Prevalence of *Helicobacter pylori* infection in Korean children: inverse relation to socioeconomic status despite a uniformly high prevalence in adults. *American Journal of Epidemiology* 143, 257-262.

Marchetti, M., Aricò, B., Burroni, D., Figura, N., Rappuoli, R. and Ghiara, P. (1995). Development of a mouse model of *Helicobacter pylori* infection that mimics human disease. *Science* 267, 1655-1658.

Marshall, B.J. and Warren, J.R. (1984). Unidentified curved bacilli in the stomach of patients with gastritis and peptic ulceration. *Lancet* i, 1311-1315.

Marshall, B.J., Armstrong, J.A., McGechie, D.B. and Glancy, R.J. (1985). Attempt to fulfil Koch's postulates for pyloric *Campylobacter*. *Medical Journal of Australia* 142, 436-439.

Marshall, D.G. (1996). Population analysis of *Helicobacter pylori*. Ph.D. Thesis, Trinity College, University of Dublin.

Marshall, D.G., Coleman, D.C., Sullivan, D.J., Xia, H., O'Morain, C.A. and Smyth, C.J. (1996). Genomic DNA fingerprinting of clinical isolates of *Helicobacter pylori* using short oligonucleotide probes containing repetitive sequences. *Journal of Applied Bacteriology* 81, 509-517.

Marshall, D.G., Dundon, W.G., Beesley, S.M. and Smyth, C.J. (1998). Helicobacter pylori - a conundrum of genetic diversity. Microbiology 144, 2925-2939.

Martin, S.L., Edbrooke, M.R., Hodgman, T.C., van den Eijnden, D.H. and Bird, M.I. (1997). Lewis X biosynthesis in *Helicobacter pylori*. Molecular cloning of an $\alpha(1,3)$ -fucosyltransferase gene. *The Journal of Biological Chemistry* 272, 21349-21356.

Martino, M.C., Stabler, R.A., Zhang, Z.W., Farthing, M.J.G., Wren, B.W. and Dorrell, N. (2001). *Helicobacter pylori* pore-forming cytolysin orthologue TlyA possesses *in vitro* hemolytic activity and has a role in colonization of the gastric mucosa. *Infection and Immunity* 69, 1697-1703.

Matic, I., Taddei, F. and Radman, M. (1996). Genetic barriers among bacteria. *Trends in Microbiology* 4, 69-73.

Maynard Smith, J., Smith, N.H., O'Rourke, M. and Spratt, B.G. (1993). How clonal are bacteria? *Proceedings of the National Academy of Sciences USA* 90, 4384-4388.

Mégraud, F. (1998). Antibiotic resistance in *Helicobacter pylori*. British Medical Bulletin 54, 207-216.

McGee, D.J. and Mobley, H.L.T. (1999). Mechanisms of *Helicobacter pylori* infection: bacterial factors. *Current Topics in Microbiology and Immunology* **241**, 155-180.

McGee, D.J., May, C.A., Garner, R.M., Himpsl, J.M. and Mobley, H.L.T. (1999). Isolation of *Helicobacter pylori* genes that modulate urease activity. *Journal of Bacteriology* 181, 2477-2484.

McGovern, K.J., Blanchard, T.G., Gutierrez, J.A., Czinn, S.J., Krakowaka, S. and Youngman, P. (2001). γ -Glutamyltransferase is a *Helicobacter pylori* virulence factor but is not essential for colonization. *Infection and Immunity* 69, 4168-4173.

McNulty, C.A.M. (1999). The discovery of Campylobacter-like organisms. Current Topics in Microbiology and Immunology, 241, 1-9.

Michetti, P., Kreiss, C., Kotloff, K.L. and 13 other authors. (1999). Oral immunization with urease and *Escherichia coli* heat-labile enterotoxin is safe and immunogenic in *Helicobacter pylori*-infected adults. *Gastroenterology* 116, 804-812.

Miehlke, S., Kibler, K., Kim, J.G., Figura, N., Small, S.M., Graham, D.Y. and Go, M.F. (1996). Allelic variation in the *cagA* gene of *Helicobacter pylori* obtained from Korea compared to the United States. *American Journal of Gastroenterology* 91, 1322-1325.

Miehlke, S., Genta, R.M., Graham, D.Y. and Go, M.F. (1999). Molecular relationships of *Helicobacter pylori* strains in a family with gastroduodenal disease. *American Journal of Gastroenterology* 94, 364-368.

Mitchell, H.M. (1999). The epidemiology of Helicobacter pylori. Current Topics in Microbiology and Immunology 241, 11-30.

Mobley, H.L.T. (2000). Ure-I mediated urea transport in *Helicobacter pylori*: an open and shut case? *Trends in Microbiology* **8**, 346-348.

Mobley, H.L.T., Cortesia, M.J., Rosenthal, L.E. and Jones, D. (1988). Characterization of urease from *Campylobacter pylori*. *Journal of Clinical Microbiology* 26, 831-836.

Montecucco, C., Papini, E., de Bernard, M. and Zoratti, M. (1999). Molecular and cellular activities of *Helicobacter pylori* pathogenic factors. *FEBS Letters* **452**, 16-21.

Montecucco, C., de Bernard, M., Papini, E. and Zoratti, M. (2001). Helicobacter pylori vacuolating cytotoxin: cell intoxication and anion-specific channel activity. Current Topics in Microbiology and Immunology 257, 113-129.

Monteiro, M.A., Appelmelk. B.J., Rasko, D.A. and 10 other authors. (2000). Lipopolysaccharide structures of *Helicobacter pylori* genomic strains 26695 and J99, mouse model *H. pylori* Sydney strain, *H. pylori* P466 carrying sialyl Lewis X, and *H. pylori* UA915 expressing Lewis B. Classification of *H. pylori* lipopolysaccharides into glycotype families. *European Journal of Biochemistry* 267, 305-320.

Moxon, E.R., Rainey, P.B., Nowak, M.A. and Lenski, R.E. (1994). Adaptive evolution of highly mutable loci in pathogenic bacteria. *Current Biology* 4, 24-33.

Mukhopadhyay, A.K., Kersulyte, D., Jeong, J.-Y. and 9 other authors. (2000). Distinctiveness of genotypes of *Helicobacter pylori* in Calcutta, India. *Journal of Bacteriology* 182, 3219-3227.

Murphy, C.A. and Belas, R. (1999). Genomic rearrangements in the flagellin genes of *Proteus mirabilis. Molecular Microbiology* 31, 679-690.

Namavar, F., Roosendaal, R., Kuipers, E.J., de Groot, P., van der Bijl, M.W., Pena, A.S. and de Graaff, J. (1995). Presence of *Helicobacter pylori* in the oral cavity, oesophagus, stomach and faeces of patients with gastritis. *European Journal of Clinical Microbiology and Infectious Diseases* 14, 234-237.

Nedenskov-Sørensen, P., Bukholm, G and Bøvre, K. (1990). Natural competence for genetic transformation in *Campylobacter pylori*. The Journal of Infectious Diseases 161, 365-366.

Negrini, R., Lisato, L., Zanella, I., Cavazzini, S., Gullini, S., Villanacci, V., Poiesi, C., Albertini, A. and Ghielmi, S. (1991). *Helicobacter pylori* infection induces antibodies cross-reacting with human gastric mucosa. *Gastroenterology* 101, 437-445.

Negrini, R., Savio, A., Graffeo, M., Rolfi, F. and Ghielmi, S. (1993). Autoantibodies and gastric *Helicobacter pylori* infection: does autoimmunity affect progression to atrophic gastritis? *European Journal of Gastroenterology and Hepatology* 5(Suppl 2), S27-S29.

Nikaido, H. (1994). Isolation of outer membranes. Methods in Enzymology 235, 225-234.

Nobusato, A., Uchiyama, I. and Kobayashi, I. (2000a). Diversity of restriction-modification gene homologues in *Helicobacter pylori*. Gene 259, 89-98.

Nobusato, A., Uchiyama, I., Ohashi, S. and Kobayashi, I. (2000b). Insertion with long target duplication: a mechanism for gene mobility suggested from comparison of two related bacterial genomes. *Gene* 259, 99-108.

Nwokolo, C.U., Bickley, J., Attard, A.R., Owen, R.J., Costas, M. and Fraser, I.A. (1992). Evidence of clonal variants of *Helicobacter pylori* in three generations of a duodenal ulcer disease family. *Gut* 33, 1323-1327.

Occhialini, A., Marais, A., Urdaci, M., Sierra, R., Muñoz, N., Covacci, A. and Mégraud, F. (2001). Composition and gene expression of the *cag* pathogenicity island in *Helicobacter pylori* strains isolated from gastric carcinoma and gastritis patients in Costa Rica. *Infection and Immunity* 69, 1902-1908.

Odenbreit, S., Till, M., Hofreuter, D., Faller, G. and Haas, R. (1999). Genetic and functional characterization of the *alpAB* gene locus essential for the adhesion of *Helicobacter pylori* to human gastric tissue. *Molecular Microbiology* 31, 1537-1548.

Odenbreit, S., Püls, J., Sedlmaier, B., Gerland, E., Fischer, W. and Haas, R. (2000). Translocation of *Helicobacter pylori* CagA into gastric epithelial cells by Type IV secretion. *Science* 287, 1497-1500.

Ohta-Tada, U., Takagi, A., Koga, Y., Kamiya, S. and Miwa T. (1997). Flagellin gene diversity among *Helicobacter pylori* strains and IL-8 secretion from gastric epithelial cells. *Scandinavian Journal of Gastroenterology* 32, 455-459.

O'Toole, G.A. and Kolter, R. (1998). Initiation of biofilm formation in *Pseudomonas fluorescens* WCS365 proceeds via multiple, convergent signalling pathways: a genetic analysis. *Molecular Microbiology* 28, 449-461.

Oudbier, J.H., Langenberg, W., Rauws, E.A.J. and Bruin-Mosch, C. (1990). Genotypical variation of *Campylobacter pylori* from gastric mucosa. *Journal of Clinical Microbiology* 28, 559-565.

Owen, R.J. (1998). *Helicobacter* – species classification and identification. *British Medical Bulletin* 54, 17-30.

Owen, R.J., Hunton, C., Bickley, J., Moreno, M. and Linton, D. (1992). Ribosomal RNA gene restriction patterns of *Helicobacter pylori*: analysis and appraisal of *Hae*III digests as a molecular typing system. *Epidemiology and Infection* 109, 35-47.

- Palmer, E.D. (1954). Investigation of the gastric spirochaetes of the human. Gastroenterology 27, 218-220.
- Pan, Z.-J., van der Hulst, R.W.M., Feller, M., Xiao, S.-D., Tytgat, G.N.J., Dankert, J. and van der Ende, A. (1997). Equally high prevalences of infection with *cagA*-positive *Helicobacter pylori* in Chinese patients with peptic ulcer disease and those with chronic gastritis-associated dyspepsia. *Journal of Clinical Microbiology* 35, 1344-1347.
- **Parsot, C. and Mekalanos, J.J. (1992).** Structural analysis of the *acfA* and *acfD* genes of *Vibrio cholerae*: effects of DNA topology and transcriptional activators on expression. *Journal of Bacteriology* 174, 5211-5218.
- Peck, B., Ortkamp, M., Diehl, K.D., Hundt, E. and Knapp, B. (1999). Conservation, localization and expression of HopZ, a protein involved in adhesion of *Helicobacter pylori*. *Nucleic Acids Research* 27, 3325-3333.
- Peek, R.M., Donahue, J.P., Tham, K.T., Thompson, S.A., Atherton, J.C., Blaser, M.J. and Miller, G.G. (1998). Adherence to gastric epithelial cells induces expression of a Helicobacter pylori gene, iceA, that is associated with clinical outcome. Proceedings of the Association of American Physicians 110, 531-544.
- Petes, T.D. and Hill, C.W. (1988). Recombination between repeated genes in microorganisms. *Annual Review of Genetics* 22, 147-168.
- Phadnis, S.H., Westblom, T.U. and Normark, S. (1993). Molecular cloning of *Helicobacter pylori* DNA: important differences between *mcrBC* deletion host strains. *Molecular Microbiology* 10, 1151.
- Phadnis, S.H., Ilver, D., Janzon, L., Normark, S. and Westblom, T.U. (1994). Pathological significance and molecular characterization of the vacuolating toxin gene of *Helicobacter pylori*. *Infection and Immunity* **62**, 1557-1565.
- **Preston, M.A. and Penner, J.L. (1987).** Structural and antigenic properties of lipopolysaccharides from serotype reference strains of *Campylobacter jejuni*. *Infection and Immunity* **55,** 1806-1812.
- Pytko-Polonczyk, J., Konturek, S.J., Karczewska, E., Bielanski, W. and Kaczmarczyk-Stachowska, A. (1996). Oral cavity as a permanent reservoir of *Helicobacter pylori* and potential source of reinfection. *Journal of Physiology and Pharmacology* 47, 121-129.
- Ramarao, N., Gray-Owen, S.D. and Meyer, T.F. (2000). *Helicobacter pylori* induces but survives the extracellular release of oxygen radicals from professional phagocytes using its catalase activity. *Molecular Microbiology* 38, 103-113.
- Rektorschek, M., Buhmann, A., Weeks, D., Schwan, D., Bensch, K.W., Eskandari, S., Scott, D., Sachs, G. and Melchers, K. (2000). Acid resistance of *Helicobacter pylori* depends on the Urel membrane protein and an inner membrane protein barrier. *Molecular Microbiology* 36, 141-152.

Reyrat, J.-M., Pelicic, V., Papini, E., Montecucco, C., Rappuoli, R. and Telford, J.L. (1999). Towards deciphering the *Helicobacter pylori* cytotoxin. *Molecular Microbiology* 34, 197-204.

Rhee, K.Y., Senear, D.F. and Hatfield, G.W. (1998). Activation of gene expression by a ligand-induced conformational change of a protein DNA complex. *The Journal of Biological Chemistry* 273, 11257-11266.

Robertson, B.D. and Meyer, T.F. (1992). Genetic variation in pathogenic bacteria. *Trends in Genetics* 8, 422-427.

Roche, R.J. and Moxon, E.R. (1995). Phenotypic variation of carbohydrate surface antigens and the pathogenesis of *Haemophilus influenzae* infections. *Trends in Microbiology* 3, 304-309.

Rota, A.C., Pereira-Lima, J.C., Blaya, C. and Beyer Nardi, N. (2001). Consensus and variable region PCR analysis of *Helicobacter pylori* 3' region of *cagA* gene in isolates from individuals with or without peptic ulcer. *Journal of Clinical Microbiology* 39, 606-612.

Roth, J.R., Benson, N., Galitski, T., Haack, K., Lawrence, J. and Miesel, L. (1996). Rearrangements of the bacterial chromosome: formation and applications. In Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology, pp. 2256-2276. Edited by F.C. Neidhardt. Washington D.C.: ASM Press.

Salama, N., Guillemin, K., McDaniel, T.K., Sherlock, G., Tompkins, L. and Falkow, S. (2000). A whole-genome microarray reveals genetic diversity among *Helicobacter pylori* strains. *Proceedings of the National Academy of Sciences USA* 97, 14668-14673.

Salomon, H. (1896). Über das spirillum des saugetiermagens und sein verhalten zu den belegzellen. Zentralblatt für Bakteriologie (Abteilung 1) 19, 433-422.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989). *Molecular Cloning: a Laboratory Manual*, 2nd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Sandkvist, M. (2001). Biology of type II secretion. Molecular Microbiology 40, 271-283.

Satin, B., Del Giudice, G., Della Bianca, V., Dusi, S., Laudanna, C., Tonello, F., Kelleher, D., Rappuoli, R., Montecucco, C. and Rissi, F. (2000). The neutrophil-activating protein (HP-NAP) of *Helicobacter pylori* is a protective antigen and a major virulence factor. *Journal of Experimental Medicine* 191, 1467-1476.

Saunders, N.J., Peden, J.F., Hood, D.W. and Moxon, E.R. (1998). Simple sequence repeats in the *Helicobacter pylori* genome. *Molecular Microbiology* 27, 1091-1098.

Saunders, N.J., Peden, J.F. and Moxon E.R. (1999). Absence in *Helicobacter pylori* of an uptake sequence for enhancing uptake of homospecific DNA during transformation. *Microbiology* **145**, 3523-3528.

- Saunders, N.J., Jeffries, A.C., Peden, J.F., Hood, D.W., Tettelin, H., Rappuoli, R. and Moxon, E.R. (2000). Repeat-associated phase variable genes in the complete genome sequence of *Neisseria meningitidis* strain MC58. *Molecular Microbiology* 37, 207-215.
- Scarlato, V., Delany, I., Spohn, G. and Beier, D. (2001). Regulation of transcription in *Helicobacter pylori*: simple systems or complex circuits. *International Journal of Medical Microbiology* **291**, 107-117.
- Schloter, M., Lebuhn, M., Heulin, T. and Hartmann, A. (2000). Ecology and evolution of bacterial microdiversity. *FEMS Microbiology Reviews* 24, 647-660.
- Schmitt, W. and Haas, R. (1994). Genetic analysis of the *Helicobacter pylori* vacuolating cytotoxin: structural similarities with the IgA protease type of exported protein. *Molecular Microbiology* 12, 307-319.
- Schmitt, W., Odenbreit, S., Heuermann, D. and Haas, R. (1995). Cloning of the *Helicobacter pylori recA* gene and functional characterisation of its product. *Molecular and General Genetics* 248, 563-572.
- Scott, D.R., Marcus, E.A., Weeks, D.L., Lee, A., Melchers, K. and Sachs, G. (2000). Expression of the *Helicobacter pylori ureI* gene is required for acidic pH activation of cytoplasmic urease. *Infection and Immunity* 68, 470-477.
- Segal, E.D., Cha, J., Lo, J., Falkow, S. and Tompkins, L.S. (1999). Altered states: involvement of phosphorylated CagA in the induction of host cellular growth changes by *Helicobacter pylori*. *Proceedings of the National Academy of Sciences USA* 96, 14559-14564.
- Shortridge, V.D., Stone, G.G., Flamm, R.K., Beyer, J., Versalovic, J., Graham, D.W. and Tanaka, S.K. (1997). Molecular typing of *Helicobacter pylori* isolates from a multicentre U.S. clinical trial by *ureC* restriction fragment length polymorphism. *Journal of Clinical Microbiology* 35, 471-473.
- Simor, A.E., Shames, B., Drumm, B., Sherman, P., Low, D.E. and Penner, J.L. (1990). Typing of *Campylobacter pylori* by bacterial DNA restriction endonuclease analysis and determination of plasmid profile. *Journal of Clinical Microbiology* 28, 83-86.
- Skouloubris, S., Thiberge, J., Labigne, A. and De Reuse, H. (1998). The *Helicobacter pylori* UreI protein is not involved in urease activity but is essential for bacterial survival in vivo. *Infection and Immunity* 66, 4517-4521.
- Slonczewski, J.L., McGee, D.J., Phillips, J., Kirkpatrick, C. and Mobley, H.L.T. (2000). pH-dependent protein profiles of *Helicobacter pylori* analyzed by two-dimensional gels. *Helicobacter* 5, 240-247.
- Smeets, L.C., Bijlsma, J.J.E., Boomkens, S.Y., Vandenbroucke-Grauls, C.M.J.E. and Kusters, J.G. (2000). *comH*, a novel gene essential for natural transformation of *Helicobacter pylori*. *Journal of Bacteriology* 182, 3948-3954.

- Sokurenko, E.V., Hasty, D.L. and Dykhuizen, D.E. (1999). Pathoadaptive mutations: gene loss and variation in bacterial pathogens. *Trends in Microbiology* 7, 191-195.
- Song, Q., Spahr, A., Schmid, R.M., Adler, G. and Bode, G. (2000). Helicobacter pylori in the oral cavity: high prevalence and great DNA diversity. Digestive Diseases and Sciences 45, 2162-2167.
- Southern, E.M. (1975). Detection of specific sequences in DNA fragments separated by gel electrophoresis. *Journal of Molecular Biology* 98, 503-517.
- Sozzi, M., Crosatti, M., Kim, S.-K., Romero, J. and Blaser, M.J. (2001). Heterogeneity of *Helicobacter pylori cag* genotypes in experimentally infected mice. *FEMS Microbiology Letters* 203, 109-114.
- Sperandio, V., Mellies, J.L., Nguyen, W., Shin, S. and Kaper, J.B. (1999). Quroum sensing controls expression of the type III secretion gene transcription and protein secretion in enterohemorrhagic and enteropathogenic *Escherichia coli*. *Proceedings of the National Academy of Sciences USA* 96, 15196-15201.
- **Spohn, G. and Scarlato, V. (1999).** Motility of *Helicobacter pylori* is coordinately regulated by the transcriptional activator FlgR, an NtrC homolog. *Journal of Bacteriology* **181**, 593-599.
- Spohn, G., Beier, D., Rappuoli, R. and Scarlato, V. (1997). Transcriptional analysis of the divergent *cagAB* genes encoded by the pathogenicity island of *Helicobacter pylori*. *Molecular Microbiology* **26**, 361-372.
- Stein, M., Rappuoli, R. and Covacci A. (2000). Tyrosine phosphorylation of the *Helicobacter pylori* CagA antigen after *cag*-driven host cell translocation. *Proceedings of the National Academy of Sciences USA* 97, 1263-1268.
- Straus, D. and Ausubel, F.M. (1990). Genomic subtraction for cloning DNA corresponding to deletion mutations. *Proceedings of the National Academy of Sciences USA* 87, 1889-1893.
- Suerbaum, S. and Achtman, M. (1999). Evolution of *Helicobacter pylori*: the role of recombination. *Trends in Microbiology* 7, 182-184.
- Suerbaum, S. (2000). Genetic variability within Helicobacter pylori. International Journal of Medical Microbiology 290, 175-181.
- Suerbaum, S., Maynard Smith, J., Bapumia, K., Morelli, G., Smith, N.H., Kuntsmann, E., Dyrek, I. and Achtman, M. (1998). Free recombination within *Helicobacter pylori*. Proceedings of the National Academy of Sciences USA 95, 12619-12624.
- Sutton, P. (2001). Progress in vaccination against Helicobacter pylori. Vaccine 19, 2286-2290.
- Sullivan, D.J., Westerneng, T.J., Haynes, K.A., Bennett, D.E. and Coleman, D.C. (1995). *Candida dubliniensis* sp. nov.: phenotypic and molecular characterization of a novel species associated with oral candidosis in HIV-infected individuals. *Microbiology* **141**, 1507-1521.

- Szczebara, F., Dhaenens, L., Armand, S. and Husson, M.O. (1999). Regulation of the transcription of genes encoding different virulence factors in *Helicobacter pylori* by free iron. *FEMS Microbiology Letters* 175, 165-170.
- Tang, X., Nakata, Y., Li, H.O., Zhang, M., Gao, H., Fujita, A., Sakatsume, O., Ohta, T. and Yokoyama, K. (1994). The optimization of preparations of competent cells for transformation of *E. coli. Nucleic Acids Research* 25, 2857-2858.
- **Taylor, D.E., Eaton, M., Chang, N. and Salama, S.M. (1992).** Construction of a *Helicobacter pylori* genome map and demonstration of diversity at the genome level. *Journal of Bacteriology* **174,** 6800-6806.
- Telford, J.L., Ghiara, P., Dell'Orco, M. and 11 other authors. (1994). Gene structure of the *Helicobacter pylori* cytotoxin and evidence of its key role in gastric disease. *Journal of Experimental Medicine* 179, 1653-1658.
- Thierry, D., Chavarot, P., Marchal, G., Le Thi, K.T., Ho, M.L., Nguyen, N.L., Le, N.V., Ledru, S., Fumox, F. and Guesdon, J.L. (1995). *Mycobacterium tuberculosis* strains unidentified using IS6110 probe can be detected by oligonucleotides derived from the Mt308 sequence. *Research in Microbiology* 146, 325-328.
- Thompson, S.A. and Blaser, M.J. (1995). Isolation of the *Helicobacter pylori recA* gene and involvement of the *recA* region in resistance to low pH. *Infection and Immunity* 63, 2185-2193.
- Tinsley, C.R. and Nassif, X. (1996). Analysis of the genetic differences between *Neisseria meningitidis* and *Neisseria gonorrhoeae*: two closely related bacteria expressing two different pathogenicities. *Proceedings of the National Academy of Sciences USA* 93, 11109-11114.
- Tomb, J.-F., White, O., Kerlavage, A.R. and 39 other authors. (1997). The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature* 388, 539-547.
- Tummuru, M.K.R., Cover, T.L. and Blaser, M.J. (1993). Cloning and expression of a high-molecular-mass major antigen of *Helicobacter pylori*: evidence of linkage to cytotoxin production. *Infection and Immunity* 61, 1799-1809.
- van Doorn, L.-J., Figueiredo, C., Sanna, R., Pena, S., Midolo, P., Ng, E.K., Atherton, J.C. and Blaser, M.J. (1998a). Expanding allelic diversity of *Helicobacter pylori vacA*. *Journal of Clinical Microbiology* 36, 2597-2603.
- van Doorn, L.-J., Figueiredo, C., Sanna, R., Plaisier, A., Schneeberger, P., de Boer, W. and Quint, W.G.V. (1998b). Clinical relevance of the cagA, vacA, and iceA status of Helicobacter pylori. Gastroenterology 115, 58-66.
- van Doorn, N.E.M., Namavar, F., Kusters, J.G., van Rees, E.P., Kuipers, E.J. and de Graaff, J. (1998). Genomic DNA fingerprinting of clinical isolates of *Helicobacter pylori* by REP-PCR and restriction fragment end-labelling. *FEMS Microbiology Letters* 160, 145-150.

- Wilson, G.G. and Murray, N.E. (1991). Restriction and modification systems. *Annual Review of Genetics* **25**, 585-627.
- Wirth, H.-P., Yang, M., Karita, M. and Blaser, M.J. (1996). Expression of the human cell surface glycoconjugates Lewis X and Lewis Y by *Helicobacter pylori* isolates is related to cagA status. *Infection and Immunity* 64, 4598-4605.
- Wirth, H.-P., Yang, M., Peek, R.M., Höök-Nikanne, J., Fried, M. and Blaser, M.J. (1999). Phenotypic diversity in Lewis expression of *Helicobacter pylori* isolates from the same host. *Journal of Laboratory and Clinical Medicine* 133, 488-500.
- Withers, H., Swift, S. and Williams, P. (2001). Quorum sensing as an integral component of gene regulatory networks in Gram-negative bacteria. *Current Opinion in Microbiology* 4, 186-193.
- Worst, D.J., Otto, B.R. and de Graaff, J. (1995). Iron-repressible outer membrane proteins of *Helicobacter pylori* involved in heme uptake. *Infection and Immunity* 63, 4161-4165.
- **Xiang, Z., Bugnoli, M., Ponzetto, A. and 8 other authors. (1993).** Detection in an enzyme immunoassay of an immune response to a recombinant fragment of the 128-kDa protein (CagA) of *Helicobacter pylori*. European Journal of Clinical Microbiology and Infectious Diseases 12, 739-745.
- Xiang, Z., Censini, S., Bayeli, P.F., Telford, J.L., Figura, N., Rappuoli, R. and Covacci, A. (1995). Analysis of expression of CagA and VacA virulence factors in 43 strains of *Helicobacter pylori* reveals that clinical isolates can be divided into two major types and that CagA is not necessary for expression of the vacuolating cytotoxin. *Infection and Immunity* 63, 94-98.
- Xu, Q., Peek, R.M., Miller, G.G. and Blaser, M.J. (1997). The *Helicobacter pylori* genome is modified at CATG by the product of *hpyIM*. *Journal of Bacteriology* 179, 6807-6815.
- Xu, Q., Stickel, S., Roberts, R.J., Blaser, M.J., Morgan, R.D. (2000a). Purification of the novel endonuclease, *Hpy*1881, and cloning of its restriction-modification genes reveal evidence of its horizontal transfer to the *Helicobacter pylori* genome. *The Journal of Biological Chemistry* 275, 17086-17093.
- Xu, Q., Morgan, R.D., Roberts, R.J. and Blaser, M.J. (2000b). Identification of type II restriction and modification systems in *Helicobacter pylori* reveals their substantial diversity among strains. *Proceedings of the National Academy of Sciences USA* 97, 9671-9676.
- Xu, Q. and Blaser, M.J. (2001). Promoters of the CATG-specific methyltransferase gene hpyIM differ between iceA1 and iceA2 Helicobacter pylori strains. Journal of Bacteriology 183, 3875-3884.
- Yamaoka, Y., Kodama, T., Kashima, K., Graham, D.Y. and Sepulveda, A.R. (1998). Variants in the 3' region of the *cagA* gene in *Helicobacter pylori* isolates from patients with different *H. pylori*-associated diseases. *Journal of Clinical Microbiology* 36, 2258-2263.

Yamaoka, Y., El-Zimaity, H.M.T., Guiterrez, O., Figura, N., Kim, J.K., Kodama, T., Kashima, K, and Graham, D.Y. (1999a). Relationship between the *cagA* 3' repeat region of *Helicobacter pylori*, gastric histology, and susceptibility to low pH. *Gastroenterology* 117, 342-349.

Yamaoka, Y., Kodama, T., Gutierrez, O., Kim, J.G., Kashima, K. and Graham, D.Y. (1999b). Relationship between *Helicobacter pylori iceA*, cagA, and vacA status and clinical outcome: studies in four different countries. *Journal of Clinical Microbiology* 37, 2274-2279.

Yamaoka, Y., Kwon, D.H. and Graham, D.Y. (2000). A M_r 34,000 proinflammatory outer membrane protein (oipA) of Helicobacter pylori. Proceedings of the National Academy of Sciences USA 97, 7533-7538.

Ye, D. and Blanke, S.R. (2000). Mutational analysis of the *Helicobacter pylori* vacuolating toxin amino terminus: Identification of amino acids essential for cellular vacuolation. *Infection and Immunity* **68**, 4354-4357.

Ziebuhr, W., Ohlsen, K., Karch, H., Korhonen, T. and Hacker, J. (1999). Evolution of bacterial pathogenesis. *Cellular and Molecular Life Sciences* 56, 719-728.