LEABHARLANN CHOLÁISTE NA TRÍONÓIDE, BAILE ÁTHA CLIATH | TRINITY COLLEGE LIBRARY DUBLIN
Ollscoil Átha Cliath | 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.

Genetic Diversity of *Helicobacter pylori* Isolates through Microevolution *in vivo*

A thesis submitted for the degree of Doctor of Philosophy

by

Ian M. Carroll

Moyne Institute of Preventive Medicine,

Department of Microbiology,

University of Dublin,

Trinity College.

October 2003

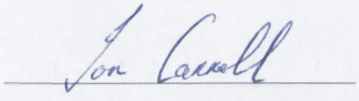
TRINITY COLLEGE
27 FEB 2004
LIBRARY DUBLIN
THESIS
7761

For Marie

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.



Ian M. Carroll

Contents

Summary. I

Acknowledgements. III

Publications and Conference Communications. V

List of figures. VI

List of tables. XI

Abbreviations. XII

Chapter 1: General introduction. 4

Chapter 2: Materials and methods. 39

Chapter 3: The population structure of Irish *Helicobacter pylori* isolates – Genotyping putative virulence genes and intergenic regions within Irish *H. pylori* isolates. 57

Chapter 4: The genetic diversity of Irish *Helicobacter pylori* isolates as revealed by fluorescent amplified length polymorphism (FAFLP) – PCR. 74

Chapter 5: Microevolution between paired *Helicobacter pylori* isolates taken from the same patients as revealed by genotyping and FAFLP-PCR. 88

Chapter 6: Intrafamilial spread of *H. pylori*: A study of two families exhibiting clonal transmission of *H. pylori*. 108

Chapter 7: Microevolution at the 3′-end of the *cagA* gene between paired *Helicobacter pylori* isolates alters the number of Tyrosine Phosphorylation Motifs (TPMs) located at this locus. 121

Chapter 8: General discussion. 141

Bibliography. 152

Summary

Chronic infection with *H. pylori* causes peptic ulcer disease and its presence is strongly linked with gastric adenocarcinoma and mucosa-associated lymphoid tissue (MALT) lymphoma. One of the most interesting characteristics of *H. pylori* is its genetic diversity. This organism has been shown to have a panmicitic or freely recombining population structure. Indeed, unrelated clinical *H. pylori* isolates display disparate DNA fingerprints. The genetic diversity of *H. pylori* isolates has been shown to vary within different parts of the world. Moreover, recent studies have demonstrated geographical partitioning of *H. pylori* isolates and have detected the presence of weakly clonal groupings. In addition, a lack of diversity has been demonstrated between the two fully genome-sequenced *H. pylori* strains (26695 and J99). Thus, the extent of genetic diversity within the *H. pylori* population structure may have been overstated. The present study aimed to characterise the population structure of clinical Irish *H. pylori* isolates and compare this structure with those of different geographical regions.

Genotyping of 74 Irish *H. pylori* isolates was preformed by polymerase chain reaction (PCR) at seven different loci [*vacA* signal sequence and mid-region, presence of the *cagA* gene and the number of repeat sequences located at the 3′-end of this gene, insertion-deletion (in–del) polymorphisms at the 3′-end of the *cag* pathogenicity island (*cag* PAI), and the non-coding regions upstream of the *ribA* and *vacA* genes]. Two *vacA* allelic combinations (s1a m1 and s1a m2) were represented in ~96% of these isolates, all isolates contained the *cagA* gene, ~90% of isolates possessed one or two repeat sequences at the 3′-end of the *cagA* gene, ~84% had either type II or IV in-del motifs and ~90% of isolates contained the RibAP1 and VacAP3 sequence variants upstream of their *ribA* and *vacA* genes, respectively. The predominant v*acA* alleles and in-del motifs suggested an ancestral relationship between Irish isolates and either specific East Asian or Northern European strains. In addition, fluorescent amplified fragment length polymorphism-PCR (FAFLP-PCR) genotyping and phylogenetic analysis of 32 representative Irish *H. pylori* isolates and 22 isolates from four different continents demonstrated that the Irish *H. pylori* isolates examined were relatively homogeneous (weakly clonal) and showed some genetic relationship with both European and Asian isolates. These genotyping techniques showed that Irish *H. pylori* isolates have distinctive features that may have evolved in this insular European population.

Chronic infection with *H. pylori* gives rise to micro-evolved strains or quasi-species. Microevolution is determined by fingerprinting paired *H. pylori* isolates obtained from the

same individual, manifesting as small differences in these DNA fingerprint profiles between strains. As *H. pylori* exhibits high mutation and recombination levels, it is believed that as environmental changes occur in the stomach, sub-clones of the founder strain will arise that are better suited to this new environment. These sub-clones exhibit a genetic makeup that is highly similar but not identical to the progenitor strain they evolved from. Understanding the genetic differences between micro-evolved *H. pylori* strains could potentially reveal how this organism persists in the gastric niche for decades.

To this end a collection of pairs of *H. pylori* isolates obtained from the same individuals were investigated from the presence of microevolution. The sequence variations located at the seven specific loci previously used to characterise a population of Irish isolates in the present study were determined by PCR in 31 pairs of antral or antral and corpus *H. pylori* isolates. Random amplified polymorphic DNA-PCR (RAPD-PCR) and FAFLP-PCR fingerprinting were also applied to these paired clinical isolates. Twenty-two pairs were demonstrated to exhibit microevolution as revealed by FAFLP-PCR. Five of these 22 pairs displayed differences in the number of repeat sequences at the 3′-end of the *cagA* gene.

Attempts to determine the tyrosine phosphorylation potential of the repeat region at the C-terminal end of the CagA protein within two pairs of *H. pylori* isolates that exhibited differing repeat sequence number at the 3′-end of the *cagA* gene were undertaken. Although the repeat domains were successfully cloned into a Semliki Forest virus mammalian expression vector and transfected into AGS cells, phosphorylation of the putatively expressed CagA polypeptides could not be confirmed. However, it was demonstrated that variation in the number of transfected repeat sequences between ancestrally related CagA proteins altered the phosphorylation status of host proteins.

In addition, analysis of *H. pylori* familial isolates revealed that clonal transmission of strains had occurred between siblings and between a mother and her daughter by FAFLP-PCR and phylogenetic analysis. An interesting finding was that two isolate pairs from siblings in one family that displayed the same RAPD-PCR and FAFLP-PCR fingerprints displayed different signal sequence and upstream variants of the *vacA* gene, strongly suggesting that a recombination event had occurred. This finding represents the first report of a change in 5′-*vacA* genotype characteristics of *H. pylori* during infection. This demonstrates an adaptive change in an *H. pylori* virulence factor gene upon colonisation of a new host.

Acknowledgements

First and foremost I would like to thank Prof. Cyril J. Smyth for giving me the opportunity to carry out a Ph.D. in his laboratory and for his support and guidance over the last four years. It has been a pleasure to work with Cyril. I could not have asked for a better supervisor.

I would also like to thank Prof. Colm Ó'Moráin of the National Children's Hospital, Tallagh for providing biopsies and financial support that were necessary for me to complete this research project. I also acknowledge the Health Research Board of Ireland for the award of studentship and research funding that were essential for completion of my Ph.D.

I do not think I can adequately express my thanks to the past and present members of the Pathogenicity Laboratory. I am indebted to Dr. Sarah Beesley who taught me everything I know about *Helicobacter pylori*. I can honestly say that I would not have lasted this long if it was not for Sarah. I know that we will always share a love for temperamental Gram-negative, spiral-shaped bacteria. I would also like to thank Jane Twohig who gave steadfast support and encouragement during times when it was really needed, thanks Jane! In the Gram-positive corner I wish to acknowledge Davida Smyth (sometimes known as David Smith). I don't think I have ever met someone who had more faith in me as a scientist; I hope you have a bright future in bacteriology, thank you. In addition, I would like to thank Dr. William Dundon and Dr. Ross Fitzgerald for their helpful suggestions throughout my Ph.D.

Many thanks to everyone in the Prep. Room for their hard work and for being great bunch of people to work with.

I would like to thank the members of my thesis committee (Prof. Charles Dorman, Prof. Peter Owen, Prof. Tim Foster and Dr. Gus Bell) for many helpful suggestions.

It is impossible to pass through the Moyne Institute without making many good friends. I would like to thank Jamie (monkey), Paul (Pablo), Brian ('the young fella'), Jennie, Ronan ('beta barrel'), Sorcha, Christopher, Alan, Padraig, Clare, Fiona, Blanca, Gowda and everyone else in the department. I would also like to thank my friends outside the department - Al, Carl (Grayzer), Tony (Mr. T), Fin, Gab, Lee and Paddy - for their support and interest over the last four years.

I must acknowledge Dr. Niyaz Ahmed and the rest of the Pathogen Evolution Group in the Centre for DNA Fingerprinting and Diagnostics (CDFD) in Hydrabad (India) for providing me with the opportunity to work with them in their lab. I received a warm welcome and had the chance the experience a new culture which I am grateful for.

I would not be where I am today if it was not for the support I received from my parents and brother. For this I give special thanks to Mam (Maura), Dad (Seán) and Paul. I hope you enjoy reading my work.

Finally, I could not finish without thanking my girlfriend. Marie's support and belief in me has never wavered even when I didn't believe in myself and for this reason I dedicate this thesis to her with all my love.

Publications

Carroll, I. M., N. Ahmed, S. M. Beesley, A. A. Khan, S. Ghousunnissa, C. A. O Moráin and C. J. Smyth. Fine Structure Molecular Typing of Irish *Helicobacter pylori* Isolates and their Genetic Relatedness to Strains from Four Different Continents. *Journal of Clinical Microbiology* **41** : **5775 - 5759**.

Carroll, I. M., N. Ahmed, S. M. Beesley, A. A. Khan, S. Ghousunnissa, C. A. O Moráin and C. J. Smyth. Microevolution between paired antral and paired antral and corpus *Helicobacter pylori* isolates recovered from individual patients. (Submitted to the *Journal of Medical Microbiology*).

Carroll, I. M., N. Ahmed and A. A. Khan. Revisiting the pestilence of *Helicobacter pylori*: insights into pathogen evolution and geographical genomics. *Infection, Genetics and Evolution* **in press.**

Conference communications

Carroll, I.M., O Moráin, C.A. and Smyth, C.J. (2002). Genotyping of coding and non-coding regions of Irish *Helicobacter pylori* isolates. *Gut* **49**, Supplement 11: A16 (abstract 3/16). [XIVth International Workshop on Gastroduodenal Pathology and *Helicobacter pylori*, European *Helicobacter pylori* Study Group, September 2001, Strasbourg, France, Poster Presentation]

Carroll, I.M., Smyth, C.J. and O Moráin, C.A. (2001). Genotyping of putative virulence factors and intergenic regions of a collection of Irish *Helicobacter pylori* isolates. *International Journal of Medical Microbiology* **291**, Supplement 31: 32 (abstract D-14). [11th International Workshop on *Campylobacter*, *Helicobacter* and Related Organisms, where, September 2001, Freiburg, Germany, Poster Presentation].

List of figures. **After page**

Fig. 1.1. Spectrum of *H. pylori*-associated diseases. 7

Fig. 1.2. Current strategy for eradication of *H. pylori* Infection based on the Maastricht 2-2000 Consensus Report. 15

Fig. 2.1. Total cellular RNA from *H. pylori* isolates run on a 1.8% agarose gel. 48

Fig. 3.1. Schematic representation of the mosaic structure of the *H. pylori vacA* gene. 60

Fig. 3.2. Diagram depicting the structural differences between European and East Asian *H. pylori* isolates at the 3′-end of the *cagA* gene. 61

Fig. 3.3. Diagram of insertion-deletion (in-del) motif types located at the right end of the *cag* PAI (adapted from Kersultye *et al.*, 2000). 61

Fig. 3.4. Diagram of intergenic diversity occurring upstream of the *vacA* and *ribA* genes. 62

Fig. 3.5. PCR products of the signal-sequence and mid-region alleles detected in this collection of Irish *H. pylori* isolates separated on a 1.8% agarose gel. 63

Fig. 3.6. Alignment and nucleotide sequence of the s1 *vacA* signal sequence allele detected in this collection of Irish *H. pylori* isolates. 64

Fig. 3.7. Alignment and nucleotide sequence of the s2 *vacA* signal sequence allele detected in this collection of Irish *H. pylori* isolates. 64

Fig. 3.8. Alignment and nucleotide sequence of the s1a *vacA* signal sequence allele detected in this collection of Irish *H. pylori* isolates. 64

Fig. 3.9. Alignment and nucleotide sequence of the m1 *vacA* mid-region allele detected in this collection of Irish *H. pylori* isolates. 64

Fig. 3.10. Alignment and nucleotide sequence of the m2 *vacA* mid-region allele detected in this collection of Irish *H. pylori* isolates. 64

Fig. 3.11. Alignment of the s1 and s2 *vacA* signal-sequence alleles detected in this collection of Irish *H. pylori* isolates. 64

Fig. 3.12. Alignment of the m1 and m2 *vacA* mid-region alleles detected in this collection of Irish *H. pylori* isolates. 64

Fig. 3.13. PCR products run on a 1.8% agarose gel for detection of the presence of *cagA* and the number of repeat sequences located at the 3′- end of the *cagA* gene. 64

Fig. 3.14. Nucleotide and amino acid sequences of a the 3′-end of the *cagA* gene of an Irish *H. pylori* isolate containing two repeat sequences. 64

After page

Fig. 3.15. Neighbour-joining phylogenetic tree displaying the relatedness of *H. pylori* isolates from various geographical locations based on the sequences of the 3′-end of the *cagA* gene. 65

Fig. 3.16. Insertion-deletion motifs located at the 3′-end of the *cag* PAI found in this collection of Irish *H. pylori* isolates. 66

Fig. 3.17. PCR products of the intergenic regions preceding the *vacA* gene run on a 1.8% agarose gel. 67

Fig. 3.18. PCR products of the intergenic regions preceding the *ribA* gene run on a 1.8% agarose gel. 67

Fig. 3.19. Partial nucleotide sequences of the sequence variants preceding the *ribA* gene found in this collection of Irish *H. pylori* isolates. 67

Fig. 3.20. Aligned nucleotide sequences of the four sequence variants found preceding the *ribA* gene in this collection of Irish *H. pylori* isolates. 67

Fig. 3.21. Geographical distribution of *H. pylori vacA* alleles throughout the world. 68

Fig. 3.22. Geographical distribution of *H. pylori* in-del motifs throughout the world. 71

Fig. 4.1. Diagram depicting the procedure for Fluorescent Amplified Length (FAFLP)-PCR fingerprinting. 81

Fig. 4.2. FAFLP-PCR profiles of 12 Irish *H. pylori* isolates. 84

Fig. 4.3. FAFLP-PCR profiles of 34 Irish *H. pylori* isolates using non-selective primer pairs. 84

Fig. 4.4. Genescan™ FAFLP-PCR profiles of three different *H. pylori* isolates. 85

Fig. 4.5. Two superimposed FAFLP-PCR profiles generated using different selective primers. 85

Fig. 4.6. Comparison of FAFLP-PCR profiles of four unrelated *H. pylori* isolates using the same selective primer. 85

Fig. 4.7. Neighbour-joining phylogenetic tree of 32 Irish *H. pylori* isolates. 85

Fig. 4.8. Comparison of FAFLP-PCR profiles of four *H. pylori* isolates from four different geographical regions using the same selective primer. 85

Fig. 4.9. Comparison of the FAFLP-PCR profiles of an Irish and Japanese *H. pylori* isolate using the same selective primer. 85

Fig. 4.10. Neighbour-joining phylogenetic tree displaying the relatedness of *H. pylori* isolates from four different continents. 85

After page

Fig. 5.1. Amplification of the 3′-end of the *cagA* gene in nine paired *H. pylori* isolates showing differences in the number of repeat sequences. 94

Fig. 5.2. RAPD-PCR fingerprint profiles of six putatively related pairs of *H. pylori* isolates based on specific loci genotyping. 95

Fig. 5.3. RAPD-PCR profiles of the nine paired *H. pylori* isolates showing variations in the number of repeat sequences at the 3′-end of the *cagA* gene within each pair and putatively deemed to be unrelated. 95

Fig. 5.4. Southern blot of *Hin*dIII-restricted genomic DNA from paired *H. pylori* isolates MI458 and MI459 using six different probes. 96

Fig. 5.5. Southern blot of *Hin*dIII-restricted genomic DNA from one ancestrally related pair of *H. pylori* isolates and one unrelated pair of isolates using a *vacA* probe. 96

Fig. 5.6. *napA* nucleotide sequences of paired *H. pylori* isolates MI355 and MI356 (s1a m2, VacAP3, in-del type IV). 96

Fig. 5.7. Alignment of *napA* sequences from paired *H. pylori* isolates MI355 and MI356. 96

Fig. 5.8. FAFLP -PCR fingerprints of 12 *H. pylori* paired isolates. 97

Fig. 5.9. Electropherogram of two paried *H. pylori* isolates with the selective primer + G. 97

Fig. 5.10. Electropherogram of paried isolates 15500A / 15500C using two different selective primers (+ G and + A). 97

Fig. 5.11. Neighbour-joining tree developed from the binary data obtained by genotypic comparisons by the Genotyper™ macro. 97

Fig. 5.12. Western immuno-blot analysis using anti-VacA antibodies. 98

Fig. 5.13. Reverse transcription polymerase chain reaction (RT-PCR) analysis of *H. pylori vacA* mRNA transcripts. 98

Fig. 6.1. RAPD-PCR fingerprints of *H. pylori* isolates from Family 1 using primer PIII fractionated on a 1.8% agarose gel. 114

Fig. 6.2. RAPD-PCR fingerprints of *H. pylori* isolates from Family 1 using primer PIV fractionated on a 1.8% agarose gel. 114

Fig. 6.3. RAPD-PCR fingerprints of *H. pylori* isolates from Family 2 using primer PI fractionated on a 1.8% agarose gel. 114

Fig. 6.4. RAPD-PCR fingerprints of *H. pylori* isolates from Family 2 using primer PIII fractionated on a 1.8% agarose gel. 114

After page

Fig. 6.5. Rectangular cladogram neighbour-joining phylogenetic tree of *H. pylori* isolates from Family 1 based on similarities of FAFLP-PCR fingerprints. 115

Fig. 6.6. Radial neighbour-joining phylogenetic tree of *H. pylori* isolates from Family 1 based on similarities of FAFLP-PCR fingerprints. 116

Fig. 6.7. Rectangular cladogram neighbour-joining phylogenetic tree of *H. pylori* isolates from Family 2 based on similarities of FAFLP-PCR fingerprints. 116

Fig. 6.8. Radial neighbour-joining phylogenetic tree of *H. pylori* isolates from Family 2 based on similarities of FAFLP-PCR fingerprints. 116

Fig. 7.1. Model for the regulation of CagA–SHP-2 signalling by Csk (adapted from Tsutsumi *et al.*, 2003). 125

Fig. 7.2. Schematic representation of the Semliki Forest Virus replicative cycle. 126

Fig. 7.3. Schematic diagram of how the pSFV1 vector can be used to produce heterologous proteins in mammalian cells. 126

Fig. 7.4. Strategy used to determine the phosphorylation potential of *cagA* repeats. 126

Fig. 7.5. Alignment of cloning primers with the 3′-end of the *cagA* gene of a European *H. pylori* isolate. 127

Fig. 7.6. Modified repeat sequences from two paired ancestrally related *H. pylori* isolates separated on a 1.8% agarose gel. 128

Fig. 7.7. Map of pSFV1 viral vector.1 128

Fig. 7.8. Digested and purified pSFV1 vector and amplified *cagA* repeats run on a 1.8% agarose gel. 128

Fig. 7.9. A. and B. Amplification of inserts from cloned and empty vectors run on a 1.8% agarose gel. 128

Fig. 7.10. PCR strategy to determine the right orientation of pSFV1 clones. 129

Fig. 7.11. PCR products run on a 1.8% agarose to confirm orientation of clones. 129

Fig. 7.12. Nucleotide and translated amino acid sequences and alignment of amino acid sequences for clones SFV512 and SFV541. 129

Fig. 7.13. Nucleotide and translated amino acid sequences and alignment of amino acid sequences for clones SFV45688A and SFV45688C. 129

Fig. 7.14. Partial digestion of pSFV1 clones with *Spe*I run on a 1% agarose gel. 131

Fig. 7.15. *in vitro* transcripts of clones separated on a 1% agarose gel. 131

After page

Fig. 7.16. Total cellular protein of AGS cells, electroporated with or without recombinant pSFV1 mRNA, fractionated on a 15% polyacrylamide gel by SDS-PAGE. 133

Fig. 7.17. Anti-PY Western blot of AGS cells transfected with and without recombinant pSFV1 mRNA. 133

Fig. 7.18. Polyprotein 1234 cleavage sites of nsP2 cysteine endopeptidases (proteinases) of alphaviruses. 139

Fig. 8.1. Graph indicating the % of Irish nationals within the total population of the Republic of Ireland from 1991–2002. 146

Fig. 8.2. Graph of the % non-Irish residents within the total population of the Republic of Ireland from 1991–2002. 146

Fig. 8.3. Histograms indicating the % of immigrants from different countries within the total population of the Republic of Ireland in 2002. 146

Fig. 8.4. Schematic representation of the probable causes of a relatively homogenous Irish *H. pylori* population. 147

List of tables. **After page**

Table 1.1. Members of the *Helicobacter* species and their natural hosts. 5

Table 2.1. Primers for genotyping used in this study. 45

Table 2.2. Genes analysed by Southern blotting and primers used to generate probes. 47

Table 2.3. RAPD-PCR fingerprinting primers used in this study. 48

Table 3.1. *H. pylori* genotypes and their distribution throughout the world. 60

Table 3.2. Primers for genotyping used in this study. 63

Table 3.3. Frequencies of genotypes or sequence variants found in this collection of Irish *H. pylori* isolates. 64

Table 3.4. Strain and genotype of each *H. pylori* isolate used in this study. 64

Table 4.1. Thermal cycler parameters for selective amplification. 81

Table 4.2. Irish *H. pylori* isolates analysed by FAFLP-PCR with *vacA* allelic and *cag* PAI types indicated. 84

Table 4.3. Strains from different geographical locations and the cluster subgroup they belonged to as revealed by FAFLP. 84

Table 5.1. Paired *H. pylori* isolates used in this study with various genotypes, ancestral relationship and presence of microevolution indicated. 93

Table 6.1. Familial *H. pylori* isolates used in this study. 113

Table 6.2. The *vacA* types and the numbers of *cagA* 3′-end repeat sequences indicated of familial isolates used in this study. 113

Abbreviations

ASY	Asymptomatic gastritis
bp	Base pair
CagA	Cytotoxin associated protein A
CLO	*Campylobacter*-like organism
CTAB	Cetyltrimethyl ammonium bromide
dATP	Deoxyadenosine triphosphate
dCTP	Deoxycytosine triphosphate
dGTP	Deoxyguanosine triphosphate
dNTP	Deoxynucleotide triphosphate
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
DU	Duodenal ulcer
EDTA	Ethylenediamine-tetraacetic acid
g	Gram
G+C%	Percentage guanosine and cytosine
GU	Gastric ulcer
h	Hour
kbp	Kilobase pairs
kDa	Kilodalton
LPS	Lipopolysaccharide
M	Molar
MALT	Mucosa-associated lymphoid tissue
Mb	Megabase
µ	Micro
µg	Microgram
µl	Microlitre
µmol	Micromole
mg	Milligram
min	Minute
ml	Millilitre
NCTC	National Collection of Type Cultures
NUD	Non-ulcer dyspepsia

PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PUD	Peptic ulcer disease
RAPD	Random amplified polymorphic DNA
RFLP	Restriction fragment length polymorphism.
RNA	Ribonucleic acid acid
r.p.m.	Revolutions per minute
RT	Room temperature
s	Seconds
SDS	Sodium dodecyl sulphate
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
TBE	Tris borate buffer
TBS	Tris-buffered saline
TE	Tris EDTA buffer
vol.	Volume
w/v	Weight per volume

Chapter 1

General introduction.

1.1. **History of *Helicobacter pylori*** 4

1.2. **Taxonomy** 5

1.3. **Bacteriology** 5

1.4. **Epidemiology** 6

1.5. ***H. pylori* reservoirs** 6

1.6. **The disease spectrum of *H. pylori*** 7

1.6.1. *Gastritis* 7

1.6.2. *Peptic ulceration* 8

1.6.2.1. Duodenal ulceration 8

1.6.2.2. Gastric ulceration 9

1.6.3. *Gastric atrophy* 9

1.6.4. *Gastric cancer* 10

1.6.4.1. Gastric adenocarcinoma 10

1.6.4.2. Mucosa-associated lymphld tissue (MALT) lymphoma 11

1.6.5. *Non-ulcer dyspepsia* 11

1.6.6. H. pylori *infection in other organ systems* 12

1.6.7. *Are* H. pylori *infections beneficial?* 12

1.6.8. in vivo *modelling of* H. pylori 14

1.6.9. *Treatment for* H. pylori *infection* 14

1.6.9.1. Antimicrobial treatment 15

1.6.9.2. *H. pylori* vaccines 15

1.7. **Pathogenicity determinants of *H. pylori*** 16

1.7.1. *Virulence factors (determinants)* 17

1.7.1.1. Flagella 17

1.7.1.2. Urease 18

1.7.1.3. *H. pylori* neutrophil-activating protein (HP-NAP) 19

1.7.1.4. γ-Glutamyltransferase 20

1.7.1.5. Iron-acquisition proteins 21

1.7.1.6. Adhesins 21

1.7.1.7. Heat shock proteins in *H. pylori* 22

1.7.2. *Virulence determinants in a subset of* H. pylori *strains* 23

1.7.2.1. The vacuolating cytotoxin (VacA) 23

1.7.2.2. The cytotoxin-associated antigen (CagA) 24

1.7.3. *Novel virulence determinants* 25

1.7.3.1. Superoxide dismutase and catalase 25
1.7.3.2. *H. pylori* pore-forming cytolysin (TlyA) 25
1.7.3.3. The *H. pylori iceA* locus 26
1.8. Genetic diversity of *H. pylori* 26
1.8.1. Bacterial population genetics 27
1.8.1.1. Population structure of *H. pylori* 27
1.8.1.2. Weakly clonal population structure 28
1.8.1.3. Geographical partitioning of *H. pylori* isolates 29
1.8.2. Mechanisms and origins of genetic diversity in H. pylori 30
1.8.2.1. Natural mutation in *H. pylori* 31
1.8.2.2. Recombination in *H. pylori* 31
1.8.2.2.1. Horizontal gene transfer 32
1.8.2.2.2. Intra-strain diversity in H. pylori 33
1.8.3. Genetic diversity in H. pylori *revealed by comparative and functional genomics* 34
1.8.4. Quasi-species development 35
1.8.5. Genetic diversity of H. pylori *via microevolution* 36
1.9. Project aims 37

1.1. History of *Helicobacter pylori.*

It has almost been twenty years since Warren and Marshall first reported the discovery of unidentified curved bacilli in the stomach of patients with gastritis and peptic ulceration in 1984. Since then a plethora of research relating to *Helicobacter pylori* and its clinical effects have been published. Although Warren and Marshall's finding in 1984 is considered to be a 'landmark discovery' for gastroenterological research, their report wasn't the first account of microorganisms living in this gastric niche. In fact it was an Italian anatomist that first reported spirochaetes inhabiting the gastric mucosa of dogs in the nineteenth century (Bizzozero, 1893). It is now thought that these spirochaetes were probably *Helicobacter canis*, *H. felis* and / or *H. heilmannii*. In 1896, as an extension of Bizzozero's work, Salomon demonstrated that these microorganisms could be propagated in mice after feeding them ground up gastric mucasa from cats and dogs (Salomon, 1896).

It was not until early in the twentieth century that pathologists noticed spiral organisms in the human gastric mucosa (Krienitz, 1906). Spiral organisms were also noted by others in the gastric mucosa of both non-human primates (Doenges, 1939) and humans (Freedberg and Baron, 1940). Indeed, Freedberg and Baron (1940) detected these spirochaetes in 40% of the samples they examined. However, in the early 1950's, Palmer challenged these findings, and found no evidence to support the presence of spirochaetes in the human stomach (Palmer, 1954), and concluded that previous observations of spirochaetes in the gastric mucosa was probably due to contamination. Palmer investigated more than 1,000 gastric biopsies, and it is odd to think that spirochaetes were not detected as it is now known that *H. pylori* infects approximately 50% of the world's population.

At the beginning of the nineteenth century clinicians discovered that the stomach produces hydrochloric acid. It was believed that excessive nervousness and other pathological nervous stimuli led either to acid overproduction or to weakening of the local stomach defences or both. An imbalance between aggressive and defence factors was thought to be responsible for peptic ulceration. This brought about the "no acid – no ulcer" theory. Regardless of all the evidence indicating the presence of microorganisms inhabiting the gastric mucosa, it was believed that the human stomach was sterile due to the harsh acidic environment. Thus, the no acid – no ulcer dogma dominated up until the 1980's. Marshall and Warren (1984) rekindled interest in microorganisms and the human gastric niche as previous histological reports not only described the presence of spirochaetes but also the presence of polymorphonuclear

leucocytes. This led to Marshall and Warren culturing Campylobacter-like organisms (CLOs) in 1984. Soon afterwards this organism was found to be strongly associated with gastroduodenal disease, and Marshall fulfilled Koch's postulates by deliberate ingestion of these CLOs (Marshall *et al.*, 1985).

1.2. Taxonomy.

The CLOs that Warren and Marshall discovered resembled *Campylobacter* in several respects, including curved morphology, growth on rich media under microaerophilic conditions, failure to ferment glucose, sensitivity to metronidazole and G+C% content (34%). Thus, these CLOs were first referred to as *Campylobacter pyloridis* (Anonymous, 1985), this being later revised to *C. pylori* in accordance with correct Latin grammar (Marshall and Goodwin, 1987). However, the multiple sheathed flagella, major protein bands and fatty acids of *C. pylori* were found to be markedly different from those of *Campylobacter* species (Pearson *et al.*, 1984; Goodwin *et al.*, 1985). Later this organism was classified as *Helicobacter pylori* based on 16S rRNA nucleotide sequences (Romaniuk *et al.*, 1987), and was the first member of the new genus *Helicobacter* (Goodwin *et al.*, 1990). The genus *Helicobacter* now contains many different species (Table 1.1).

1.3. Bacteriology.

H. pylori organisms are spiral, Gram-negative, microaerophilic (5% O_2, 5–10% CO_2) bacteria. In gastric specimens *H. pylori* are 2.5–5 μm long, 0.5–1 μm wide and have four to six sheathed unipolar flagella (Goodwin *et al.*, 1990). *H. pylori* strains exhibit a rod-like morphology when cultured on solid medium and its spiral form is rarely seen (Goodwin *et al.*, 1990). After prolonged culture of *H. pylori* coccoid structures usually predominate (Bode *et al.*, 1993). Coccoid forms are metabolically active. However, they cannot be cultured *in vitro* (Bode *et al.*, 1993). It is not yet known whether the coccoid forms adopted by *H. pylori* after prolonged culturing are viable or dead forms of the bacterium (Kusters *et al.*, 1997; Wang *et al.*, 1997). The *H. pylori* genome has an average size of 1.6 Mb (Alm *et al.*, 1999), with a G+C% composition of 34–39 mol%. Presently, the full genome sequences of two unrelated *H. pylori* strains are available (Tomb *et al.*, 1997; Alm *et al.*, 1999). Plasmids are found in approximately 40% of isolates. However, no virulence determinants have been found to be encoded by them (Owen, 1999).

Table 1.1. Members of the *Helicobacter* species[a] and their natural hosts

Taxon	Natural host
H. acinonychis	Cheetah
H. bilis	Mouse
H. bizzozeronii	Dog
H. canadensis	Human
H. canis	Dog
H. cholecystus	Hamster
H. cinaedi	Human
H. felis	Cat
H. fennelliae	Human
H. hepaticus	Mouse
H. muridarum	Mouse / Rat
H. mustelae	Ferret
H. nemestrinae	Monkey
H. pametensis	Tern
H. pullorum	Human / Chicken
H. pylori	Human
H. rappini	Sheep / Dog / Mouse / Human
H. rodentium	Rat
H. salomonis	Cat
H. succinogenes	Cow
H. suncus	Shrew
H. trogontum	Rat
H. typhlonicus	Mouse
H. westmeadii	Human

[a] Sourced from Versalovic and Fox (2001) and Solnick and Vandamme (2001).

1.4. Epidemiology.

H. pylori infection rates vary throughout the world. In developing countries between 70–90% of the population can be infected, whilst in developed countries between 25–50% of the population are infected (Dunn *et al.*, 1997). In addition, *H. pylori* equally infects both males and females. It is believed that the differences in prevalence of infection between developed and developing worlds are attributable to the rates of acquisition of *H. pylori* in childhood (Mitchell *et al.*, 1992). In China the prevalence of infection of children less than 10 years of age was 27% while, in contrast, the infection rate of children under 10 in Australia was 4%. An explanation for this observation is that all persons are infected in childhood and the decreased levels of *H. pylori* infection associated with younger age groups are due to gradual improvements in medical care, sanitation, and living conditions (Cullen *et al.*, 1993; Banatvala *et al.*, 1993). Socio-economic status and ethnic grouping also play an important role in *H. pylori* epidemiology, as there is a higher infection rate in children from poorer countries and members of different ethnic groups have different levels of infection even within developed countries (Frenck and Clemens, 2003). Examination of poorer racial groups living in developed countries clarifies this point. A study examining the relationship between the socio-economic status in childhood and the prevalence of *H. pylori* infection within African-American and Hispanic populations resident in the United States indicated that the prevalence of *H pylori* infection was inversely related to social class, with the lowest social class exhibiting the highest infection rate (80%) infection rate and the highest social class exhibiting the lowest rate of infection (11%) (Malaty and Graham, 1994).

1.5. *H. pylori* reservoirs.

Apart form the human stomach no substantial reservoir for *H. pylori* has been reported, although different animal reservoirs have been suggested. A study of Sardinian shepherds revealed a positive association between *H. pylori* prevalence in these shepherds and contact with sheep and sheepdogs (Dore *et al.*, 1999a). The discovery of *H. pylori* in sheep's milk led to the speculation that sheep may be the ancestral host of *H. pylori* (Dore *et al.*, 1999b). A similar study demonstrated that *H. pylori* infection in Polish shepherds and their family members originated from sheep, and it was argued that *H. pylori* infection might be considered as zoonosis (Papiez *et al.*, 2003). *H. pylori* has also been isolated from rhesus monkeys (Handt *et al.*, 1997). However the relatively low levels of interaction between humans and monkeys rules them out as being a substantial reservoir. Other studies have failed

to associate *H. pylori* infection in people that own pets, although a study demonstrated that domestic cats may be involved (Handt *et al.*, 1994).

1.6. The disease spectrum of *H. pylori*.

Since Warren and Marshall's discovery in 1984, *H. pylori* and its clinical consequences have been subject of intense research. Although the majority of individuals colonised by *H. pylori* develop chronic active gastritis, most people remain asymptomatic (Kuipers and Blaser, 1998). This suggests that additional factors are necessary for the development of *H. pylori*-associated diseases. It was soon appreciated that there was a connection between chronic gastritis and diseases such as peptic ulceration, gastric carcinoma and lymphoma. It is also now well established that *H. pylori* is the aetiological agent responsible for chronic gastritis. Thus, it is reasonable to assume that *H. pylori* is in some way responsible for these other diseases. To date *H. pylori* has been implicated in a wide range of gastric illnesses. However, recent studies have demonstrated that this organism can colonise other organs, such as the heart (Chmiela *et al.*, 2003), and it has been suggested this pathogen may somehow be involved in heart disease. Fig. 1.1 outlines the current views of *H. pylori*-related disease progression from initial colonisation of the gastric mucosa to diseases such as gastritis, peptic ulceration and gastric cancer.

1.6.1. Gastritis.

There have been a number of studies carried out that demonstrate *H. pylori's* role in gastritis. The first piece of evidence came when Marshall and Warren (1984) observed that *H. pylori* was associated with areas of inflamed tissue in the stomach. In addition, they observed that individuals that were not colonised by *H. pylori* did not have gastritis. Later Blaser (1990) demonstrated that *H. pylori* could be found in 100% of individuals with type B gastritis. Other studies have shown that anti-microbial therapy that eliminates *H. pylori* infection also eliminates gastritis (Blaser, 1990), and that intentional ingestion of *H. pylori* results in gastritis (Marshall *et al*, 1985).

Gastritis may be defined as a heterogeneous group of gastric mucosal disorders characterised by widespread injury at the gross and / or microscopic level, and usually associated with an acute, chronic or mixed inflammatory response (Blaser, 1990). Once ingested *H. pylori* penetrates the viscous gastric mucosa of the stomach and replicates in close proximity to the surface epithelial cells. The epithelium responds to infection by mucin depletion, cellular

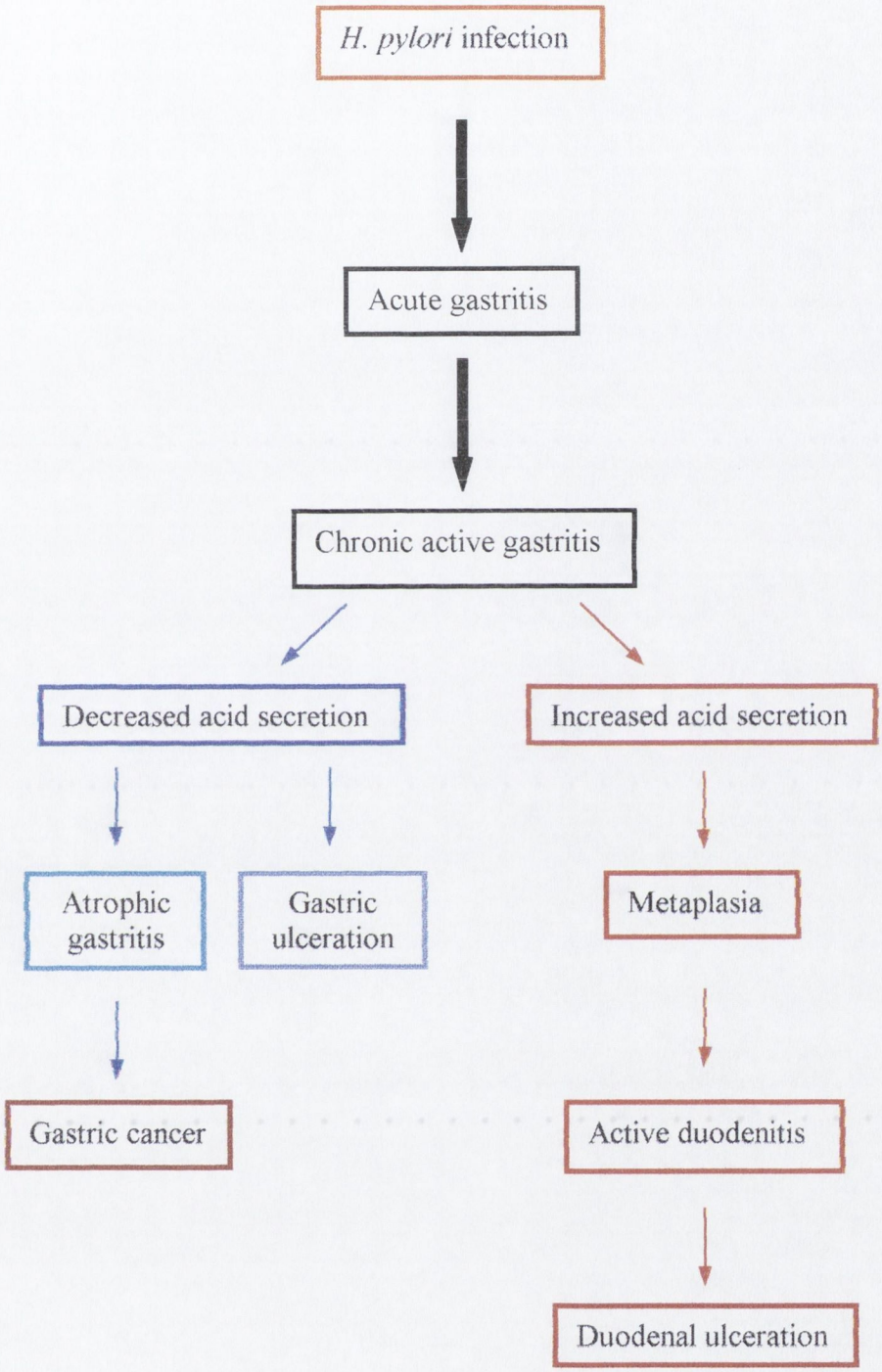


Fig. 1.1. Spectrum of *H. pylori*-associated diseases.

exfoliation, compensatory regenerative changes and an immunological response. The initial stage of infection results in acute gastritis, which is short lived. If the immune response fails to clear the infection within 3–4 weeks there is a gradual accumulation of chronic inflammatory cells. If upon histological examination degeneration of mucosal epithelial cells, a mucosal infiltration by chronic inflammatory cells (lymphocytes, plasma cells and eosinophils) and an active infiltrate of neutrophils are detected then the condition is referred to as type B or chronic active gastritis (Bodger and Crabtree, 1998).

1.6.2. Peptic ulceration.

Although the histology of an ulcer is basically the same regardless of location in the stomach or duodenum and although the pathogenic mechanisms are similar, the healing rates of ulcers differ depending upon their location (Peterson, 1991). Therefore, ulcer location is strictly defined with gastric ulcers designated as ulcers which are more than 2–3 cm in length proximal to the pylorus and duodenal ulcers as those ulcers confined to the duodenal bulb. Ulcers are differentiated from erosions by the presence of depth. A significant advance in the understanding of gastroduodenal disease has been the realisation that infection with *H. pylori* is a significant cause of gastroduodenal ulcers. Approximately 90–95% of duodenal ulcers and 70–75% of gastric ulcers are attributable to infection with *H. pylori* (Ernst and Gold, 2000). Evidence of *H. pylori's* involvement in peptic ulceration comes from the finding that the recurrence rate of duodenal ulceration is markedly reduced after successful treatment of *H. pylori* infection.

1.6.2.1. Duodenal ulceration.

Patients with duodenal ulceration tend to have higher acid secretion than do normal healthy individuals. The passing of acid from the stomach into the duodenum has adverse consequences. The most significant consequence is gastric metaplasia. Metaplasia is defined as a potentially reversible change in which a fully differentiated cell type is replaced with another differentiated cell type, and usually represents a change to cells better able to withstand an adverse environment (Dixon, 2001). In the duodenum gastric metaplasia presents as gastric-type mucus-secreting cells in the surface spithelium (Walker and Dixon, 1996). This phenotypic change of the duodenal epithelium is probably an adaptive response to the increased acid environment associated with duodenal ulceration.

It is believed that the natural history of *H. pylori* infection of the stomach involves a mechanism that predisposes one to duodenal inflammation and ulceration (Lacy and Rosemore, 2001). Initial infection by *H. pylori* incurs an inflammatory response predominantly in the antrum of the stomach with the body of the stomach being relatively spared. The resultant gastritis leads to increased stimulation of G-cells that release gastrin. Gastrin causes the release of acid from parietal cells via the induction of histamine from terochromaffin-like cells. Over time this increased load of acid delivered to the dudenum causes mucosal irritation. As previously stated, high acid loads to the duodenum causes gastric metaplasia. Since *H. pylori* can colonise gastric epithelial cells, the presence of gastric metaplasia in the duodenum allows a *H. pylori* infection to occur in this location. Colonisation of gastric metaplastic cells in the duodenum by *H. pylori* leads to active duodenitis and ultimately leads to ulceration (Blaser, 1990). Thus, two risk factors are involved in duodenal ulceration, namely, the presence of gastric metaplasia in the duodenum and *H. pylori* infection.

1.6.2.2. Gastric ulceration.

Most of the evidence that proves *H. pylori's* involvement in duodenal ulceration also applies to gastric ulceration. As with gastritis and duodenal ulcers, eradication of *H. pylori* infection significantly lowers the chances of gastric ulcer recurrence (Dixon, 2001). Gastric ulceration occurs less commonly than duodenal ulceration in most populations (Ernst and Gold, 2000). It has been demonstrated that gastric ulceration occurs more frequently among individuals that become colonised by *H. pylori* at an early age, whilst duodenal ulceration is more common in individuals that are colonised later on in life (Kuipers and Blaser, 1998).

1.6.3. Gastric atrophy.

Atrophy in the stomach is conventionally defined as loss of glandular tissue from repeated or continuing mucosal injury and is a common denominator in all pathological processes causing progressive mucosal damage, including long-standing *H. pylori* infection (Dixon, 2001). Unlike the pattern of increased acid secretion seen in *H. pylori*-induced duodenal ulcers, a chronic *H. pylori* infection can result in a down-regulation of gastric acid secretion (acid hyposecretion). A *H. pylori* infection in these cases is usually more extensive and produces an inflammation within the whole body of the stomach (pan-gastritis). An extensive inflammation of the stomach induced by a *H. pylori* infection can lead to destruction of acid-secreting parietal cells, thus lowering the level of acid in the stomach. Atrophy in the stomach

is generally followed by the replacement of glandular tissue with intestinal metaplastic cells. Chronic atrophic gastritis is believed to be a precursor state in the pathway to the development of gastric cancer. It is interesting to note the attempts to isolate *H. pylori* from a cancerous stomach usually fail due to the replacement of gastric epithelial cells with atrophic mucosa and intestinal metaplasia (Blaser and Parsonnet, 1994). Thus, *H. pylori* induce the generation of an environment that they are unable to inhabit.

1.6.4. Gastric cancer.

Parsonnet *et al.* (1991) first described the association of *H. pylori* and gastric cancer. *H. pylori* is presently classified as a type I carcinogen by the World Health Organisation (WHO). Evidence for *H. pylori's* association with gastric cancer comes from *Helicobacter*-positive serological tests on samples taken years before the development of gastric cancer (Parsonnet *et al.*, 1991), the presence of *H. pylori* adjacent to gastric carcinoma upon histological testing, long-term *H. pylori* infection within animal models showing induction of adenocarcinoma (Watanabe *et al.*, 1998) and the decline of gastric cancer in the Western world over the last fifty years mirroring a decrease in the prevalence of *H. pylori* and atrophic gastritis. A recent study examined 1,246 and 280 *H. pylori*-positive and -negative patients, respectively, over an eight-year period (Uemura *et al.*, 2001). Thirty-six (2.9%) of the *H. pylori*-positive patients developed gastric cancer whilst none of the *H. pylori*-negative patients developed this disease, strongly indicating an association of *H. pylori* in the development of gastric cancer. It is important to note that although *H. pylori* has a definite association with gastric cancer, any one individual infected with *H. pylori* has a 1:40,000 chance of developing this disease (Lacy and Rosemore, 2001).

1.6.4.1. Gastric adenocarcinoma.

Carcinoma of the stomach is the second most common cancer worldwide (Parkin *et al.*, 1988), and adenocarcinoma of the stomach accounts for nearly 95% of all gastric cancers. The association between gastric cancer and *H. pylori* may be explained by two possible mechanisms, one based on a carcinogenesis-promoting effect of *H. pylori* itself and the second based on the establishment of a carcinogenic environment due to long-term infection with *H. pylori*. In the latter proposed mechanism it is believed that *H. pylori* infection produces an active chronic gastritis, leading to chronic atrophic gastritis, which leads to intestinal metaplasia followed by dysplasia, ultimately leading to gastric cancer (Lacy and Rosemore, 2001). Correa (1992) suggested that malignant transformation is caused by the

persistent oxidative stress attributable to the inflammatory response that occurs during gastritis. However, the *cagA* gene has been described as a marker for gastric cancer (Blaser *et al*, 1995). A recent study suggests that tyrosine phosphorylation of the CagA antigen may play a role in gastric carcinogenesis (Lai *et al*., 2003).

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

MALT lymphomas account for approximately 5% of gastric cancers. Lymphoid tissue is absent from the stomach. However, lymphoid follicles appear in the stomach after *H. pylori* infection with resultant gastritis (Genta *et al*., 1993). The accumulation of lymphoid tissue attracted by *H. pylori* in the stomach forms the basis from which MALT lymphomas emerge. MALT lymphomas are ultimately composed of B-cells. The evidence for *H. pylori's* association with MALT lymphoma comes from studies showing that more than 90% of patients with this disease are infected with *H. pylori* (Wotherspoon *et al*., 1991). In addition, eradication of *H. pylori* infection eliminated MALT lymphomas in 76% of patients (Stolte, 1992).

1.6.5. Non-ulcer dyspepsia.

Non-ulcer dyspepsia (NUD) is a disorder of the upper gastrointestinal tract for which a biochemical or structural cause cannot be found. NUD symptoms include upper abdominal pain, nausea, heartburn, acid regurgitation, excessive burping, and bloating. Unlike the previous diseases described, the role of *H. pylori* infection in NUD is relatively unclear and highly controversial. Given the spectrum of disease attributable to *H. pylori* it is clearly important to know if a patient with frequent upper gastrointestinal symptoms has a *H. pylori* infection. Abnormal gastric emptying is considered one of the possible mechanisms that lead to NUD. However, is there clinical evidence for *H. pylori* induced changes in gastric emptying? A study demonstrated that *H. pylori*-positive and -negative patients displayed no difference in stomach emptying rates, but eradication of infection improved patients' conditions (Murakami *et al*., 1995). In addition, another study demonstrated that severely perturbed gastric emptying normalised after *H. pylori* eradication in NUD patients (Miyaji *et al*., 1999). In contrast, a question has been posed as to whether the overlap of NUD and *H. pylori* is casual or coincidental as both conditions are common (Kalantar *et al*., 1999).

1.6.6. H. pylori *infection in other organ systems.*

The disease spectrum of *H. pylori* has mainly been associated with the stomach and duodenum. However recent research has associated *H. pylori* with illnesses in unexpected areas of the body. Recent focus has centred on *Helicobacter* species and their role in biliary disease and primary liver cancer (Megraud, 2000). Another unusual area of *H. pylori* clinical research has focused on the skin, with a study showing acne reduction with the elimination of *H. pylori* (Utas *et al.*, 1999). Moreover, it has been suggested that *H. pylori* is an aetiological agent in the pathogenesis of atherosclerosis. Danesh *et al.* (1999) showed a weak association between coronary artery disease (CAD) and *H. pylori*, with 54% of CAD patients seropositive for *H. pylori* as opposed to 46% of control subjects being seropositve. *H. pylori* is not the first bacterium to be implicated in CAD, as *Chlamydia* spp. are also believed to have a role in this disease. Possible links between atherosclerosis and inflammation have been proposed (Chmiela *et al.*, 2003). It has also been suggested that *Chlamydia pneumoniae* and *H. pylori* may be responsible for this inflammatory reaction, which is a major step in the disease process. It is also believed that an infection can initiate or maintain the atherosclerotic process. *H. pylori* is capable of binding platelets (Byrne *et al.*, 2003) which contributes to the association between *H. pylori* infection and cardiovascular disease. Indeed, *H. pylori* DNA has been detected by PCR in human atherosclerotic plaques (Farsak *et al.*, 2000). Although there are many reports implicating *H. pylori* with CAD, no strong conclusions have been drawn as yet.

1.6.7. Are H. pylori *infections beneficial?*

There is no doubt that *H. pylori* is the causative agent of certain gastric illnesses. Despite the fact that approximately 50% of the world's population is infected with this organism, only a subset of infections cause disease (approximately 10%) and out of this subset approximately 2% cause gastric cancer. This implies that there is a large *H. pylori* population that does not cause harm to their hosts. This raises the question as to whether *H. pylori* can be beneficial to humans and whether eradication of this organism is necessary.

With improved sanitation and the widespread use of antibiotics *H. pylori* infection is in steep decline in many parts of the world. This may seem good news to many biologists, but recent studies have shown that having a *H. pylori* infection may be advantageous. How long humans have been infected by *H. pylori* is still controversial. However, it is accepted that this organism has infected humans for many thousands of years, and the successful persistence of

H. pylori in the human stomach for such a long period may be because this organism offers some advantages to the host. A study has shown that *H. pylori* produce a cecropin-like peptide (antibacterial peptide) with high antimicrobial properties (Pütsep *et al.*, 1999). It is postulated that when a *H. pylori* cell lyses, its contents, along with this cecropin-like peptide, are released in the stomach killing other microbial organisms, which in turn prevents any other organism from colonising the stomach. A study carried out within a German school revealed that children infected with *H. pylori* were less likely to have diarrhoea than children without an infection (Rothenbacher *et al.*, 2000), implying that *H. pylori* may have beneficial properties to human hosts.

Interestingly, there has been a marked decline in the instances of peptic ulcer disease and gastric cancer in the 20th century. Paralleling this decline is a dramatic increase in the incidences of gastro-oesophageal reflux disease (GERD), Barrett's oesophagus and adenocarcinoma of the oesophagus in Western countries (Richter *et al.*, 1998). This observation led to the hypothesis that *H. pylori* may in some way be associated with the low prevalence of these diseases and perhaps may be capable of preventing their onset. Studies have also shown that *cagA*$^+$ *H. pylori* strains have a more protective effect than *cagA*$^-$ strains (Chow *et al.*, 1998, Vicari *et al.*, 1998). The presence of *cagA*$^+$ *H. pylori* strains can reduce the acidity of the stomach. It has been suggested that the raising of the pH by *H. pylori* prevents GERD, Barrett's oesophagus and adenocarcinoma of the oesophagus. Moreover, it has been reported that oesophagitis developed in healthy patients after eradication of *H. pylori* infection (Labenz *et al.*, 1997).

Conversely, arguments have been made that, although *H. pylori* may prevent these reflux-associated diseases, the risks of acquiring gastric cancer via *H. pylori* infection far outweigh any possible benefits it may provide (Graham, 1999). However, it has been stated that, if *H. pylori* does provide protection from GERD, the notion of restriction of anti-*H. pylori* treatment to only a few cases (peptic ulcer disease and MALT lymphoma) could be justified (Vieth, 1999). In spite of this controversy recent reports have demonstrated a protective role for *H. pylori* in erosive reflux oesophagitis (Koike *et al.*, 2001, Warburton-Timms *et al.*, 2001, Yamaji *et al.*, 2001). However, as safe and potent anitsecretory drugs to prevent gastro-oesophageal reflux are available (Graham, 2003) it seems unjustified to use a dangerous organism that has been associated with gastric diseases as a protective strategy. Opinions as to whether *H. pylori* should not be eradicated within some patients are still divergent, and

perhaps further analysis regarding how *H. pylori* isolates elicit their diverse pathogenic effects is needed before this debate can be resolved.

1.6.8. in vivo *modelling of* H. pylori.

The clinical association of *H. pylori* with several gastric illnesses led to the study of how this organism exerts its pathogenic effects *in vivo*. Full comprehension of the mechanisms and genetic factors needed for *H. pylori* colonisation of the gastric mucosa or how it mediates its virulence could potentially lead to treatments or eradication therapies. Initial attempts to establish *H. pylori* infection models in laboratory animals such as mice, rats and rabbits were unsuccessful (Cantorna and Balish, 1990). Thus, alternative *Helicobacter* models were developed using animals that were either already naturally infected with closely related *Helicobacter* spp. [ferrets (Fox *et al.*, 1990) and non-human primates (Dubois *et al.*, 1994)], or with those that could be experimentally infected with *H. pylori* [gnotobotic piglets (Krakowka *et al.*, 1987) and dogs (Radin *et al.*, 1990)]. The discovery that rodents were susceptible to infection with *H. felis* paved the way for affordable animal model studies that did not require specialised facilities or expertise (Lee *et al.*, 1990). It was not until 1995 that immunocompetent mice were experimentally infected with *H. pylori* (Marchetti *et al.*, 1995), which led the way to establishing a mouse model permitting long-term high bacterial density colonisation of *H. pylori* (Lee *et al.*, 1997).

A variety of bacterial factors have now been tested for the roles in gastric colonisation of animal models. Isogenic mutants have demonstrated that strains deficient in urease or lacking flagellar-mediated motility cannot colonise animal models (Eton and Krakowka, 1994; Eton *et al.*, 1996). Interestingly, single mutations in the *flaA* or *flaB* flagillin genes of *H. pylori* yielded an avirulent strain. However, identical mutations in *H. mustelae* enabled a low level of colonisation in a ferret model (Andrutis *et al.*, 1997). Indeed, conflicting findings have arisen from studies in which *H. pylori* mutants were tested in different animal models (Blanchard *et al.*, 1999; Chevalier *et al.*, 1999). Thus, although animal models have played a critical role in many of the important discoveries in the *Helicobacter* field, there is no single model that is the best for all applications.

1.6.9. Treatment for H. pylori *infection.*

Eradication of *H. pylori* infection cures most ulcer disease (Malfertheiner *et al.*, 2000), MALT lymphoma (Sackmann *et al.*, 1997) and other gastric illnesses. However as *H. pylori* can

remain asymptomatic in most individuals is global eradication really necessary? This is an interesting argument, yet the total eradication of *H. pylori* is economically unrealistic, so this question seems irrelevant. Primary prevention of infection may eventually be feasible when effective vaccines become available but at the moment treatment recommendations will target those most likely to benefit.

1.6.9.1. Antimicrobial treatment.

Antimicrobial treatment of *H. pylori* is now based on antibiotic triple-therapy regimes for complete eradication. These regimes are based on the Maastrict 2-2000 Consensus Report and recommends the use of proton pump inhibitors (PPI's) / ranitidine bismuth citrate (RBC) with clarithromycin, and amoxycillin or metronidazole as first-line treatment in primary care (Qasim and O'Moráin, 2002). Second line treatment after primary treatment failure includes the use of PPI's, bismuth, metronidazole and tetracycline for a minimum of seven days. Fig. 1.2 displays the strategy used in the treatment of *H. pylori* infection. Although antimicrobial therapy is highly successful some problems with antibiotic treatment have arisen. It is now known that antibiotic resistance is a major cause of treatment failure. Large USA- and Australia-based studies have demonstrated an overall prevalence of metronidazole resistance from 30–36.9% (Meyer *et al.*, 2002) and a prevalence of clarithromycin resistance from 10.1–15.7% (Grove and Koutsouridis, 2002). It was also demonstrated that clarithromycin resistance varies geographically. Incomplete eradication usually results in recrudescence of the infection. Such situations require a second line of treatment for *H. pylori* eradication, which is not ideal when patient compliance and cost is considered (Qasim and O'Moráin, 2002). Since the elucidation of two unrelated *H. pylori* genomes new targets for antimicrobial treatment can be identified. One such target is the *ureI* gene (Jenks, 2001), which is part of the urease gene cluster and is necessary for internal pH homeostasis and for colonisation of the gastric mucosa. However, it will be some time before new anti-*H. pylori* drugs will be clinically available.

1.6.9.2. *H. pylori* vaccines.

Although novel antimicrobial agents may facilitate treatment in the future, only a vaccine will have an impact on the global prevalence of *H. pylori* infection. Immunisation has been successful in animal models and there have been encouraging results from studies with a urease-based vaccine in human subjects (Michetti *et al.*, 1999). Genome searches have identified suitable candidate vaccines for *Mycoplasma* infections (Barry *et al.*, 1995), and

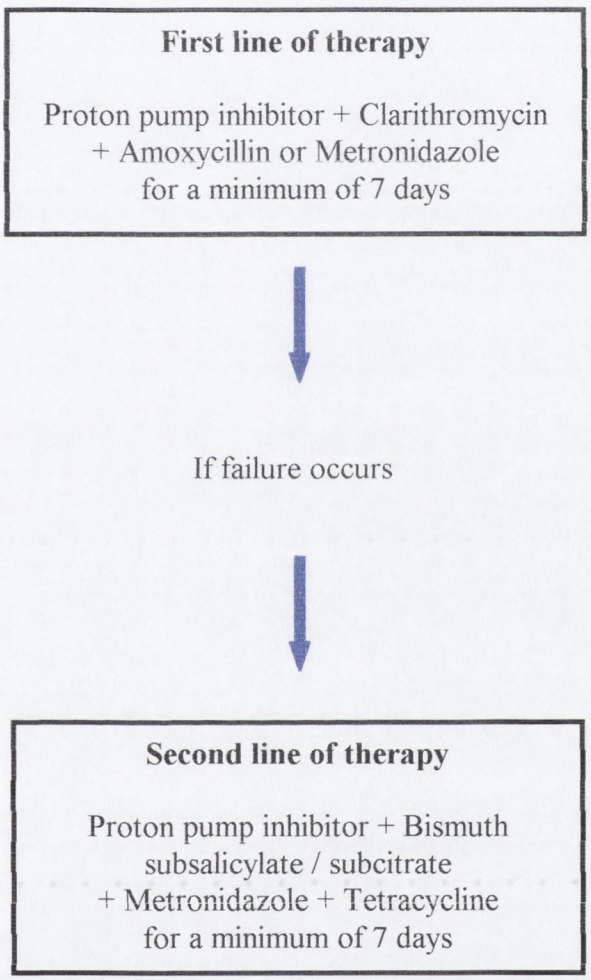


Fig. 1.2. Current strategy for eradication of *H. pylori* Infection based on the Maastricht 2-2000 Consensus Report.

application of such approaches may identify novel *H. pylori* vaccines. It has been argued that since the natural host immune response in ineffective in clearing *H. pylori* infection, a vaccine approach might not work. However, numerous animal model studies have demonstrated protection against *H. pylori* after vaccination (Michetti *et al.*, 1994; Corthésy-Theulaz *et al.*, 1995).

There are many advantages to using animal models in vaccine studies. However, caution must be applied when comparing effects on animals (rat / mouse) to what may occur in humans. Thus, clinical trials have been carried out on humans in attempts to develop *H. pylori* vaccines. One of the first of the human trials encompassed administering *H. pylori*-infected volunteers with *Helicobacter* urease and heat-labile toxin from *Escherichia coli* (Michetti *et al.*, 1994). Unfortunately this approach did not have the desired effect and the adjuvant induced diarrhoea. A different approach used live *Salmonella* vector vaccines on human volunteers (DiPetrillo *et al.*, 1999). Volunteers were administered *Salmonella typhimurium* that expressed urease subunits of *H. pylori*. The volunteers developed a good mucosal immune response to the *Salmonella* antigens but not to the *H. pylori* urease. More encouraging results were obtained by Angelakopoulos and Hohmann (2000) using a similar system. In this study three out of six volunteers did develop antibodies to *H. pylori* urease. Vaccine development for *H. pylori* infections is still in its infancy. However, similarly to novel drug development, the two published *H. pylori* genome sequences have allowed the identification of potential vaccine targets, mostly encompassing families of outer membrane proteins (Alm *et al.*, 2000).

1.7. Pathogenicity determinants of *H. pylori*.

H. pylori infection is the cause of a variety of gastric diseases. *H. pylori*-associated diseases are dependent on the virulence of the infecting strain, the genetic susceptibility of the host and environmental factors. Of these, bacterial virulence factors have received the most focus. Virulence factors or determinants are defined as molecules or traits that are necessary for pathogenesis, or whose loss significantly reduces pathogenesis (O'Toole *et al.*, 2000). As approximately 10% of *H. pylori* infected individuals acquire a *H. pylori*-associated disease, researchers have tried to uncover the factors in *H. pylori* that are responsible for bringing about disease. Many characteristics of *H. pylori* are necessary for disease and thus are potential therapeutic targets and important subjects for research.

1.7.1. Virulence factors (determinants).

1.7.1.1. Flagella.

A prominent characteristic of all gastric *Helicobacter* species is their flagella-driven motility. *Helicobacter* species exhibit much higher motility in viscous media than other enteric bacteria such as *E. coli* (Hazell *et al.*, 1986). Motility is essential for *H. pylori* colonisation, as non-motile mutants cannot colonise the stomach (Yoshiyama *et al.*, 1999). Thus, *H. pylori* flagella may be more correctly described as a colonisation factor. Since *H. pylori* is not an acidophile, flagella are necessary to move *H. pylori* out of the acidic gastric lumen and into the less acidic gastric mucus overlaying the gastric mucosa. Motility is also necessary for *H. pylori* to maintain a constant reservoir of bacteria, as the gastric mucus is subject to rapid turnover (Blaser, 1994; Suerbaum, 1995), and to encounter the gastric epithelial cell surface. *H. pylori* are guided towards the gastric mucosa by chemotactic factors (Yoshiyama *et al.* 1999). This chemotactic response involves over 40 genes (Tomb *et al.*, 1997). It was believed that as the membranous sheath that surrounds *H. pylori* flagella is composed of LPS, the flagella might be involved in adhesion as well as motility. However, Clyne *et al.* (2000) demonstrated that the flagella of *H. pylori* and *H. mustelae* do not play a role in adherence of the bacterium to the gastric epithelium.

H. pylori cells have 4–6 unipolar sheathed flagella that are composed of flagellin subunits encoded by the *flaA* and *flaB* genes (Suerbaum *et al.*, 1995). These genes are transcribed by different sigma factors (σ^{28} and σ^{54}, respectively). This supports the model in which *H. pylori* varies the ratio of FlaA and FlaB in the flagellum in response to changes in the environment (Josenhans and Suerbaum, 2001). These flagellins share ~50% amino acid identity and may have evolved from a common ancestral flagellin. It is believed that both FlaA and FlaB are present throughout the length of the filament while FlaB is only found in the hook of the flagella (Josenhans and Suerbaum, 2001). In addition to the *flaA* and *flaB* genes, the genes *flgE* (encoding hook protein), *fliD* (encoding flagellar cap protein), *flhA* / *flhB* / *fliQ* (encoding export-associated proteins) and *fliI* (encoding the energy-providing ATPase of the flagellar type III system) were shown to be necessary for synthesis of flagella by mutagenesis (Josenhans and Suerbaum, 2001). Analysis of the *H. pylori* genome sequence shows that the genetic organisation of flagella-related genes lacks clustering. This is in contrast to the concentration of flagellar genes in well-defined regions in most other bacteria such as *S. typhimurium* (O'Toole *et al.*, 2000). A *H. pylori* flagella regulatory gene was identified, namely, *flgR*. The FlgR regulator was found to be necessary for transcription of flagellar

genes encoding the basal body-hook complex (including *flaB*), however it also repressed transcription of the *flaA* gene. This regulatory process is reminiscent of the *Enterobacteriaceae* and *Caulobacter* genus (Spohn and Scarlato, 1999).

1.7.1.2. Urease.

Urease is a protein that is expressed in high quantity by *H. pylori*. Urease catalyses the hydrolysis of urea to yield ammonia and carbamate. With regard to bacterial pathogens, a successful infection requires that the bacterium colonises the host, avoids host defences and damages host tissues. It is believed that *H. pylori* urease contributes to all three stages, and therefore may be considered central to the pathogenesis of *H. pylori*-associated diseases.

H. pylori urease is a nickel-containing high molecular weight enzyme composed of UreA and UreB subunits and been shown to account for 5–10% of the total cellular protein of *H. pylori* (Evans *et al.*, 1991; Mobley *et al.*, 1995). As stated previously *H. pylori* is not an acidophile and is sensitive to low pH. It is believed that *H. pylori* is only transiently exposed to low pH on initial entry to the gastric lumen. It is at these early stages that urease is necessary for colonisation. The ammonia produced by urease-associated hydrolysis of urea neutralises the gastric acid and provides time for the bacteria to traverse the mucus layer and to colonise the gastric epithelium. Originally it was thought that urease was located on the cell surface and created a cloud of ammonia around the cell. Thus, the interaction of H^+ ions with urease-produced ammonia would be expected to increase the local pH allowing *H. pylori* to survive (Evans, 1991). As urease is known to be a cytoplasmic enzyme (Scott *et al.*, 1998), the so-called 'altruistic autolysis mechanism' was proposed (Stingl *et al.*, 2002). This hypothesis proposed that a considerable proportion of the cells lyse and that the released urease binds to the surface of non-lysed cells allowing them to survive in an acidic environment by producing a protective cloud of amonia. However, a second hypothesis has been proposed that suggests that urease activity in the cytoplasm is essential for acid survival. The ammonia produced in the cytoplasm is thought to enter the periplasm in its non-charged state and to bind protons thereby elevating the periplasmic pH and preventing acidification of the cytoplasm (Scott *et al.*, 1998). The exact mechanism by which *H. pylori* survives acid challenge *in vivo* remains a point of debate.

The functional part of *H. pylori* urease is composed of two structural subunits encoded by the *ureA* and *ureB* genes. All genes encoding *H. pylori* urease are located as a single 6.13-kb gene

cluster on the bacterial chromosome (Clayton *et al.*, 1990; Labigne *et al.*, 1991; Cussac *et al.*, 1992). Seven contiguous genes, all transcribed in the same direction, are necessary for synthesis of an active enzyme (Cussac *et al.*, 1992). These genes have been designated *ureABIEFGH*. The genes *ureE*, *ureF*, *ureG* and *ureH* function as accessory genes required for nickel incorporation to produce the catalytically active metalloenzyme. The u*reI* gene is the only gene that does not share sequence identity with urease genes from other species. This gene was shown not to be involved in the formation of active urease but its absence inhibited acid survival of *H. pylori in vivo* (Skouloubris *et al.*, 1998). Weeks *et al.* (2000) found that UreI is an integral cytoplasmic membrane protein that might form a urea-specific pore. This study demonstrated that the UreI pore can open as the surrounding pH drops below pH 6.5, allowing urea to reach cytoplasmic urease. As the acid pH is neutralised, the enzyme is denied access to the substrate, thus, avoiding excess ammonia. Rektorschek *et al.* (2000) demonstrated that UreI mediates acid-dependent uptake of urea. Thus, as the pH drops urea is made accessible to cytoplasmic urease via UreI, neutralising the pH and allowing survival of *H. pylori*. An active mechanism that regulates acid tolerance in *H. pylori* is in accordance with the theory that *H. pylori* does not survive in acidic environments by release of urease via autolysis.

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

Since the primary consequence of *H. pylori* infection is gastritis, research has tried to find the mechanisms involved in the gastric inflammatory response. To this end, it was uncovered that the N-terminus of the large subunit of urease was chemotactic for neutrophils (Mai *et al.*, 1992). Moreover, activation of neutrophils by cell-free water extract of *H. pylori* was reported (Yoshida *et al.*, 1993). Later it was found that the neutrophil-activating factor in these water extracts is a 150-kDa protein that has been termed HP-NAP (*H. pylori* neurtophil activating protein) and is encoded by the *napA* gene (Evans *et al.*, 1995). Purified HP-NAP has been shown to be chemotactic for human neutrophils and monocytes, suggesting that this protein plays a role in the accumulation of these cells at the site of *H. pylori* infection (Satin *et al.*, 2000).

Structural analysis of HP-NAP showed that this protein shares similarities with the *E. coli* DNA-binding protein Dps, a bacterial stress protein that is induced during periods of nutrient limitation. However, unlike Dps, HP-NAP has a high affinity for binding iron (Dundon *et al.*, 2000). This fact and the demonstration that the *napA* gene bears sequence identity to bacterial

ferritin genes, suggest that HP-NAP is important in iron uptake. However, the Lyme disease pathogen *Borrelia burgdorferi* does not require iron as a cofactor for its proteins and has a *napA*-like homologue. This suggests that HP-NAP may not be a bacterioferritin. Thus, it has been hypothesised that the original function of HP-NAP was as an iron-binding protein. However, this function was superseded by activation of neutrophils during the course of *H. pylori* infection. By activating neutrophils and inducing a moderate inflammatory reaction leading to alteration of the epithelial tight junctions and basal membranes, HP-NAP possibly promotes the release of nutrients from the mucosa to support growth of *H. pylori* (Blaser, 1993; Montecucco *et al.*, 1999).

Satin *et al.* (2000) demonstrated that HP-NAP is highly immunogenic in mice and humans. As *H. pylori* is genetically diverse, variation in the HP-NAP protein could rule it out as a vaccine target. However, nucleotide sequences of the *napA* gene in twelve strains isolated from different geographical regions indicated strong conservation of this gene (Dundon *et al.*, 2000). Moreover, it was predicted that the non-conserved amino acid residues in HP-NAP did not affect its secondary structure. Thus, HP-NAP could be a possible candidate for an anti-*H. pylori* vaccine.

1.7.1.4. γ-Glutamyltransferase.

To identify traits that allow *H. pylori* to colonise a diverse niche, Chevalier *et al.* (1999) sought to characterise γ-glutamyltransferases (GGTs) as very little is known regarding the physiological role of this enzyme in bacteria. GGTs play a major role in glutathione metabolism in mammalian tissues. GGTs are peptidases, enzymes that catalyse the hydrolytic cleavage of peptides to form amino acids or smaller peptides. They catalyse the transfer of the γ-glutamyl moiety from peptides to an acceptor peptide. In bacteria it is thought that GGTs might play a role in the transport of amino acids across cell membranes. Chevalier *et al.* (1999) found that *H. pylori* strains that lacked GGT were unable to colonise the gastric mucosa of mice and suggested that it was essential for infection. However, it was later reported that GGT was not essential for *H. pylori* colonisation of pigs and mice but did provide an advantage for colonisation (McGovern *et al.*, 2001). It was suggested that the contrasting results obtained were due to different experimental designs. As a mechanism or role in virulence was not determined by either of these studies, it was unknown as to whether GGT was a true virulence factor. However, a recent report demonstrated an apoptopic role for GGT *in vitro* (Shibayama *et al.*, 2003). Shibayama *et al.* (2003) demonstrated that *H. pylori*

deficient in GGT were unable to induce apoptosis in AGS cells, whereas the wild-type strain was. Thus, *H. pylori* GGT may as yet have a role in pathogenesis.

1.7.1.5. Iron-acquisition proteins.

The requirement for iron as a critical component for cellular processes has long been appreciated. Bacteria require iron for incorporation into haeme groups and cytochromes. Without the ability to obtain iron from a relatively iron-free environment most bacteria would die or stop growing. Iron is poorly soluble and is scarce in the stomach where *H. pylori* reside. To overcome this problem *H. pylori* have developed systems that contribute to the ability of the organism to assimilate iron that is sequestered by host cell proteins. Lactoferrin is found in the gastric mucosal surfaces and within human neutrophils and would thus be an accessible iron source for *H. pylori*. *H. pylori* was found to produce a 70-kDa outer membrane protein (Lbp) which specifically binds human lactoferrin (Dhaenens *et al.*, 1997) and it is now thought that *H. pylori* use lactoferrin as its sole iron source.

Another iron-acquisition system has been described in *H. pylori* that implies the use of siderophores (Illingworth *et al.*, 1993). Siderophores are small iron-chelating molecules synthesised and released into the environment to scavenge iron. In contrast, Husson *et al.* (1993) were unable to detect the presence of siderophores in *H. pylori* and suggested that lactoferrin-binding proteins were the sole iron acquisition system in this organism. Nevertheless, analysis of the *H. pylori* genome sequence revealed genes homologous to the *E. coli* ferric citrate (*fec*) uptake system (Marais *et al.*, 1999). In addition, sequence analysis of the *H. pylori* genome has revealed orthologues to a ferric uptake regulator (*fur*) periplasmic iron-binding proteins (*ceuE*) and iron-storing proteins (ferritins, *pfr*). Interestingly, *H. pylori's* ability to scavenge iron may contribute significantly to its virulence, i.e., *H. pylori* infections have been linked to iron deficiency anaemia (Barabino, 2002).

1.7.1.6. Adhesins.

Bacterial adhesins are surface structures that are involved in the initial stages of colonisation by mediating the interaction between the bacterium and the host cell surface. Adhesins are generally regarded as virulence factors due to their role in host colonisation. *H. pylori* has evolved adherence mechanisms to maintain itself in an environment with constant cell turnover and peristalsis. Adhesins are not only important for attachment but are important for communication with the host, a phenomenon known as 'cross talk'.

It was demonstrated that *H. pylori* can bind to fucosylated glycoconjugates containing Lewis b structures, a histo-blood group antigen present on some gastric epithelial cells (Falk *et al.*, 1993; Borén *et al.*, 1993). Falk *et al.* (1995) also demonstrated that *H. pylori* could colonise the gastric epithelium of transgenic mice expressing human Lewis b antigens, whereas normal mice could not be colonised. The adhesin responsible for this binding was found to be an outer membrane protein termed BabA. Other outer membrane proteins have been implicated in *H. pylori* adhesion to epithelial cells, namely AlpA and AlpB. Odenbreit *et al.* (1999) demonstrated that isogenic mutants lacking *alpA* and *alpB* genes showed greatly reduced adherence to surface epithelial cells on human gastric tissue sections. Similarly, *H. pylori* lacking the outer membrane protein HopZ displayed reduced adherence to epithelial cells (Peck *et al.*, 1999). AlpA, AlpB, BabA, HopZ and Hops A–E all belong to a paralogous family of 32 outer membrane proteins. As four members of this family are involved in adherence (AlpA, AlpB, BabAand HopZ), it may be possible that this entire family of proteins share this trait. Moreover, as phase variation has been observed for some of these outer membrane proteins, a fine specificity of *H. pylori* adherence to specific host cells under different conditions *in vivo* may be achieved.

1.7.1.7. Heat shock proteins in *H. pylori*.

Heat shock proteins (HSPs) are associated with an adaptive response of bacteria to an elevation in temperature. HSP's include chaperones, proteases, alternative sigma factors and other regulatory and structural proteins. Some HSP's are themselves virulence factors while others affect pathogenesis indirectly, by increasing bacterial resistance to host defences or regulating virulence genes. Homologues of the *E. coli* heat shock proteins GroEs and GroEl have been identified in *H. pylori* (HspA and HspB, respectively, Evans *et al.*, 1992; Homuth *et al.*, 2000). Evans *et al.* (1992) revealed that HspB belonged to the HSP60 family of chaperonins and was involved in the assembly or protection of urease. Thus, HspB has an important role in virulence. The *hspB* gene is part of a bicistronic operon (*hspA-hspB*). The *hspA* gene, located upstream of *hspB*, encodes a protein with nickel-binding properties. Expression of HspA and HspB with urease increases urease activity. Thus, it is believed that HspA plays a role in the integration of nickel into the functional urease molecule. *H. pylori* heat shock proteins have also been shown to be involved in adherence (Testerman *et al.*, 2001). Thus, the participation of heat shock proteins in *H. pylori* pathogenesis implies an important function of these factors in virulence.

1.7.2. Virulence determinants in a subset of H. pylori *strains.*

1.7.2.1. The vacuolating cytotoxin (VacA).

Initial analysis of *H. pylori* supernatants demonstrated that over half of the clinical isolates investigated could induce the formation of large cytoplasmic vacuoles in cultured eukaryotic cells (Leunk *et al.*, 1988). The protein responsible for these vacuoles was later purified and found to be a 95-kDa protein and was named VacA (Cover and Blaser, 1992). Subsequently, it was demonstrated that *H. pylori* strains expressing the VacA toxin (Tox^+) were more commonly isolated from patients with peptic ulcers or atrophic gastritis than from patients without these conditions. It was demonstrated that purified VacA or sonicates from Tox^+ strains (and not from Tox^- strains) induced gastric epithelial cell damage in a mouse (Ghiara *et al.*, 1995; Marchetti *et al.*, 1995), and that Tox^+ strains specifically inhibited epithelial cell proliferation (Ricci *et al.*, 1996). However, as Tox^- and Tox^+ strains could be isolated from patients with and without disease, respectively, the presence of VacA alone did not explain disease.

Following isolation of VacA the gene that encoded this toxin (*vacA*) was cloned and sequenced. Similarly to other *H. pylori* genes the *vacA* gene is characterised by considerable diversity. The *vacA* gene displays no sequence similarity to other known genes. Four different domains were identified in the VacA toxin, namely, a 33 residue N-terminal signal sequence (permits the export of the toxin from the cytosol to the periplasm), a 37-kDa region predicted to be rich in pleated β-sheets, a 58-kDa region and a C-terminus rich in amphipathic pleated beta segments. The VacA toxin is a secreted and cleaved protein. Its secretion and cleavage is analogous to that for the IgA 1 proteases of pathogens such as *Neisseria gonorrhoeae* (Schmitt and Haas, 1994). VacA is believed to bind to gastric epithelial cells whereupon it becomes internalised within an acidic vacuole and inhibits some aspect of intracellular vesicle trafficking, although the exact mechanism of VacA-induced toxicity remains unknown.

It was discovered that the *vacA* gene is present in all *H. pylori* isolates yet only ~50% express the VacA toxin (Cover and Blaser, 1992). Atherton *et al.* (1995) demonstrated that the *vacA* gene occurred as a mosaic structure, with diversity between strains arising within the signal-sequence and mid-region of the gene. The diversity between VacA proteins from different strains was believed to account for the fact that some strains of *H. pylori* expressing the VacA toxin do not cause disease. Four signal-sequence allelic subtypes (s1a, s1b, s1c and s2), and

two mid-region allelic subtypes were found (m1 and m2). Atherton *et al.* (1995) found that strains of the *vacA* s1 m1 type produced higher levels of the vacuolating cytotoxin *in vitro* than strains with any of the other *vacA* allelic type combinations. The s1 m1 genotype was found to be associated with increased gastric epithelial damage, enhanced gastric inflammation and duodenal ulceration. The s1 allele was also shown to be associated with peptic ulcer disease and the allelic type s1 m1 was shown to correlate with gastric carcinoma in Germany. Characterisation of the s2 allele has shown that a naturally occurring 12-amino acid hydrophilic N-terminal sequence blocks the vacuolating activity and that removal of this sequence confers an s1 allele-like activity (McClain *et al.*, 2001; Letley *et al.*, 2003). Although VacA research has predominantly been related to the mechanism of action of the toxin and the diseases associated with certain *vacA* alleles, recent reports have shown that *vacA* alleles are associated with particular geographical regions (Ito *et al.*, 1997; Pan *et al.*, 1998; Van Doorn *et al.*, 1999a; Kersulyte *et al.*, 2000; Kim *et al.*, 2002). Thus, the *vacA* gene can be used to characterise *H. pylori* populations in different parts of the world and indicate geographical relationships. This aspect is investigated in Chapter 3.

1.7.2.2. The cytotoxin-associated antigen (CagA).

The cytotoxin-associated gene (*cagA*) was originally thought to be necessary for VacA activity or expression based on its strong correlation with the *vacA* gene in clinical isolates. However, isogenic *cagA* mutants demonstrated that CagA is not necessary for the expression or the cytotoxic activity of VacA (Tummuru *et al.*, 1994; Xiang *et al.*, 1995). Further analysis of the *cagA* gene revealed that this gene was located within a 40-kb pathogenicity island [*cag* PAI (Odenbreit *et al.*, 2000)]. It is now known that the CagA protein is transferred to gastric epithelial cells via a type IV secretion apparatus (encoded by the *cag* PAI) whereupon it becomes tyrosine phosphorylated and induces a 'hummingbird' morphology in the epithelial cells (Stein *et al.*, 2000). This recently defined role of CagA has provided much awaited information relating to *H. pylori*-host interactions. In addition, the 3′-end of the *cag* PAI has proved to be an informative locus for studying geographic partitioning of *H. pylori* isolates and human migrations (Kersulyte *et al.*, 2000). A more in-depth description of CagA is outlined in Chapter 7.

1.7.3. Novel virulence determinants.

1.7.3.1. Superoxide dismutase and catalase.

H. pylori is known to induce acute and chronic inflammatory responses in the gastric mucosa. The inflammatory response consists largely of phagocytic cells (monocytes and neutrophils). Phagocytes engulf foreign material and release oxidative bursts, which are toxic to microorganisms, consisting of hydrogen peroxide (H_2O_2), superoxide anions (O_2^-) and hydroxyl radicals ($\cdot OH$). To combat this most microbes produce enzymes to detoxify these metabolites, namely, catalase ($2\ H_2O_2 \rightarrow 2H_2O + O_2$) and superoxide dismutase ($2O_2^- + 2H^+ \rightarrow H_2O_2 + O_2$). These enzymes can aid in the survival of microbes within neutrophils. Both superoxide dismutase (SodB) and catalase (KatA) have been detected in *H. pylori* (Odenbreit *et al.*, 1996b). Mutations in the *katA* gene have resulted in complete loss of catalase activity. Moreover, *H. pylori* has been shown to induce and survive phagocytic extracellular oxygen radicals via catalase (Ramarao *et al.*, 2000). As reactive oxygen metabolites are known to induce injury to the gastric mucosa, it is possible the *H. pylori* causes tissue damage via its catalase enzyme. Interestingly, a second copy of a catalase-like protein is present in the genome sequences of strains 26695 and J99 that is closely related to the catalase-like protein of *Synechococcus* sp. However, the function of this gene in *H. pylori* is as yet unknown. In addition, *sodB*-deficient mutants could not be generated suggesting that this enzyme is essential for survival of *H. pylori*.

1.7.3.2. *H. pylori* pore-forming cytolysin (TlyA).

Haemolysins are defined as bacterial toxins that lyse erythrocytes in *in vitro* tests by cell membrane disruption but are correctly referred to as cytolysins, as these toxins damage a variety of cells *in vivo* such as phagocytic cells, fibroblasts, lymphocytes and platelets. Cytolysins are produced in a number of different pathogens, including *E. coli*, *Staphylococcus aureus*, *Streptococcus pyogenes* and *Listeria monocytogenes*. Cytolysins are assumed to be important virulence factors. As *H. pylori* exhibits haemolytic activity when grown on unlysed blood agar, Martino *et al.* (2001) investigated the role of this cytolysin in the virulence of *H. pylori*. Analysis of the genome of the *H. pylori* strain 26695 revealed the presence of a gene with sequence identity to that of a pore-forming cytolysin (TlyA, HP1086). The introduction of the gene encoding this cytolysin (*tlyA*) conferred haemolytic activity to an *E. coli* haemolysin-deficient strain. It was also discovered that TlyA was necessary for *H. pylori* colonisation of a mouse model. Thus, TlyA can be considered a novel virulence factor in *H. pylori*.

1.7.3.3. The *H. pylori iceA* locus.

A potential virulence gene was detected in two adherent *H. pylori* isolates following attempts to uncover additional virulence loci within the *H. pylori* genome (Peek *et al.*, 1998). It was discovered that the expression of this gene was upregulated upon contact with gastric epithelial cells and was thus termed *iceA* (induced by contact with epithelium). Sequence analysis of this gene revealed two distinct variants, designated *iceA1* and *iceA2*. Only the *iceA1* gene was found to be induced by contact with epithelial cells. The *iceA1* gene possesses significant sequence identity to a type II restriction endonuclease from *Neisseria lactamica*, whereas the *iceA2* gene bears no sequence identity to known genes (Figueiredo *et al.*, 2000).

Peek *et al.* (1998) initially showed an association between the presence of the *iceA1* gene and disease, as ~50% of duodenal ulcer patients harboured the *iceA1* allelic variant and its presence was rare in non-ulcer patients. Another study was carried out using a slightly larger sample size. In this study an association between duodenal ulceration and the presence of the *iceA1* gene was confirmed. However, ~50% of the non-ulcer patients also harboured isolates with this allelic variant (van Doorn *et al.*, 1998b). To definitively determine whether *iceA* alleles had an association with disease a large study (over 400 strains) investigated the presence of the *iceA1* gene in duodenal ulcer, gastric ulcer, gastric cancer and gastritis (Yamaoka *et al.*, 1999b). This study encompassed different geographical regions to overcome the problem of regional variation of *H. pylori* isolates. The results found no association between the presence of the *iceA1* gene and clinical outcome in any of the countries tested (Korea, Japan, USA and Colombia). As there is currently no known biological or epidemiological evidence for a role for the *iceA* gene as a virulence factor, it is difficult to determine whether this locus is a true virulence determinant. However, as a report has demonstrated that the *iceA1* and *iceA2* gene regulate the transcription of the *hpyIM* locus (encoding a DNA methylase), future work may demonstrate a role for the *iceA* locus in virulence (Xu and Blaser, 2001).

1.8. Genetic diversity of *H. pylori*.

The genetic diversity observed among pathogenic bacteria mirrors their different lifestyles and physiological versatilities and evolves from adaptation to their niches or growth conditions. Indeed, one of the most distinctive features of *H. pylori* is the genetic diversity it displays between clinical isolates. The nucleotide sequence diversity of *H. pylori* exceeds that of many other bacterial species. Such genetic diversity allows for single clinical *H. pylori*

isolates to be discriminated from others by DNA fingerprinting techniques (Wang *et al.*, 1993, Marshall *et al.*, 1995, Marshall *et al.*, 1996) or sequencing of corresponding genes [a high degree of sequence divergence can be seen between orthologues (3–5%) and identical sequences are rare] (Achtman *et al.*, 1999, Owen and Xerry, 2003). In addition, *H. pylori* has been shown to have a panmictic or freely recombining population structure (Go *et al.*, 1996) and is naturally competent (Israel *et al.*, 2000). These characteristics can allow for intra-strain rearrangements and horizontal transfer of foreign genetic sequences that is a common occurrence in *H. pylori* isolates.

1.8.1. Bacterial population genetics

Since pathogenic bacteria exist as populations, the integration of population genetic, evolutionary and epidemiological studies can provide important insights into the origins and spread of bacterial diseases (Musser, 1996). Population genetics is defined as the study of the natural genetic diversity, its causes and consequences, within bacterial populations (Smith, 1995). Three main types of bacterial populations have been described (Smith *et al.*, 1993), namely, clonal, panmictic and epidemic [these population structures are described in greater detail in Chapter 3]. The relative contribution of recombination in the generation of new bacterial genotypes varies among these bacterial population types. Increasing the amount of recombination within a bacterial population decreases its clonality. Consequently, a spectrum of population structures exists. Few species conform to a highly clonal or completely non-clonal population structure, with most bacterial populations containing elements of both population structure types. The evolutionary history can easily be estimated by phylogenetic analysis in clonal bacteria as *de novo* mutations arising in a population are spread vertically (mutations only passed on to descendants of cells in which they arose). In contrast, the evolutionary history of panmictic bacteria is difficult to predict as genetic sequences acquired from external sources (horizontally) can be found that do not share the same ancestral history. Thus, as most bacterial populations do not fall under a completely clonal or completely panmicitc category, the ability to predict the evolutionary histories of bacterial populations can vary between species.

1.8.1.1. Population structure of *H. pylori*.

As stated previously *H. pylori* exhibits extensive genetic diversity between clinical isolates as revealed by DNA fingerprinting techniques. It has been demonstrated that every isolate from unrelated individuals has a unique multilocus genotype (Go *et al.*, 1996). In addition,

homoplasy tests have identified frequent recombination within *H. pylori* and identified it as a panmictic organism (Suerbaum *et al.*, 1998). It is believed that the diverse nature of the *H. pylori* genome is related to a lack of DNA mismatch-repair systems (Wang *et al.*, 1999) and the natural transformability of this organism (Hofreuter *et al.*, 2000). However, it has been argued that the genetic diversity associated with *H. pylori* is due to unusually high levels of recombination (Achtman and Suerbaum, 2000).

Suerbaum *et al.* (1998) also found identical *H. pylori* isolates in three members of the same family and suggested that *H. pylori* is clonal over a short period of time. Evidence of clonal transmission of *H. pylori* within families has also been detected in subsequent studies (Han *et al.*, 2000; Owen and Xerry, 2003). In addition, distinct *H. pylori* populations that correlate with geographical location within Dutch and Chinese *H. pylori* strains have been found upon examination of the *cagA* gene (van der Ende *et al.*, 1998). However, analysis of a housekeeping gene exhibited no clonal associations. Another study analysing the signal-sequence and mid-region of the *vacA* gene demonstrated geographical separation of *H. pylori* strains in East Asia and strains obtained from the Western hemisphere (Pan *et al.*, 1998). These apparent clonal groupings of *H. pylori* strains are based on virulence genes (*cagA* and *vacA*). Thus, it is likely that these clonal groupings are a result of the selection of antigenic variants that provide a selective advantage in these particular geographical regions.

Conversely, remnants of clonal decent from distant common ancestors (weakly clonal groupings) within *H. pylori* strains have been detected when analysing housekeeping genes by multi-locus sequence typing (MLST, Achtman *et al.*, 1999). Usually frequent recombination will obliterate evidence of clonal relationships within DNA sequences. This report implies that clonal decent can still be detected even within frequently recombining *H. pylori* strains and probably within other panmictic species.

1.8.1.2. Weakly clonal population structure.

Neisseria meningitidis causes meningitis and septicaemia worldwide. Meningococcal populations are highly diverse but only a small number of lineages cause disease (van Belkum, 2003). As a result of the concentration of attention on *N. meningitidis* isolates responsible for disease, the genetic diversity of this organism has been understated (Spratt and Maiden, 1999). MLST has shown strong linkage disequilibrium (clonal structure) in *N. meningitidis* (Caugant *et al.*, 1987). However, the linkage disequilibrium associated with this

organism decreases when a sample bias is corrected for. Moreover, unstable clones have been detected by MLST (Maiden *et al.*, 1998). The meningococcal population thus provides an example of a non-clonal population that exhibits clonality upon the emergence of strains with increased virulence. These clonal strains with increased virulence do not represent the entire *N. meningitidis* population. Thus, this organism has a mixed or weakly clonal population structure. In addition, *Streptococcus pneumoniae* and *Escherichia coli* have also been shown to have a clonal population structure that is influenced by recombination (Spratt and Maiden, 1999).

As clonal descent from distant common ancestors in *H. pylori* has been detected and as this organism has always been considered to have a panmictic population structure, it is likely that *H. pylori* may have a mixed or weakly clonal population structure. However, as disease-associated isolates are usually examined, the question as to whether these isolates are truly representative of the entire *H. pylori* species arises. Thus, the population structure of *H. pylori* may yet remain to be determined.

1.8.1.3. Geographical partitioning of *H. pylori* isolates.

It is widely accepted that the family is the core unit for *H. pylori* transmission. The restriction of transmission of *H. pylori* within families (parents to siblings, spouse to spouse or partner to partner, and parents to relatives) implies that the spread of infection occurs slowly as humans migrate between geographical locations. Thus, the biogeographical diversity of *H. pylori* isolates would reflect the original founder populations that were first introduced into particular areas (Suerbaum and Achtman, 2001). Biogeographical diversity has been reported for viruses (JC virus) and *Haemophilus influenzae* (Musser *et al.*, 1990; Agostini *et al.*, 1997) but has not yet been extensively reported for bacteria. It has been suggested that the biogeographical diversity of bacteria might also reflect differences in the genetic susceptibility or diet of the human hosts they reside in (Suerbaum and Achtman, 2001).

Initial geographical differences in *H. pylori* strains were observed in the sequences of the *cagA* genes between isolates from the Netherlands and China (van der Ende *et al.*, 1998). Subsequent analysis of the *cagA* gene revealed that East Asian *H. pylori* strains differed from strains obtained from East Asia, Europe, the Americas, Africa and India (Achtman *et al.*, 1999; van Doorn *et al.*, 1999b). Geographical partitioning of *H. pylori vacA* alleles was also revealed (van Doorn *et al.*, 1998a, 1999a). Further analysis of the *vacA* gene in different

geographical regions has shown that this locus could differentiate *H. pylori* strains from Europe / America, East Asia and India (Mukhopadhyay *et al.*, 2000). Similar, geographical partitioning has been detected upon analysis of the *hspA* gene (Ng *et al.*, 1999).

Caution must be taken when interpreting geographical partitioning based on the virulence loci described above (*cagA*, *vacA* and *hspA*) as region-associated diversity may arise due to different selection pressures associated with different human populations. However, geographical diversity of *H. pylori* housekeeping genes that are not prone to selection pressures could potentially reflect the original founder population. Initial analysis of the housekeeping *glmM* gene (phosphoglucosamine mutase) failed to reveal geographical partitioning of *H. pylori* strains (van der Ende *et al.*, 1998). However, MLST analysis of twenty *H. pylori* strains from twelve different geographical regions revealed partitioning of strains according to their sources (Achtman *et al.*, 1999). Achtman *et al.* (1999) demonstrated that Asian isolates segregated from isolates from the remainder of the world and were dubbed the 'Asian clone'. In addition an African American isolate and African isolates formed a distinct group (clone 2) and strains from Europe, the Americas and Australia formed a heterogeneous third group. It was suggested that these biogeographical groupings probably reflect founder populations that were physically isolated from each other until recently.

Although MLST reveals highly related isolates that contain identical alleles at seven different distinct loci, this typing technique cannot be extensively applied to *H. pylori* as it is a highly diverse organism that exhibits unique sequence types for each isolate. However, the above study has demonstrated that this technique is useful for revealing geographical partitioning of *H. pylori* isolates and demonstrating weak clonal lineages. It may prove to be useful in the future in determining the evolutionary history of this organism.

1.8.2. Mechanisms and origins of genetic diversity in H. pylori.

The availability of two compete *H. pylori* genome sequences has provided information regarding the genetic diversity of *H. pylori* isolates and the possible mechanisms responsible for its diversity. Generation of genetic diversity within a non-clonal bacterial species usually results from natural mutation, exchange of DNA within a particular bacterial cell or lineage (intra-strain rearrangements) or exchange of DNA between different strains (horizontal transfer), and in rare cases horizontal transfer occurs between different species.

1.8.2.1. Natural mutation in *H. pylori*.

Mutations arise from errors in copying normal template DNA and misreplication of DNA damage. Guanine depurination, guanine oxidation and cytosine deamination are leading examples of endogenous DNA damage that can elevate mutagenesis if unrepaired. Most bacteria have evolved a variety of DNA damage-repair mechanisms that act together to remove almost all of the DNA damage before replication. Analysis of the two sequenced *H. pylori* genomes shows that these strains lack genes homologous to DNA damage repair genes found in *E. coli* (Tomb *et al.*, 1997; Alm *et al.*, 1999). However, *H. pylori* does seem to encode the equivalent of the *uvrABC* nucleotide excision repair system (Wang *et al.*, 1999). Nevertheless, this lack of mismatch repair systems indicates that *H. pylori* is less effective at repairing DNA lesions than *E. coli*.

Although mutation rates are high in *H. pylori*, most of the nucleotide diversity is mainly due to synonymous substitutions (Alm *et al.*, 1999). Synonymous substitution occurs when a nucleotide change has no effect on the amino acid sequence. Analysis of the *flaA* and *flab* genes revealed that 15–21% of nucleotides at synonymous positions are different between pairs and only 0.3–2.5% at non-synonymous positions are divergent (Wang *et al.*, 1999). Thus, although high levels of divergence at the nucleotide sequence level are present in the *H. pylori* genome, proteins sequences are relatively conserved.

1.8.2.2. Recombination in *H. pylori*.

The sequence diversity in *H. pylori* is only partially due to extensive nucleotide variation. Evidence for extremely high levels of recombination affecting three different loci (*vacA*, *falA* and *falB*) has been shown (Suerbaum *et al.*, 1998). Shuffling of sequence diversity by extensive intragenic recombination is so frequent in *H. pylori* that it can generate a much larger number of unique sequences than in species where recombination is less frequent. DNA uptake is facilitated by the naturally transformable nature of *H. pylori* (Smeets and Kusters, 2002). Kersulyte *et al.* (1999) provided evidence for *in vivo* exchange of genetic sequences between two unrelated *H. pylori* strains. As approximately 10% of people in developed countries and 34% of people in developing countries are colonised by more than one *H. pylori* strain, ample opportunity exists for inter-strain recombination. In addition, transient *H. pylori* strains unable to colonise an infected host may provide DNA sequences to the resident *H. pylori* population as it passes through the host. Recombination in *H. pylori* may seem random. However, a recent study has demonstrated that integration of foreign genetic elements

acquired via natural transformation is limited by the presence of restriction-modification systems (Aras *et al.*, 2002).

1.8.2.2.1. Horizontal gene transfer.

Horizontal gene transfer is defined as the acquisition and incorporation of genetic material from another organism (Israel, 2001). Horizontal transfer is believed to be a significant factor in the evolution of microbes, resulting in genetic diversity. Indeed, this method of genetic exchange has a significant role in the evolution of *H. pylori* (Odenbreit and Haas, 2002). Horizontal transfer of genetic material can have an impact on the virulence of a strain as virulence factors, antigenic determinants and antibiotic resistance genes can be acquired by an organism via this mechanism (Salyers and Amabile-Cuevas, 1997; Maiden, 1998).

Evidence for genetic exchange between *H. pylori* strains *in vivo* has been detected (Kersulyte *et al.*, 1999). Indeed, the complete genome sequences of two *H. pylori* strains suggest that this organism may have incorporated genetic material from other organisms during the course of its existence. There are numerous regions of genetic sequences with different G+C% contents than the rest of the genome implying that these sequences were obtained from unrelated species (Tomb *et al.*, 1997; Alm *et al.*, 1999). Moreover, some predicted genes in *H. pylori* are more homologous to Gram-positive than Gram-negative bacterial species (Solnick *et al.*, 1997). Putative restriction-modification systems are also present in the sequenced *H. pylori* genomes that have strong sequence identities to those of other organisms (Xu *et al.*, 1997). These data strongly imply that the *H. pylori* genome is collectively made up of genetic sequences received from its own and other species.

Horizontal transfer of genetic sequences can occur via transformation, conjugation and tranduction. The primary mechanism responsible for horizontal gene transfer in *H. pylori* is transformation. Transformation is defined as the mechanism by which exogenous DNA is taken up by bacteria and the DNA becomes heritable (Israel, 2001). As *H. pylori* is naturally transformable or competent, DNA from its environment can be incorporated without special *in vitro* treatment. Indeed, it has been reported that up to 75% of *H. pylori* strains tested were naturally competent (Haas *et al.*, 1993; Tsuda *et al.*, 1993). For some time the mechanism involved in the natural competence of *H. pylori* remained unknown. However, a gene cluster encoding a type IV secretion system that is independent of the *cagA* secretion apparatus has been identified, namely, the *comB* gene cluster (Hofreuter *et al.*, 2001). This gene cluster has

been linked to natural transformation and competence in *H. pylori*. Subsequently, the genes essential for natural competence in *H. pylori* were identified (Odenbreit *et al.*, 1996a; Hofreuter *et al.*, 1998). Homology searches revealed that the ComB-associated proteins are homologues to the VirB proteins of *Agrobacterium tumefaciens* (Smeets *et al.*, 2002). The VirB proteins of *A. tumefaciens* are responsible for transfer of a tumour-inducing (Ti) plasmid into plant cells via a type IV secretion apparatus.

1.8.2.2.2. Intra-strain diversity in H. pylori.

Major rearrangements of the genetic material of the cell may be accompanied by translocation of fragments from one position of the genome to another. Intra-genomic rearrangements such as deletions, tandem duplication, inversions and translocations can occur in bacterial genomes. Many of these elements may arise as a result of the presence of repetitive DNA sequences (Marshall *et al.*, 1998). The two sequenced *H. pylori* genomes reveal a large number of repetitive DNA sequences (Tomb *et al.*, 1997; Alm *et al.*, 1999). In addition, several copies and partial copies of the insertion elements IS*605* and IS*606* have been found throughout the genome. Strain 26695 contains five complete copies of IS*605* while strain J99 lacks a complete copy (Alm and Trust, 1999). The number of copies and the presence of insertion elements throughout the *H. pylori* genome suggest that these elements may have led to recombinational events resulting in chromosomal rearrangements.

The identification of homopolymeric tracts and dinucleotide repeats in several genes in strains 26695 and J99 implies that 'slipped-strand repair mechanisms' modulate gene expression. It is believed that this mechanism may be involved in antigenic variation and adaptive evolution. Almost half of the genes that possess these repeat sequences encode outer cell envelope components of *H. pylori* (Alm and Trust, 1999). Thus, *H. pylori* could potentially alter the appearance of its outer membrane via this mechanism and thereby avoid the host immune system. The best known example of this phenomenon relates to the expression of Lewis X (Le^X) and Lewis Y (Le^Y) antigens on *H. pylori's* lipopolysaccharide (LPS). It has been reported that the expression of these antigens are not stable (Appelmelk *et al.*, 1998). This study demonstrated that the Le^X on the LPS of a *H. pylori* strain could change generating LPS variants in one bacterial cell population.

1.8.3. Genetic diversity in H. pylori *revealed by comparative and functional genomics.*

Although *H. pylori* is responsible for certain gastric diseases, less than 10% of people infected with *H. pylori* exhibit clinical symptoms. This fact suggests that there may be specific *H. pylori* strains that are responsible for its pathogenic effects. Indeed, several reports have demonstrated that certain genotypes of the *vacA* gene and the presence of the *cagA* gene are associated with an increased risk of peptic ulceration and gastric cancer, respectively (Atherton *et al.*, 1995, 1997; Atherton, 1997). However, these correlations vary depending the geographical region studied and attempts to correlate other *H. pylori* alleles with clinical diseases have failed.

So how may some *H. pylori* strains exhibit higher virulence than other strains? Current advances in functional and comparative genomics are helping to uncover this conundrum. Functional genomics uses the expression profiling of mRNA to provide a condition-dependent and time-specific genome scale portrait of an organism's transcriptome (Ferra and Brown 1999; Harrington *et al.*, 2000). Comparative genomics, on the other hand, contrasts two or more genomes at the open reading frame (ORF) content level (Ochman and Jones 2000; Ochman and Moran 2001). Both approaches apply *in silico* analysis to determine genetic functions.

As mentioned previously, *H. pylori* was the first bacterial species to have the complete genome of two unrelated clinical isolates [J99 and 26995] fully sequenced and compared (Tomb *et al.*, 1997; Alm *et al.*, 1999; Alm and Trust, 1999). As *H. pylori* is a panmictic organism (Go *et al.*, 1996) the question of whether isolates J99 and 26695 would accurately represent the genetic diversity found in clinical isolates was posed. Since these strains were isolated ten years apart on two different continents and from different gastric disorders, it has been argued that the genetic diversity exhibited by clinical isolates would be expected to be portrayed by them. Comparison of the two completely sequenced organisms revealed a significant genetic variation between genomes. For example, the genome of strain J99 is smaller (1,643,831 bp) than that of strain 26695 (1,667,867 bp) and has 57 less predicted ORFs (Tomb *et al.*, 1997, Alm *et al.*, 1999). Strains 26695 and J99 contain 110 and 52 strain-specific genes, respectively (Janssen *et al.*, 2001), approximately half of which reside within a locus termed the plasticity zone. A revised annotation and comparison of the two sequenced *H. pylori* genomes was recently carried out (Boneca *et al.*, 2003). This study led to the reclassification of coding sequences and the total number of hypothetical proteins was

reduced from 40% to 33%. A significant amount of size variation was also discovered between orthologous genes mostly due to natural polymorphisms.

Functional genomics utilising microarray technology has provided researchers with a powerful tool to investigate the genetic diversity of clinical isolates (Salama *et al.*, 2000), the transcriptional profiles of isolates grown under different conditions (Ang *et al.*, 2001), the identification of strain-specific and species-specific genes (Israel *et al.*, 2001a, Janssen *et al.*, 2001) and the diversity between strains giving rise to differing clinical illnesses (Maeda *et al.*, 2001). One of the most interesting revelations using microarray genotyping was the discovery that *H. pylori* isolates can undergo 'microevolution' and give rise to sub-species during prolonged colonisation of a single host (Covacci and Rappuoli; 1998, Bjorkholm *et al.*; 2001, Israel *et al.*, 2001b). The presence of stable sub-species within a single individual suggests an adaptation of a *H. pylori* population to specific host niches, facilitated by the plasticity of its genome. Bjorkholm *et al.* (2001) demonstrated that several loci differed within two genetically related isolates from the same host, one major difference being the presence of the *cag* pathogenicity island (*cag* PAI) in one isolate and not the other. As the *cagA* gene and *cag* PAI are markers for elevated virulence within strains, the excision of the *cag* PAI within a strain may indicate that reducing the virulence of a strain could be a favourable adaptation.

1.8.4. Quasi-species development.

As stated previously *H. pylori* isolates exhibit extensive genetic diversity. Although many explanations have been proposed to account for this phenomenon, the diversity exhibited by unrelated *H. pylori* isolates may also reflects variation and changes in host niches. These changes include environmental changes encountered upon entry into a new host, upon colonisation of different regions in the stomach (discussed in Chapter 5) and environmental changes associated with chronic infection. For instance, when an isolate leaves a host and enters a new one differences in host genetics and diets imply that an altered environment will be encountered. In addition, some individuals that carry a *H. pylori* infection develop an atrophic gastric condition that is associated with a marked reduction in pH (Kuipers *et al.*, 1995). Thus, an environmental change associated with a chronic *H. pylori* infection arises. All these factors may influence a *H. pylori* strain and cause it to modify its genome in order to adapt to a changing environment. Thus, a newly genetically modified strain can arise that is better suited to its new environment but still displays a genetic makeup that is highly similar to the founder strain it evolved from. The ability of *H. pylori* to rapidly acquire mutations and

accept foreign DNA sequences could provide a pool of variants that can be selected for optimal colonisation of particular gastric niches.

Kuipers *et al.* (2000) hypothesised that mutations frequently develop in the *H. pylori* genome over the course of prolonged colonisation of a host that can be detected during long-term assessment. This group demonstrated that *H. pylori* isolates taken from the same individual at 7 to 10 years apart displayed highly similar, but not identical, DNA fingerprints. This study employed randomly amplified polymorphic DNA (RAPD)-PCR and amplified fragment length polymorphism (AFLP)-PCR fingerprinting. Minor band differences between genetic fingerprints of *H. pylori* isolates obtained from the same stomach years apart were detected implying that mutations can occur in *H. pylori* strains over time. Another study using microarray analysis demonstrated that 10 recent *H. pylori* isolates obtained from the same individual were related to an archival J99 strain. However, each isolate was unique. Thus, within an apparently homogeneous *H. pylori* population genetic differences are present within single isolates. These observations are consistent with a model of continuous microevolution of *H. pylori* within a single host.

1.8.5. Genetic diversity of H. pylori *via microevolution.*

Initial colonisation of the gastric mucosa by a founder *H. pylori* strain leads to persistent infection (Marshall *et al.*, 1998). As a founder strain spreads throughout the stomach it encounters different selection pressures. Thus to survive and to meet these challenges *H. pylori* is able to generate genetic diversity within a bacterial population and to adapt to the new environment it encounters on entering a new host.

The occurrence of spontaneous mutations after serial passage in gnotobiotic piglets indicates that *H. pylori* can undergo host-specific adaptation (Akopyants *et al.*, 1995). This suggests that a founder strain of *H. pylori* may need to alter its genome to adapt to a new environment. This theory was confirmed in a study where RAPD analysis showed that genetic rearrangements arise from a founder strain during chronic infection (Kersulyte *et al.*, 1999). In addition, various DNA fingerprinting methods applied to two or more *H. pylori* isolates taken from the same patient have shown that the fingerprint profiles of such strains are highly similar with minor band differences (Marshall *et al.*, 1998; Kuipers *et al.*, 2000). This implies that two or more isolates recovered from one patient may share an ancestral relationship with

a founder strain but have undergone independent genomic alterations. This phenomenon has been termed microevolution (Marshall *et al.*, 1995; Marshall *et al.*, 1998).

It is well known that *H. pylori* is transmitted within the family unit (Dominici *et al.*, 1999), either within siblings (Goodman *et al.*, 2000), between mother and child (Elistur *et al.*, 1999) or between father and child (Bamford *et al.*, 1993; Wang *et al.*, 1993). This mode of transmission along with the knowledge of quasi-species development in *H. pylori* allows for a model for microevolution to be put forward. As a *H. pylori* strain is passed through the family unit it encounters slightly different environments upon entry into a new host. As the *H. pylori* strain adapts to its new environment its genome becomes modified due to its high mutation rate, natural competence and high level of recombination. Once a member of this family starts a new family the infection will be passed to spouse and siblings. Once again the *H. pylori* strain adapts uniquely to each new host. Following several generations of infection within families the genetic makeups of the original strain will have been modified considerably. As the adaptation of *H. pylori* to each new host is random and unique the genetic makeup of *H. pylori* strains derived from an ancestral strain will be different. Thus, *H. pylori* strains from different individuals will exhibit a high level of genetic diversity.

1.9. Project aims.

The present study was aimed at characterising the population structure of Irish *H. pylori* isolates and comparing Irish *H. pylori* populations with those from different geographical regions. This study also aimed to detect the presence of microevolution within paired *H. pylori* isolates obtained from the same individual and to analyse the genetic differences between microevolved isolates. Finally, this study set out to determine the path of transmission of *H. pylori* within the family unit.

Specific investigations were carried out:

- To characterise a collection of Irish *H. pylori* isolates at specific genetic loci using PCR.
- To determine the level of heterogeneity among Irish *H. pylori* isolates using fluorescent amplified fragment length polymorphism (FAFLP)-PCR and to compare their genetic fingerprints to fingerprints from isolates from different geographical regions.

- To detect microevolution or the presence of quasi-species, using PCR, random amplified polymorphic DNA (RAPD)-PCR and FAFLP-PCR, between paired *H. pylori* isolates obtained from the same individuals.
- To assess whether variations at the 3′-end of the *cagA* gene between microevolved *H. pylori* isolates affects the phosphorylation of the CagA protein.
- To determine the route of transmission of *H. pylori* within families using RAPD-PCR and FAFLP-PCR fingerprinting.

Chapter 2

Materials and Methods.

2.1. ***Helicobacter pylori* strains** ... 42
2.1.1. Nomenclature of H. pylori *isolates* ... 42
2.1.2. Primary isolation of H. pylori *from gastric biopsies and storage of isolates* ... 42
2.1.3. Culture of H. pylori ... 43
2.2. ***E. coli* strains** ... 43
2.3. **Routine molecular techniques** ... 43
2.4. **DNA isolation and analysis** ... 43
2.4.1. Extraction of genomic DNA from H. pylori ... 43
2.4.2. Small-scale preparation of plasmid DNA by alkaline lysis ... 44
2.4.3. PCR ... 45
2.4.4. Agarose gel electrophoresis ... 46
2.4.5. Southern hybridisation ... 46
2.4.5.1. Capillary blotting ... 46
2.4.5.2. Probe hybridisation ... 46
2.4.6. Oligonucleotides and nucleotide sequencing ... 47
2.5. **RNA analysis** ... 47
2.5.1. RNA extraction ... 48
2.5.2. Reverse transcriptase-polymerase chain reaction (RT-PCR) ... 48
2.6. **Fingerprinting of *H. pylori* isolates...** ... 48
2.6.1. Random amplified polymorphic DNA-PCR (RAPD-PCR) fingerprinting ... 48
2.6.2. Fluorescent amplified fragment length-PCR (FAFLP-PCR) fingerprinting analysis ... 49
2.7. **Cloning of genomic *H. pylori* fragments** ... 49
2.7.1. Preparation and transformation of competent bacteria ... 49
2.7.2. Cloning of H. pylori cagA *repeat sequences* ... 50
2.8. **Animal cell culture** ... 50
2.8.1. Aseptic technique ... 50
2.8.2. AGS cell culture ... 51
2.9. **AGS cell transfection** ... 52
2.9.1. Spe*I linearisation of plasmid DNA* ... 52
2.9.2. in vitro *SP6 RNA transcription* ... 52
2.9.3. Electroporation of AGS cells ... 52

2.10. Protein analysis 53
2.10.1. H. pylori cell lysate preparation 53
2.10.2. AGS cell lysate preparation 53
2.10.3. Sodium dodecyl-sulphate polyacrylamide gel electrophoresis (SDS-PAGE) 54
2.10.4. Transfer of separated proteins to nitrocellulose membranes 54
2.10.5. Western immunoblotting 55
2.10.5.1. VacA analysis 55
2.10.5.2. Tyrosine phosphorylated proteins 55

2.1. *Helicobacter pylori* strains.

2.1.1. Nomenclature of H. pylori *isolates.*

Helicobacter pylori strains that were isolated in the Moyne Institute were designated by the letters MI (for Moyne Institute) followed by a three or four digit number arbitrarily chosen from the hospital number accompanying the biopsy. Isolates lacking a MI prefix were isolated at the Adelaide / Meath Hospital. Where the letters A or C follow this number, they indicate that the isolate in question was obtained 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 biopsy sites. The study herein describes the analysis of 74 antral *H. pylori* isolates.

Thirty-one of these 74 isolates had a corresponding paired isolate, either another antral isolate or an isolate from the corpus. Within these thirty-one, the six antrum and corpus pairs 577A / 577C, 9613A / 9613C, MI610 / MI611, 218A / 218C, 605A / 605C and 222A / 222C represent isolates from two different families (Chapter 7). In addition, paired isolates MI506 / MI520, MI571 / MI575, MI517 / MI566 and MI512 / MI541 were obtained from antral biopsies from four patients 3, 6, 14 and 14 months apart, respectively.

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

Primary isolation of *H. pylori* strains was carried out by Dr. Sarah Beesley. Briefly, strains of *H. 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 was streaked, with the blood vessels facing upwards, using a sterile inoculating loop onto the surface of two plates containing Campylobacter Agar Base enriched with lysed or defibrinated horse blood (7% v/v) and containing vancomycin (10 μg/ml) and amphotericin B (5 μg/ml). The plates were then incubated in a microaerophilic 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 re-streaked 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., U.K.) and stored at –70 °C.

2.1.3. Culture of H. pylori.

H. pylori was routinely cultured on Campylobacter Agar Base 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 μg/ml 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 using a sterile inoculating loop. Agar plates were incubated at 37 °C in a microaerophilic atmosphere (5% O_2, 10% CO_2), created using a gas-generating kit for Campylobacters (Oxoid, Hampshire, U.K.) 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.

2.2. *Escherichia coli* strains.

Strains XL-1 Blue (Stratagene, Amsterdam, The Netherlands) and TOP10 (Invitrogen life technologies, Paisley, U.K.) were used as transformation recipient strains. The genotypes of these strains are as follows: XL-1 Blue, *supE44 hsdR17 recA1 endA1 gyrA46 thi relA1 lac* F'[*proAB*$^+$ *lacI*^{q} *lacZ*ΔM15 Tn*10* (Tet^r)], and TOP10, F$^-$ *mcrA* Δ(*mrr-hsdRMS-mcrBC*) Φ80*lacZ*ΔM15 Δ*lacX74 recA1 deoR araD139* Δ(*ara-leu*)7697 *galU galK rpsL* (Str^R) *endA1 nupG*. *E. coli* strains were cultured in L-broth and on L-agar 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, U.K.) at a concentration of 16 μg/ml.

2.3. Routine 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 manufacturer's instructions. Commercially purchased kits were also used for some applications and used according to the accompanying specifications.

2.4. DNA isolation and analysis.

2.4.1. Extraction of genomic DNA from H. pylori.

H. pylori strains, from which genomic DNA was isolated, were lawned on the surface of Campylobacter Agar plates containing horse blood and antibiotics. Genomic DNA was

extracted after incubation for 72 h at 37 °C using a modification of the method of Ausubel *et al.* (1987). The cells were harvested into 1 ml phosphate buffered saline (PBS) [137 mM NaCl, 2.7 mM KCl, 4.3 mM $Na_2HPO_4.7H_2O$, 1.4 mM KH_2PO_4, pH ~7.3] using sterile cotton swabs. Cells were pelleted by centrifugation at maximum speed (~15,000 r.p.m.) in a bench-top centrifuge 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 37 °C water bath 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% CTAB in 0.7 M NaCl] was then added, again followed by thorough mixing of the suspension after which a 10-min incubation at 65 °C took place.

To obtain the nucleic acid from the cell lysate an equal volume (~750 µl) of chloroform was then added. Vigorous shaking then mixed organic and aqueous phases. The organic phase containing precipitated polysaccharide and protein was separated from the nucleic-acid-containing aqueous phase by centrifugation at maximum speed in a bench-top centrifuge for 8 min. The upper aqueous phase was removed to a clean 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 maximum speed for 10 min. The pellet was washed with 1 ml 70% (v/v) ethanol followed by a further 5 min centrifugation. The ethanol was poured off carefully and the isolated DNA was routinely dried at room temperature and resuspended in 50 µl TE buffer. The DNA concentration was determined by agarose gel electrophoresis against lambda bacteriophage DNA of known concentration.

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

E. coli strains containing plasmids with cloned *H. pylori cagA* fragments were cultured in L-broth containing 10 µg ampicillin/ml. 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 15,000 r.p.m. for 30 s. The medium was 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 H_2O (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 15,000 r.p.m. 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 ethanol followed by vortexing. The mixture was held at room temperature for 2 min followed by centrifugation at maximum speed for 5 min. After removal of the supernatant fluid the pellet was rinsed with 1 ml 70% 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 µl TE buffer (pH 8.0) containing DNAase-free RNAase (20 µg/ml).

2.4.3. PCR.

PCR was performed on chromosomal DNA isolated from three-day-old agar cultures of *H. pylori*. The sequences of the oligonucleotides used as primers and their corresponding PCR product sizes are listed in Table 2.1. The signal sequences and mid-regions of the *vacA* gene, the *cagA* gene, the 3′-end of the *cagA* gene, the in-del motifs at the 3′-end of the *cag* PAI and the regions upstream of the *vacA* and *ribA* genes were amplified in separate reactions using *Taq* polymerase (Promega Corp., Madison, USA). The reactions were carried out in a final volume of 25 µl containing 1 µl DNA template (100 ng/µl), 2 mM $MgCl_2$, 40 pmol of each primer, 2.5 U *Taq* polymerase, 80 µM each of dATP, dCTP, dGTP, dTTP in 10 mM Tris-HCl buffer (pH 8.3), 50 mM KCl, 0.001% (v/v) gelatin and overlaid with 25 µl of mineral oil. PCR was preformed in a PTC-100 thermocycler (MJ Research, Nevada, USA) under the following conditions: 5 min at 95 °C, then 35 cycles of [1 min at 94 °C, 1 min at 50–52 °C (depending on primers used) and 1 min at 72 °C]. This was followed by 10 min at 72 °C. The reaction products were then stored at 4 °C. Representative PCR products for each allele or sequence variant were sequenced for confirmation of their identities (MWG-Biotech, Milton Keynes, UK), and sterile distilled water. DNA from *E. coli* strain XL-1 Blue and sterile distilled water were used as negative controls for PCR.

Table 2.1. Primers for genotyping used in this study.

Gene or DNA region amplified	Primer	Primer sequence (5′ - 3′)	PCR product size (bp)
cagA	CAGAF[a] CAGAR[a]	GATAACAGGCAAGCTTTTGAGG CTGCAAAAGATTGTTTGGCAGA	349
3′-end of *cagA*	CAG1[b] CAG2[b]	ACCCTAGTCGGTAATGGGTTA GTAATTGTCTAGTTTCGC	0 = 550 1 = 650 2 = 750 3 = 850
Upstream of *vacA*	VACAP-L1[c] VACAP-R1[c]	ACAAGGCACGATCTGGG GTGTGTTTGTTGTATTTCCAT	VacAP3 = 290 VacAP4 = 340
Upstream of *ribA*	RIBAP-L1[c] RIBAP-R1[c]	CATCGCTCTTATGGATC GCCTCAAGGGATACT	RibAP1 = 220 RibAP3 = 300
Signal sequence of *vacA*	VA1-F[d] VA1-R[d]	ATGGAAATACAACAAACACAC CTGCTTGAATGCGCCAAAC	s1 = 259 s2 = 286
Signal sequence s1a	SS1-F[d] VA1-R[d]	GTCAGCATCACACCGCAAC CTGCTTGAATGCGCCAAAC	190
Mid-region of *vacA*	VAG-F[a] VAG-R[a]	CAATCTGTCCAATCAAGCGAG GCGTCTAAATAATTCCAAGG	m1 = 570 m2 = 645
Extreme right end of the *cag* PAI:			
Total *cagA* RJ region	cagF4584[e] cagR5280[e]	GTTATTACAAAAGGTGGTTTCCAAAAATC GGTTGCACGCATTTTCCCTTAATC	IV =1200 IIIa=1000 Ia = 1000 IIIa = 700 II=700
Type IV specific	Xins.R[e] cagF4584[e]	CGCTCTCTAATTGTTCTAGGA same as above	520
Type II specific	IS606-1692[e] cagR5280[e]	CTAACAATTTGCCATTATGCTGT same as above	400
Type III specific	fcn unk[e] cagR5280[e]	TGGATTAAATCTTAATGAATTATCG same as above	320
Type I and IV specific	CagF4856[e] cagR5280[e]	GCGATGAGAAGAATATCTTTAGCG same as above	I = 350 IV = 400

[a] Primer sequences from Yamaoka *et al.* (1999)
[b] Primer sequences from Yamaoka *et al.* (1998)
[c] Primer sequences from Bereswill *et al.* (2000)
[d] Primer sequences from Atherton *et al.* (1995)
[e] Primer sequences from Kersulyte *et al.* (2000)

2.4.4. Agarose gel electrophoresis.

Genomic DNA, PCR products, RT-PCR products and digested genomic DNA were electrophoresed through agarose gels according to standard protocols (Sambrook *et al.*, 1989) using Tris-Borate-EDTA buffer at a 0.5 × final concentration [1 litre 10 × stock, 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, 2 M Tris base, 57.1 ml glacial acetic acid, 0.1 M $Na_2EDTA.2H_2O$, pH ~8.5] where appropriate. Agarose gels were routinely cast with agarose 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 by exposure of the gel to ultra-violet light. Horizontal gel tanks (Gibco BRL, Life Technologies) of various sizes were routinely used.

2.4.5. Southern hybridisation.

2.4.5.1. Capillary blotting.

Total genomic DNA (10 μg), cleaved with restriction endonuclease *Hin*dIII, was fractionated through a 1.8% agarose gel. Gels were then treated as follows prior to transfer. Each gel was 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 H_2O 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 maintaining gentle shaking.

The digested DNA was then transferred to a positively charged nylon membrane (Hybond-N+, Amersham) by capillary blotting using 20 × SSC [3 M NaCl, 0.3 M Sodium Citrate], as previously described (Southern, 1975). Each gel was turned upside down so that the restricted DNA fragments present in the gel would only travel a short distance, as the DNA was located at the bottom of each gel. Blotting took place overnight after which time the transfer system was disassembled. The transferred DNA was fixed to the membrane by heating at 80 °C for 2 h.

2.4.5.2. Probe hybridisation.

DNA probes labelled with Digoxigenin (DIG, Roche Diagnostics, Mannheim, Germany) were used throughout these studies. Probes were labelled by PCR DIG-labelling in PCR reactions which incorporated DIG-dUTP into the synthesised DNA. Primer sequences and PCR product

size used for probe generation are shown in Table 2.2. Properly labelled probes migrated at slower rates through agarose gels than the native unlabelled PCR product; this characteristic was used to confirm adequate labelling of probes. Hybridisation of the probe to the membrane was carried out in Standard Hybridisation Buffer [5 × SSC, 0.1% (w/v) *N*-lauroylsarcosine, 0.02% (w/v) SDS, 1% (w/v) Blocking reagent (Roche Diagnostics, Mannheim, Germany)] when high stringency temperatures were utilised (65–68 °C). Hybridisation took place overnight following prehybridisation for 2 h at the same temperature and in the same buffer used for hybridisation. Detection of the chemiluminescent signal from labelled hybridised probes was carried out after washing of the membrane twice in 2 × SSC-0.1% SDS for 10 min each, followed by two washes of 20 min each at 65–68 °C in 0.5 × SSC-0.1% 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 antibodies 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 CDP-Star chemiluminescent substrate (Roche Diagnostics, Mannheim, Germany) 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.

2.4.6. Oligonucleotides and nucleotide sequencing

Oligonucleotides were synthesised by MWG-Biotech, Milton Keynes, England. Table 2.1 shows a list of the oligonucleotides used in this study. Automated nucleotide sequencing of PCR products was carried out at MWG-Biotech.

2.5. RNA analysis.

For RNA analysis certain precautions were taken since ubiquitous RNAase enzymes easily degrade RNA. RNAase enzymes are resistant to autoclaving but can be inactivated by diethylpyrocarbonate (DEPC; 0.1 % (v/v), Sigma-Aldrich, Dorset, UK). Thus all solutions used for RNA work were treated with DEPC. Solutions containing amines such as Tris cannot be DEPC-treated directly as DEPC is inactivated by these chemicals. These solutions were prepared in DEPC-treated water. Gloves and disposable sterile plastics were used at all times and any glassware used was baked overnight at 200 °C. Following the extraction procedure

Table 2.2. Genes analysed by Southern blotting and primers used to generate probes.

Gene analysed and encoded protein	Primer sequence (5′ - 3′)	Approximate PCR product size (bp)
flhF (flagellar biosynthesis protein)	F-GGAAACGGCCGCTGAA R-GAGGGATATTGACAGAATTA	500
hsdR (type I restriction enzyme)	F-GCCTGAAAGCATGGCTA R-TGATGGAATCTGGCGCAT	600
Omp / HP0492 (outer membrane protein)	F-TGCGAGCGTGGGTTTTGA R-TGCTTATACCCTACGCTAA	450
gmd (GDP-mannose dehydratase)	F-GGCAAGACGGGAGCT R-TGCCATTAACCGCAAACA	450
napA (neutrophil activating protein A)	F-CGCGGATCCTTAAGCTAAATGGGCTTC R-CCCAAGCTTGGATAGCGATCAAGGTTT	450
galE (UDP-galactose 4-epimerase)	F-ATGGCATTATTATTCACAGG R-GCTCCATAAGGATTAATGGG	400
vacA (vacuolating cytotoxin)	VA1-F-ATGGAAATACAACAAACACAC VAG-R-GCGTCTAAATAATTCCAAGG	2000

manipulations with RNA were carried out quickly and on ice to prevent degradation of RNA by endogenous RNAases.

2.5.1. RNA extraction.

Total *H. pylori* RNA was isolated using TRI reagent® (Sigma-Aldrich, Dorset, UK) and the manufacturer's instructions were followed. Bacterial lawns were grown and harvested as described for Southern blotting. The TRI reagent (1 ml) was added to the cells and left at room temperature for 5 min. Chloroform (200 µl) was then added to the solution and mixed. The mixture was centrifuged for 15 min at 4 °C. The upper aqueous layer was removed to a fresh microfuge tube. Isopropanol (500 µl) was added and the solution was left at room temperature for 10 min to precipitate the RNA. The RNA was then pelleted by centrifugation for 10 min at 4 °C. The RNA pellet was washed with 1 ml of 75% (v/v) diethylpyrocarbonate (DEPC)-treated ethanol. The pellet was air-dried and resuspended in 50 µl of DPEC-treated water. The RNA sample was fractionated on a 1% agarose gel to detect the presence of intact RNA (Fig. 2.1). The presence of 5S, 16S and 23S ribosomal RNA bands on an agarose gel indicated an un-degraded RNA sample.

2.5.2. Reverse transcriptase-polymerase chain reaction (RT-PCR).

Contaminating DNA was removed from RNA preparations by DNAase treatment using RNAase-free DNAase I (Roche Diagnostics, Mannheim, Germany). *H. pylori* RNA (1 µg) was uncoiled by incubation at 65 °C for 5 min. First strand cDNA synthesis of the mRNA was carried out using AMV reverse transcriptase (Finzymes, Finland). The reactions were completed in a final volume of 20 µl containing 1 µg RNA, 40 pmol of each reverse primer concerned, 1 × reaction buffer and 10 mM dithiothreitol (DTT). The reaction was incubated at 42 °C for 1 h for cDNA synthesis and then at 75 °C for 10 min to inactivate the reverse transcriptase. The PCR was carried out as above using primers P1 / P2 as a reference and VAG-F / VAG-R and VA1-F / VA1-R (Table 2.1) to detect the presence of *vacA* mid-region and signal-sequence cDNA, respectively.

2.6. Fingerprinting of *H. pylori* isolates.

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

H. pylori isolates were fingerprinted by RAPD-PCR (Akopyanz *et al.*, 1992) using different RAPD primers (Table 2.3). RAPD-PCR reactions took place in a final volume of 25 µl that included 10 × reaction buffer, 3 mM $MgCl_2$, 10 pmol each primer, 0.4 mM each of dATP,

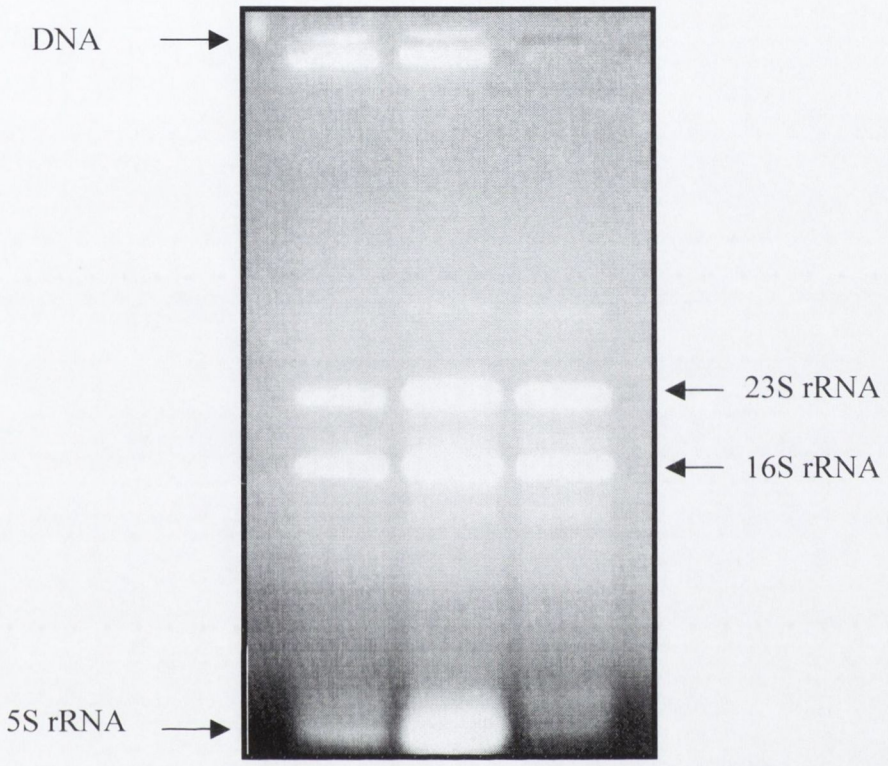


Fig. 2.1. Total cellular RNA from *H. pylori* isolates run on a 1.8% agarose gel.

Total cellular RNA was extracted using TRI^{TM} solution. Three bands representing 5S, 16S and 23S rRNA are indicative of an intact RNA preparation. Contaminating DNA must be removed in order to carry out RT-PCR.

Table 2.3. RAPD-PCR fingerprinting primers used in this study.

Primer	Primer sequence (5′ - 3′)
PI	CCGCAGCCAA
PII	AGTGAATTCGCGGTGAGATGCCA
PIII	AACGCGCAAC
PV	GAGCGGCCAAAGGGAGCAGAC

dGTP, dCTP, dTTP, 0.5 units of *Taq* DNA polymerase (Promega Corp., Madison, USA) and used 1 µl genomic DNA as template (~50–100 ng/µl). The cycling parameters routinely employed for RAPD primers >17 bases long (PII and PV, Table 2.3) 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. For RAPD primers 10 bases long (PI and PIII, Table 2.3) the cycling parameters were as follows: 4 cycles of 94 °C for 5 min, 36 °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, 36 °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. When primers >17 bases and 10 bases in length were used in combination (PIII and PV) the cycling parameters for the >17 base long primers were used. RAPD-PCR products were electrophoresed through 1.8% (w/v) agarose gels containing ethidium bromide (0.5 µg/ml) and fingerprints visualised by exposure of the gel to ultraviolet light.

2.6.2. Fluorescent amplified fragment length polymorphism-PCR (FAFLP-PCR) fingerprinting analysis.

FAFLP-PCR fingerprinting was carried out as described by Ahmed *et al.* (2002). Details of this method can be found in Chapter 4.

2.7. Cloning of genomic *H. pylori* fragments.

2.7.1. 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 ~16 h at 37 °C. After overnight incubation, 500 µl of this culture was inoculated into 100 ml pre-warmed L-broth and incubated at 37 °C with shaking until an OD_{600nm} of 0.4 was reached. After brief centrifugation (3,000 r.p.m. for 3 min) the cell pellets were resuspended in 10 ml of an ice-cold solution containing 80 mM $CaCl_2$ and 50 mM $MgCl_2$ 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 $CaCl_2$ and 1 ml 50% glycerol and stored as 80-µl aliquots at –70 °C. For transformation, 80 µl of 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. Of SOC

medium [2% (w/v) tryptone, 0.5% (w/v) yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, 20 mM glucose] 400 µl 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 contained on the vector.

2.7.2. Cloning of H. pylori cagA *repeat sequences.*

The repeat sequences of two paired *H. pylori* isolates were amplified by PCR. Each amplicon was digested with *Bam*HI and ligated into a SFV-1 vector. This procedure and strategy for cloning of these fragments is described in greater detail in Chapter 7. The SFV-1 viral vector was a gift form Prof. Gregory Atkins (Moyne Institute, Dublin). Since site-directed cloning was not used, shrimp alkaline phosphatase (SAP, Promega Corp., Madison, USA) was used to prevent the SFV-1 vector from religating without accepting a *H. pylori* DNA sequence. Briefly, 1 µg of *Bam*HI-digested SFV-1 vector was purified using a GENECLEAN® kit (BIO 101, CA). This template vector was incubated with 1 U of SAP (Promega Corp., Madison, USA) in a 1 × reaction buffer [50 mM Tris-HCl (pH 9.0 at 37 °C), 10 mM $MgCl_2$] at a final volume of 50 µl at 37 °C for 15 min. The reaction mixture was then heated to 65 °C for 15 min to inactivate the SAP. The reaction mixture was centrifuged for 1 min at 15,000 rpm. Approx. 1–2 µl from the top of the reaction mix was used for cloning *H. pylori* DNA sequences.

Purified and *Bam*HI-digested PCR products (*cagA* repeats) were incubated with 1 µg of the linearised dephosphorylated SFV-1 vector in a ratio of 10:1. Ligation of vector and PCR products was carried out using T4 DNA ligase. Briefly, 1 µg of PCR product was incubated with 0.1 µg of SFV-1 vector and 1 U of T4 DNA ligase (Promega Corp., Madison, USA) in a 1 × reaction buffer [30 mM Tris-HCL (pH 7.8), 10 mM $MgCl_2$, 10 mM DTT, 1 mM ATP] at a final volume of 10 µl at room temperature overnight. The ligation reactions were then transformed into an *E. coli* XL-1 blue or TOPO10 strain. The transformation mixture was plated out on L-agar plates containing the appropriate antibiotic and incubated at 37 °C overnight. Any colonies that grew were screened for SFV-1 vectors containing inserts.

2.8. Animal cell culture.

2.8.1. Aseptic technique.

The use of sterile technique in cell cultivation is essential to prevent bacterial and fungal infection. The principle of aseptic technique is to keep sterile the internal areas of culture

flasks, bottles and any implements/plastics that cells may be exposed to. The following aseptic technique procedures were adhered to for all cell culture manipulations.

All glassware, pipette tips, dissection instruments, deionised H_2O and microfuge tubes (Sarstedt, Leicester, England) were wrapped in aluminium foil and autoclave tape and then autoclaved at 121 ^{o}C for 20 min (Priorclave Ltd., Model EH150, London, England) before being wiped down with 70 % ethanol and placed in the laminar flow hood. All equipment used in the dissection procedure was oven baked at 200 ^{o}C for 1 h prior to usage, ensuring sterility.

All cell culture work was carried out in a laminar flow hood (laminar flow workstation, Astec-Microflow, Florida, U.S.A). Air passes through HEPA (high efficiency particle air) filters at the top of the flowhood and flows downwards. The airflow creates a downward barrier in front of the open portion of the hood, the strength of this down-draft of air preventing entry of external airborne contaminants into the laminar flow hood thus providing a sterile work area. Before using the laminar flow hood, the interior was sterilized by wiping down all accessible surfaces with 70% ethanol, followed by a 10-min exposure to ultraviolet (UV) light. Disposable latex gloves (sprayed with 70% ethanol) were worn at all times when cell manipulations were being carried out in the hood. Gloves were changed regularly to avoid contamination.

2.8.2. AGS cell culture.

Animal cells used for heterologous protein expression were human gastric adenocarcinoma (AGS) cells. These cells were originally obtained from American Type Culture Collection (ATCC, Rockville, MD, USA). AGS cells were grown in 75-cm^2 plastic tissue culture bottles and incubated in Dulbecco's Modified Eagle Medium (DMEM, Invitrogen life technologies, Paisley, U.K.) supplemented with 10% (v/v) foetal calf serum (FCS [Life Technologies, Inc., Paisley, Scotland]), penicillin (100 U/ml) and streptomycin (100 U/ml) (Life Technologies, Inc., Paisley, Scotland) at 37 °C with 5% CO_2. AGS cells were cultured until confluent and then passaged at a 1:3 ratio using trypsin to detach cells.

2.9. AGS cell transfection.

2.9.1. Spe*I linearisation of plasmid DNA.*

For the production of recombinant SFV viral RNA, the recombinant pSFV-1 plasmid was first linearised with an *Spe*I restriction enzyme. This is a unique restriction site in the infectious clone preceding the non-structural protein genes. A total of 10 μg SFV-1 plasmid DNA was linearised in final reaction volumes of 20 μl, containing 2 μl *Spe*I buffer (50 mM NaCl, 10 mM Tris-HCL, 10 mM $MgCl_2$, 1 mM DTT, pH 7.9), and 15 U *Spe*I. As the *Spe*I site could also be found within the cloned fragments a partial digestion of each clone was carried out. This entailed digesting each clone for three different lengths of time (10, 20 and 30 min) at 37 °C. These differently timed reactions were then pooled. Linearised plasmid DNA was brought to a final volume of 100 μl. The DNA was cleaned using a Qiagen nucleotide extraction kit (Qiagen, Germany), and plasmid DNA was resuspended in a final volume of 40 μl sterile dH_2O. Plasmid DNA concentration was assessed by mixing a 1-μl aliquot with 1 μl loading buffer and running this on a 0.8% (w/v) agarose gel. The concentration of plasmid was estimated from this gel by comparing band intensity with that of 5 and 10 μl lambda (*Hin*dIII- and *Eco*RI-digested) DNA molecular weight markers.

2.9.2. In vitro *SP6 RNA transcription.*

For the production of recombinant SFV viral RNA, SP6 RNA polymerase was used to initiate *in vitro* RNA transcription from *Spe*I-linearised recombinant SFV-1 plasmid DNA. Standard reaction mixtures contained 0.5 μg DNA template, 1 x SP6 buffer [40 mM *N*-2-hydroxyethyl-piperazine-*N'*-2-ethansulphonic acid-KOH (Hepes-KOH), pH 7.4, 6 mM MgOAc, 2 mM spermidine-HCl], 1 mM $m^7G(5')ppp(5')G$, 5 mM dithiothreitol, 1 mM each rATP, rCTP, rUTP, 500 μM rGTP, and 10 U recombinant Rnasin (Promega Corp., Madison, USA). The reactions were incubated at 37 °C for 1 h and transcripts were analysed by electrophoresis on a 0.6% (w/v) agarose gel.

2.9.3. Electroporation of AGS cells.

AGS cells were grown in 75-cm^2 tissue culture flasks until ~80% confluent. Cell monolayers were washed twice with PBS, and then incubated with 0.25% (w/v) trypsin / EDTA at 37 °C until detachment was evident. The flask tapped to complete detachment and 6 ml medium added to terminate trypsinisation and to resuspend the cells. Cells were centrifuged (170*g*, 10 min), and pellet resuspended in 6 ml PBS. The cells were centrifuged (170*g*, 10 min) and resuspended in a final volume of 700 μl PBS. *In vitro* transcribed RNA was then added to

suspended cells, and the mixture incubated on ice for 15 min in a 0.4 μm electroporation cuvette. The cuvette was then electroporated at 0.85 kV and 25 μF capacitance, using a BioRad Gene Pulser 11. Following 2 pulses, cells were replaced on ice for a further 15 min, before resuspension in 20 ml DMEM (Invitrogen life technologies, Paisley, U.K.) at 37 °C. The cells were immediately placed cell culture flasks, and maintained at 37 °C in a humidified atmosphere of 5% CO_2 for the indicated timepoints.

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 (Complete™, Roche Diagnostics, Mannheim, Germany) 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 15,000 r.p.m. 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 individual cellular protein molecules. Unused lysates were stored at –20 °C.

2.10.2. AGS cell lysate preparation.

Following electroporation of AGS cells each flask containing transfected cells was washed twice with PBS. Then 1 ml of trypsin was added to the flask followed by incubation at 37 °C with 5% CO_2 for 5 min to detach the cells. To the flask of cells 5 ml PBS was added and then the PBS and the cells were transferred to a universal tube. The cells were spun at 1,500 r.p.m. for 5 min. The supernatant fraction was poured off leaving a pellet of cells. The cells were washed again with 5 ml PBS as before. Once the cells had been washed they were kept on ice at all times. To the cell pellet 200 μl lysis buffer (25 mM HEPES, 5 mM $MgCl_2$, 5 mM DTT, 5 mM EDTA [pH 7.4] and Complete™ cocktail of protease inhibitors) was added. The cells in lysis buffer were passed through a 0.25-μm syringe to physically break up the cells. The lysate was spun at 15,000 r.p.m. to remove cell debris. The supernatant fluid was transferred to a fresh Eppendorf tube and stored at –70°C. The protein concentration of each sample was

assessed using a Bradford™ assay (Sigma-Aldrich, Dorset, UK). Of each sample 20 μg of protein was run on SDS gels.

2.10.3. 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'*-tetramethylethylenediamine (TEMED). Stacking gels, consisting of 5% (w/v) 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:H_2O (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.

2.10.4. Transfer of separated proteins to nitrocellulose membranes.

When the polypeptides separated by SDS-PAGE were to be analysed by Western immunoblotting, 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). When transferring AGS cell protein Immobilon™-P Polyvinylidene Difluoride (PVDF) filters (Sigma-Aldrich, Dorset, UK) were used. PVDF filters needed to be soaked in 100% methanol for 15 s in distilled H_2O for 2 min and then equilibrated in transfer buffer as before. After transfer was completed, any proteins remaining in the gel were stained with Coomassie brilliant blue.

2.10.5. Western immunoblotting.

2.10.5.1. VacA analysis.

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 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), 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 murine monoclonal anti-VacA antibodies (Dundon, Italy). The anti-VacA primary antibody was diluted 1:500 in a solution of 0.5% (w/v) Marvel in PBS.

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^{TM} dried milk poweder in PBS. Rabbit anti-mouse IgG was used as the secondary antibody to detect the murine monoclonal antibodies, or goat anti-rabbit IgG as secondary to the anti-CagA serum, both supplied by Dako). 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% 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, Illinois, USA) on filters exposed to X-Ray film.

2.10.5.2. Tyrosine phosphorylated proteins.

Probing PVDF membranes with anti-phosphotyrosine (PY) antibodies was carried out in much the same way with the following exceptions. Dried milk could not be used as a blocking agent as milk contains phospholipids. Thus, the anti-PY antibodies would attach to the phospholipids and not to the membrane; therefore, bovine serum albumin (BSA) was used. Primary and secondary antibodies (anti-PY) were diluted 1:500 in 0.2% BSA. Overnight blocking took place in 2% BSA.

Similar to dried milk, PBS contains phosphate that the primary anti-PY antibody would attach to. Thus, Tris buffer saline (TBS) was used for all washing and dilution steps.

Chapter 3

The population structure of Irish Helicobacter pylori *isolates –*
Genotyping putative virulence genes and intergenic regions within Irish H. pylori *isolates.*

3.1. Introduction 59

3.2. Methods 63

3.3. Results 63

3.3.1. vacA *Genotyping* 63

3.3.2. cagA *analysis* 64

3.3.2.1. Presence of the *cagA* gene 64

3.3.2.2. Detection of the number of repeat sequences at the 3′-end of the *cagA* gene 64

3.3.2.3. Phylogenetic analysis of the repeat sequences at the 3′-end of the *cagA* gene 65

3.3.2.4. Insertion-deletion (in-del) analysis within the *cag* PAI 65

3.3.3. Genetic Diversity within intergenic regions in Irish H. pylori *isolates* 66

3.3.3.1. Diversity preceding the *vacA* gene 66

3.3.3.2. Diversity preceding the *ribA* gene 67

3.4. Discussion 68

3.4.1. vacA *genotyping* 68

3.4.2. cagA analysis 69

3.4.2.1. Frequency data 69

3.4.2.2. Repeat sequences at the 3′-end of the *cagA* gene 70

3.4.2.3. Phylogenetic analysis of the repeats located at the 3′-end of the *cagA* gene 71

3.4.2.4. Analysis of the motifs located at the 3′-end of the *cag* PAI 71

3.4.3. Sequence variations preceding the ribA *gene* 72

3.1. Introduction.

Bacterial population genetics is the study of the natural genetic diversity, its causes and consequences, within bacterial populations (Smith, 1995). Understanding the genetic variability among bacterial species can potentially explain the diverse clinical diseases between bacterial isolates of the same species (Atherton, 1997). Different isolates of the same bacterial species have shown significant variation in their ability to cause infection and disease (Joyce *et al.*, 2002). The genetic variability of different bacterial species using differing fingerprinting techniques, such as ribotyping, RFLP-PCR, MLST, AFLP and RAPD-PCR, has been well documented (Marshall, 1995; Rasmussen *et al.*, 1999; Schouls *et al.*, 2003).

Genetic diversity within different bacterial species can vary, with strains within certain species being genetically similar whilst strains within other species being genetically dissimilar. Genetically different strains within the same species can be easily distinguished by simple fingerprinting methods. The variability within bacterial population structures has be categorised into three types (Smith *et al.*, 1993): clonal, panmictic and epidemic. Clonal species show similar genetic profiles between clinical isolates and their evolutionary history can easily be estimated by phylogenetic analysis. An example of a clonal organism would be *Mycobacterium tuberculosis*. Panmictic species are highly genetically diverse due to recombination and horizontal genetic exchange. It is difficult to trace the evolutionary history of panmictic organisms due to the integration of foreign DNA into their chromosomes. *Neisseria gonorrhoeae* and *Helicobacter pylori* are examples of panmictic organisms. Epidemic population structures are cases where a clone from a panmictic organism can infect a host and spread so quickly through a host population that it displays a clonal population structure. *Neisseria meningitidis* is an example of an organism with an epidemic population structure.

As previously mentioned *H. pylori* has been shown to display a panmictic population structure (Go *et al.*, 1996). The genetic diversity between *H. pylori* isolates is a result of recombination, horizontal gene transfer (Go *et al.*, 1996) and the lack of DNA mismatch repair systems (Alm *et al.*, 1999). DNA fingerprints vary considerably between *H. pylori* isolates, and in fact, unrelated *H. pylori* clinical isolates display very different genetic fingerprints (Marshall, 1995). Although genetic typing methods can demonstrate the variability between isolates, the location and identity of this variability cannot be determined.

Therefore, studying the genetic diversity of isolates at specific locations within an organism's genome can be more informative regarding the virulence (Atherton *et al.*, 1995) or origin (Achtman *et al.*, 1999) of the isolate. Tying *H. pylori* at specific genetic loci has been extensively studied (Blaser *et al.*, 1995; Atherton *et al.*, 1997; Yamaoka *et al.*, 1998; Kersulyte *et al.*, 2000).

The genetic diversity within the *H. pylori* genome is partly due to recombination. Evidence of this can be seen in the mosaic structure of the *vacA* gene (Atherton *et al.*, 1995). Mosaicism within a gene is where the genetic diversity occurs in blocks and in many bacterial species mosaicism is a result of recombination, which can be recognised by inspection of sequences. Mosaic genes, comprising of DNA derived from the host and other bacteria (often members of a different species), can express variant proteins that confer resistance to antibiotics, i.e., penicillin-resistant penicillin-binding proteins (PBPs) in *Streptococcus pneumoniae* and the pathogenic *Neisseria* species (Maiden *et al.*, 1998).

The *vacA* gene, which encodes the vacuolating cytotoxin, is present in all *H. pylori* strains, but only 50% of strains produce an active toxin (Cover and Blaser, 1992). The divergent regions that are indicative of a mosaic structure occur within the signal sequence and mid-region of the *vacA* gene (Atherton *et al.*, 1995). The signal-sequence presents as four allelic subtypes (s1a, s1b, s1c and s2), and the mid-region as two allelic subtypes (m1 and m2, Fig. 3.1). Atherton *et al.* (1995) found that strains of the *vacA* s1 m1 type produced higher levels of the vacuolating cytotoxin *in vitro* than strains with any of the other *vacA* allelic type combinations. The s1 m1 genotype was associated with increased gastric epithelial damage, enhanced gastric inflammation and duodenal ulceration. The s1 allele has also been shown to be associated with peptic ulcer disease (Strobel *et al.*, 1998) and the allelic type s1 m1 has been shown to correlate with gastric carcinoma in Germany (Miehlke *et al.*, 2000). Characterisation of the s2 allele has shown that a naturally occurring 12-amino acid hydrophilic N-terminal sequence blocks the vacuolating activity and that removal of this sequence confers an s1 allele-like activity (McClain *et al.*, 2001; Letley *et al.*, 2003). Although the alleles within the *vacA* locus have been shown to predict clinical outcome in some studies, correlation of *vacA* allelic type and clinical outcome cannot be applied in certain geographical regions (Pan *et al.*, 1998; Fernando *et al.*, 2002; Smith *et al.*, 2002b). In addition, predominant *vacA* alleles have also been shown to distinguish *H. pylori* isolates in different geographical regions (Table 3.1).

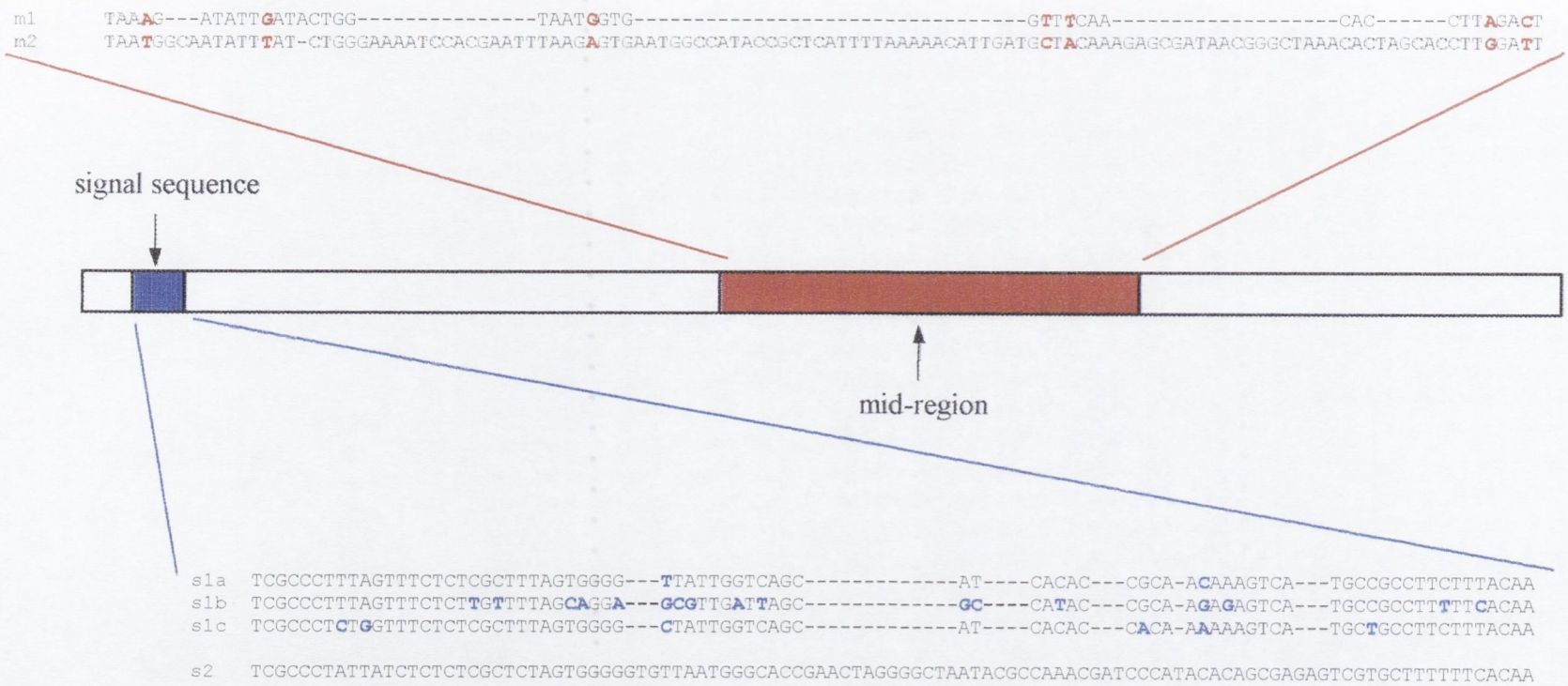


Fig. 3.1. Schematic representation of the mosaic structure of the *H. pylori vacA* gene.

The signal sequence occurs as four allelic variants, s1a, b, c and s2, and the mid-region occurs as two main allelic variants, m1 and m2. The nucleotide diversity between the signal sequence and mid-region alleles is indicated (nucleotide variations are shown in bold). Strains with the *vacA* allelic signal sequence and mid-region combination s1 m1 produce higher levels of the vacuolating cytotoxin *in vitro* and are associated with increased gastric epithelial damage, enhanced gastric inflammation and duodenal ulceration (Atherton *et. al*, 1997).

Table 3.1. *H. pylori* genotypes and their distribution throughout the world.

Gene	Predominant genotype	Geographic distribution (%)
vacA	s2 m2	France[a] (22%) Italy[a] (22%)
	s1a m2	Northern Europe[a] (32%) Eastern Europe[a] (32%)
	s1b m1	Portugal[a] (39%) Spain[a] (39%)
	s1a m2	China[b] (81%) Taiwan[c] (46%) Ireland (69%)
	s1a m1	Japan[d] (96%)
	s1c m1	Korea[e] (62%)
cag PAI (in – del)	I	Spain[f] (92%) Peru[f] (91%) Africa[f] (100%) Northern Europe[f] (42%)
	II	Japan[f] (96%) China[f] (95%) Ireland (49%) Northern Europe[f] (42%)
	III	India[f] (85%)
	IV	England[f] (100%)

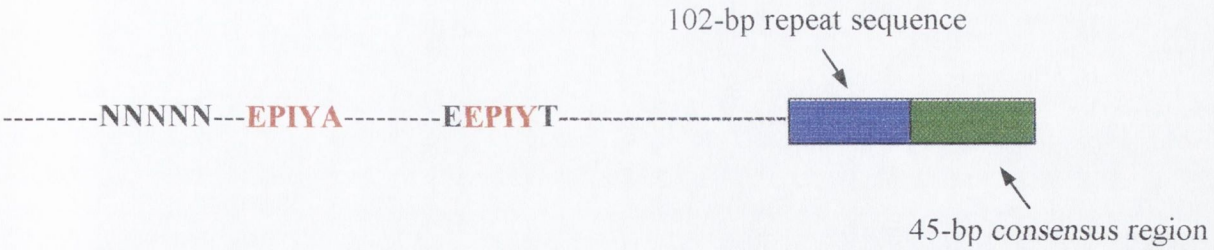
[a] Van Doorn *et al.* (1999a)
[b] Pan *et al.* (1998)
[c] Lin *et al.* (2000)
[d] Ito *et al.* (1997)
[e] Kim *et al.* (2001)
[f] Kersulyte *et al.* (2000)

The *cagA* gene (cytotoxin-associated gene) is another well-characterised putative virulence locus. Isolates carrying the *cagA* genetic locus are more commonly recovered from patients with duodenal ulcers and gastric tumours (Blaser *et al.*, 1992). Although there have been reports showing that the presence of the *cagA* gene is associated with some clinical diseases, it has also been suggested that the *cagA* gene does not predict clinical outcome due to its high prevalence (75–95%) in strains in several geographical regions with a relatively low incidence of clinical disease (Hacker and Kaper, 2000). Studies in Japan, Korea and China have shown that more than 90% of *H. pylori* strains contained the *cagA* gene irrespective of clinical outcome (Miehlke *et al.*, 1996, Pan *et al.*, 1997 and Shimoyama *et al.*, 1997). Therefore, the presence of the *cagA* gene may not alone be an informative marker of virulence. However, the 3′-end of the *cagA* gene contains 102-bp repeats, which have been shown to vary between clinical isolates (Evans *et al.*, 1998). The variation in the number of these repeat sequences within isolates has been shown to be a result of homologous recombination and / or misaligned replication (Hoshino *et al.*, 2000). Determining the frequency of repeats in this region could potentially be indicative of virulence and also determine the geographical relationship of different strains. Furthermore, amino acid sequence analysis of the *cagA* 3′-region has been shown to differentiate between Eastern (East Asia, Japan) and Western (Europe, France and Italy) *H. pylori* isolates (Covacci *et al.*, 1993; Tummuru *et al.*, 1993; Yamaoka *et al.*, 1998, Fig. 3.2 A and B).

The *cagA* gene is part of the 40-kb *cag* pathogenicity island (*cag* PAI) (Covacci *et al.*, 1993). The function of the *cagA* gene and the *cag* PAI remained unclear for years. However, a major breakthrough occurred in 2000 with the demonstration that the CagA protein is transferred via a type IV secretary apparatus (coded for by the *cag* PAI) into epithelial cells where it becomes tyrosine phosphorylated and involved in cytoskeletal rearrangements (Stein *et al.*, 2000). The 3′-end of the *cag* PAI (the region between the 3′-end of the *cagA* gene and the 3′-end of the glutamate racemase gene [*glr*]) contains different motifs that have been used previously for studying relationships between *H. pylori* isolates from different geographical regions (Kersulyte *et al.*, 2000, Fig. 3.3). Characterisation of this region revealed that eight different insertion-deletion (in-del) motifs can be present (designated as; Ia, Ib, Ic, II, IIIa, IIIb, IV and V). The geographical distribution of these motifs can be seen in Table 3.1.

Diversity within the non-coding DNA in *H. pylori* preceding the *vacA* and *ribA* genes and the intergenic region at the *cag* PAI insertion site has been reported (Bereswill *et al.*, 2000, Fig.

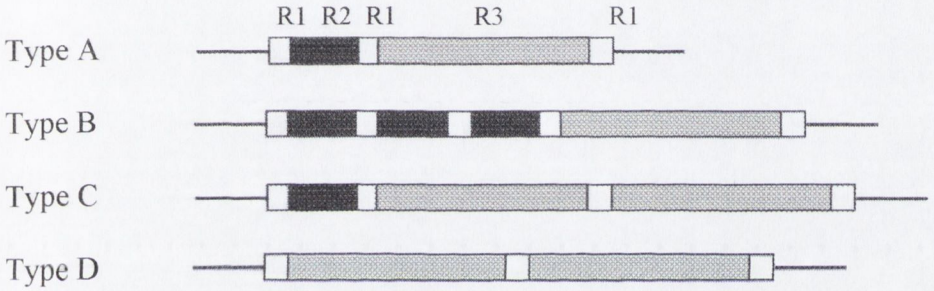
(**A**) 3′-CagA amino acid structure and sequence of Irish strains:



Amino acid sequence of 102-bp repeat region:
FPLKRHSKVDDLSKVGLSANHEPIYATIDDLGGP

Amino acid sequence of 45-bp consensus region:
FPLKRHSKVDDLSKV

(**B**) 3′-CagA structure of Japanese strains:



Amino acid sequence of R3 repeat region:
QVARKVSAKIDQLNEATSAINRKIDRIN**KIASAGKGVGGFSGA**GRSANP

Fig. 3.2. Diagram depicting the structural differences between European and East Asian *H. pylori* isolates at the 3′-end of the *cagA* gene.

(A). The 3′-CagA amino acid structure and sequence of Irish *H. pylori* isolates (adapted from Beesley, 2002). The poly-N motif, EPIYA motifs, the 102-bp repeat sequence and the consensus sequence are indicated. The number of 102-bp repeats varies between clinical isolates.

(B). The 3′-CagA structure of Japanese strains (adapted from Yamaoka *et al.*, 1998) The four different structures and repeat sequences are indicated (types A–D and repeats R1–R3, respectively). The deduced amino acid sequence of the R3 region of the CagA protein differs with respect to a 15-amino acid sequence [KIASAGKGVGGFSGA] from central European and Irish isolates [FPLKRHDKVDDLSKV] (Covacci *et al.*, 1993).

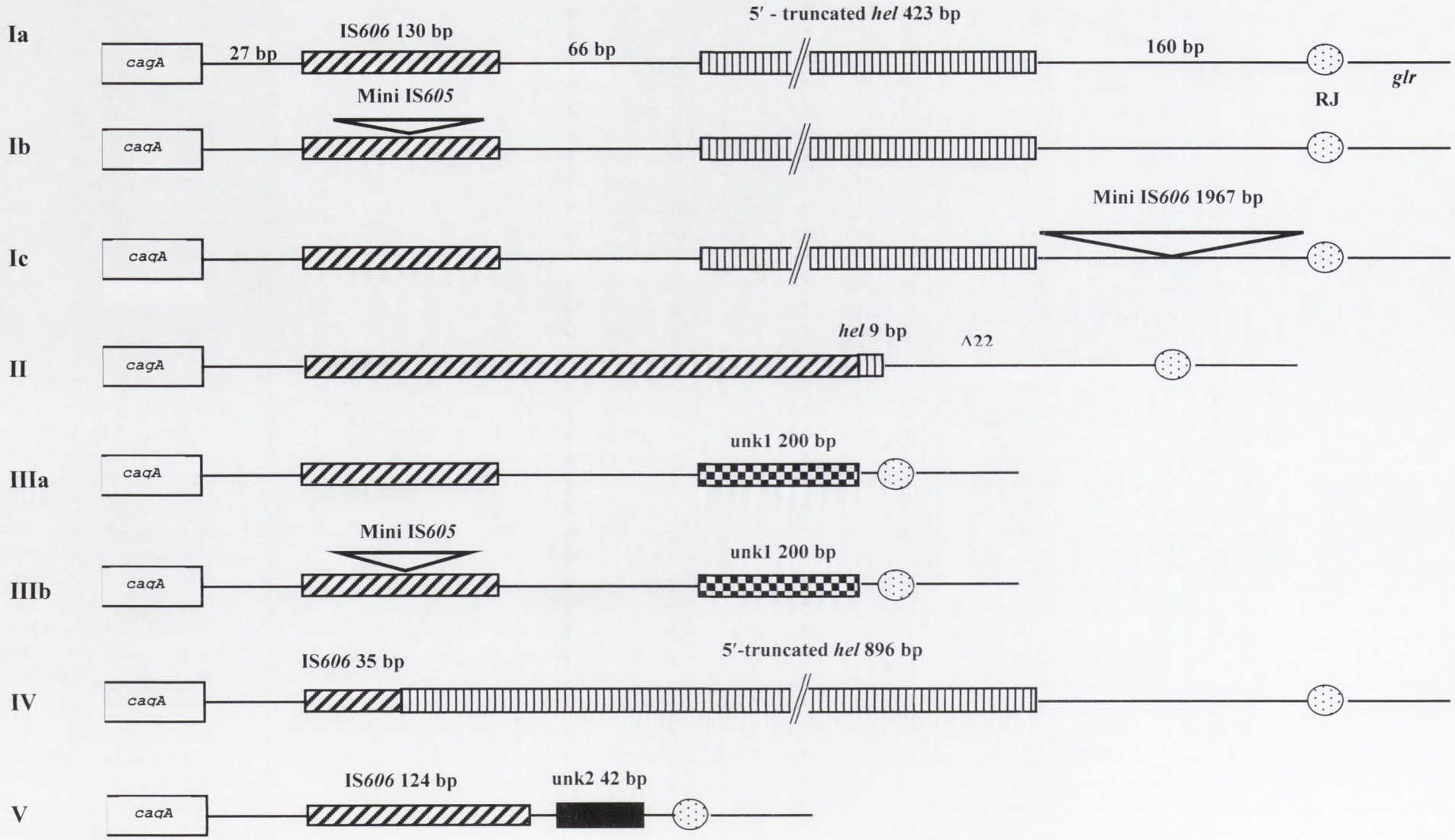

Ia
cagA
27 bp
IS606 130 bp
66 bp
5′ - truncated hel 423 bp
160 bp
glr
RJ
Ib
Mini IS605
Ic
Mini IS606 1967 bp
II
hel 9 bp
Λ22
IIIa
unk1 200 bp
IIIb
Mini IS605
unk1 200 bp
IV
IS606 35 bp
5′-truncated hel 896 bp
V
IS606 124 bp
unk2 42 bp

Fig. 3.3. Diagram of insertion-deletion (in-del) motif types located at the right end of the *cag* PAI (adapted from Kersultye *et al.*, 2000).

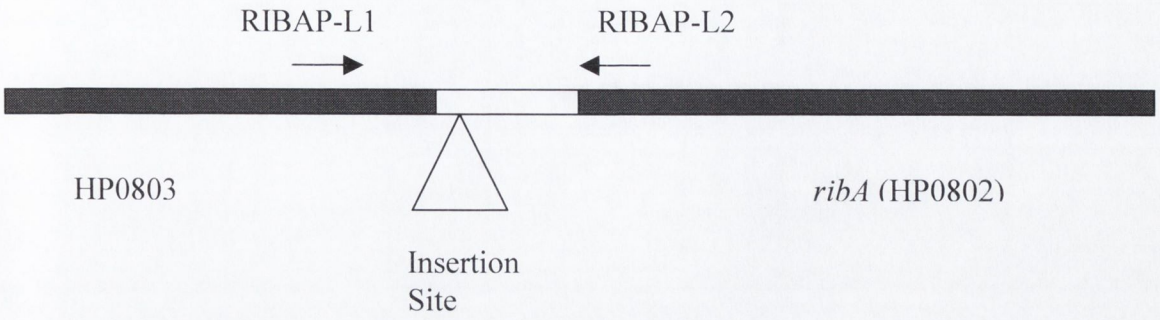
RJ indicates the 31 bp that contains the 3′-end of the glutamate racemase gene (*glr*). Truncated forms of the helicase gene (*hel*), the insertion sequence IS*606* and a new insertion sequence designated Mini IS*605* are indicated. Two different sized DNA sequences (200 bp and 42 bp) of unknown function or origin (unk1 and unk2) is present in type III and V motifs, respectively. Different historical insertion and deletion events have lead to these various motifs.

3.4). These intergenic polymorphisms contribute to the genetic diversity of *H. pylori*. The diversity within such genetic regions has not as yet been extensively investigated, although polymorphisms in these regions have proved to be useful for typing *H. pylori* (Bereswill *et al.*, 2000).

Sequence variations within the *H. pylori* genome have been shown not only to have associations with specific diseases (Strobel *et al.*, 1998, Blaser *et al.*, 1995), but also can illustrate the relationship of various clinical isolates from different geographical regions (Van Doorn *et al.*, 1999a). Although *H. pylori* is a panmictic organism (Go and Graham, 1996), a study encompassing isolates from different regions of the world identified weakly clonal groupings within this bacterium (Achtman *et al.*, 1999). There have also been indications of non-random geographical distributions of *H. pylori* genotypes that may aid in the understanding of the evolution of *H. pylori* and its clinical consequences in different parts of the world (Kersulyte *et al.*, 2000 and Van Doorn *et al.*, 1999b).

In the present study unrelated clinical Irish *H. pylori* isolates were typed by PCR to determine their genetic diversity. The polymorphisms present within coding (*vacA* and *cagA*) and non-coding regions (sequences preceding the *vacA* and *ribA* genes and the motifs located at the right end of the *cag* PAI) within the *H. pylori* genome were determined. These different alleles and sequence variants were used to determine the genetic diversity of Irish *H. pylori* isolates, and also to determine their geographical relationship with isolates from different countries based on the predominance of these different sequence variants.

(A) *ribA* intergenic diversity



(B) *vacA* intergenic diversity

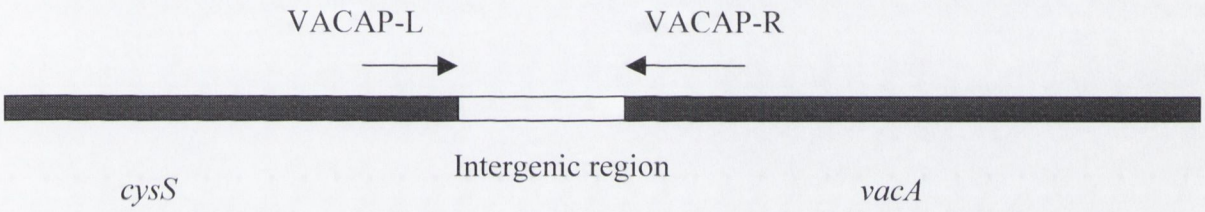


Fig. 3.4. Diagram of intergenic diversity occurring upstream of the *vacA* and *ribA* genes.

(**A**). DNA sequence diversity upstream of the *ribA* gene. Diversity in this intergenic region is a result of inserted sequences generating five different sequence variants (RibAP1–RibAP5). Black arrows indicate the annealing sites of the primers (RIBAP-L1and RIBAP-L2) used to amplify the different variants.

(**B**) DNA sequence diversity upstream of the *vacA* gene. Diversity in the intergenic region results in five different sequence variants (VacAP1–VacAP5). Black arrows indicate the annealing sites of the primers (VACAP-L and VACAP-R) used to amplify the different variants.

3.2. Methods.

The genetic diversity within this collection of Irish *H. pylori* isolates was assessed using PCR. Eight different loci throughout the *H. pylori* genome were amplified in single PCR's using the primers listed in Table 2.1. This table displays the loci investigated, the sequences of the primers used and the PCR product sizes for each allele, motif or sequence variant amplified.

The phylogenetic tree used in this study (Fig. 3.15) was constructed using Clustal X and Treeview™. The accession numbers for all non-Irish sequences were deposited in GenBank by Yamaoka *et al.* (2002). The accession numbers are as follows: ACTC1407 = AB057003, Japan = AB057004–AB057039, Korea = AB057040–AB057044, Taiwan = AB057045–AB057049, Hong Kong = AB057050–AB057054, Vietnam = AB057055–AB057058, Thailand = AB057059–AB057063, Alaska = ABO57064–AB057073, Arizona = AB057074 + AB057075, Columbia = AB57076–AB057093 and AB057101–AB057105, Italy = AB057094, USA = AB057095 + AB057099 + AB057100, Kazakhstan = AB057096–AB067098 and UK = NC_000915 (complete genome).

3.3. Results.

3.3.1. vacA *Genotyping.*

The signal sequences and mid-regions of the *vacA* gene of 74 single antral Irish *H. pylori* isolates were amplified using PCR and the different signal sequence and mid-region genotypes were distinguished by differences in the PCR product sizes [Fig. 3.5 A and B (s1 = 259 bp, s2 = 286 bp, m1 = 570 bp and m2 = 645 bp)]. The signal sequences were amplified using primers VA1-F and VA1-R, and the mid-regions were amplified using primers VAG-F and VAG-R (Table 2.1). The s1-positive isolates were further characterised as the s1a signal sequence variant using primers specific for this allele (Fig. 3.5 A [s1a = 190 bp]).

Of the isolates examined 71 (95.9%) contained the s1 sequence variant of the *vacA* signal sequence, and 3 (4.1%) contained the s2 variant. All of the 71 isolates that possessed the s1 allele were found to be of the s1a type. No s1b or s1c allelic types were found in spite of attempts to amplify the signal sequence using primers specific for these types. Fifty-four (73.0%) of the 74 isolates examined were found to contain the m2 allelic variant, and 20 (27.0%) contained the m1 allelic variant.

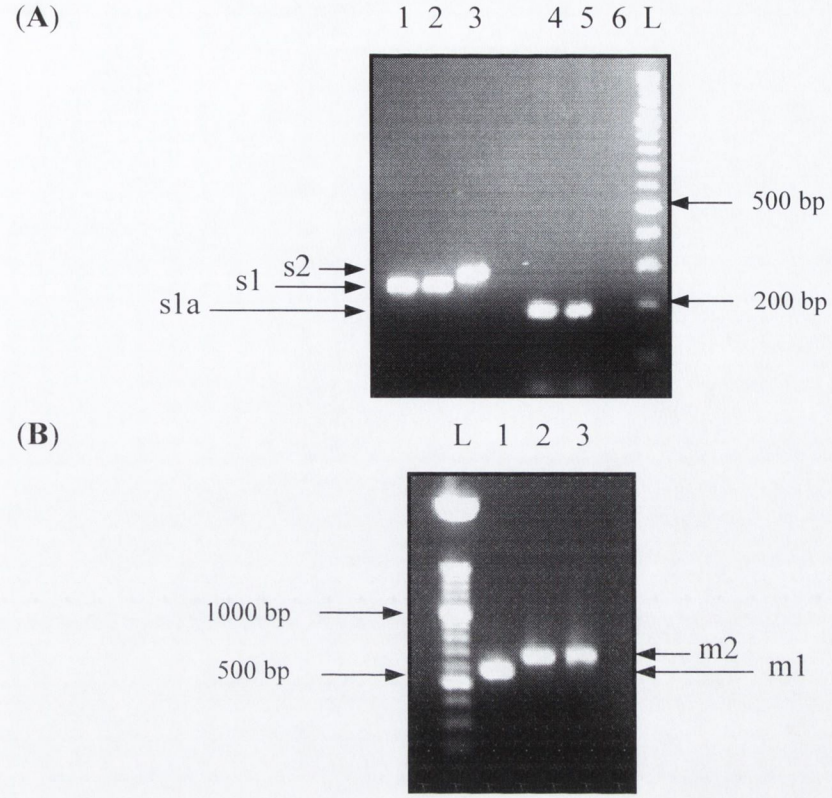


Fig. 3.5. PCR products of the signal-sequence and mid-region alleles detected in this collection of Irish *H. pylori* isolates separated on a 1.8% agarose gel.

(A). Agarose gel electrophoresis (1.5%) of PCR products generated by amplifying the signal sequence of the *vacA* gene of three clinical isolates. Differences in PCR product sizes discern the s1 (259 bp) and s2 (286 bp) allelic variants, and specific amplification of a 190 bp product indicates the presence of the s1a variant. Lanes 1 –3 contain PCR products using primers VA1-F and VA1-R, which amplify the signal sequence of the *vacA* gene [L = 1000-bp Ladder, 1 = 167233A (s1), 2 = 577A (s1), 3 = 9613A (s2)], and lanes 4 and 5 contain PCR products generated using primers SS1-F and VA1-R, which amplify s1a signal sequence [4 = 167233A (s1a), 5 = 577A (s1a)]. No fragment is visible in lane 6 as isolate 9613A is of allelic type s2 and will not generate a PCR product using s1a specific primers.

(B). Agarose gel electrophoresis (1%) of PCR products generated by amplifying the mid-region of the *vacA* gene of three clinical isolates. Differences in PCR product sizes discern the m1 (570 bp) and m2 (645 bp) allelic variants. Lanes 1–3 contain PCR products generated using primers VAG-F and VAG-R [L = 1000 bp Ladder, 1 = 167233A (m1), 2 = 577A (m2), 3 = 9613A (m2)]. On the basis of these results isolate167233A is s1a m1, isolate 577A is s1a m2 and isolate 9613A is s2 m2.

Representative PCR products of the s1, s2, s1a, m1 and m2 allelic variants were sequenced for confirmatory purposes (Fig. 3.6–3.10 A and B, respectively). Blast alignments confirmed the identity of each sequenced PCR product. Figures 3.11 and 3.12 display the alignments of the s1 / s2 and m1 / m2 alleles, respectively. The genetic diversity between these alleles can be clearly seen.

Out of the 74 Irish *H. pylori* isolates studied 51 (68.9%) were of *vacA* type s1a m2, 20 (27.0%) were of *vacA* type s1a m1 and 3 (4.1%) were of *vacA* type s2 m2 (Tables 3.2 and 3.3). The commonest signal sequence and mid-region alleles were s1a and m2, respectively, which is in accordance with previous studies on Irish *H. pylori* isolates (Ryan *et al.*, 2000).

3.3.2. cagA *analysis.*

3.3.2.1 presence of the *cagA* gene.

All the 74 Irish *H. pylori* isolates examined were found to contain the *cagA* gene using PCR. Primers CAGAF and CAGAR were used to detect the presence of the *cagA* gene (Table 3.1, Fig. 3.13 A).

3.3.2.2. Detection of the number of repeat sequences at the 3′-end of the *cagA* gene

The number of repeat sequences at the 3′-end of the *cagA* gene could be determined in 65 (87.8%) of the 74 Irish *H. pylori* isolates used in this study. Nine (12.2%) of these isolates did not yield PCR products despite repeated attempts. The repeat sequences were amplified using primers CAG1 and CAG2 (Table 2.1). The number of repeats located in this region was determined by the size of the PCR products generated, i.e., one repeat = 650 bp, 2 repeats = 750 bp, 3 repeats = 850 bp, 4 repeats = 950 and 5 repeats = 1050 bp. Out of the 65 isolates that yielded usable PCR products 48 (73.8%) possessed 1 repeat, 11 (16.9%) possessed 2 repeats, 4 (6.2%) possessed 0 repeats, 1 (1.5%) had 5 repeats and 1 (1.5%) yielded a PCR product that was smaller than the product for zero repeats (Tables 3.2 and 3.3, Fig. 3.13 B), which may indicate a deletion in this region. Efforts made to obtain the nucleotide sequence of this PCR product were unsuccessful despite a repeated attempt.

Sequence analysis of the 3′-end of the *cagA* gene of a representative isolate with two 102-bp repeats located in this region can be seen in Fig. 3.14. A and B. Four different putative tyrosine phosphorylation motifs (TPM's) can be seen in this region. Two are located outside the 102-bp repeat sequences and fit the consensus sequence for a TPM ([R,K]-X(2,3)-[D,E]-

(A)

Isolate 167233A:

GCCCTCTGGTTTCTCTTGCTTTACTAGGAGCATTAGTCAGCATCACACCGCAACAAAGTC
ATGCCGCCTTTTTCACAACCGTGATCATTCCAGCCATTGTTGGGGGTATCGCTACAGGCG
CTGCTGTAGGAACGGTCTCAGGGCTTCTTAGCTGGGGGCTCAAACAAGCCGAAGAAGCCA
ATAAAACCCCAGATAAACCCGATAAAGTTTGGGCCATTCAAG

(B)

```
F-78 (s1)       ATGGAAATACAACAAACACACCGCAAAATCAATCGCCCTCTGGTTTCTCTCGCTTTAGTA
167233A         ----------------------------------GCCCTCTGGTTTCTCTTGCTTTACTA
                                                  *************** ****** **

F-78 (s1)       GGAGCATTGGTCAGCATCACACCGCAACAAAGTCATGCCGCCTTTTTCACGACCGTGATC
167233As1       GGAGCATTAGTCAGCATCACACCGCAACAAAGTCATGCCGCCTTTTTCACAACCGTGATC
                ******** ****************************************** *********

F-78 (s1)       ATTCCAGCCATTGTTGGGGGCATCGCTACAGGCGCTGCTGTAGGAACGGTCTCAGGGCTT
167233A         ATTCCAGCCATTGTTGGGGGTATCGCTACAGGCGCTGCTGTAGGAACGGTCTCAGGGCTT
                ******************** ****************************************

F-78 (s1)       CTTAGCTGGGGGCTCAAACAAGCCGAAGAAGCCAATAAAACCCCAGATAAACCCGATAAA
167233A         CTTAGCTGGGGGCTCAAACAAGCCGAAGAAGCCAATAAAACCCCAGATAAACCCGATAAA
                ************************************************************

F-78 (s1)       GTTTGGCGCATTCAAGCAGGAAGAGCTTTAATGAATTCCTCTAACAAGGAATACGACTTA
167233A         GTTTGGGCCATTCAAG----------------------------------------------
                 ******* ********
```

Fig. 3.6. Alignment and nucleotide sequence of the s1 *vacA* signal sequence allele detected in this collection of Irish *H. pylori* isolates.

(A). Partial nucleotide sequence of the signal sequence of *H. pylori* strain 167233A generated by PCR using primers VA1-F and VA1-R (Table 3.1). Sequencing was performed by MWG-Biotech. The 222-bp sequence is shown.

(B). Alignment of putative s1 type signal sequence of strain 167233A with that of the s1 type reference strain F-78 (AF049638) using CLUSTAL W for confirmatory purposes. These sequences have over 97% nucleotide identity, and therefore the signal sequence of isolate 167233A is of type s1.

(A)

Isolate 9613A:

GCCCTATTATCTCTCTCGCTCTAGTGGGGGTGTTAATGGGTACCGAACTAGGGGCTAACA
CGCCAAACGATCCCATACACAGCGAGAGTCGCGCCTTTTTTACAACCGTGATCATTCCAG
CCATTGTTGGGGGTATCGCTACAGGCGCTGCTGTAGGAACGGTTTCAGGGCTTCTTAGCT
GGGGGCTCAAACAAGCCGAAGAAGCCAATAAAACCCCAGATAAACCCGATAAAGTTTGGG
CCATTCAAGCA

(B)

```
SS1 (s2)        ATGGAAATACAACAAACACACCGCAAAATCAATCGCCCTATTATCTCTCTCGCTCTAGTG
9613A           ----------------------------------GCCCTATTATCTCTCTCGCTCTAGTG
                                                  *************************

SS1 (s2)        GGGGTGTTAATGGGCACCGAACTAGGGGCTAATACGCCAAACGATCCCATACACAGCGAG
9613A           GGGGTGTTAATGGGTACCGAACTAGGGGCTAACACGCCAAACGATCCCATACACAGCGAG
                ************** ***************** **************************

SS1 (s2)        AGTCGTGCTTTTTTCACAACCGTGATCATTCCAGCCATTGTTGGAGGTATCGCTACAGGC
9613A           AGTCGCGCCTTTTTTTACAACCGTGATCATTCCAGCCATTGTTGGGGGTATCGCTACAGGC
                ***** ** ***** ***************************** ***************

SS1 (s2)        GCTGCTGTAGGAACGGTCTCAGGGCTTCTTAGCTGGGGGCTCAAACAAGCCGAACAAGCC
9613A           GCTGCTGTAGGAACGGTTTCAGGGCTTCTTAGCTGGGGGCTCAAACAAGCCGAAGAAGCC
                ***************** ************************************ *****

SS1 (s2)        AATAAAGCCCCGGATAAACCCGATAAAGTTTGGCGCATTCAAGCAG
9613A           AATAAAACCCCAGATAAACCCGATAAAGTTTGGGCCATTCAAGCA-
                 ****** **** **********************  **********
```

Fig. 3.7. Alignment and nucleotide sequence of the s2 *vacA* signal sequence allele detected in this collection of Irish *H. pylori* isolates.

(A). Partial nucleotide sequence of the signal sequence of *H. pylori* strain 9613A generated by PCR using primers VA1-F and VA1-R (Table 3.1). Sequencing was performed by MWG-Biotech. The 251-bp sequence is shown.

(B). Alignment of putative s2 type signal sequence of strain 9613A with that of the s2 type reference strain SS1 (AY049006) using CLUSTAL W, for confirmatory purposes. These sequences have over 96% nucleotide identity, and therefore the signal sequence of isolate 9613A is of type s2.

(A)

Isolate 556:

CCGCCTTTTTCCACCGTGATCATTCCAGCCATTGTTGGGGGTATCGCTACAGGCGCTGCT
GTAGGAACAGTTTCAGGGCTTCTTAGCTGGGGGCTCAAACAAGCCGAAGAAGCCAATAAA
ACTCCGGATAAACCCGATAAAGTTTGGCGCATTCAAGCA

(B)

```
F-78 (s1a)      GGAGCATTGGTCAGCATCACACCGCAACAAAGTCATGCCGCCTTTTTCACGACCGTGATC
556             -------------------------------------CCGCCTTTTTC--CACCGTGATC
                                                     ***********   *********

F-78 (s1a)      ATTCCAGCCATTGTTGGGGGCATCGCTACAGGCGCTGCTGTAGGAACGGTCTCAGGGCTT
556             ATTCCAGCCATTGTTGGGGGTATCGCTACAGGCGCTGCTGTAGGAACAGTTTCAGGGCTT
                ******************** ************************** ** *********

F-78 (s1a)      CTTAGCTGGGGGCTCAAACAAGCCGAAGAAGCCAATAAAACCCCAGATAAACCCGATAAA
556             CTTAGCTGGGGGCTCAAACAAGCCGAAGAAGCCAATAAAACTCCGGATAAACCCGATAAA
                ****************************************** ** ***************

F-78 (s1a)      GTTTGGCGCATTCAAGCAGGAAGAGCTTTAATGAATTCCTCTAACAAGGAATACGACTTA
556             GTTTGGCGCATTCAAGCA------------------------------------------
                ******************
```

Fig. 3.8. Alignment and nucleotide sequence of the s1a *vacA* signal sequence allele detected in this collection of Irish *H. pylori* isolates.

(A). Partial nucleotide sequence of signal sequence of the *H. pylori* strain J556 generated by PCR using primers SS1-F and VA1-R (Table 3.1). Sequencing was preformed by MWG-Biotech. The 159-bp sequence is shown.

(B). Alignment of putative s1a type signal-sequence of strain 556 with that of the s1a type reference strain F-78 (AF049638) using CLUSTAL W for confirmatory purposes. These sequences have over 95% nucleotide identity, and therefore the signal sequence of isolate 9613A is of type s1a.

Fig. 3.9. Alignment and nucleotide sequence of the m1 *vacA* mid-region allele detected in this collection of Irish *H. pylori* isolates.

(A). Partial nucleotide sequence of the mid-region of the *H. pylori* strain 167233A generated by PCR using primers VAG-F and VAG-R (Table 3.1). Sequencing was performed by MWG-Biotech. The 501-bp sequence is shown.

(B). Alignment of putative m1 type mid-region of strain 167233A with that of the m1 type reference strain NCTC 11637 (AF049653) using CLUSTAL W for confirmatory purposes. These sequences have over 93% nucleotide identity, and therefore the mid-region of isolate 167233A is of type m1.

(A)

Isolate 167233A:

TAACCGGAAATATCACCGTTGATGGGCCTTTAAGAGTGAATAATCAAGTGGGTGGTTATG
CTTTGGCAGGATCAAGCGCGAATTTTGAGTTTAAGGCTGGTACGGATACCAAAAACGGCA
CAGCCACTTTTAATAACGATATTAGTTTGGGAAGATTTGTGAATTTAAAAGTGGATGCTC
ATACAGCTAATTTTAAAGGTATTGATACTGGTAATGGTGGTTTCAACACTTTAGATTTTA
GTGGAGTTACAGACAAAGTCAATATCAACAAGCTCATTACAGCTTCCACTAATGTGGCCA
TTAAAAACTTCAACATTAATGAATTGTTGGTTAAGACCAATGGGGTGAGTGTGGGGGAAT
ACACTCATTTTAGCGAAGATATAGGCAGTCAATCGCGCATCAACACCGTGCGTTTAGAAA
CTGGCACTAGGTCAATCTATTCTGGCGGTGTTAAATTTAAAGGCGGTGAAAAACTGGTTA
TTAATGATTTTTACTACGCCC

(B)

```
NCTC 11637(m1)  CAATCTGTCCAATCAAGCGAGCGGGCGCTCTCTTTTAGTGGAAAATCTAACCGGGAATAT
167233A         ----------------------------------------------TAACCGGAAATAT
                                                              ******* *****

NCTC 11637(m1)  CACCGTTGATGGGGCTTTAATGGTGAATAATCAAGTGGGCGGTTATGCTTTGGCAGGCTC
167233A         CACCGTTGATGGGCCTTTAAGAGTGAATAATCAAGTGGGTGGTTATGCTTTGGCAGGATC
                ************* ******  ***************** ***************** **

NCTC 11637(m1)  AAGCGCAAATTTTGAGTTTAAAGCTGGTGTGGATACCAAAAACGGCACAATCGCTTTTAA
167233A         AAGCGCGAATTTTGAGTTTAAGGCTGGTACGGATACCAAAAACGGCACAGCCACTTTTAA
                ****** ************** ******  ******************  * *******

NCTC 11637(m1)  TAACAATATCAGTCTGGGAAGATTTGTGAATTTAAAAGCGAGCGCTCATACGGTCAATTT
167233A         TAACGATATTAGTTTGGGAAGATTTGTGAATTTAAAAGTGGATGCTCATACAGCTAATTT
                **** **** *** ************************ *   ******** *  *****

NCTC 11637(m1)  TAAAGATATTGATACTGGTAATGGTGGTTTCAACACCTTAGACTTTAGTGGTGTTACAAA
167233A         TAAAGGTATTGATACTGGTAATGGTGGTTTCAACACTTTAGATTTTAGTGGAGTTACAGA
                ***** ****************************** ***** ******** ****** *

NCTC 11637(m1)  CAAGGTCAATATCAACAAGCTCATTACAGCTTCCACTAATGTGGCCATTAAAAACTTCAA
167233A         CAAAGTCAATATCAACAAGCTCATTACAGCTTCCACTAATGTGGCCATTAAAAACTTCAA
                *** ************************************************************

NCTC 11637(m1)  CATTAATGAATTGTTGGTTAAAACCAATGGGATAAGTGTGGGGGAATACACTAATTTTAG
167233A         CATTAATGAATTGTTGGTTAAGACCAATGGGGTGAGTGTGGGGGAATACACTCATTTTAG
                ********************** ********* * ****************** *******

NCTC 11637(m1)  CGAAGATATAGGCAATCAATCGCGCATCAATACCGTGCGTTTAGAAACTGGCACTAGGTC
167233A         CGAAGATATAGGCAGTCAATCGCGCATCAACACCGTGCGTTTAGAAACTGGCACTAGGTC
                ************** *************** ******************************

NCTC 11637(m1)  AATCTATTCTGGGGGTGTTAAATTTAAAGGCGGTGAAAAACTGGTTATCAATGATTTTTA
167233A         AATCTATTCTGGCGGTGTTAAATTTAAAGGCGGTGAAAAACTGGTTATTAATGATTTTTA
                ************ ************************************ ***********

NCTC 11637(m1)  CTACGCCCCTTGGAATTATTTTGACGC
167233A         CTACGCCC--------------------
```

Fig. 3.10. Alignment and nucleotide sequence of the m2 *vacA* mid-region allele detected in this collection of Irish *H. pylori* isolates.

(A). Partial nucleotide sequence of the mid-region of the *H. pylori* strain 577A generated by PCR using primers VAG-F and VAG-R (Table 3.1). Sequencing was preformed by MWG-Biotech. The 581-bp sequence is shown.

(B). Alignment of putative m2 type mid-region of strain 577A with that of the m2 type reference strain India90 (AF220119) using CLUSTAL W, for confimatory purposes. These sequences have over 94% nucleotide identity, and therefore the mid-region of 577A is of type m2.

(A)

Isolate 577A

GGAAATCTAACCGGCAATATCACCGTTGAGGGGACTTTAAGGGTGAATAATCAAGTGGGC
GGTGCTGCTGTGGCAGGCTCAAGCGCGAATTTTGAGTTTAAGGCTGGCACTGATACCAAC
AACGGCACAGCCACTTTTAATAACGATATCCATCTAGGAAAAGCGGTGAATTTAAGAGTG
GATGCTCATACAGCTTATTTTAATGGCAATATTTATCTGGGAAAATCCACGAATTTAAGA
GTGAATGCCCTAAGCGCTCATTTTAAAAATATTGATGCCACAAAGAGCGATAACGGGCTA
AACACTAGCGCTTTGGATTTTAGCGGCGTTACAGATAAAGTCAATATCAACAAGCTCACT
ACATCTGCCACTAACGTGAACATTAAAAATTTTGACATTAAGGAATTGGTGGTTACAACC
CGTGTTCAAAGTTTTGGGCAATACACTATTTTTGGCGAAAATATAAGCGATAAGTCTCGC
ATTGGTGTCGTGAGTTTGCAAACGGGATATAGCCCGGCCTATTCTGGGGGGCGTTACTTTT
AAAAGCGGTAAAAAACTGGTTATAGATGAAATTTACCATGC

(B)

```
India90 (m2)    CAATCTGTCCAATCAAGCGAGCGGGCGCACCCTTTTAGTGGAAAATCTAACCGGGAATAT
577A            ----------------------------------------GGAAATCTAACCGGCAATAT
                                                        * ************ *****

India90 (m2)    CACCGTTGAGGGGACTTTAAGGGTGAATAATCAAGTGGGCGGTGCTGCTATAGCAGGTTC
577A            CACCGTTGAGGGGACTTTAAGGGTGAATAATCAAGTGGGCGGTGCTGCTGTGGCAGGCTC
                ************************************************ * ***** **

India90 (m2)    AAGCGCGAATTTTGAGTTTAAGGCTGGTGAGGATACCAACAACGCCACAGCCACTTTCAA
577A            AAGCGCGAATTTTGAGTTTAAGGCTGGCACTGATACCAACAACGGCACAGCCACTTTTAA
                **************************    ************* ************ **

India90 (m2)    TAACGATATTCATCTGGGAAAAGCGGTGAATTTAAGAGTGGATGCCCATACGGCTAATTT
577A            TAACGATATCCATCTAGGAAAAGCGGTGAATTTAAGAGTGGATGCTCATACAGCTTATTT
                ********* ***** ****************************** ***** *** ****

India90 (m2)    TAATGGCAATATTTATCTGGGAAAATCCACGAATTTAAGAGTGAATGGCCATACCGCTCA
577A            TAATGGCAATATTTATCTGGGAAAATCCACGAATTTAAGAGTGAATGCCCTAAGCGCTCA
                ********************************************** **  * ******

India90 (m2)    TTTTAAAAACATTGATGCTACAAAGAGCGATAACGGGCTAAACACTAGCACCTTGGATTT
577A            TTTTAAAAATATTGATGCCACAAAGAGCGATAACGGGCTAAACACTAGCGCTTTGGATTT
                ********* ******** ****************************** * ********

India90 (m2)    CAGTGGCGTTACAGACAAAGTCAATATCAACAAGCTCACTACGGCTGCCACTAATGTGAA
577A            TAGCGGCGTTACAGATAAAGTCAATATCAACAAGCTCACTACATCTGCCACTAACGTGAA
                 ** *********** ************************** ********** *****

India90 (m2)    TATTAAAAACTTTGACATTAAGGAATTGGTGGTTACAACCCGTGTTCAGAGTTTTGGGCA
577A            CATTAAAAATTTTGACATTAAGGAATTGGTGGTTACAACCCGTGTTCAAAGTTTTGGGCA
                 ******** ************************************ ***********

India90 (m2)    ATACACTATTTTTGGCGAAAATATAGGCGATAAGTCTCGCATTGGTGTCGTGAGTTTGCA
577A            ATACACTATTTTTGGCGAAAATATAAGCGATAAGTCTCGCATTGGTGTCGTGAGTTTGCA
                ************************* ********************************

India90 (m2)    AACGGGATATAGCCCGGCCTATTCTGGGGGCGTTACTTTTAAAGGCGGTAAAAAACTGGT
577A            AACGGGATATAGCCCGGCCTATTCTGGGGGCGTTACTTTTAAAAGCGGTAAAAAACTGGT
                ****************************************** ****************

India90 (m2)    TATAGATGAAATTTACCATGCCCCTTGGAATTATTTTGACGC
577A            TATAGATGAAATTTACCATGC----------------------
                *********************
```

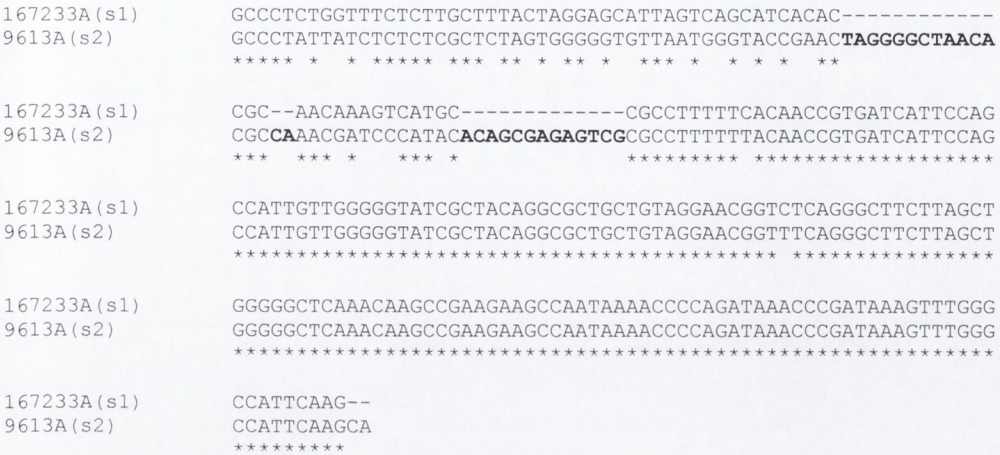


Fig. 3.11. Alignment of the s1 and s2 *vacA* signal-sequence alleles detected in this collection of Irish *H. pylori* isolates.

Sequence alignment of the s1 (strain 167233A) and s2 (strain 9613A) alleles of the *vacA* gene using CLUSTAL W. PCR products obtained from the amplification of the s1 and s2 alleles yielded nucleotide sequences of 222 bp and 251 bp, respectively. Sequence diversity between these two alleles is mainly due to insertions of extra nucleotides in the s2 allele (bold).



Fig. 3.12. Alignment of the m1 and m2 *vacA* mid-region alleles detected in this collection of Irish *H. pylori* isolates.

Sequence alignment of the m1 (167233A) and m2 (577A) alleles of the *vacA* gene using CLUSTAL W. PCR products obtained from the amplification of the m1 and m2 alleles (167233A and 577A, respectively) yielded nucleotide sequences of 501 bp and 581 bp, respectively. Sequence diversity between these two alleles is mainly due to insertions of extra nucleotides in the m2 allele (bold).

Table 3.2. Frequencies of genotypes or sequence variants found in this collection of Irish *H. pylori* isolates.

Gene	Genotype	Frequency (%)	Predominant Genotype
vacA			
signal sequence:	s1a	95.5%	
	s2	4.5%	
			s1a m2
mid-region:	m1	27.0%	
	m2	73.0%	
cagA			
3′-end repeat	unknown	1.5%	
	0	6.2%	
	1	73.8%	1 repeat
	2	16.9%	
	5	1.5%	
cag PAI	Ia	9.5%	
	II	49.2%	
	IIIa	4.8%	II
	IIIb	1.6%	
	IV	34.9%	
Intergenic region			
preceding *ribA*	RibAP1	93.2%	
	RibAP2	1.4%	RibAP1
	RibAP3	4.0%	
	RibAP5	1.4%	
Intergenic region			
preceding *vacA*	VacAP2	1.4%	
	VacAP3	89.2%	VacAP3
	VacAP4	9.4%	

Table 3.3. Strain and genotype of each *H. pylori* isolate used in this study.

Strain	*vacA* type	In-del motif	*cagA* 3′ repeats	*ribA* motif	*vacA motif*
J48	s1a m1	II	1	RibAP1	VacAP3
J063	s1a m1	II	2	RibAP1	VacAP3
J81	s1a m2	II	1	RibAP1	VacAP3
J126	s1a m2	M[b]	1	RibAP1	VacAP3
J128	s1a m2	IV	1	RibAP1	VacAP3
J131	s1a m2	Ia	0	RibAP1	VacAP3
J136	s1a m2	II	1	RibAP1	VacAP3
J159	s1a m2	II	1	RibAP5	VacAP4
J171	s1a m2	II	2	RibAP1	VacAP3
J225	s1a m1	II	1	RibAP1	VacAP3
J248	s1a m2	II	2	RibAP1	VacAP3
J260	s1a m2	Ia	1	RibAP1	VacAP3
J268	s1a m2	Ia	-[a]	RibAP1	VacAP3
J283	s1a m2	IV	1	RibAP1	VacAP3
J290	s1a m2	II	unknown	RibAP1	VacAP3
J295	s1a m2	IV	0	RibAP1	VacAP3
J298	s1a m2	IV	1	RibAP1	VacAP3
J323	s1a m2	II	1	RibAP1	VacAP3
J325	s1a m2	-[a]	-[a]	RibAP1	VacAP3
J442	s1a m1	II	1	RibAP1	VacAP3
J517	s1a m1	IIIa	5	RibAP1	VacAP3
J536	s1a m2	-[a]	1	RibAP1	VacAP3
J556	s1a m2	IV	1	RibAP1	VacAP4
J561	s1a m2	-[a]	1	RibAP1	VacAP3
J563	s1a m2	IIIa	1	RibAP1	VacAP3
J612	s1a m2	II	-[a]	RibAP1	VacAP3
J613	s2 m2	IV	-[a]	RibAP1	VacAP4
J631	s1a m2	IV	1	RibAP1	VacAP3
J671	s1a m1	II	1	RibAP1	VacAP3
J714	s1a m2	IV	1	RibAP1	VacAP3
J734	s1a m1	II	1	RibAP1	VacAP3
J755	s1a m2	IV	1	RibAP1	VacAP3
J780	s1a m2	IV	-[a]	RibAP1	VacAP3
J803	s1a m2	IV	2	RibAP1	VacAP3
FR138	s1a m1	II	1	RibAP1	VacAP2
FR147	s1a m2	M[b]	1	RibAP3	VacAP3
FR150	s1a m2	IV	1	RibAP1	VacAP3

Strain	*vacA* type	In-del motif	*cagA 3′* repeats	*ribA* motif	*vacA motif*
FR191	s1a m1	II	1	RibAP1	VacAP3
FR205	s1a m1	II	1	RibAP1	VacAP3
FR207	s1a m1	-[a]	1	RibAP1	VacAP3
FR217	s1a m2	IIIa	1	RibAP1	VacAP3
FR222	s1a m1	IV	1	RibAP3	VacAP3
FR229	s1a m2	II	1	RibAP1	VacAP3
FR235	s1a m2	II	1	RibAP1	VacAP3
FR296	s1a m2	Ia	1	RibAP1	VacAP3
FR298	s1a m1	II	1	RibAP2	VacAP4
563C	s1a m2	IIIb	1	RibAP1	VacAP3
112886A	s1a m2	II	2	RibAP1	VacAP3
15500A	s1a m2	Ia	2	RibAP1	VacAP3
167233A	s1a m1	IV	2	RibAP1	VacAP4
218A	s1a m2	II	1	RibAP1	VacAP3
222A	s1a m2	-[a]	1	RibAP1	VacAP3
34683A	s1a m1	II	1	RibAP1	VacAP3
45688A	s1a m2	II	0	RibAP1	VacAP3
50334A	s1a m2	II	1	RibAP1	VacAP3
577A	s1a m2	IV	1	RibAP1	VacAP3
58745A	s1a m2	IV	1	RibAP1	VacAP3
59393A	s1a m1	II	1	RibAP1	VacAP3
605A	s2 m2	M[b]	-[a]	RibAP1	VacAP4
68545A	s1a m2	IV	1	RibAP1	VacAP3
79862A	s1a m2	IV	1	RibAP1	VacAP3
87212A	s1a m2	IV	2	RibAP1	VacAP3
9613A	s2 m2	IV	1	RibAP1	VacAP4
MI227	s1a m1	-[a]	-[a]	RibAP1	VacAP3
MI355	s1a m2	IV	2	RibAP1	VacAP3
MI458	s1a m2	II	1	RibAP1	VacAP3
MI496	s1a m1	II	2	RibAP1	VacAP3
MI506	s1a m2	II	1	RibAP3	VacAP3
MI512	s1a m2	IV	0	RibAP1	VacAP3
MI517	s1a m1	II	1	RibAP1	VacAP3
MI568	s1a m2	-[a]	-[a]	RibAP1	VacAP3
MI571	s1a m1	II	1	RibAP1	VacAP3
F431	s1a m2	-[a]	-[a]	-[a]	-[a]
F610	s1a m2	Ia	2	RibAP1	VacAP3

[a] refers to no PCR product obtained.
[b] refers to multiple PCR products obtained.

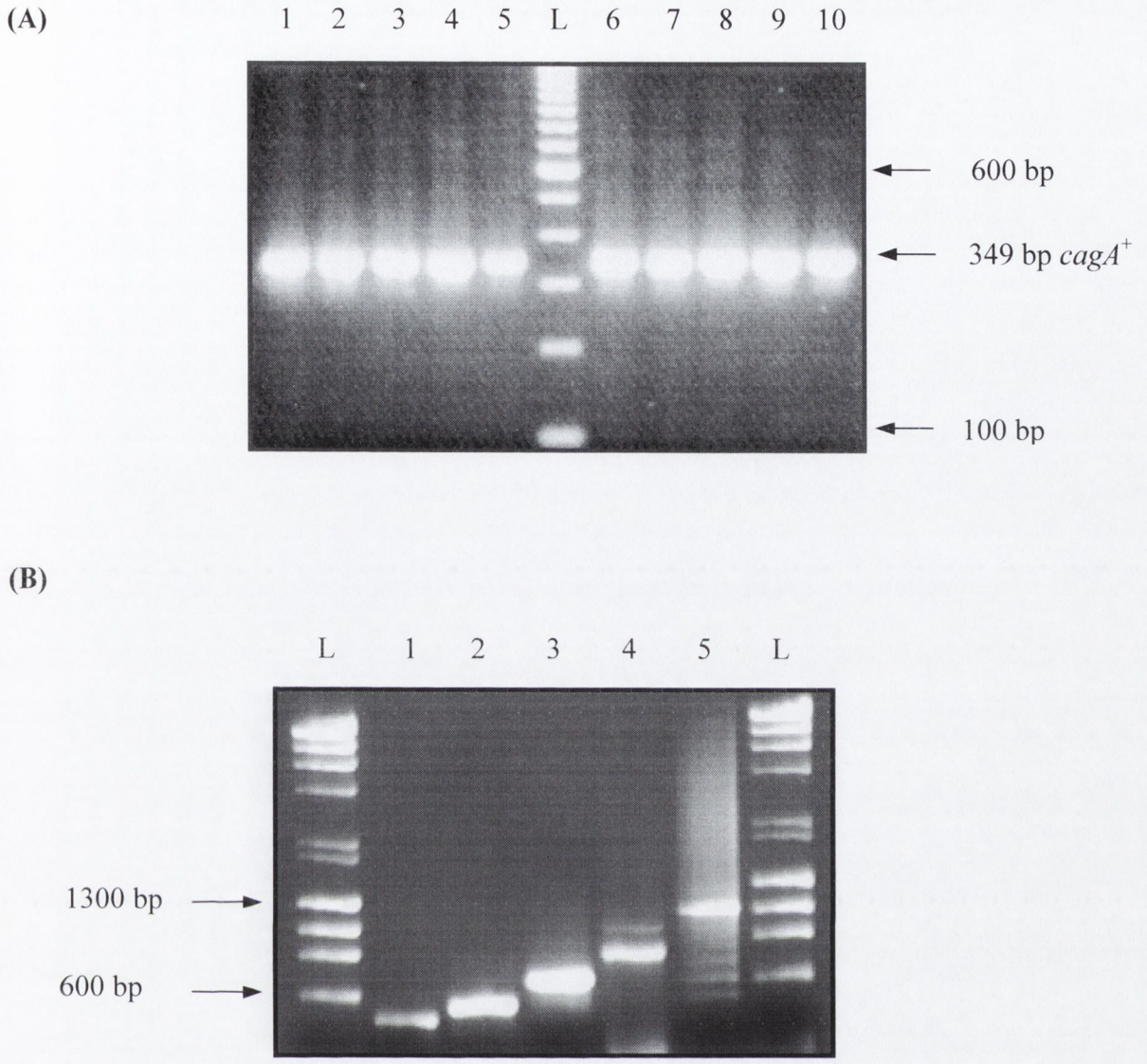


Fig. 3.13. PCR products run on a 1.8% agarose gel for detection of the presence of *cagA* and the number of repeat sequences located at the 3′- end of the *cagA* gene.

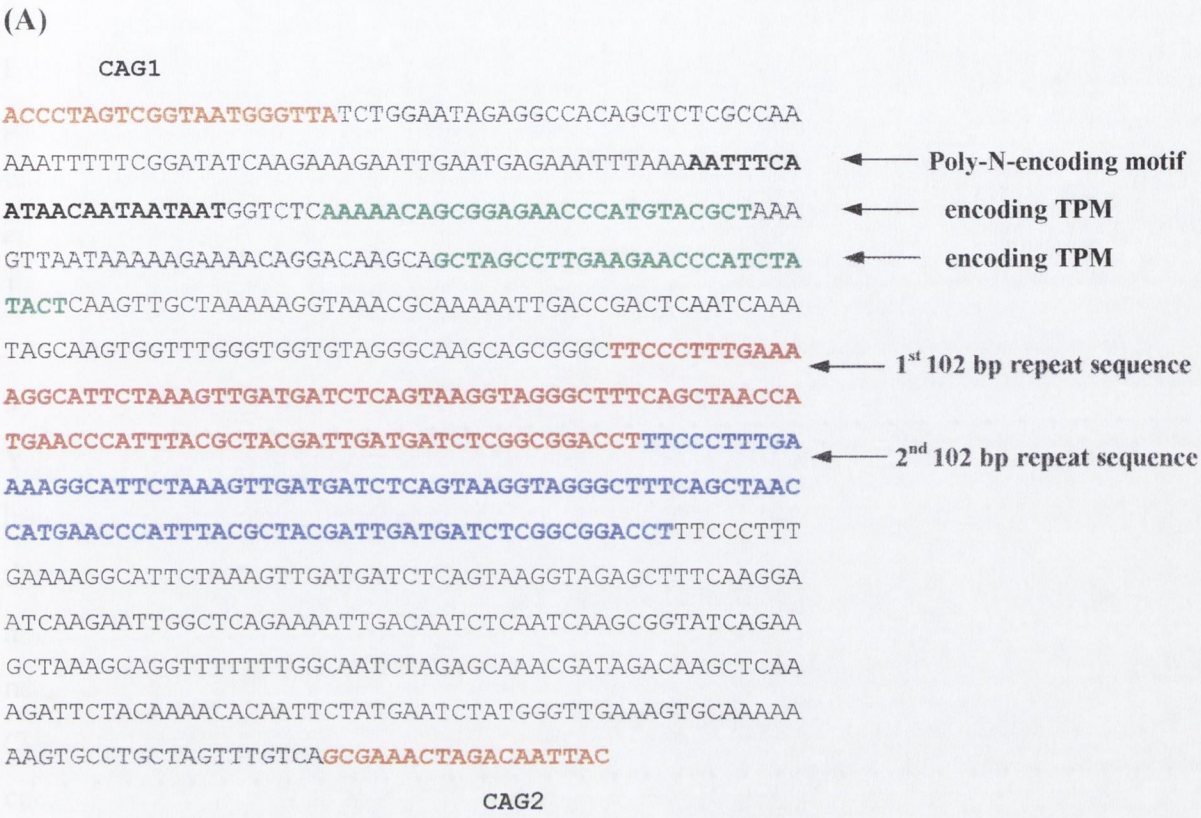
(A). PCR products indicating the presence of the *cagA* gene within ten Irish *H. pylori* isolates using primers CAGAF and CAGAR. Presence of the *cagA* gene results in a 349-bp fragment; all isolates tested contained this gene. Lanes 1–10 show PCR products from isolates J755, J714, J48, J517, J612, J671, J81, J734, J556 and J536, respectively. L = 1000-bp ladder.

(B). PCR products generated after amplification of the 3′-end of the *cagA* gene using primers CAG1 and CAG2. Five different numbers of repeats were found (L = 1000-bp ladder, 1 = unknown, 2 = 0 repeats, 3 = 1 repeat, 4 = 2 repeats and 5 = 5 repeats) in isolates within this collection. Lanes 1–5 contain isolates J209, J159, J248 and J517, respectively.

Fig. 3.14. Nucleotide and amino acid sequences of a the 3′-end of the *cagA* gene of an Irish *H. pylori* isolate containing two repeat sequences.

(A). Nucleotide sequences of the 3′-end of the *cagA* gene of Irish *H. pylori* isolate MI566. Primers used to amplify this region (CAG1 and CAG2) are indicated (orange). This region encodes a poly-N motif (bold), two consensus putative tyrosine phosphorylation motifs (green) and contains two 102-bp repeat regions (red and blue).

(B). Translated (amino acid) sequences of the 3′-end of the *cagA* gene. TPM's are indicated as above. Although the 102-bp repeat sequences (red and blue) do not contain any consensus TPM's they do, however, contain EPIYA amino acid motifs (black boxes), which could potentially be tyrosine phosphorylated. If these motifs are functional their presence could theoretically have an effect on the biological activity of the CagA protein.



(B)

T L V G N G L S G I E A T A L A K N F S D I K K E L N E
K F K N F N N N N N G L K N S G E P M Y A K V N K K K T G Q A A S L
E E P I Y T Q V A K K V N A K I D R L N Q I A S G L G G V G Q A A G
F P L K R H S K V D D L S K V G L S A N H E P I Y A T I D D L G G P
F P L K R H S K V D D L S K V G L S A N H E P I Y A T I D D L G G P
F P L K R H S K V D D L S K V E L S R N Q E L A Q K I D N L N Q A V
S E A K A G F F G N L E Q T I D K L K D S T K H N S M N L W V E S A
K K V P A S L S A K L D N

X(2,3)-[Y], where amino acids are given in single letter codes and X can be any amino acid). The last two putative TPM's are located within the 102-bp repeat sequences (one putative TPM within each repeat). These latter two putative TPM's do not fit the consensus sequence for a TPM. However, they are considered putative TPM's as they contain EPIYA motifs. EPIYA motifs within the *cagA* gene have previously been shown to be tyrosine phosphorylated (Backert *et al.*, 2001). Furthermore, the amino acid sequence of Irish isolates within this repeat region is more like Europen sequences [FPLKRHDKVDDLSKV, Covacci *et al.*, (1993)] rather than Japanese sequences [KIASAGKGVGGFSGA, Yamaoka *et al.*, 1998].

3.3.2.3 Phylogenetic analysis of the repeat sequences at the 3′-end of the *cagA* gene.

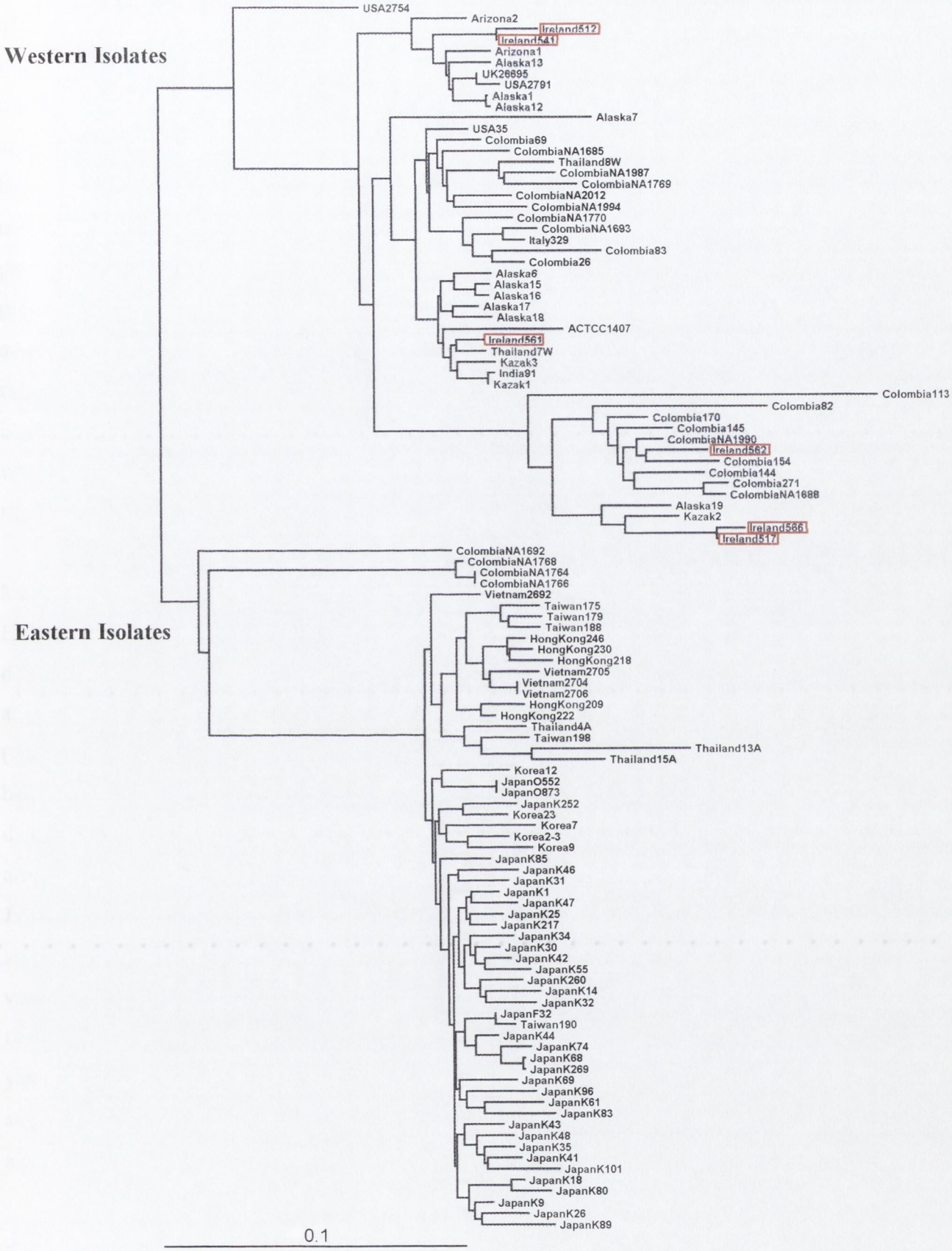
Yamaoka *et. al.* (2002) demonstrated how Eastern and Western isolates can be differentiated based on sequences located within the 3′-end of the *cagA* gene. The repeat sequences located the 3′-end of the *cagA* gene of six representative isolates were amplified using PCR and the nucleotide sequences of these PCR products were determined for each isolate. These nucleotide sequences were then used in a similar phylogenetic study to Yamaoka *et al.*, (2002). The six Irish *H. pylori* isolates included MI512, MI571 and J561 (Table 3.3) plus a corresponding paired isolate (MI541, MI566 and J562, respectively, shown in chapter 5) for each of these isolates. The sequences within the 3′-end of the *cagA* gene of these six Irish isolates were phylogenetically compared to 104 sequences (obtained from GenBank) from isolates from different geographical regions (15 – USA [2 – Arizona, 10 – Alaska], 1 – UK, 23 – Columbia, 1 – Italy, 3 – Kazakhstan, 1 – ACTCC1407 (reference), 5 – Thailand, 5 – Hong Kong, 4 – Vietnam, 5 – Korea, 36 – Japan, 5 - Taiwan). The clustering of the isolates in this study was in agreement with the observations of Yamaoka *et al.* (2002) (Fig. 3.15) where Eastern (East Asian) and Western (European, USA, South America) isolates clustered separately. The Irish *H. pylori* sequences used in this phylogenetic analysis clustered within the Western region of the tree, indicating that Irish *H. pylori* isolates are more like Western strains at this locus.

3.3.3. Insertion-deletion (in-del) analysis within the cag *PAI.*

Firstly, in-del analysis was performed on all isolates using primers cagF4584 and cagR5280 (Table 3.1), which span the in-del region. PCR products using these primers cannot specifically type each in-del motif, as in some cases different in-del types will yield the same size products, i.e., type II and IIIa both yield a PCR product of approx. 700 bp (Kersultye *et*

Fig. 3.15. Neighbour-joining phylogenetic tree displaying the relatedness of *H. pylori* isolates from various geographical locations based on the sequences of the 3′-end of the *cagA* gene.

Eastern (East Asian) and Western (European, USA and South America) isolates are clearly divergent at this locus. Irish *H. pylori* isolates (red boxes) are more similar to Western isolates based on 3′-*cagA* sequences. The clustering of Eastern and Western isolates separately is in agreement with previous studies demonstrating different structures at the 3′-end of the *cagA* between these two regions. The accession numbers for all non-Irish isolate sequences can be found in Section 3.2. The scale at the bottom indicates 0.1 nucleotide substitutions per site.


Western Isolates
Eastern Isolates
USA2754
Arizona2
Ireland512
Ireland541
Arizona1
Alaska13
UK26695
USA2791
Alaska1
Alaska12
Alaska7
USA35
Colombia69
ColombiaNA1685
Thailand8W
ColombiaNA1987
ColombiaNA1769
ColombiaNA2012
ColombiaNA1994
ColombiaNA1770
ColombiaNA1693
Italy329
Colombia83
Colombia26
Alaska6
Alaska15
Alaska16
Alaska17
Alaska18
ACTCC1407
Ireland561
Thailand7W
Kazak3
India91
Kazak1
Colombia113
Colombia82
Colombia170
Colombia145
ColombiaNA1990
Ireland562
Colombia154
Colombia144
Colombia271
ColombiaNA1688
Alaska19
Kazak2
Ireland566
Ireland517
ColombiaNA1692
ColombiaNA1768
ColombiaNA1764
ColombiaNA1766
Vietnam2692
Taiwan175
Taiwan179
Taiwan188
HongKong246
HongKong230
HongKong218
Vietnam2705
Vietnam2704
Vietnam2706
HongKong209
HongKong222
Thailand4A
Taiwan198
Thailand13A
Thailand15A
Korea12
JapanO552
JapanO873
JapanK252
Korea23
Korea7
Korea2-3
Korea9
JapanK85
JapanK46
JapanK31
JapanK1
JapanK47
JapanK25
JapanK217
JapanK34
JapanK30
JapanK42
JapanK55
JapanK260
JapanK14
JapanK32
JapanF32
Taiwan190
JapanK44
JapanK74
JapanK68
JapanK269
JapanK69
JapanK96
JapanK61
JapanK83
JapanK43
JapanK48
JapanK35
JapanK41
JapanK101
JapanK18
JapanK80
JapanK9
JapanK26
JapanK89
0.1

al., 2000). Therefore, these primers were used for initial amplification of the in-del region and each PCR product was assumed to indicate one of two types. If a 1500-bp product was obtained, it was assumed to be either type Ib or IV, if a 1000-bp product was obtained, it was assumed to be either type Ia or IIIb and if a 700-bp product was obtained, it was assumed to be either type II or IIIa. Each isolate was then characterised to its corresponding in-del type using specific primers (Fig. 3.16 A–E, Table 2.1). Using this strategy in–del analysis was performed on all 74 isolates, of which 63 (85.1%) yielded useable PCR products, 4 (5.4%) produced multiple PCR products (which made it impossible to determine the different motifs at this locus), and 7 (9.5%) did not yield PCR products under the test conditions used despite repeated attempts. In this study 5 of the possible 8 in-del motifs were detected. Of the 63 isolates that yielded in-del PCR products, 31 (49.2%) were found to be of type II, 22 (34.9%) were of type IV, 6 (9.5%) were of type Ia, 3 (4.8%) were of type IIIa and 1 (1.6%) was of type IIIb (Tables 3.2 and 3.3).

Most PCRs yielded amplicons representative of the in-del type the isolate belonged to. However, when using primers cagF4584 and cagR5280, which span the in-del region, three different sized PCR products were found for type IV motifs. In this case, all isolates generated a product of 1500 bp, while strain J714 generated a PCR product that was slightly bigger (1600 bp) and strain 167233A generated a PCR product that was considerably smaller (1200 bp) (Fig. 3.15 A). These differences in sizes could potentially be due to insertions or deletions of nucleotides. Using primers specific for primers for type IV and I motifs that amplify the extreme right end of the in-del region (Fig. 3.15 E), the PCR product for isolate J714 was larger, indicating that this is the region where this strain has a longer nucleotide sequence (insertion). However, the PCR product for isolate 167233A showed no size variation using these primers, indicating that if sequences have been deleted from this strain they are not deleted at the far right of the in-del region. These were interesting observations, yet these variations in type IV motif sequence were very low in occurrence. Efforts to sequence the in-del regions in these isolates failed. Thus, the exact location of these insertion and deletion events could not be determined.

3.3.4. Genetic Diversity within intergenic regions in Irish H. pylori *isolates.*

3.3.4.1. Diversity preceding the *vacA* gene.

The non-coding regions of DNA upstream of the *vacA* gene were amplified using primers VACAP-L1 and VACAP-R1 (Table 2.1). The different sequence variants were differentiated

Fig. 3.16. Insertion-deletion motifs located at the 3′- end of the *cag* PAI found in this collection of Irish *H. pylori* isolates.

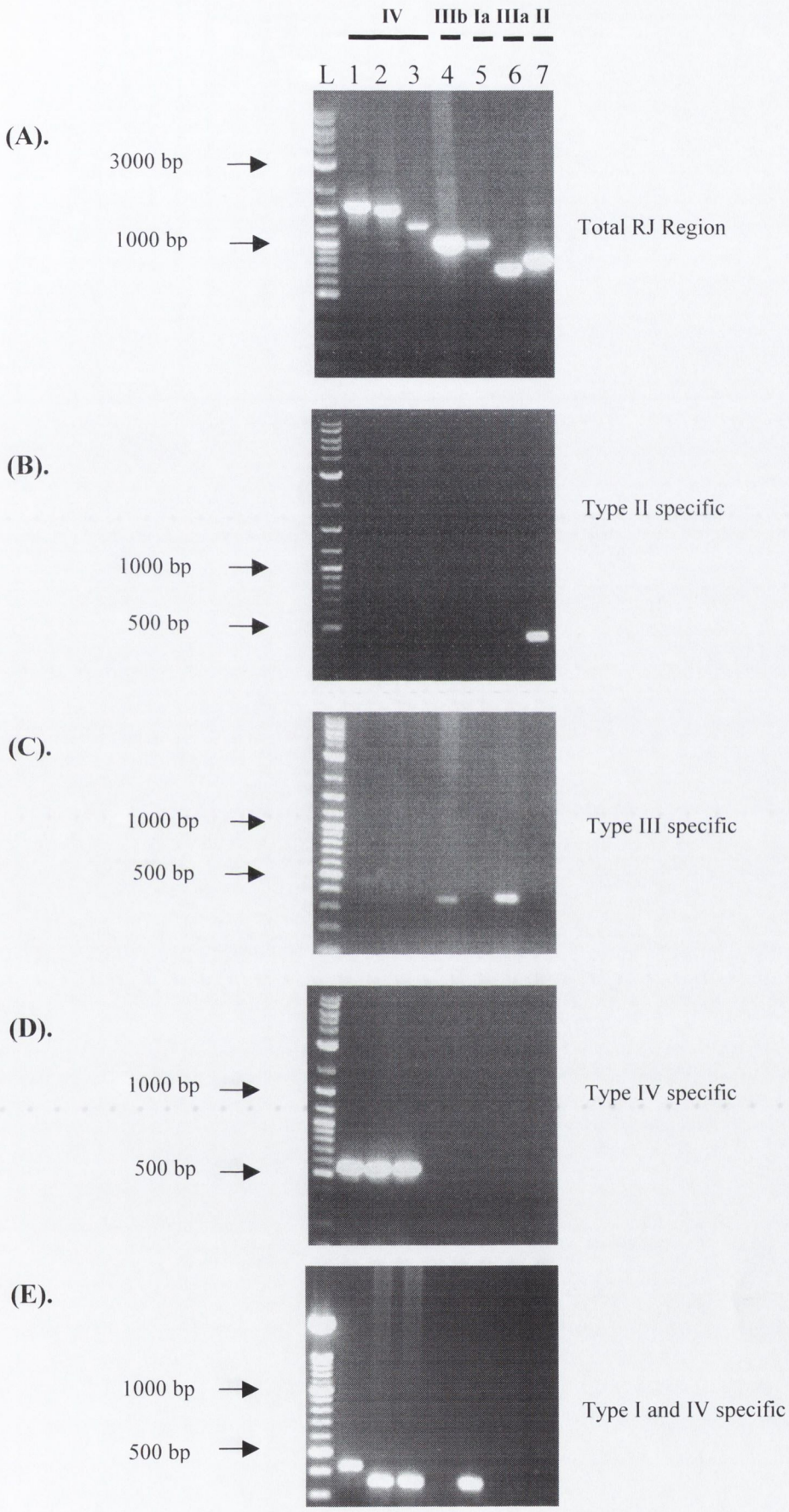
(A). Insertion-deletion (in-del) typing of Irish *H. pylori* isolates. L =1000-bp ladder, isolates 1 = J714, 2 = 79862A, 3 = 167233A, 4 = 563C, 5 = FR296, 6 = J517, 7 = J063. Total right junction (RJ) was amplified using primers cagF4584 and cagR5280 (Table 3.2). The different motifs cannot be fully typed using these primers; therefore, specific amplification of the motif types was used (B–E).

(B). Amplification of type II specific motifs using primers IS606-1692 and cagR5280 (Table 3.2). Only isolate J063 produced a PCR product indicating that it is of type II.

(C). Amplification of type III specific motifs using primers fcn unk(6) and cagR5280 (Table 3.2). Isolates 563C and J517 both produced PCR products indicating that they are of type III. These isolates can be further characterised into types IIIb and IIIa (563C and J517, respectively) based on the different size PCR products they yield when the total in-del region is amplified (3.16 A).

(D). Amplification of type IV specific motifs using primers cagF4584 and Xins.R (Table 3.2). Isolates J714, 79862C and 167233A produced PCR products indicating that they are of type IV.

(E). Amplification of type I and IV specific motifs using primers cagF4856 and cagR5280 (Table 3.2). Isolates J714, 79862A, 167233A and FR296 produced PCR products confirming that J714, 79862A and 167233A are of type IV, and that FR296 is of type Ia. Isolate J714 shows a larger PCR product than other type IV isolates. This can also be seen in gel A. As type I and IV specific primers amplify the right end of the motifs, isolate J714 probably has an insertion of sequences at this location resulting in these larger PCR products.


IV
IIIb
Ia
IIIa
II
L 1 2 3 4 5 6 7
(A).
3000 bp
1000 bp
Total RJ Region
(B).
1000 bp
500 bp
Type II specific
(C).
1000 bp
500 bp
Type III specific
(D).
1000 bp
500 bp
Type IV specific
(E).
1000 bp
500 bp
Type I and IV specific

based on the size of the PCR products generated. Three of the 5 variants previously described (VacAP1–5, Bereswill *et al.*, 2000) were found in this collection of 74 Irish *H. pylori* isolates, namely, VacAP2 (1.4%), VacAP3 (89.2%) and VacAP4 (9.4%) (Fig. 3.17, Tables 3.2 and 3.3). Efforts to sequence these sequence variants failed. Interestingly, 73.8% of VacAP3 isolates were of *vacA* type s1a m2 and 26.2% were of type s1a m1. Most of the VacAP4 sequence variants were of *vacA* allelic type s2 m2 (42.8%) with the remaining VacAP4 types divided evenly between *vacA* types s1a m1 and s1a m2 (28.6% each). The only isolate with a VacAP2 sequence variant exhibited a *vacA* type of s1a m1. All isolates with a *vacA* allelic type s2 m2 exhibited the VacAP4 sequence variant.

3.2.4.2. Diversity preceding the *ribA* gene.

The non-coding regions of DNA upstream of the *ribA* gene were amplified using primers RIBAP-L1 and RIBAP-R1 (Table 2.1). Four of the 5 sequence variants previously described (RibAP1–5, Bereswill *et al.*, 2000) were detected in this collection of isolates, namely, RibAP1 (93.2%), RibAP2 (1.4%), RibAP3 (4.0%) and RibAP5 (1.4%) were detected in this collection of isolates (Fig. 3.18, Tables 3.2 and 3.3). Fig. 3.19 and 3.20 show the sequences of the variants found and the diversity between each isolate can be seen. Each sequence variant contained a relatively conserved region (5′-and 3′-ends) and a region of diversity located in the centre of the sequence. The diverse sequences for each sequence variant are aligned in Fig. 3.20. Each sequence variant has a different sized region of diversity which contributes to the differences in PCR product size. From Fig. 3.20 it can be seen that the region of diversity in RibAP2 is a truncated version of diverse sequence in RibAP3, whilst the other sequences show little identity.

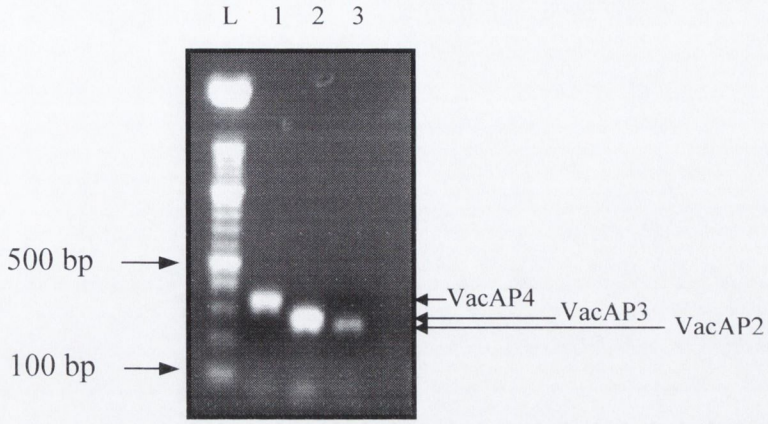


Fig. 3.17. PCR products of the intergenic regions preceding the *vacA* gene run on a 1.8% agarose gel.

Amplification of the intergenic regions preceding the *vacA* gene, using primers VACAP-L1 and VACAP-R1 (Table 3.1.). Three different sequence variants were found in this collection of isolates (VacAP2, VacAP3 and VacAP4). L = 1000-bp ladder, 1 = FR138 (VacAP2, 230 bp), 2 = 577A (VacAP3, 290 bp) and 3 = 167233A (VacAP4, 340 bp).

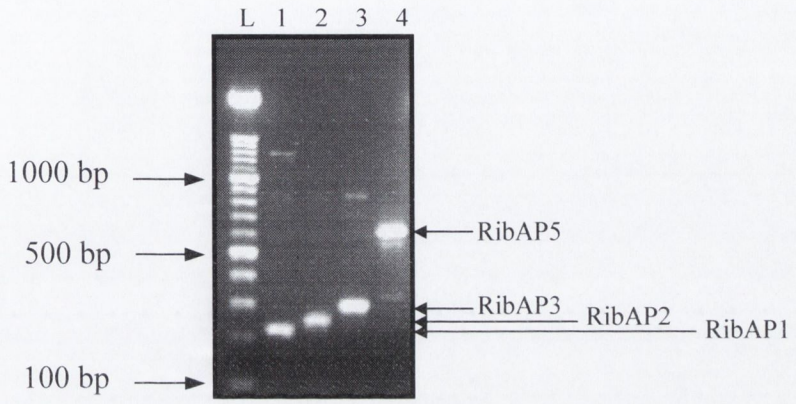


Fig. 3.18. PCR products of the intergenic regions preceding the *ribA* gene run on a 1.8% agarose gel.

Amplification of the intergenic regions preceding the *ribA* gene, using primers RIBAP-L1 and RIBAP-R1 (Table 3.2). Four different sequence variants were found in this collection of isolates (RibAP1, RibAP2, RibAP3 and RibAP5). L = 1000-bp ladder, 1 = 167233A (RibAP1, 220 bp), 2 = FR298 (RibAP2, 250 bp), 3 = MI506 (RibAP3, 300 bp) and 4 = J159 (RibAP5, 600 bp).

Isolate 167233A(RibAP1)

GGCTTGGGGAGCTTTAGGGATTTATAATAAACATGAAATCCGTTCCAGCTTAAGCCCCAA
TGGTTTGAATTTGACTTACAAATTTTAA**GAATTTAAAATTTAAGAA**AATTATAAAGAGTT
TTGATAGAATACCTTTTTTTAAAAGACGCGTTTTAAAACTAATTTTAG

Isolate FR298(RibAP2)

CTTGGGGAAGCTTTAGGGATTTATAATAAACATGAAATCCGTTCTAATTTAAGCCCTAAT
GGTTTGAATTTGACTTACAAATTTTAA**AGGAGTAACCCGCATCAAGCGATTTTAAAAACA**
TTTTTATCGGTGCAAATGATAAAGAGTTTTGATAGAATACCTTTTTTTAAAAGATGCGTT
TTAAAACTAATTTTAG

Isolate MI506(RibAP3)

GGGCTTGGGGAAGCTTTAGGGATTTATAATAAACATGAAATCCGTTCCAGCTTAAGCCCC
AATGGTTTGAATTTGACTTACAAATTTTAA**AGGAGTAGCCCGCATCAAGCGATTTTAAAA**
ACATTTTTATCGGTGAGGTTTTTAAAAGATAGGGTTGTTCGCTTAGTAACTTTTTAAGAA
AATTATAAAGAGTTTTGATAGAATACCTTTTTTTAAAAGATGCGTTTTAAAACTAATTTT
A

Isolate J159(RibAP5)

GGCTTGGGGAAGCTTTAGGGATTTATAATAAACATGAAATCCGTTCCAGCTTAAGCCCCA
ATGGTTTGAATTTGACTTACAAATTTTAA**AAAAAGACCCTATGTCAAGTGATTTTAAAAA**
CCTTTTCATCGGTGCGGATATAATTCTAAACGCACCAAGTCTAACTTATTGGATAGCGAA
AAAGGTCCTATCCATTCTCATTCGTTATCATCTCTCATTGATGCCCGCCTTTTAAGGAAT
ACAATGAAGCTAAAAGCAAATACCTCTCTTCAAAATCTCTGTCAATTCTAACCAACCCCT
ATCCGCCCAACTAAACGAGTATGATAAGAAAAAGTCAGCGTTAAGAGCCGTTTTAAACAC
AAGGTGAGATGATAAATCTAAATCCTCTAGGTTTGGGTTCTCCACCTATCTATAACGGAG
CGTATAGTTTTAGGGGTGGTGAGTGTGTTAGAGATTCTTAAAAGATTAGGTTGTTCGTTC
CATAACTTTTTAAGAAAATTATAAAGAGTTTTGATAGAATACCTTTTTTTAAAAGACGCG
TTTTAAAACTAATTTTAG

Fig. 3.19. Partial nucleotide sequences of the sequence variants preceding the *ribA* gene found in this collection of Irish *H. pylori* isolates.

Partial nucleotide sequences of the intergenic sequence variants (RibAP1, RibAP2, RibAP3 and RibAP5) preceding the *ribA* gene found in this collection of Irish *H. pylori* isolates. Genetic diversity in this region is the result of inserted sequences (red). These insertions vary in length and sequence and result in the different sized PCR products.

```
RibAP1      ---------- ---------- ---------- ---------- ---------- ----------
RibAP2      ---------- ---------- ---------- ---------- ---------- ----------
RibAP3      ---------- ---------- ---------- ---------- ---------- ----------
RibAP5      CAAGGTGAGATGATA AATCTAAATCCTCTA GGTTTGGGTTCTCCA CCTATCTATAACGGA GCGTATAGTTTTAGG GGTGGTGAGTGTGTT


RibAP1      ---------- ---------- ---------- --
RibAP2      ---------- ---------- ---------- --
RibAP3      ---------- ---------- ---------- --
RibAP5      AGAGATTCTTAAAAG ATTAGGTTGTTCGTT CCATAACTTTTTAAG AA
```

Fig. 3.20. Aligned nucleotide sequences of the four sequence variants found preceding the *ribA* gene in this collection of Irish *H. pylori* isolates.

(A). Nucleotide sequence alignment of the four different sequence variants (RibAP1, RibAP2, RibAP3 and RibAP5) preceding the *ribA* gene lacking insertions. The inserted sequences are integrated between the bolded residues TAA and AAT (black sequences Fig. 3.19). The nucleotide sequences lacking the inserted sequences display a high level of sequence identity.

(B). Nucleotide sequence alignment of the four different inserted sequences (red sequences Fig. 3.19) found within the four different sequence variants (RibAP1, RibAP2, RibAP4 and RibAP5) preceding the *ribA* gene detected in this collection of Irish *H. pylori* isolates. The genetic diversity in these inserted sequences is extensive, with the exception of RibAP2, which is a truncated version of RibAP3.

(A)

```
167233A(RAP1)      -GGCTTGGGGA-GCT TTAGGGATTTATAAT AAACATGAAATCCGT TCCAGCTTAAGCCCC AATGGTTTGAATTTG ACTTACAAATTTTAA
FR298(RAP2)        ---CTTGGGGAAGCT TTAGGGATTTATAAT AAACATGAAATCCGT TCTAATTTAAGCCCT AATGGTTTGAATTTG ACTTACAAATTTTAA
MI506(RAP3)        GGGCTTGGGGAAGCT TTAGGGATTTATAAT AAACATGAAATCCGT TCCAGCTTAAGCCCC AATGGTTTGAATTTG ACTTACAAATTTTAA
J159(RAP5)         -GGCTTGGGGAAGCT TTAGGGATTTATAAT AAACATGAAATCCGT TCCAGCTTAAGCCCC AATGGTTTGAATTTG ACTTACAAATTTTAA

167233A(RAP1)      AATTATAAAGAGTTT TGATAGAATACCTTT TTTTAAAAGACGCGT TTTAAAACTAATTTT AG
FR298(5RAP2)       AATGATAAAGAGTTT TGATAGAATACCTTT TTTTAAAAGATGCGT TTTAAAACTAATTTT AG
I506(RAP3)         AATTATAAAGAGTTT TGATAGAATACCTTT TTTTAAAAGATGCGT TTTAAAACTAATTTT A-
J159(RAP5)         AATTATAAAGAGTTT TGATAGAATACCTTT TTTTAAAAGACGCGT TTTAAAACTAATTTT AG
```

(B)

```
RibAP1         GAATTTAAAATTTAA GAA------------ --------------- --------------- --------------- ---------------
RibAP2         AGGAGTAACCCGCAT CAAGCGATTTTAAAA ACATTTTTATCGGTG CA------------- --------------- ---------------
RibAP3         AGGAGTAGCCCGCAT CAAGCGATTTTAAAA ACATTTTTATCGGTG AGGTTTTTAAAAGAT AGGGTTGTTCGCTTA GTAACTTTTTAAGAA
RibAP5         AAAAAGACCCTATGT CAAGTGATTTTAAAA ACCTTTTCATCGGTG CGGATATAATTCTAA ACGCACCAAGTCTAA CTTATTGGATAGCGA

RibAP1         --------------- --------------- --------------- --------------- --------------- ---------------
RibAP2         --------------- --------------- --------------- --------------- --------------- ---------------
RibAP3         --------------- --------------- --------------- --------------- --------------- ---------------
RibAP5         AAAAGGTCCTATCCA TTCTCATTCGTTATC ATCTCTCATTGATGC CCGCCTTTTAAGGAA TACAATGAAGCTAAA AGCAAATACCTCTCT

RibAP1         --------------- --------------- --------------- --------------- --------------- ---------------
RibAP2         --------------- --------------- --------------- --------------- --------------- ---------------
RibAP3         --------------- --------------- --------------- --------------- --------------- ---------------
RibAP5         TCAAAATCTCTGTCA ATTCTAACCAACCCC TATCCGCCCAACTAA ACGAGTATGATAAGA AAAAGTCAGCGTTAA GAGCCGTTTTAAACA
```

3.4. Discussion

3.4.1. vacA *genotyping.*

As stated in Section 3.3.1, typing of the signal sequences and mid-regions of the *vacA* gene of this collection of Irish *H. pylori* isolates revealed that the predominant allele was s1a m2, which is in accordance with a previous Irish study (Ryan *et* al., 2000). The lack of occurrence of the s2 m1 allelic combination in this collection of isolates also confirms earlier observations that this combination is extremely rare (Atherton *et al.*, 1995). This collection of isolates was obtained from patients suffering from peptic ulcer disease or who had duodenal ulcers, illnesses that the s1 allele has been associated with (Strobel *et al.*, 1998). Furthermore, it has been demonstrated that strains possessing the s1 m1 allele of the *vacA* gene produce higher levels of the vacuolating cytotoxin than strains with the s1 m2 and s2 m2 allelic combinations (Atherton *et al.*, 1995). This indicates that the majority of Irish *H. pylori* isolates (69%) would appear to produce less VacA toxin. However, previous findings indicated that there was no association with *vacA* genotype and virulence among Irish *H. pylori* isolates (Marshall *et al.*, 1999).

The sequences of the PCR products that were generated while typing the signal sequences and mid-regions of the *vacA* genes of these Irish isolates and their alignment to *H. pylori vacA* sequences previously characterised can be seen in Figs. 3.6–3.10 A and B. These PCR products were sequenced and aligned for confirmatory purposes. The signal sequence and mid-regions of the four strains sequenced (167233A, 9613A, J556 and 577A) displayed between 93–97 % identity to those of previously characterised strains.

Many studies have shown an apparent geographical distribution of *H. pylori* putative virulence genes (Achtman *et al.*, 1999; Van Doorn *et al.*, 1999a; Kersulyte *et al.*, 2000). The geographical differences in the distributions of *vacA* alleles strongly suggest a geographical partitioning of *H. pylori* populations throughout the world (Table 3.1, Fig. 3.21). The frequency of the predominant *vacA* allele found in this study [s1a m2, 69%] is similar to those found in Taiwan [82%] (Lin *et al.*, 2000) and China [85%] (Pan *et al.*, 1998). Although Northern Europe *H. pylori* isolates share the same predominant *vacA* allelic combination (s1a m2), the frequency of occurrence of this allelic combination is much lower than those of Irish H. pylori isolates [32% (van Doorn *et al.*, 1999)] and are thus dissimilar. Although these results suggest that Irish *H. pylori* isolates are similar to Eastern Asian isolates with respect to this locus, *vacA* alleles have been shown to differ in type and frequency among East Asian

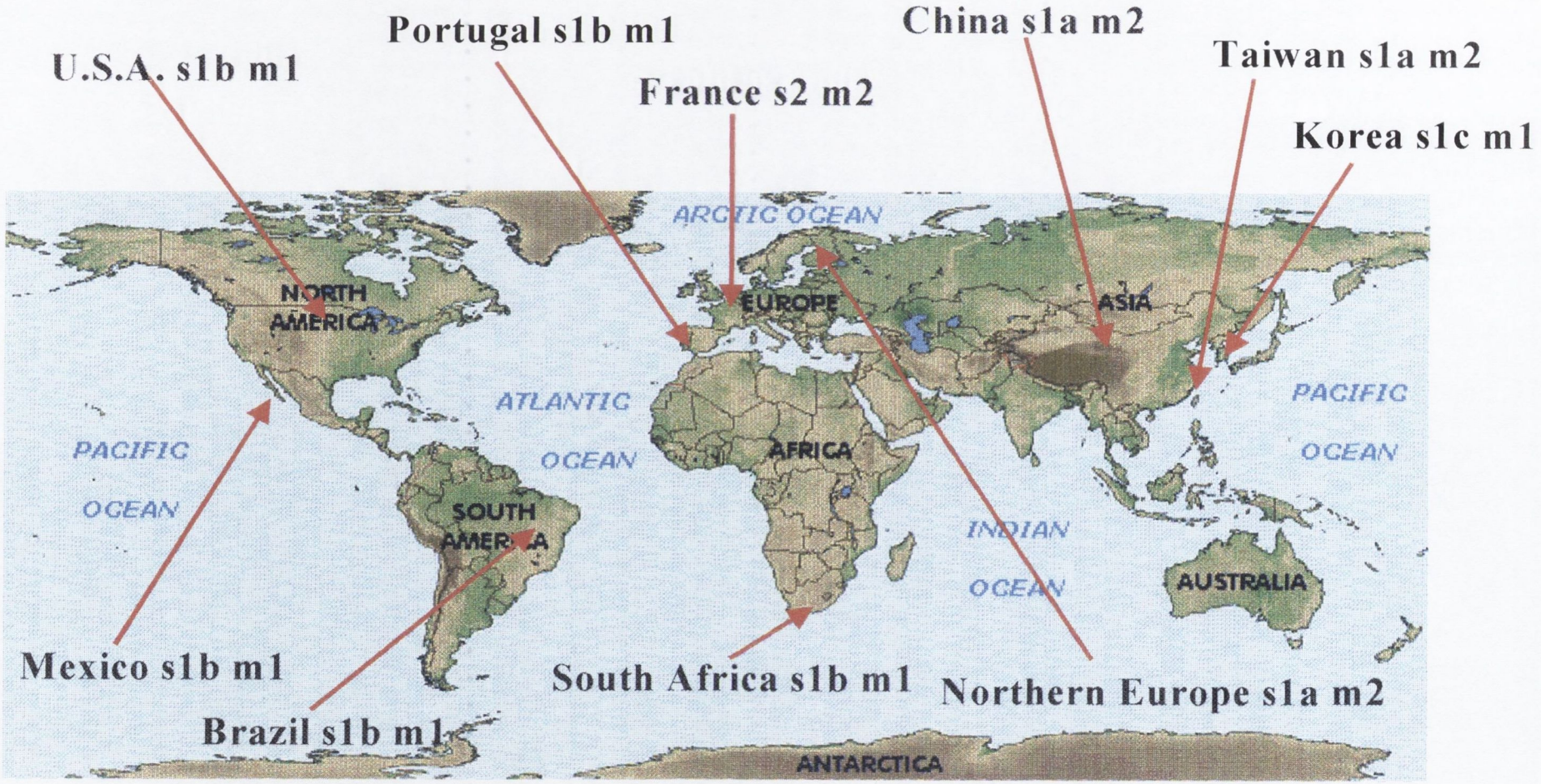


Fig. 3.21. Geographical distribution of *H. pylori vacA* alleles throughout the world.

isolates (Ito *et al.*, 1997; Pan *et al.*, 1998; Lin *et al.*, 2000). Furthermore, a recent study has shown that *vacA* alleles not only vary within Eastern Asian countries, but they also differ regionally (China: Beijing – s1a m2, Shanghai – s1c m2 and Guangxi – s1c m1, Wang *et al.*, 2003). Thus, Irish *H. pylori* isolates may only share genetic similarities with isolates from certain countries or regions within Eastern Asia.

It is also interesting to note that there seems to be a restriction of *vacA* alleles within this collection of Irish *H. pylori* isolates, as the majority of isolates exhibited only two allelic combinations of this gene (95.9% were s1a m1 / m2). In contrast *vacA* alleles from different geographical regions exhibit more s and m allelic combinations with varying frequencies of occurrence (Van Doorn *et al.*, 1999a). The limited number of *vacA* alleles suggest that this collection of Irish *H. pylori* isolates may be weakly clonal with respect to this locus, as numerous distinct alleles are generated within panmictic organisms as a result of repeated genetic exchange between divergent lineages (Achtman *et al.*, 1999). Evidence of genetic homogeneity within *H. pylori* isolates has been found where the human population is traditionally an isolated one (Campbell *et* al., 1997 and Evans *et* al., 1999). It must, however, be noted that neither the ethnic nor geographical origins of the individuals from where the isolates used in this study were obtained are known. Thus, it cannot be stated for certain that these isolates are solely from native Irish individuals as they may encompass both Irish and non-Irish (emigrated) individuals. However, as these isolates were obtained at time when emigration into Ireland was low (before 1997), it is assumed that the *H. pylori* isolates used in this study represent strains isolated from a native Irish population.

3.4.2. cagA analysis.

3.4.2.1. Frequency data.

All isolates investigated were found to possess the *cagA* gene. The isolates used in this study were obtained from patients with severe clinical diseases. It has been previously documented that the *cagA* locus is associated with increased severity of disease (Blaser *et al.*, 1992). The uniform presence of the *cagA* gene in this collection of isolates is high when compared to other European countries [France, Italy, Northern and Eastern Europe = ~75% $cagA^+$] (Atherton *et* al., 1995, Van Doorn *et* al., 1995), and seems to resemble the Asian situation where a much higher percentage of $cagA^+$ isolates are found [China = 98%, East Asia = 90.1%] (Pan *et* al., 1997; Van Doorn 1999a). However, a previous Irish study included isolates from asymptomatic patients and showed a lower prevalence of the *cagA* gene [69%] (Dundon *et al.*, 2000). Thus, the *cagA* prevalence in these other European studies could

potentially be lower due to the inclusion of asymptomatic patients. However, the clinical symptoms of the patients included in these other European studies were not stated. It has been suggested that the *cagA* gene may not a useful marker to predict clinical outcome (Miehlke *et* al., 1996; Ryan *et* al., 2000). As the *cagA* gene was present in all the Irish isolates examined and all isolates were obtained from patients with duodenal or gastric ulcers the ability to predict clinical outcome in Irish patients harbouring such *H. pylori* isolates possessing this locus alone may be useful. However, inclusion of asymptomatic patients needs to be carried out to confirm this observation. In addition, the high frequency of this locus further suggests a restriction of genetic diversity within virulence-associated loci within this collection of Irish *H. pylori* isolates. However, as the repeat sequence number at the 3′-end of the *cagA* gene is known to vary (Evans *et al.*, 1998) the genetic restriction of the *cagA* gene cannot be determined based on just its presence.

3.4.2.2. Repeat sequences at the 3′-end of the *cagA* gene.

Considerable variability is located at the 3′-end of the *cagA* gene. This region contains a 102-bp repeat region that can vary in clinical isolates (Evans *et al.*, 1998; Fig.3.2 A). The amino acid sequence of the CagA C-terminal repeats discriminates between Eastern (East Asian) and Western (USA and Europe) *H. pylori* isolates (Covacci *et al.*, 1993; Tummuru *et al.*, 1993; Yamaoka *et al.*, 1998; Fig. 3.2 A and B). The nucleotide and amino acid sequences of an Irish *H. pylori* isolate containing two 102-bp repeats is shown in Fig. 3.14 A and B, respectively. The poly-N motif, two EPIYA motifs and the two 102-bp repeats are indicated. The amino acid sequence of the C-terminus of the CagA protein of this representative Irish isolate is similar to the sequences of Western (USA and Europe) isolates [FPLKRHDKVDDLSKV] rather than Eastern (East Asian) isolates [KIASAGKGVGGFSGA]. This is an informative locus as it definitively differentiates between Eastern and Western isolates. In the present analysis of Irish *H. pylori* isolates there is an indication that Irish isolates are more similar to isolates from certain parts of East Asia based on the basis of the *vacA* allele frequency data and the presence of the *cagA* gene. However, sequence analysis of an Irish *cagA* locus points to the contrary. This indicates that, although the *cagA* gene in Irish *H. pylori* isolates display s a similar frequency of occurrence to those in certain parts of East Asia, this locus in Irish isolates is not ancestrally related to the locus in East Asian isolates due to sequence dissimilarity.

The sequence of the 3′-end of the *cagA* gene is of great interest, as recently the CagA protein has been shown to be translocated into gastric epithelial cells where it becomes tyrosine phosphorylated and adopts a role in eukaryotic signal transduction pathways (Odenbreit *et al.*, 2000; Stein *et al.*, 2000). The 3′-end of this gene contains a sequence that fits the consensus sequence for a tyrosine phosphorylation motif (TPM). In addition, the repeat sequence located in this region contains an EPIYA motif. This motif contains a conserved tyrosine residue (Y), which recurs in each repeat sequence (Fig. 3.14 B). If this EPIYA motif is in fact a TPM, increasing the number of repeat sequences in this region would potentially increase the biological activity of the protein.

3.4.2.3. Phylogenetic analysis of the repeats located at the 3′-end of the *cagA* gene.

Phylogenetic analysis of the repeat region located at the 3′-end of the *cagA* gene confirmed that Irish *H. pylori* isolates are more similar to Western isolates with respect to this locus (Fig. 3.15). Six Irish *H. pylori* sequences were used in the construction of this tree. The Irish isolates clustered with Western isolates. Previously, it was shown that the predominant *vacA* alleles from Ireland, China and Taiwan are similar, yet in this tree the Chinese (Hong Kong) and Taiwanese isolates are far removed from the Irish isolates, indicating no ancestral linkage between these Asian and the Irish *H. pylori* isolates. The present analysis of the repeat region indicates that one repeat is the predominant number of *cagA* 3′-repeats (73.8%) within Irish *H. pylori* isolates, again demonstrating a low level of diversity within this genetic locus. There is, unfortunately, very little data on the distribution of *cagA* repeats throughout the world. However, this typing method has aided in further characterising Irish isolates.

3.4.2.4. Analysis of the motifs located at the 3′-end of the *cag* PAI.

The eight different motifs that can occur at the right end of the *cag* PAI (Fig. 3.3) show strong geographical partitioning (Table 3.1, Fig. 3.22). The predominant motifs found in the present study were types II (49.2%) and IV (34.9%). Type II motifs have been found to predominate in Japan and China, and type IV motifs were found in an English strain and strains from West Virginia, USA (Kersulyte *et al.*, 2000). Although type II motifs were found predominately in Japan and China, these motifs were also found in strains from Northern Europe (Sweden and Lithuania). However, these Northern European strains also displayed an abundance of type I and III motifs.



Fig. 3.22. Geographical distribution of *H. pylori* in-del motifs throughout the world

The present study demonstrates that these motifs occur at a relatively high level in this collection of Irish *H. pylori* isolates. The typing of the in-del motifs within Irish *H. pylori* isolates shows that they have a rather characteristic population structure. These Irish isolates are unlike Northern and Southern European strains as they lack type I and III motifs, and although Irish isolates share the same predominant motif as Eastern Asian strains (type II), the frequencies of occurrence of this motif differ (Ireland = 49.2%, East Asia = 95%), and Asian isolates also lack the type IV motif that occurs regularly within the Irish isolates. As no particular geographical region so far shares similarities with Irish *H. pylori* isolates at this locus, this novel in-del combination may be characteristic of Western European isolates.

3.4.3. Sequence variations within intergenic regions.

The results obtained from this study show that intergenic polymorphisms upstream of the *vacA* and *ribA* genes contribute to the genetic diversity of *H. pylori*. Analysis of non-coding regions of the *H. pylori* genome has been previously carried out (Bereswill *et al.*, 2000). Five different sequence variants were found for regions upstream of both the *vacA* and *ribA* gene. Herein three sequence variants were found in front of the *vacA* gene and four sequence variants were found in front of the *ribA* genes. The commonest sequence variants were RibAP1 and VacAP3. These sequence variants were also the most frequent found by Bereswill *et al.* (2000). Diversity within the intergenic regions preceding the *vacA* gene is due to the presence or absence of small stretches of DNA. However, the diversity found preceding the *ribA* gene is due to the insertions of sequences of varying sizes. The sequences inserted into the DNA preceding the *ribA* gene in Irish isolates can be seen in Figs. 3.19 and 3.20. The four insertion sequences found within the four different sequence variants in Irish *H. pylori* isolates are aligned in Fig. 3.20 B. Similarity can be seen between the RibAP2 and RibAP3 variants, where RibAP2 seems to be a truncated version of RibAP3. This may represent an insertion event of a particular sequence (RibAP3) or a partial excision event of that sequence generating RibAP2. Types RibAP3 and RibAP5 show some sequence identity although the sequence of RibAP5 is much larger. Sequence type RibAP1 bears little identity to the other variants. In addition, the high prevalence of one intergenic sequence variant preceding both the *vacA* and *ribA* genes again suggests a restriction of diversity at these loci within this collection of Irish *H. pylori* isolates.

Bereswill *et* al. (2000) suggested that certain non-coding sequence variants are associated with distinct geographical regions. The RibAP2 variant was found in Turkish and Chinese

isolates but not in isolates from Germany or other Western European countries. Interestingly the RibAP2 sequence variant was detected in only one of the Irish isolates studied. Polymorphisms in front of the *vacA* and *ribA* genes could potentially contribute to variations in transcriptional levels of these genes. Variations in transcriptional levels would be under a strong selection pressure as the *ribA* and *vacA* genes are housekeeping and virulence genes, respectively.

Initially it seemed that this collection of Irish *H. pylori* isolates did not show any similarities with other European isolates, and on the basis of the uniform presence of the *cagA* gene and the allelic combinations of the *vacA* gene it appeared that isolates from certain regions of East Asia may be more similar to Irish isolates. These similarities may reflect independent specific host adaptation in these particular regions or they could be due to an ancestral relationship. Sequence and phylogenetic analysis of the 3′-end of the *cagA* gene showed that this collection of Irish isolates displayed a Western structure, with no indication of an ancestral relationship of Irish and East Asian *H. pylori* isolates. However, the in-del motifs revealed that, although the Irish *H. pylori* population shared some similarities to Asian isolates with respect to prevalence of motifs, it appeared that these isolates may be unique with regard to their population structure and represent a Western European *H. pylori* structure.

Finally each typing scheme indicated a certain amount of diversity between isolates, but not as much as one would expect with a panmictic organism. This could indicate that Irish *H. pylori* isolates maybe weakly clonal and could also be indicative of an insular country.

Chapter 4

The genetic diversity of Irish* Helicobacter pylori *isolates as revealed by fluorescent amplified length polymorphism (FAFLP) – PCR.

4.1. Introduction 76

4.2. Methods 81

4.2.1. Fluorescent amplified fragment length (FAFLP)-PCR fingerprinting analysis 81

4.2.1.1. Generation of FAFLP-PCR fingerprints 81

4.2.1.2. Automated fingerprinting 82

4.2.1.3. Genotypic analysis 82

4.3. Results 84

4.3.1. FAFLP-PCR fingerprinting of Irish H. pylori *isolates* 84

4.3.2. Phylogenetic analysis of Irish H. pylori *FAFLP – PCR fingerprints* 85

4.3.3 Phylogenetic anlaysis of Irish H. pylori *isolates with respect to isolates from different geographic regions* 85

4.4. Discussion 86

4.3.1. Phylogenetic analysis of Irish H. pylori *FAFLP-PCR fingerprints* 86

4.3.2 Phylogenetic anlaysis of Irish H. pylori *isolates with respect to isolates from different geographic regions* 86

4.1. Introduction.

Microbial typing is a useful tool in clinical epidemiology for defining the source and route of infection and for studying the persistence and re-infection rates, clonal selection in the host and bacterial evolution. Reproducibility, discriminatory power, ease of performance and interpretation, cost, and toxic procedures associated with each method must be taken into consideration when applying a particular typing technique. A variety of discriminatory fingerprinting techniques have been applied to *H. pylori*, most of which are PCR based. These techniques include Random Amplified Polymorphic DNA (RAPD)-PCR (Marshall *et al.*, 1995), Repetitive Extragenic Palindromic (REP)-PCR fingerprinting (Li *et al.*, 2002b), Oligofingerprinting (Marshall *et al.*, 1996), Ribotyping (Owen *et al.*, 1994), PCR - Restriction Fragment Length Polymorphism (RFLP) (Clayton *et al.*, 1993), PCR of specific loci (Atherton *et al.*, 1995), Pulsed Field Gel Electrophoresis (PFGE) (Han *et al.*, 2000), Multi-locus Sequence Typing (MLST) (Achtman *et al.*, 1999) and Amplified Fragment Length Polymorphism (AFLP)-PCR (Gibson *et al.*, 1998; Owen *et al.*, 2001). A key feature of *H. pylori* is the high degree of genetic diversity it exhibits. This diversity is thought to be a result of frequent recombination and *in vivo* point mutations (Suerbaum *et al.*, 1998; Wong *et al.*, 1999). Two strains isolated from individual stomachs can easily be distinguished using one of the previously mentioned techniques.

PCR-RFLP is a technique where a specific gene or part of a gene is amplified by PCR and then the PCR product is digested with a restriction enzyme to generate a specific restriction pattern. This technique has been successfully applied to *H. pylori* for epidemiological purposes (Clayton *et al.*, 1993). However, although specific genes from different strains may accidentally be matched in restriction sites for one or two enzymes, they may differ at many other sites throughout the entire genome, reflecting the limitation of the technique. The urease gene cluster has been used for typing *H. pylori* isolates. Restriction of an amplified 2.4-kb fragment (*ureA* and *ureB*) with *Hae*III or *Hae*III-*Mbo*I verified the diversity of *H. pylori* isolates from different patients (Akopyanz *et al.*, 1992; Foxall *et al.*, 1992; Hurtado and Owen, 1994; Owen *et al.*, 1994). Other studies have used PCR-RFLP to assess clarithromycin resistance within certain populations (Kim *et al.*, 2002; Piana *et al.*, 2002; Eun *et al.*, 2003). A major drawback of using PCR-RFLP is that it can only detect a single base-pair change at a particular site and it is unclear whether the single base-pair change detected by this technique results in an amino acid change.

REP-PCR is a method that amplifies DNA segments using primers corresponding to naturally occurring interspersed repetitive elements in bacteria, such as repetitive palindromic sequences. The PCR products generated are run out on an agarose gel yielding a banding display or fingerprint. Applying REP-PCR to *H. pylori* has shown that geographical differences among strains exist in the conserved non-coding regions throughout the entire bacterial genome (Li *et al.*, 2002b). Li *et al.* (2002b) not only showed that REP-PCR is a useful fingerprinting technique but it can also be an informative technique for studying the geographical relationship of *H. pylori* isolates. REP-PCR has also been used to investigate *H.* pylori-associated diseases, with contrasting results. Kwon *et al.* (1998) described REP-PCR as being useful in identifying disease-specific strains, however, van Doorn *et al.* (1998c) found no association between REP-PCR fingerprints and disease-specific strains.

RAPD-PCR is a technique that has been applied to a number of bacterial species for many diagnostic and epidemiological reasons. RAPD-PCR works on the basis of non-specific primers randomly annealing to different parts of an organism's genome. These primers are then used to amplify random segments of the organism's genome. The PCR products generated are run out on an agarose gel and the band pattern obtained represents a genetic fingerprint of the isolate. This technique samples the genome in its entirety. In the case of *H. pylori* RAPD-PCR has been used to show that isolates from a single patient are homogeneous at a single gastric site (Marshall *et al.*, 1995). Marshall *et al.* (1995) also showed that *H. pylori* isolates from different patients displayed noted heterogeneity, a characteristic of panmictic organisms. This technique has also been used to assess the effectiveness of patients undergoing triple-therapy (Cellini *et al.*, 2003) and to identify *H. pylori* clonal transmission within a family unit (van der Ende, *et al.*, 1996). Studies have used RAPD-PCR in conjunction with other typing methods. RAPD-PCR used in conjunction with REP-PCR to identify disease-specific strains demonstrated that these techniques were partially complementary, with REP-PCR being more useful to identify disease-specific strains (Kidd *et al.*, 2001). RAPD-PCR used in conjunction with PCR-RFLP demonstrated that RAPD-PCR has a greater discriminatory power (Yakoob *et al.*, 2001; Smith *et al.*, 2002a). The ease of use and the high discriminatory power of RAPD-PCR make it an attractive choice when typing *H. pylori* isolates.

Oligofingerprinting is a fingerprinting technique based on the ability of oligonucleotide probes containing short repetitive sequence motifs [$(GACA)^4$, $(GT)^8$, $(GTG)^5$ and $(GGAT)^4$]

to anneal to restricted enzyme digests of an organism's genome. This technique is an extension of Southern blotting. This technique proved to be a very powerful tool for fingerprinting *H. pylori* isolates (Marshall *et al.*, 1996).

Amplification of specific loci throughout the *H. pylori* genome has been applied to both coding regions (virulence genes) (Atherton *et al.*, 1995) and non-coding regions (Bereswill *et al.*, 2000). This approach can differentiate between unrelated strains but it has the disadvantage of the need to amplify many different loci in different reactions to find differences. Moreover, two strains can generate the same allelic combinations even though they are unrelated. However, a recent study demonstrated that loci-specific PCR of virulence genes can be carried out in a multiplex reaction (Monstein and Ellnebo-Svedlund, 2002) thereby considerably reducing the time taken to type specific loci.

PFGE is a technique whereby genomic DNA digested with a restriction enzyme with few restriction sites is run on an agarose gel using alternating electric fields. This technique resolves large fragments of DNA of different sizes. PFGE is a very sensitive fingerprinting method and has been used to detect epidemiological relationships in families infected with *H. pylori* (Han *et al.*, 2000). PFGE is technically demanding. Although their method has become the 'gold standard' for epidemiological typing of some bacterial species, e.g., *Staphylococcus aureus* it has been not frequently used in the analysis of *H. pylori* isolates.

MLST uses sequences of housekeeping genes to differentiate bacterial isolates. This method has been successful in identifying weakly clonal groups in *Neisseria meningitidis* (Maiden *et al.*, 1998) and *H. pylori* (Achtman *et al.*, 1999), two organisms that exhibit extensive genetic diversity between clinical isolates. Housekeeping genes are used because they are more selectively neutral than genes associated with virulence (*vacA* or *cagA*). Thus, tracing the historical or geographical lineage in panmictic organisms may be possible if genes undergoing extensive recombination are excluded. However, a study encompassing isolates from different geographical regions demonstrated that *H. pylori* strains could only be geographically grouped at one locus (*vacA*) and even then only East Asian strains (Maggi Solca *et al.*, 2001). This study also demonstrated that South African strains could be geographically grouped, but again only at one locus (*atpD*). It was suggested that frequent recombination at the other loci studied (*cagA*, *iceA*, *glnA*, *scoB*, *recA* and IS*606*) masked the evolutionary relationship between strains. In contrast, Owen and Xerry (2003) demonstrated

that clonal lineages within family members can be detected between *H. pylori* isolates using MLST. *H. pylori* transmission within family members would not allow enough time for a high frequency of recombination to occur within a strain's genome, which is probably why this typing scheme was useful.

Ribotyping is a method in which only the restricted fragments of ribosomal RNA genes are visualised. The rRNA gene restriction patterns are investigated by Southern blotting followed by hybridisation with a labelled rRNA gene probe. The major disadvantage of this method is that it is time consuming. However, it is highly discriminatory and easy to analyse. Ribotyping has been useful in determining sources of infection and transmission of *H. pylori* within families (Tee *et al.*, 1992).

Assessing the interstrain diversity within a *H. pylori* population has also been perfomed using the fingerprinting technique termed amplified fragment length polymorphism (AFLP)-PCR (Gibson *et al.*, 1998 and Owen *et al.*, 2001). This is a robust high-resolution marker technique in which adapter molecules are ligated to a subset of restriction fragments that are subsequently selected as target sits for selected primers in PCR amplification. Application of AFLP to *H. pylori* isolates has been useful in indexing *H. pylori* strain diversity, identification of clonal or sub-clonal variants, and assessing diversity within metronidazole-resistant isolates (Owen *et al.*, 2001). The benefits of AFLP-PCR are its sensitivity, reproducibility and efficiency in generating genetic fingerprints.

A modified version of the AFLP technique termed fluorescent amplified fragment length (FAFLP)-PCR has also been applied to *H. pylori*. A fluorescent probe is used to visualise the amplified fragments in this fingerprinting technique. An outline of the FAFLP-PCR procedure is given in Fig. 4.1. FAFLP-PCR has been used successfully to investigate outbreaks of infection by *Streptococcus pyogenes, Staphylococcus aureus* and *Neisseria meningitidis* (Desai *et al.*, 1998; Goulding *et al.*, 2000; Grady *et al.*, 2000). FAFLP-PCR analysis generates a specific profile for each strain under examination. These profiles can be grouped together resulting in generalised patterns. These patterns can then be phylogenetically analysed, allowing the relatedness of *H. pylori* isolates to be assessed by clustering of strains based on the genetic similarities in their fingerprint profiles. Phylogenetic clustering of isolates depends on the presence of monomorphic (unique) and polymorphic (variable) bands in the FAFLP-

PCR profiles. This scheme was recently used to examine the geographical partitioning of *H. pylori* isolates (Ahmed *et al.*, 2001).

Recent data displaying the genetic diversity of *H. pylori* isolates by genotyping specific loci from strains obtained from different regions show that the frequencies of different allelic combinations can vary, suggesting that some geographical regions display higher or lower genetic diversity than others (Van Doorn *et al.*, 1999a). Furthermore, MLST has shown that weak clonal groupings of *H. pylori* populations can be detected despite frequent recombination (Achtman *et al.*, 1999). These studies show that in some regions of the world *H. pylori* populations can display less genetic heterogeneity than in others, and that in spite of high levels of recombination the geographical separation of isolates can also be determined.

The present study applied the highly discriminatory fingerprinting technique FAFLP-PCR to a collection of Irish *H. pylori* isolates with a view to determining (i) the genetic heterogeneity of this population and (ii) the relationship of Irish *H. pylori* isolates to isolates from different parts of the world based on the relatedness of their FAFLP-PCR profiles.

4.2 Methods

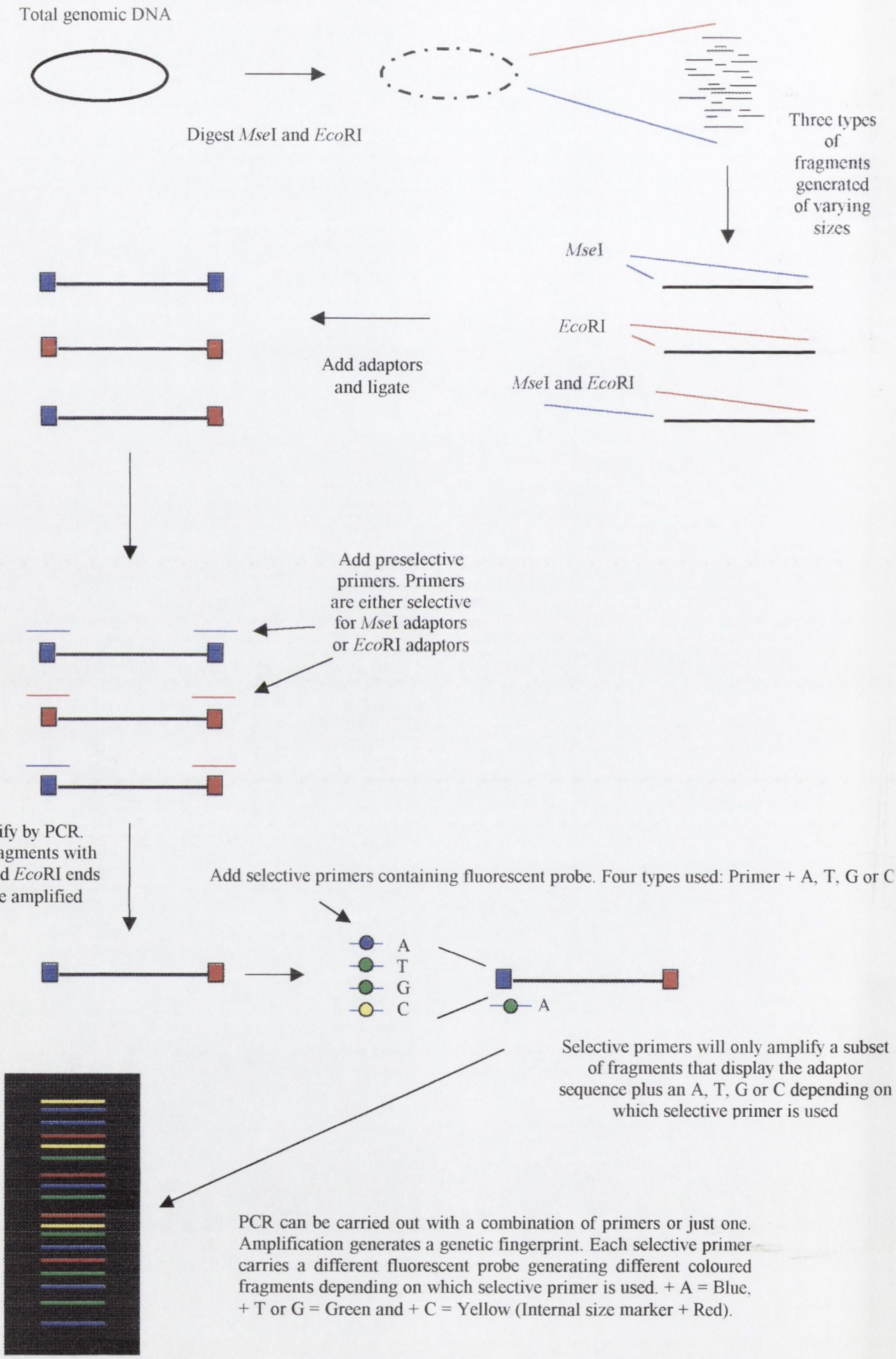
4.2.1. Fluorescent amplified fragment length (FAFLP)-PCR fingerprinting analysis.

4.2.1.1. Generation of FAFLP-PCR fingerprints.

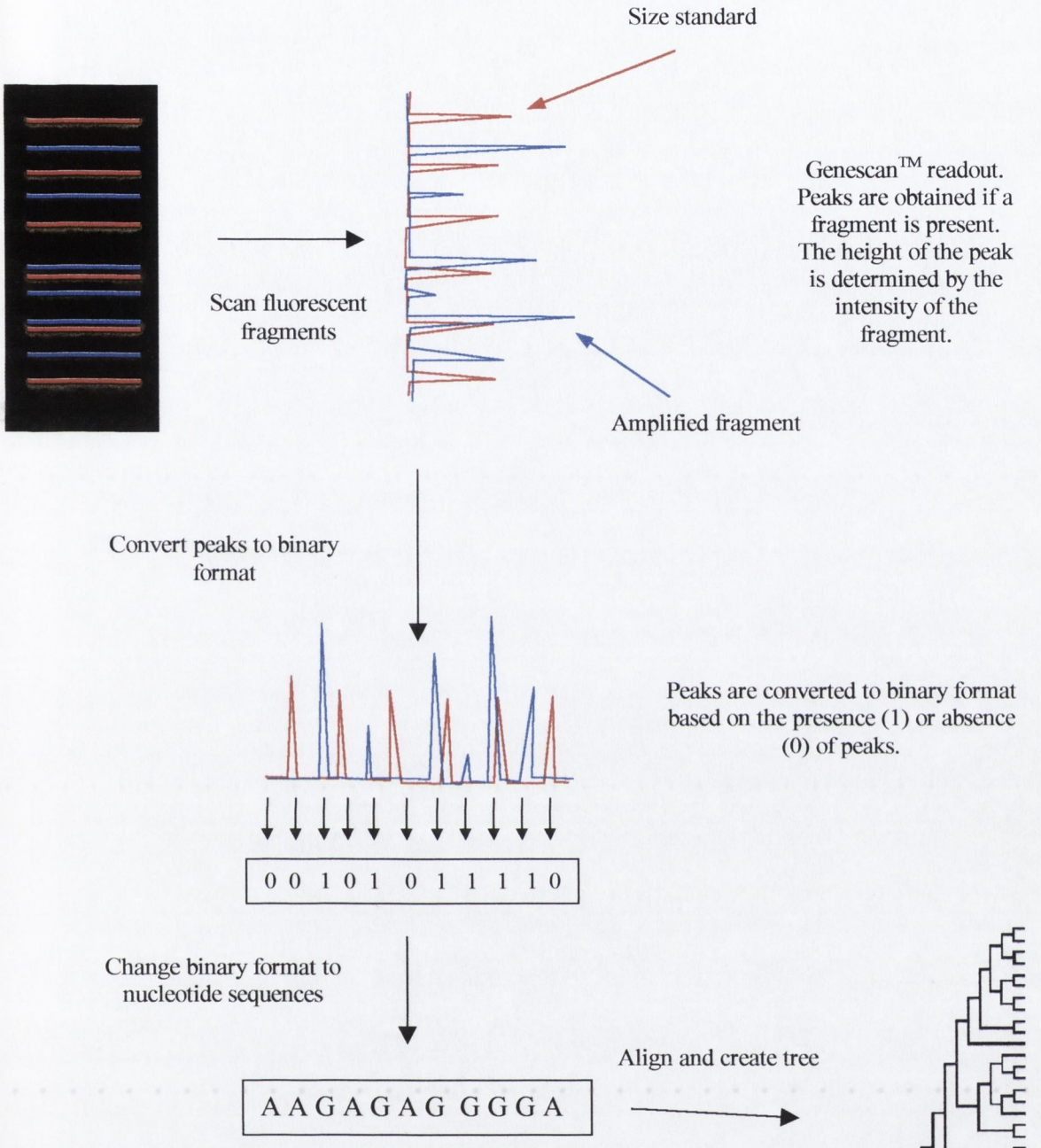
In FAFLP-PCR the total genomic DNA of a bacterial strain is digested with a frequent cutter (*Eco*RI) and an infrequent cutter (*Mse*I) restriction endonuclease. *Eco*RI and *Mse*I adaptor molecules are then ligated to the digested DNA fragments. Pre-selective amplification is carried out with two primers that recognise either an *Eco*RI or *Mse*I adaptor sequence. Selective amplification is carried out with primers that either recognise an *Mse*I adaptor sequence or recognise an *Eco*RI adaptor sequence plus an extra nucleotide (*Mse*I adaptor sequence + A, T, G or C). Selective amplification amplifies a subset of the PCR products generated by pre-selective amplification. This procedure is outlined in Fig. 4.1 Stage 1.

DNA (10 ng) was digested with the restriction enzymes *Eco*RI and *Mse*I, and the resulting fragments were ligated to double-stranded adapters, one containing an *Eco*RI site (5′-CTCGTAGACTGCGTACC-3′ plus 3′-CTGACGCATGGTTAA-5′) and one an *Mse*I site (5′-GACGATGAGTCCTGAG-3′ plus 3′-TACTCAGGACTCAT-5′). The digested and ligated products were then diluted 20 fold. Of these products, 1.5 µl was used for pre-selective amplification using an amplification core mix (Perkin Elmer, USA) and pre-selective primers (5′-GTAGACTGCGTACCAATTC-3′ [*Eco*RI adaptor sequence] and 5′-GACGATGAGTCCTGAGTAA-3′ [*Mse*I adaptor sequence]) in a final volume of 10 µl. This PCR reaction selectively amplifies restricted fragments containing both *Eco*RI and *Mse*I adaptor molecules only. PCR conditions for pre-selective amplification were: 20 cycles of [94 °C for 20 s, 60 °C for 30 s and 72 °C for 2 min]. The reaction products were then stored at 4°C. The PCR product obtained was further diluted 20 fold and 1.5 µl of the diluted product was used for selective amplification with a primer that recognises the *Eco*RI adaptor site plus either A, T, G or C (5′-GTAGACTGCGTACCAATTC-A/T/G/C-3′) and contains a fluorescent label (6FAM or HEX or TET or JOE). This primer was used in conjunction with a primer that recognises the *Mse*I adaptor molecule site only (5′-GACGATGAGTCCTGAGTAA-3′) in a 10 µl reaction. PCR conditions for selective amplification were as described previously (Ahmed *et al.*, 2001, 2002, 2003a, 2003b; Table 4.1).

Stage 1


Total genomic DNA
Digest *Mse*I and *Eco*RI
Three types of fragments generated of varying sizes
*Mse*I
*Eco*RI
*Mse*I and *Eco*RI
Add adaptors and ligate
Add preselective primers. Primers are either selective for *Mse*I adaptors or *Eco*RI adaptors
Amplify by PCR. Only fragments with *Mse*I and *Eco*RI ends will be amplified
Add selective primers containing fluorescent probe. Four types used: Primer + A, T, G or C
A
T
G
C
A
Selective primers will only amplify a subset of fragments that display the adaptor sequence plus an A, T, G or C depending on which selective primer is used
PCR can be carried out with a combination of primers or just one. Amplification generates a genetic fingerprint. Each selective primer carries a different fluorescent probe generating different coloured fragments depending on which selective primer is used. + A = Blue, + T or G = Green and + C = Yellow (Internal size marker + Red).

Stage 2



Once the nucleotide formats of several isolates have been obtained, they can be aligned by CLUSTAL W / X, and then a neighbour-joining phylogenetic tree can be constructed. Clustering in this tree will be based on how related the fingerprints of isolates are. The benefit of this technique is that the relatedness of isolates is based on the identities of the genomes rather than on specific parts.

Fig. 4.1. Diagram depicting the procedure for Fluorescent Amplified Length (FAFLP)-PCR fingerprinting.

The amplified fragments can be analysed by one of two sequencing machines (ABI 377, ABI3100). When using the ABI 377 sequencer the PCR products are run on a sequencing gel, and then each fingerprint is scanned to produce Genescan™ readout of the fingerprint. The ABI 3100 is a capillary sequencer, therefore a sequencing gel is not needed. This sequencer also produces Genescan™ readouts.

Table 4.1. Thermal cycler parameters for selective amplification.

Hold	PCR cycle conditions			Number of cycles
94 °C 2 min.	94 °C 20 sec.	66 °C 30 sec.	72 °C 2 min.	1
–	94 °C 20 sec.	65 °C 30 sec.	72 °C 2 min.	1
–	94 °C 20 sec.	64 °C 30 sec.	72 °C 2 min.	1
–	94 °C 20 sec.	63 °C 30 sec.	72 °C 2 min.	1
–	94 °C 20 sec.	62 °C 30 sec.	72 °C 2 min.	1
–	94 °C 20 sec.	61 °C 30 sec.	72 °C 2 min.	1
–	94 °C 20 sec.	60 °C 30 sec.	72 °C 2 min.	1
–	94 °C 20 sec.	59 °C 30 sec.	72 °C 2 min.	1
–	94 °C 20 sec.	58 °C 30 sec.	72 °C 2 min.	1
–	94 °C 20 sec.	57 °C 30 sec.	72 °C 2 min.	1
–	94 °C 20 sec.	56 °C 30 sec.	72 °C 2 min.	20
60 °C 30 min.		–		1
4 °C forever		–		1

4.2.1.2. Automated fingerprinting.

The following describes the visualisation of the generated FAFLP-PCR products on one of two sequencing machines (Fig. 4.1, Stage 1-bottom).

FAFLP-PCR products were separated on two different sequencing machines, namely (i) an ABI 377 automated sequencer (PE, Applied Biosystems) and (ii) an ABI 3100 capillary sequencer (PE, Applied Biosystems). In (i) FAFLP-PCR products were loaded on to an ABI Prism 377 XL-96 DNA sequencer (PE Biosystems) with formamide loading dye (1.5 µl final volume) and with an internal lane standard GS-500 Rox (PE, Biosystems). Fragment separation was allowed to continue for 5 h on a 4% (w/v) denaturing polyacrylamide gel. Fragments were detected and compiled by the ABI Data Collection™ software (PE, Biosystems). A gel image was generated and each lane was scanned to make individual electropherograms. Fragment analysis was performed with the Genescan Analysis™ 3.1 package (PE, Applied Biosystems). In (ii) separation and analysis was carried out the same way except that the FAFLP-PCR products were fractionated in a capillary tube.

4.2.1.3. Genotypic analysis.

A genotyping program (Genotyper™ 2.5) was used to analyse the FAFLP-PCR electropherograms (Genescan™) obtained from each sequencing machine and convert them into binary data. The binary data obtained are representative of each FAFLP-PCR fingerprint generated and can be used to construct a phylogenetic tree based on *H. pylori* fingerprints (Fig. 4.1, Stage 2).

The Genescan™ data of individual isolates were exported to the Genotyper™ 2.5 package (PE, Applied Biosystems) for digitization and genotyping. Based on the repertoire of amplified FAFLP-PCR products, different FAFLP profiles were identified as 'amplitypes'. These amplitypes were color coded and overlaid to identify traces, fluorescence intensity (peak height), data points on the gel and the frequency of the monomorphic bands.

Amplified products were sized in base pairs with respect to defined marker sizes (Ahmed *et al.*, 2001, 2002, 2003a, 2003b). Presence or absence of amplicons was scored by the Genotyper™ macro. 'Allele scores' (presence or absence of amplicons) were converted into binary format [1, 0] and were used in establishing a similarity matrix needed for constructing a neighbour-joining network. A neighbour-joining tree was generated from the allele scores to

delineate genetic similarities among the *H. pylori* fingerprints. Phylogenetic trees constructed were generated using the Genotyper™ 2.5 package (percentage heterogeneity of fingerprints, see Fig. 4.7) or using Clustal X™ and Treeview™ packages (relatedness, see Fig. 4.10).

4.3. Results

4.3.1. FAFLP-PCR fingerprinting of Irish H. pylori *isolates.*

Initially FAFLP-PCR fingerprints were generated using four selective primers (*Mse*I adaptor sequence + 0 / *Eco*RI adaptor sequence + A, T, G and C). However, the T selective primer (*Eco*RI adaptor sequence + T) did not yield any fingerprints. Generation of fingerprints was carried out using pairs of selective primers (i.e., C and T selective primers used in the same PCR). It is possible that the C selective primer (*Eco*RI adaptor sequence + C) could have overpowered the T selective primer in this reaction. Thus, the T selective primer could not generate fingerprints. Therefore, an isolate was included in this analysis if it yielded fingerprints for three selective primers (A, C and G).

FAFLP-PCR analysis was applied to 45 Irish *H. pylori* isolates, which genotypically represented the larger set of isolates. Of these 45 isolates 32 (71.1%) yielded informative fingerprints. The remaining 13 isolates either did not yield profiles with all three selective primers (usually due to degraded DNA). The 32 isolates that yielded informative profiles were still genotypically representative of the larger set of isolates (Table 4.2). Twenty-five of the latter 32 isolates (78.1%) yielded in-del motifs (type II, 13, 52%; type IV, 9, 36%; type Ia, 2, 8%; type IIIa, 1, 4%; c. f., Table 3.4), while the remaining isolates did not type by in-del analysis (NT, 7, 21.9% of the 32 isolates; c.f., 11 out of 74 = 14.9%). The distribution of the *vacA* alleles in this selected group of isolates was s1a m2 (19, 59.4%), s1a m1 (10, 31.3%) and s2 m2 (3, 9.3%), c. f., Table 3.4.

Selective amplification was carried out using two selective primers in each FAFLP-PCR. One of two combinations of selective primer pairs was used, either [*Mse*I adaptor sequence + 0 / *Eco*RI adaptor sequence + A and + G] or [*Mse*I adaptor sequence + 0 / *Eco*RI adaptor sequence + C and + T]. The amplification reactions yielded unique profiles for each isolate. Fig. 4.2. shows the PCR products for 12 isolates using selective primers with A and G selectivity (i.e., *Mse*I adaptor sequence + 0 / *Eco*RI adaptor sequence + A and + G).

Generation of FAFLP-PCR fingerprints using non-selective primer pairs (*Mse*I adaptor sequence + 0 / *Eco*RI adaptor sequence + 0) was initially carried out on 34 isolates (Fig. 4.3.). However, this method of generating fingerprints was abandoned as using selective primers (*Mse*I adaptor sequence + 0 / *Eco*RI adaptor sequence A, G or C) provided more informative fingerprints.

Table 4.2. Irish *H. pylori* isolates analysed by FAFLP-PCR with *vacA* allelic and *cag* PAI types indicated.

Isolate	*vacA* type	*cag* PAI (in-del) type
FR138	s1a m1	II
FR147	s1a m2	M[a]
FR150	s1a m2	IV
FR165	s1a m1	-[b]
FR191	s1a m1	II
FR205	s1a m1	II
FR207	s1a m1	-[b]
FR222	s1a m2	IV
FR229	s1a m2	II
FR296	s1a m2	Ia
FR298	s1a m1	II
112886C	s1a m2	II
15500A	s1a m2	Ia
167233A	s1a m1	IV
218A	s1a m2	II
222A	s1a m2	-[b]
45688A	s1a m2	II
577A	s1a m2	IV
605A	s2 m2	M[a]
79862A	s1a m2	IV
87212A	s1a m2	IV
9613A	s2 m2	IV
68545A	s1a m2	IV
MI458	s1a m2	II
MI496	s1a m1	II
MI506	s1a m2	II
MI517	s1a m1	II
MI571	s1a m1	II
J561	s1a m2	-[b]
J563	s1a m2	IIIa
431	s1a m2	-[b]
613	s2 m2	IV

[a] Refers to when multiple PCR products were generated.
[b] Refers to when the no PCR product was generated.

1 2 3 4 5 6 7 8 9 10 11 12

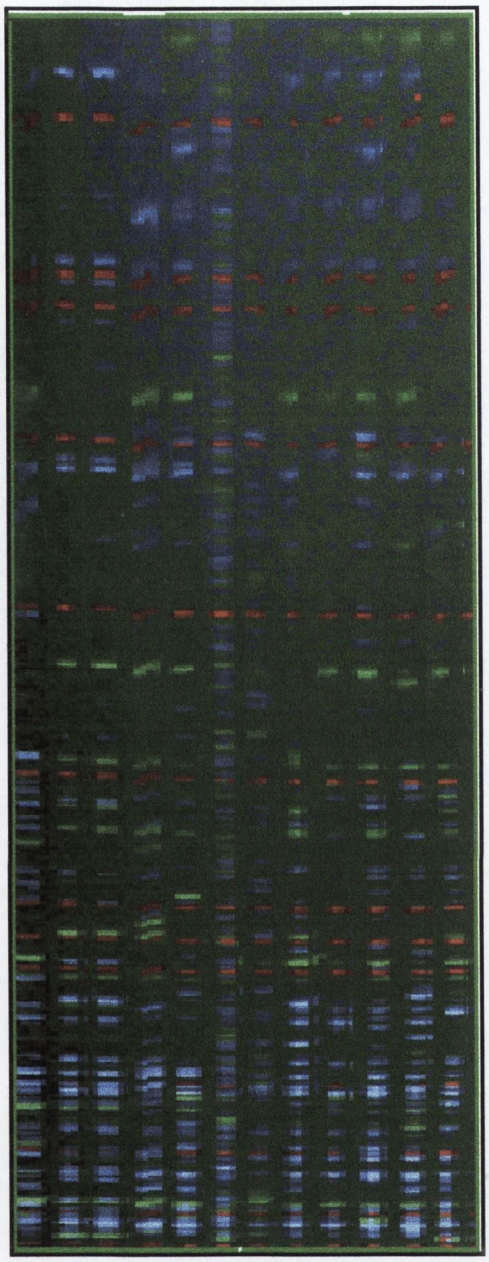


Fig. 4.2. FAFLP-PCR profiles of 12 Irish *H. pylori* isolates.

Each fingerprint was run on a sequencing gel. The FAFLP-PCR procedure was carried out using selective primer pairs *Mse*I + 0 / *Eco*RI + A-FAM (Blue bands) and *Mse*I + 0 / *Eco*RI + G-JOE (Green bands). Red bands indicate the internal lane standard (Rox 500). Lane 1 = 167233A, 2 = 577A, 3 = 9613A, 4 = MI571, 5 = 34683A, 6 = 112886A, 7 = 79862A, 8 = MI512, 9 = 218A, 10 = 58745A, 11 = 15500A, 12 = 59393A. Diversity between profiles can clearly be seen within isolates.

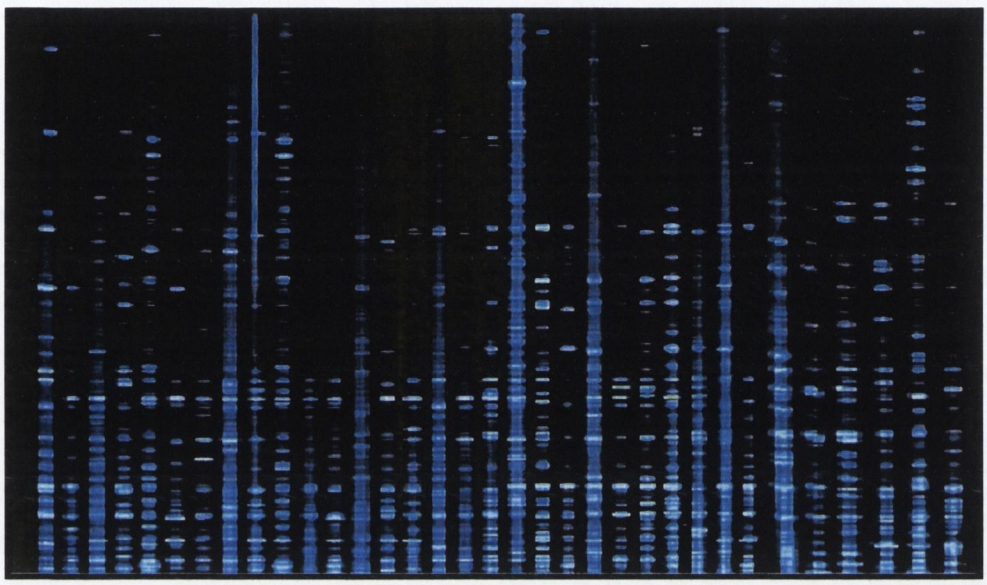


Fig. 4.3. FAFLP-PCR profiles of 34 Irish *H. pylori* isolates using non-selective primer pairs.

The primer pairs used were *Mse*I + 0 / *Eco*RI + 0. Each fingerprint is unique to each isolate, a characteristic of panmictic organisms. Non-selective amplification yields numerous bands, therefore generation of fingerprints with selective primers makes the analysis less complicated.

Genescan™ fingerprint profiles of all 32 isolates were obtained for selective primers + A, + G and + C. Fig. 4.4 A, B and C shows the Genescan™ profiles of isolates with A, G and C selectivity, respectively. As mentioned previously no fingerprints were obtained using selective primers with + T. The Genescan™ profiles of the same isolates can be overlaid (Fig. 4.5). Fig. 4.6 displays the genetic diversity detected between isolates using one selective primer.

4.3.2. Phylogenetic analysis of Irish H. pylori *using FAFLP-PCR fingerprints.*

The presence [1] or absence [0] of peaks (amplified fragments) within Genescan™ profiles was used to construct a binary sequence for each isolate using the Genotyper™ 2.5 macro (sequence is based on all three selective primers + A, G and C). The binary sequences for each isolate were then converted into a nucleotide sequence (1 → G and 0 → C). The nucleotide sequences, which are representative of each isolate's FAFLP-PCR fingerprints, were aligned using CLUSTAL X. A neighbour-joining tree was then constructed based on the alignment of each of these representative sequences (Fig. 4.7). Analysis of this tree shows that the majority (28, 87.5%) of isolates in this group differ in percentage heterogeneity by approximately 7.8% with respect to their FAFLP profiles.

4.3.3 Phylogenetic analysis of Irish H. pylori *isolates with respect to isolates from different geographic regions.*

The FAFLP-PCR profiles of various *H. pylori* isolates from different parts of the world (6 isolates from India, 5 from the UK, 5 from Peru, 4 from Spain, one from Japan and one from Africa) were generated in the same way as the profiles for the Irish isolates. The Genescan™ profiles of Irish, UK, Peruvian, and Spanish isolates are compared in Fig. 4.8. In this figure some conserved fragments in all four representative isolates can be seen, yet the Irish and UK isolates look similar, as do Spanish and Peruvian isolates. In addition, it was found that the profiles of an Irish and Japanese strain showed little in common (Fig. 4.9).

Six random Irish *H. pylori* isolates were chosen to be used in phylogenetic analysis with the isolates from different parts of the world. A phylogenetic tree was generated in the same way as previously described (Fig. 4.10). This tree generated 6 main groupings (A–F). The groups into which each isolate clustered are listed in Table 4.3. Most isolates from a specific region did not cluster in the same group. However, Irish isolates remained in the same group, indicating a relatively homogeneous population structure.

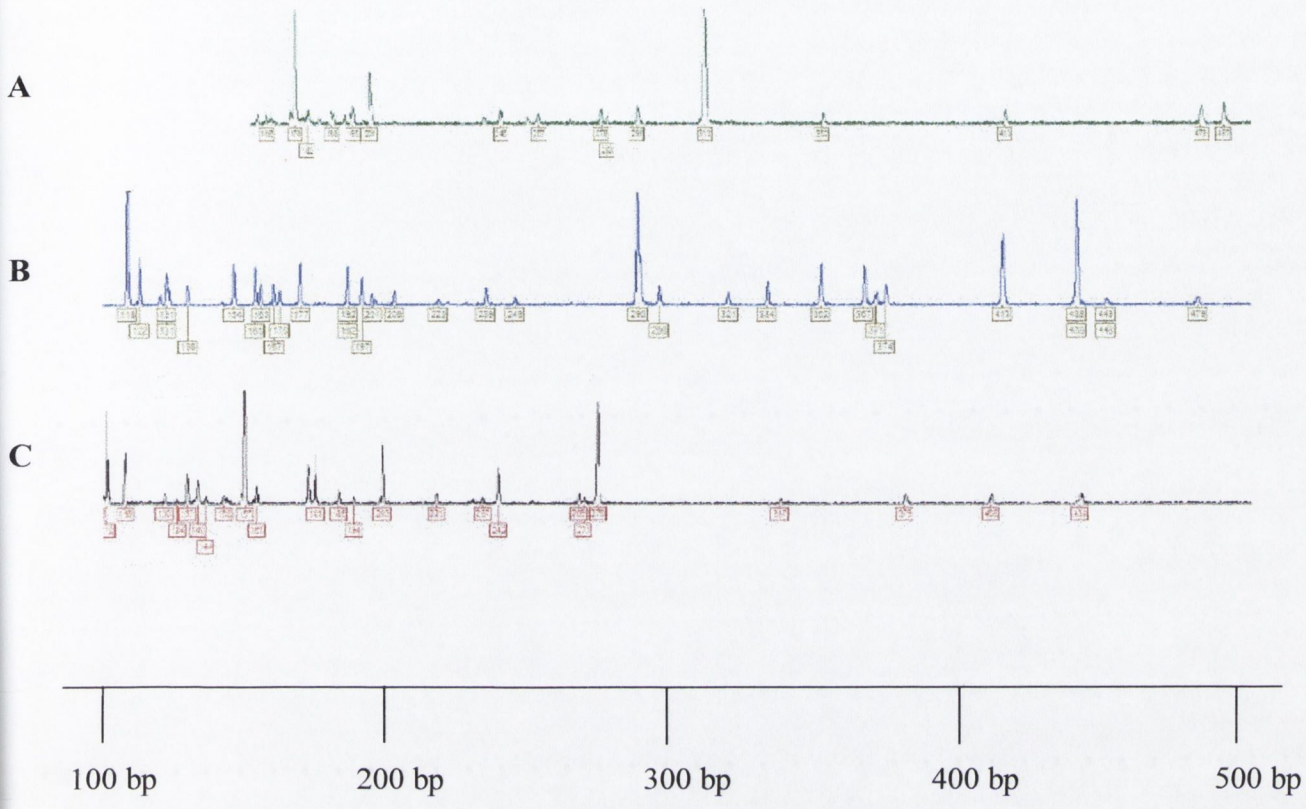


Fig. 4.4. Genescan™ FAFLP-PCR profiles of three different *H. pylori* isolates.

Comparison of the FAFLP-PCR profiles of three unrelated *H. pylori* isolates (A = 167233A, B = FR217 and C = MI512) using selective primer combinations of (**A**) *Mse*I + 0 / *Eco*RI + G, (**B**) *Mse*I + 0 / *Eco*RI + A and (**C**) *Mse*I + 0 / *Eco*RI + C. Fragments generated by the FAFLP-PCR technique are represented by peaks. The height of each peak is reflective of the intensity of the fragment produced. Fragment sizes are indicated on the x-axis with a sensitivity of ±2 bp. As the scale generated from the Genotyper™ program is difficult to visualise an approximate scale (bp) is shown at the bottom of the profiles.

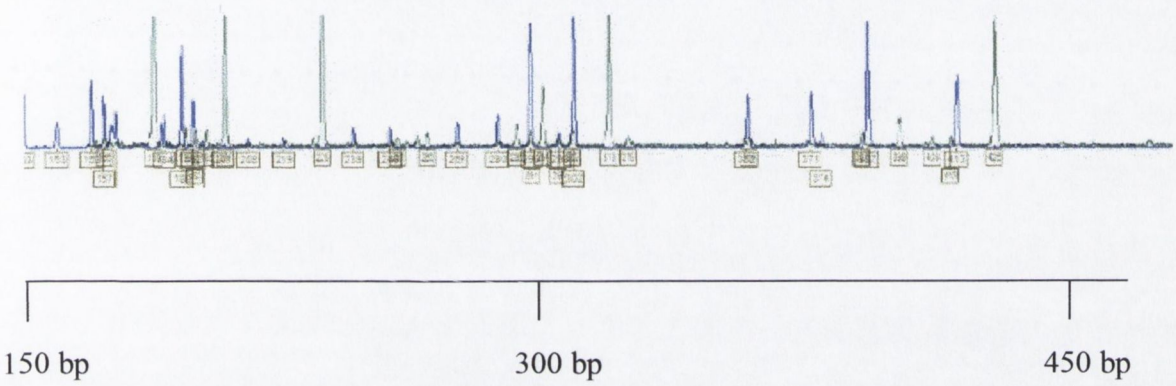


Fig. 4.5. Two superimposed FAFLP-PCR profiles generated using different selective primers.

Overlaid Genescan™ profiles of isolate 15500A using selective primers + A and + G. Fragments generated by one selective primer are not generated by the other selective primer. The bottom scale indicates approximate sizes (bp) of peaks.

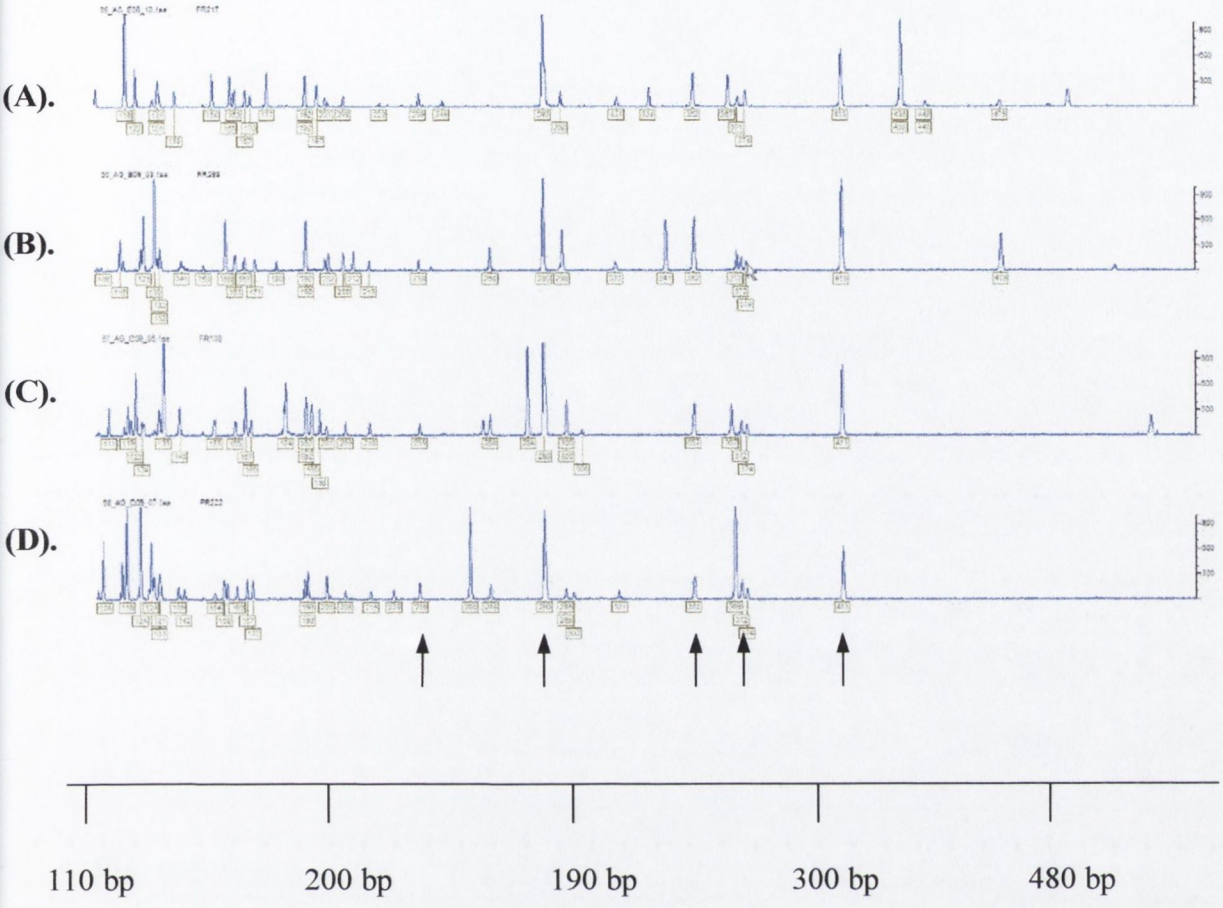


Fig. 4.6. Comparison of FAFLP-PCR profiles of four unrelated *H. pylori* isolates using the same selective primer.

Genescan™ profiles of 4 different *H. pylori* isolates using selective primers *Mse*I + 0 / *Eco*RI + A. These four isolates exhibit diversity between their profiles, a characteristic of a panmictic organism. However, some conserved fragments are present (black arrows). The bottom scale displays the approximate sizes (bp) of peaks. A = isolate FR217, B = isolate FR298, C = isolate FR138 and D = isolate FR222.

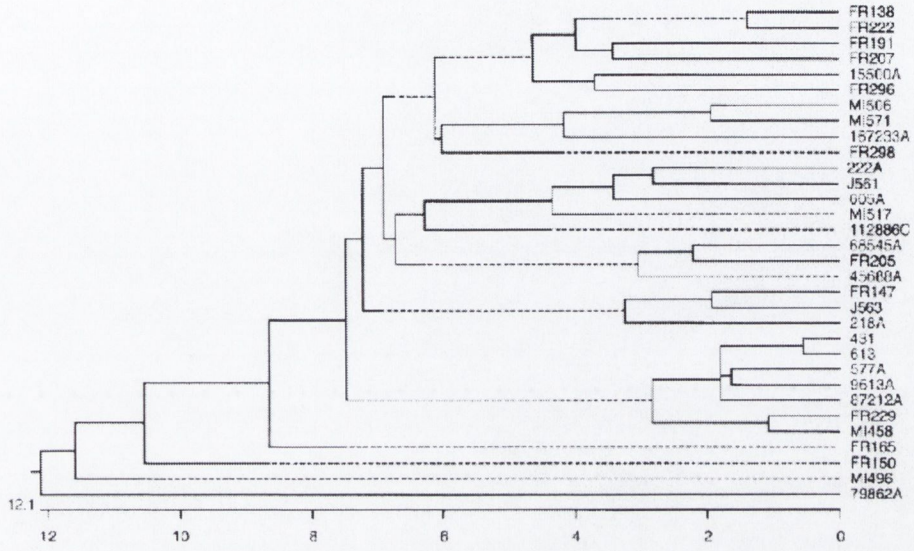


Fig. 4.7. Neighbour-joining phylogenetic tree of 32 Irish *H. pylori* isolates.

Relatedness of isolates is based on similarities of the FAFLP-PCR fingerprints of each isolate. The bottom scale indicates percentage heterogeneity within these fingerprints.

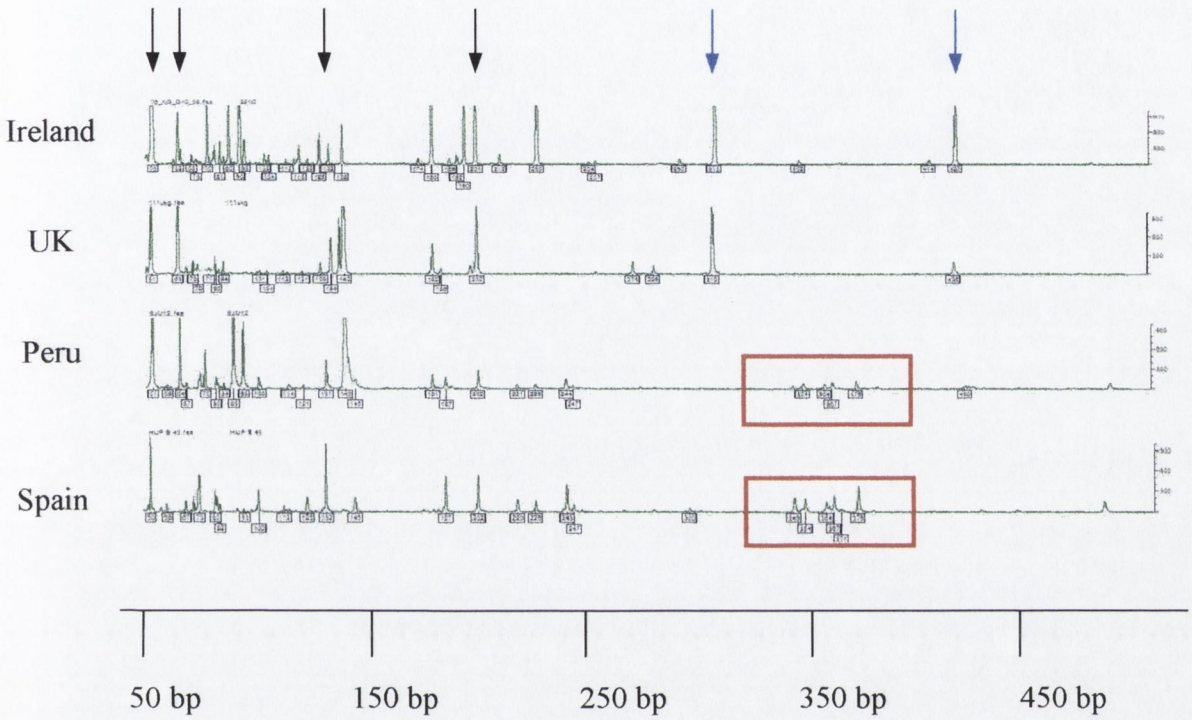


Fig. 4.8. Comparison of FAFLP-PCR profiles of four *H. pylori* isolates from four different geographical regions using the same selective primer.

Genescan™ profiles of four *H. pylori* isolates from different regions of the world (Ireland, UK, Peru and Spain) using selective primers *Mse*I + 0 / *Eco*RI + G. Black arrows indicate peaks conserved in all four isolates. The profiles of Irish and UK isolates show more similarity to each other than they do to those of the Peruvian and Spanish isolates. Likewise, the profiles of Peruvian and Spanish isolates share more similarities than they do to Irish and UK isolates. Blue arrows indicate two predominant peaks that are found in Irish and UK isolates but not in the Spanish and Peruvian isolates. There is a region of clustered peaks that are present in the Spanish and Peruvian isolates (red box) but absent in the Irish and UK isolates. The bottom scale displays the approximate sizes (bp) of peaks.

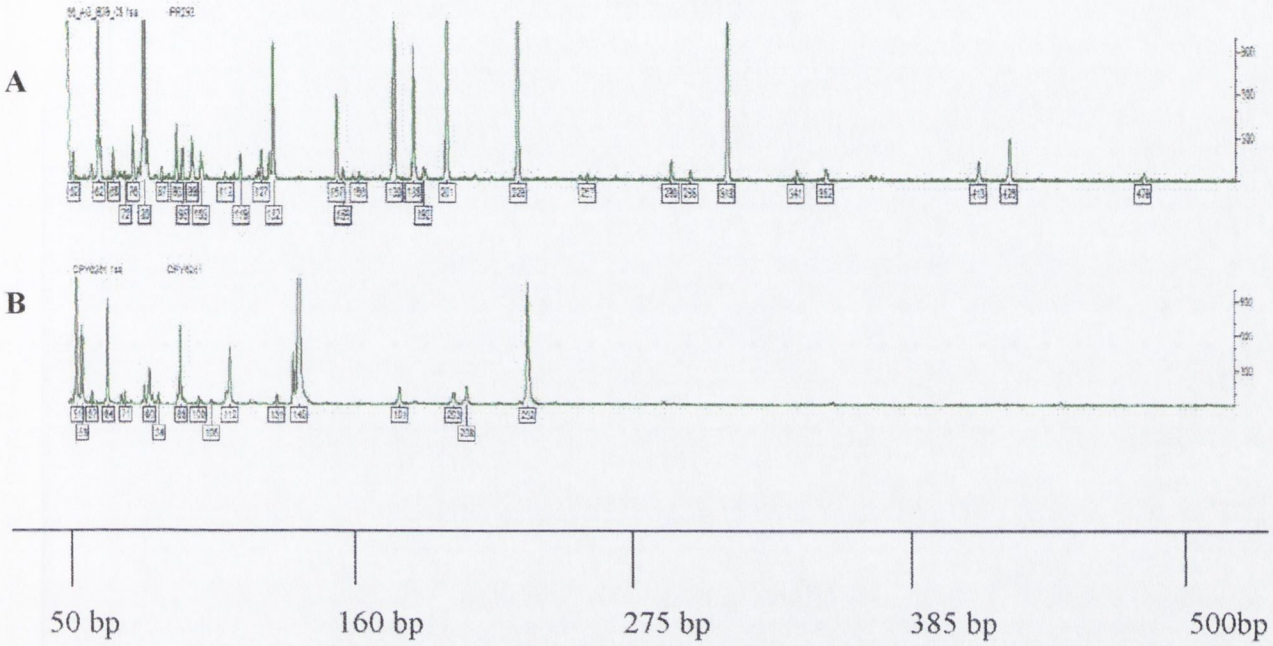


Fig. 4.9. Comparison of the FAFLP-PCR profiles of an Irish and Japanese *H. pylori* isolate using the same selective primer.

Genescan™ profiles of (**A**) an Irish *H. pylori* isolate and (**B**) a Japanese isolate, using primer pairs *Mse* I + 0 / *Eco* RI + G. It was postulated in chapter 3 that Irish *H. pylori* isolates and isolates form certain parts of East Asia have a distant ancestral relationship. The profiles of these isolates show little similarity. However, these isolates may not be indicative of what other isolates in these regions are like. The bottom scale displays the approximate (bp) sizes of peaks.

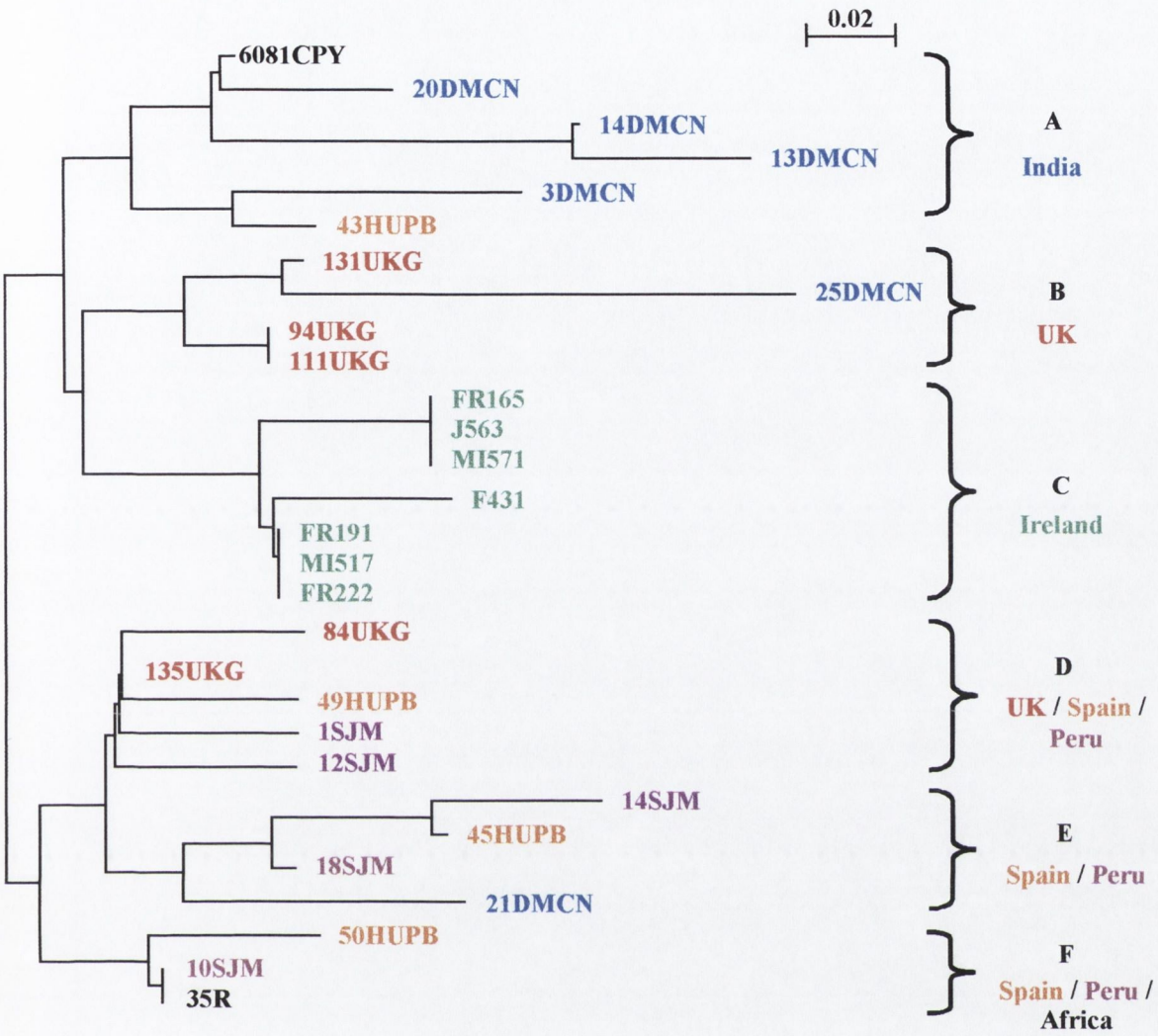


Fig. 4.10. Neighbour-joining phylogenetic tree displaying the relatedness of *H. pylori* isolates from four different continents.

Relatedness of isolates is based on similarities of the FAFLP-PCR fingerprints of each isolate. Irish isolates cluster exclusively in their own group (C), a characteristic not shared with isolates from other geographical regions. The scale at the top right displays the number of fingerprint band substitutions per site.

Table 4.3. Strains from different geographical locations and the cluster subgroup they belonged to as revealed by FAFLP.

Country	Strain	Subgroup
Ireland:		
	F431	C
	FR191	C
	FR222	C
	MI517	C
	FR165	C
	MI571	C
	J563	C
India:		
	25DMCN	B
	3 DMCN	A
	20 DMCN	A
	13 DMCN	A
	14 DMCN	A
	21DMCN	E
UK:		
	131UKG	B
	111 UKG	B
	94 UKG	B
	84 UKG	D
	135 UKG	D
Peru:		
	1SJM	D
	12SJM	D
	10SJM	F
	18SJM	E
	14SJM	E
Spain:		
	49HUPB	D
	43HUPB	A
	50HUPB	F
	45HUPB	E
Japan:		
	6081CPY	A
Africa:		
	35R	F

4.4. Discussion.

4.3.1. Phylogenetic analysis of Irish H. pylori *FAFLP-PCR fingerprints.*

One of the most striking observations from the phylogenetic tree generated for Irish *H. pylori* isolates (Fig. 4.7) is the lack of diversity between isolates. Although, there is a certain degree of heterogeneity between these isolates [a characteristic of panmictic organisms (Go *et al.*, 1996)], it is not typical of the diversity reported for different geographical regions using similar and other fingerprinting approaches (Ahmed *et al.*, 2001). This relative lack of diversity is based on the fingerprint profiles that are representative of the structures of the whole genomes of these isolates, and not on the exact sequences of the genomes. However, FAFLP-PCR fingerprinting is a very sensitive technique, is capable of demonstrating the diversity within the population of many different organisms (Vos *et al.*, 1995), and is capable of detecting relatively high levels of diversity within a population if it exists.

The benefit of analysing the phylogenetic behaviour of a population's genetic fingerprints is that whole genome structures are compared rather than specific sequences. Phylogenetic analysis of specific sequences within an organism's genome is extremely informative for clonal bacteria (Feil *et al.*, 2001), but not for genetically diverse organisms, such as *H. pylori* (Maggi *et al.*, 2001), that contain DNA sequences that have come from external sources via horizontal transfer and that therefore do not properly reflect the ancestral or geographical relationships of the population under study (Achtman *et al.*, 1999). This lack of diversity further confirms the observations in Chapter 3 that Irish *H. pylori* isolates may be weakly clonal.

4.3.1. Phylogenetic analysis of Irish H. pylori *isolates with respect to isolates from different geographical regions.*

Phylogenetic comparison of Irish *H. pylori* isolates with isolates from four different continents revealed that Irish isolates may have a unique genotype, an observation made in Chapter 3, as all the Irish isolates cluster together but not with isolates from other geographical regions (Fig. 4.10).

Fig. 4.10 illustrates that, although Irish isolates are distinct from isolates from other geographical regions, some non-Irish isolates cluster in close proximity to Irish isolates, i.e., isolates from the UK, India and Japan. It was previously postulated that Irish isolates were similar to isolates from these regions based on the predominance of certain alleles, although

the Asian isolates that displayed similar genotypes as Irish *H. pylori* isolates were from China and Taiwan. Furthermore, Spanish and Peruvian isolates cluster in close proximity to each other in Fig. 4.10 suggesting an ancestral relationship, an observation reported previously based on in-del typeing (Kersulyte, *et al.*, 2000). These observations are further confirmed when comparing the Genescan™ profiles of representative isolates from different geographical regions.

In summary, the results obtained in Chapter 3 suggested that Irish *H. pylori* isolates have associations with Northern European isolates as revealed by in-del analysis. Yet, the *vacA* allelic predominance among these Irish isolates displayed similarities with isolates from China and Taiwan. Furthermore, FAFLP-PCR genotyping revealed that Irish isolates are weakly clonal, displaying more homogeneous fingerprint profiles compared to isolates from other geographical locations, but with some associations with both European and Asian isolates. These three genotyping techniques show that Irish *H. pylori* isolates have distinctive features that may have evolved in this insular European population.

Chapter 5

Microevolution between paired Helicobacter pylori *isolates taken from the same patients as revealed by genotyping and FAFLP-PCR.*

5.1. Introduction.....90

5.2 Results.....93

3.2.1. vacA *genotyping*.....93

5.2.2. Sequences preceding vacA *and* ribA.....93

5.2.3. Insertion-deletion (in-del) analysis.....94

5.2.4. cagA *genotyping*.....94

5.2.5. RAPD-PCR analysis.....95

5.2.6. Restriction fragment length polymorphism (RFLP) and sequence analysis of paired H. pylori *isolates*.....95

5.2.6.1. RFLP analysis of paired *H. pylori* isolates at the *napA*, *ampA*, *hsd*, *galE*, *gmd*, and *fln* loci.....95

5.2.6.2. RFLP analysis of the *vacA* locus.....96

5.2.6.3. Sequence analysis of paired *H. pylori* isolates.....96

5.2.7. FAFLP-PCR genotyping.....96

5.2.8. Phylogenetic analysis of FAFLP-PCR profiles.....97

5.2.9. VacA analysis of microevolved H. pylori *isolates*.....98

5.2.10. vacA *transcriptional analysis of microevolved* H. pylori *isolates*.....98

5.3 Discussion.....100

5.3.1. Genotyping of paired isolates at 5 different loci.....100

5.3.2 cagA *genotyping of paired isolates*.....100

5.3.3. RAPD -PCR analysis of paired H. pylori *isolates*.....101

5.3.4. Southern blotting of related and unrelated* H. pylori *isolates.....101

5.3.4. Sequence analysis of paired H. pylori *isolates*.....101

5.3.5. FAFLP -PCR analysis of paired H. pylori *isolates*.....102

5.3.6. Phylogenetic analysis of paired H. pylori *isolates*.....103

5.3.7. Analysis of familial isolates.....104

5.3.8. Phenotypic analysis of paired H. pylori *isolates*.....105

5.3.9. Transcriptional analysis of paired H. pylori *isolates that display phenotypic differences*.....106

5.3.10. Model for Microevolution.....106

5.1. Introduction.

Helicobacter pylori has been characterised as a panmictic organism exhibiting high levels of genetic diversity between clinical isolates (Go *et al.*, 1996; Salaun *et al.*, 1998). This diversity is a result of point mutations within conserved genes (e.g., the *ureC* gene, Kansau *et* al., 1996), the presence of non-conserved genes (e.g., the *cag* region, Censini *et al.*, 1996; Akopyonts *et al.*, 1998), the presence of insertion elements (Akopyants *et al.*, 1998; Höök-Nikanne *et* al., 1998), the presence of plasmids (Minnis *et al.*, 1995) and gene mosaics (e.g., the *vacA* gene, Atherton *et al.*, 1995), and the occurrence of chromosomal rearrangements (Jiang *et al.*, 1996). A long evolutionary history, a high level of mutation and frequent horizontal gene transfer are among the theories put forward to account for the high level of diversity within *H. pylori* isolates (Kuipers *et al.*, 1999).

The non-clonal nature of *H. pylori* allows for distinct isolates to be identified using discriminatory DNA fingerprinting techniques (Marshall *et al.*, 1995; Thoreson *et al.*, 2000; Miehlke *et al.*, 1999; Cellini *et al.*, 2003; Roma-Giannikou *et al.*, 2003). On this basis DNA fingerprinting techniques have been useful in identifying individuals who harbour two or more *H. pylori* strains (Fantry *et al.*, 1996). However, there have been several reports where *H. pylori* isolates taken from the same individual have shown highly similar fingerprint profiles with minor differences (Langenberg *et al.*, 1986; Goodwin, 1988; Majewski and Oudbier *et al.*, 1990, Covacci and Rappuoli, 1998; Gibson *et al.*, 1998; Enroth *et al.*, 1999; Kersulyte *et al.*, 1999; Kuipers *et al.*, 1999; Kuo *et al*, 1999; Israel *et al.*, 2001b; van der Ende *et al.*, 2001). Due to *H. pylori* being a highly genetically diverse organism, any isolates displaying highly similar, but not identical, genetic fingerprints would indicate that they are co-existing clonal variants of each other.

The generation of sub-species within a single gastric niche can possibly be the result of direct DNA exchange between strains occupying the same niche or by intra-strain chromosomal recombination. This hypothesis was confirmed in a study where RAPD-PCR fingerprinting demonstrated that DNA was exchanged between unrelated strains during infection of the same host (Kersulyte *et al.*, 1999). However, the difference in fingerprint profiles could also be due to host-specific adaptation. The occurrence of spontaneous mutations after serial passage in gnotobiotic piglets demonstrates that *H. pylori* can undergo host-specific adaptation (Akopyants *et al.*, 1995). This implies that a *H. pylori* founder strain can alter its genome to adapt to a new environment, resulting in a founder strain and a clonal variant

occupying the same host. The occurrence of isolates with highly similar genetic profiles that contain minor differences has been termed microevolution (Marshall *et al.*, 1995, 1998; Israel *et al.*, 2001b).

Microevolution has not only been detected by fingerprinting techniques. In a recent phylogenetic study it was demonstrated that strains from the same patient (that were isolated either from different gastro-intestinal sites six years apart or after passage through various animal models) were more similar to each other than strains that originated from different patients (Joyce *et al.*, 2002). Additionally, microarray analysis has shown that minor genetic differences between isolates taken from the same individual can occur over time (Kuipers *et al.*, 2000; Israel *et al.*, 2001b) or exhibit both genetic and phenotypic differences (Bjorkholm *et al.*, 2001).

Although there is no doubt that microevolution does occur during *H. pylori* infection, the mechanistic basis, the genetic sequences affected and the benefit of these rearrangements remain unclear. A recent study was able to define some of the loci affected during microevolution as the *cag* PAI was present in one isolate and absent in another, which otherwise possessed nearly identical gene profiles, obtained from the same (Bjorkholm *et al.*, 2001). Israel *et al.* (2001b) discovered that *H. pylori* isolates taken six years after the original biopsy had not only lost several genes but had also gained some genetic sequences. However, although the isolates investigated in the latter study were found to be derived from a common founder strain, each isolate was found to be unique implying that the rearrangements acquired during infection do not follow a set pattern, and could possibly be random. The fact that microevolution can have genetic and phenotypic effects on *H. pylori* isolates that can alter the virulence of an isolate (Bjorkholm *et* al., 2001), could account for the difficulty in correlating a particular bacterial genotype with clinical outcome of the disease (Li *et al*, 2002a; Smith *et al*, 2002b).

Identical strains of *H. pylori* usually colonise discrete sites within the gastric mucosa. It has been suggested that bacteria colonising different sites within the stomach of an individual come under slightly different selection pressures, e.g., the pH levels within different regions of the stomach have been shown to vary considerably (Karita & Blaser, 1998; Marshall *et al.*, 1998; Bijlsma *et al.*, 2000). Upon entering a host, bacteria can use their flagella to spread through the viscous gastric mucus and encounter new gastric sites with different

microenvironments that could possibly select for genetic rearrangements. Obtaining biopsies from regions of the gastric mucosa that are far removed from each other (such as the antrum and corpus) may have a greater potential for recovering *H. pylori* isolates that have undergone microevolution, than biopsies taken at the same site. In addition, it has also been documented that micro-evolved isolates can be obtained in biopsies taken from the same gastric site but at some time apart (Kuipers *et al.*, 1999, Israel *et al.*, 2001b), demonstrating that genetic rearrangements occur over time.

The present study describes the characterisation of paired *H. pylori* isolates that were isolated from the same (antrum) or different (antrum / corpus) gastric sites. These isolates were either obtained at the same endoscopic session or at some time apart. Each isolate was genotypically characterised at seven different loci and then fingerprinted and phylogenetically analysed using FAFLP-PCR. The aims of this study were (i) to compare the ability of conventional genotyping methods with a more sensitive fingerprinting technique (FAFLP-PCR) for discriminating unrelated, identical and micro-evolved *H. pylori* isolates, and (ii) to determine whether the degree of microevolution varies between isolates obtained from the same or different gastric sites, and isolates obtained at the same endoscopic session or isolates taken some time apart. The analysis described herein demonstrates the detection of subtypic variants of *H. pylori* isolates that occur *in vivo* during infection.

5.2 Results.

Genotyping of multiple loci, RAPD-PCR and FAFLP-PCR fingerprinting were applied to 31 paired antral or paired antrum and corpus *H. pylori* isolates. Six of these pairs represented paired isolates from familial studies, three pairs from Family 1 (F610 / F611 [father], 577A / 577C [sibling], 9613A / 9613C [sibling]) and 3 pairs from Family 2 (218A / 218C [mother], 605A / 605C [son], 222A / 222C [daughter's spouse; a single isolate available for the daughter was not included in this study]. Twenty-seven of the 31 pairs were obtained at the same endoscopic session, whereas, four were sequential samples, i.e., an initial biopsy was taken and a second biopsy was taken between 3–14 months later. Isolates MI506 / MI520, MI571 / MI575, MI517 / MI566 and MI512 / MI541 were obtained from antral biopsies from four patients 3, 6, 14 and 14 months apart, respectively.

5.2.1. vacA *genotyping.*

Out of the 31 paired isolates examined no differences in *vacA* signal sequence and mid-region alleles were detected within pairs, e.g., if one isolate had a *vacA* type of s1a m2 its corresponding paired isolate was also s1a m2 (Table 5.1). Two of the three pairs of isolates from members of Family 1 displayed the same *vacA* genotype (577A and 577C [sibling], s1a m2; 610 and 611 [father], s1a m2), whereas the third pair within this family displayed a different *vacA* type (9613A and 9613C [sibling], s2 m2). Two of the three pairs of isolates within Family 2 displayed the same *vacA* type (218A and 218C [mother], s1a m2; 222A and 222C [daughter's spouse], s1a m2), whereas, the third pair displayed a different *vacA* type (605A and 605C [son], s2 m2). Out of the 31 pairs of isolates examined 21 (67.7 %) were found to be of *vacA* type s1a m2, 8 (25.8 %) were of *vacA* type s1a m1 and 2 (6.5 %) were of *vacA* type s2 m2.

5.2.2. Sequences preceding vacA *and* ribA.

The intergenic sequence variants that precede the *vacA* and *ribA* genes were typed in all 31 paired isolates. Again there were no differences in sequence variants preceding the *vacA* genes within pairs (Table 5.1). Two of the three pairs of isolates obtained from Family 1 displayed identical sequence variants preceding the *vacA* gene (577A and 577C, VacAP3; F610 and F611, VacAP3), while, the third pair of isolates within this family displayed a different sequence variant (9613A and 9613C, VacAP4).

Table 5.1. Paired *H. pylori* isolates used in this study with various genotypes, ancestral relationship and presence of microevolution indicated.

Pairs	Origin of strain	Isolate name	*vacA* type	No. of *cagA* 3′-repeats	RibAP type	VacAP type	in-del type	RAPD analysis	FAFLP analysis	Diversity between related pairs – branch lengths (cm)
1	Antrum	577A	s1a m2	1	RibAP1	VacAP3	IV	related	related	1.5
	Corpus	577C	s1a m2	1	RibAP1	VacAP3	IV			0.2
2	Antrum	9613A	s2 m2	1	RibAP1	VacAP4	IV	related	related	1.2
	Corpus	9613C	s2 m2	1	RibAP1	VacAP4	IV			0.3
3	Antrum	F610	s1a m2	1	RibAP1	VacAP3	Ia	related	NA[d]	NA[d]
	Corpus	F611	s1a m2	1	RibAP1	VacAP3	Ia			NA[d]
4	Antrum	218A	s1a m2	1	RibAP1	VacAP3	II	related	related	0.25
	Antrum	218C	s1a m2	1	RibAP1	VacAP3	II			1.2
5	Antrum	605A	s2 m2	-[a]	RibAP1	VacAP4	-[a]	related	related	1.05
	Corpus	605C	s2 m2	-[a]	RibAP1	VacAP4	-[a]			1.2
6	Antrum	222A	s1a m2	1	RibAP1	VacAP3	IV	unrelated	unrelated	NA[d]
	Corpus	222C	s1a m2	1	RibAP1	VacAP3	IV			NA[d]
7	Antrum	112886A	s1a m2	2	RibAP1	VacAP3	II	unrelated	NA[d]	NA[d]
	Corpus	112886C	s1a m2	2	RibAP1	VacAP3	II			NA[d]
8	Antrum	15500A	s1a m2	**1**	RibAP1	VacAP3	Ia	related	related	1.8
	Corpus	15500C	s1a m2	**3**	RibAP1	VacAP3	Ia			0.4
9	Antrum	167233A	s1a m1	2	RibAP1	VacAP4	IV	related	related	0.1
	Corpus	167233C	s1a m1	2	RibAP1	VacAP4	IV			0.6
10	Antrum	34683A	s1a m1	**1**	RibAP1	VacAP3	II	unrelated	NA[d]	NA[d]
	Corpus	34683C	s1a m1	**2**	RibAP1	VacAP3	II			NA[d]
11	Antrum	45688A	s1a m2	**1**	RibAP1	VacAP3	II	related	related	0.2
	Corpus	45688C	s1a m2	**2**	RibAP1	VacAP3	II			1.2
12	Antrum	50334A	s1a m2	1	RibAP1	VacAP3	II	Related	NA[d]	NA[d]
	Corpus	50334C	s1a m2	1	RibAP1	VacAP3	II			NA[d]
13	Antrum	58745A	s1a m2	1	RibAP1	VacAP3	IV	related	NA[d]	NA[d]
	Corpus	58745C	s1a m2	1	RibAP1	VacAP3	IV			NA[d]
14	Antrum	59393A	s1a m1	1	RibAP1	VacAP3	II	Related	NA[d]	NA[d]
	Corpus	59393C	s1a m1	1	RibAP1	VacAP3	II			NA[d]
15	Antrum	68545A	s1a m2	1	RibAP1	VacAP3	IV	related	related	0.85
	Corpus	68545C	s1a m2	1	RibAP1	VacAP3	IV			2.1

16	Antrum	79862A	s1a m2	**1**	RibAP1	VacAP3	IV	unrelated	unrelated	NA[d]
	Corpus	79862C	s1a m2	**2**	RibAP1	VacAP3	IV			NA[d]
17	Antrum	87212A	s1a m2	**2**	RibAP1	VacAP3	IV	related	related	0.4
	Corpus	87212C	s1a m2	**1**	RibAP1	VacAP3	IV			0.4
18	Antrum	MI355	s1a m2	**2**	RibAP1	VacAP3	IV	related	NA[d]	NA[d]
	Antrum	MI356	s1a m2	**1**	RibAP1	VacAP3	IV			NA[d]
19	Antrum	MI458	s1a m2	1	RibAP3	VacAP3	II	related	related	0.05
	Antrum	MI459	s1a m2	1	RibAP3	VacAP3	II			0.3
20	Antrum	MI496	s1a m1	2	RibAP1	VacAP3	II	related	related	1.1
	Antrum	MI497	s1a m1	2	RibAP1	VacAP3	II			0.4
21	Antrum	MI506[e]	s1a m2	1	RibAP3	VacAP3	II	related	related	0.8
	Antrum	MI520[e]	s1a m2	1	RibAP3	VacAP3	II			1.8
22	Antrum	MI517[e]	s1a m1	1	RibAP1	VacAP3	II	unrelated	unrelated	NA[d]
	Antrum	MI566[e]	s1a m1	1	RibAP1	VacAP3	II			NA[d]
23	Antrum	MI512[e]	s1a m2	**0**	RibAP1	VacAP3	IV	related	NA[d]	NA[d]
	Antrum	MI541[e]	s1a m2	**1**	RibAP1	VacAP3	IV			NA[d]
24	Antrum	MI571[e]	s1a m1	**1**	RibAP1	VacAP3	II	unrelated	unrelated	NA[d]
	Antrum	MI575[e]	s1a m1	**2**	RibAP1	VacAP3	II			NA[d]
25	Antrum	FR147	s1a m2	1	RibAP3	VacAP3	M[c]	related	related	1
	Antrum	FR148	s1a m2	1	RibAP3	VacAP3	M[c]			0.4
26	Antrum	FR150	s1a m2	**1**	RibAP1	VacAP3	IV	unrelated	unrelated	NA[d]
	Antrum	FR151	s1a m2	**2**	RibAP1	VacAP3	IV			NA[d]
27	Antrum	FR205	s1a m1	1	RibAP1	VacAP3	II	related	related	0.2
	Antrum	FR206	s1a m1	1	RibAP1	VacAP3	II			0.5
28	Antrum	FR207	s1a m1	1	RibAP1	VacAP3	-[a]	related	related	0.9
	Antrum	FR208	s1a m1	1	RibAP1	VacAP3	-[a]			10.3
29	Antrum	FR229	s1a m2	1	RibAP1	VacAP3	II	unrelated	unrelated	NA[d]
	Antrum	FR230	s1a m2	1	RibAP1	VacAP3	II			NA[d]
30	Antrum	FR294	s1a m2	1	RibAP2	VacAP2	M[c]	related	NA[d]	NA[d]
	Antrum	FR295	s1a m2	1	RibAP2	VacAP2	M[c]			NA[d]
31	Antrum	FR296	s1a m2	1	RibAP1	VacAP3	Ia	related	related	0.9
	Antrum	FR297	s1a m2	1	RibAP1	VacAP3	Ia			0.5

[a] Unable to obtain genotype despite repeated attempts.
[b] Microevolution is not applicable between unrelated isolates.
[c] Multiple PCR products obtained.
[d] Not applicable.
[e] Sequential isolates.

Within each pair isolates possessed the same sequence variant preceding the *ribA* gene (RibAP1). Within Family 2, all three pairs of isolates displayed the same sequence variant preceding the *ribA* gene (RibAP1). However, paired isolates 218A / 218C and 222A / 222C possessed the sequence variant VacAP3 preceding the *vacA* gene, whilst the third pair of isolates 605A / 605C displayed the sequence variant type VacAP4.

Out of the 31 pairs, 28 (90.3 %) were of type RibAP1, 2 (6.5 %) were of type RibAP3, 1 was of type RibAP2 (3.2 %), with respect to the sequence variants preceding the *vacA* gene, 27 pairs (87.1 %) were of type VacAP3 and 4 pairs (12.9 %) were of type VacAP4.

5.2.3. Insertion-deletion (in-del) analysis.

The in-del motifs located at the right end of the *cag* PAI were examined. No differences were detected within pairs of isolates at this site in the genome (Table 5.1). Within Family 1 sibling strains 577A / 577C and 9613A / 9613C displayed an in-del motif of type IV. However, the pair of isolates F610 / F611 from their farther displayed a motif of type Ia. The pairs of isolates from members of Family 2 (218A /218C, 222A / 222C and 605A / 605C) displayed different in-del motifs (Table 5.1).

Out of the 31 pairs examined, 27 pairs (87.1 %) yielded PCR products that could be used to determine which motif was located at the right end of the *cag* PAI. Of the 27 typable isolates 13 (48.2 %) were of type II, 11 (40.7%) were of type IV and 3 (11.1%) were of type Ia.

5.2.4. cagA *genotyping.*

All 31 pairs of isolates were found to possess the *cagA* gene. Further examination at the 3′-end of the *cagA* gene revealed that there was no variation in the number of repeat sequences within isolates for 22 pairs, while 9 pairs displayed variation in the number of repeats in this region (Table 5.1). The 9 pairs included isolates 45688A (one repeat) / 45688C (two repeats), 15500A (one repeat) / 15500C (three repeats), 87212A (2 repeats) / 87212C (one repeat), MI355 (two repeats) / MI356 (one repeat), MI512 (no repeats) / MI541 (one repeat), 79862A (one repeat) / 79862C (two repeats), MI571 (one repeat) / MI575 (two repeats), FR150 (one repeat) / FR 151 (two repeats) and 34683A (one repeat) / 34683C (two repeats) (Fig. 5.1). There were no differences in the number of repeats among paired isolates within both families, except that the number of repeats could not be determined within isolates F610 / F611 (Family 1).

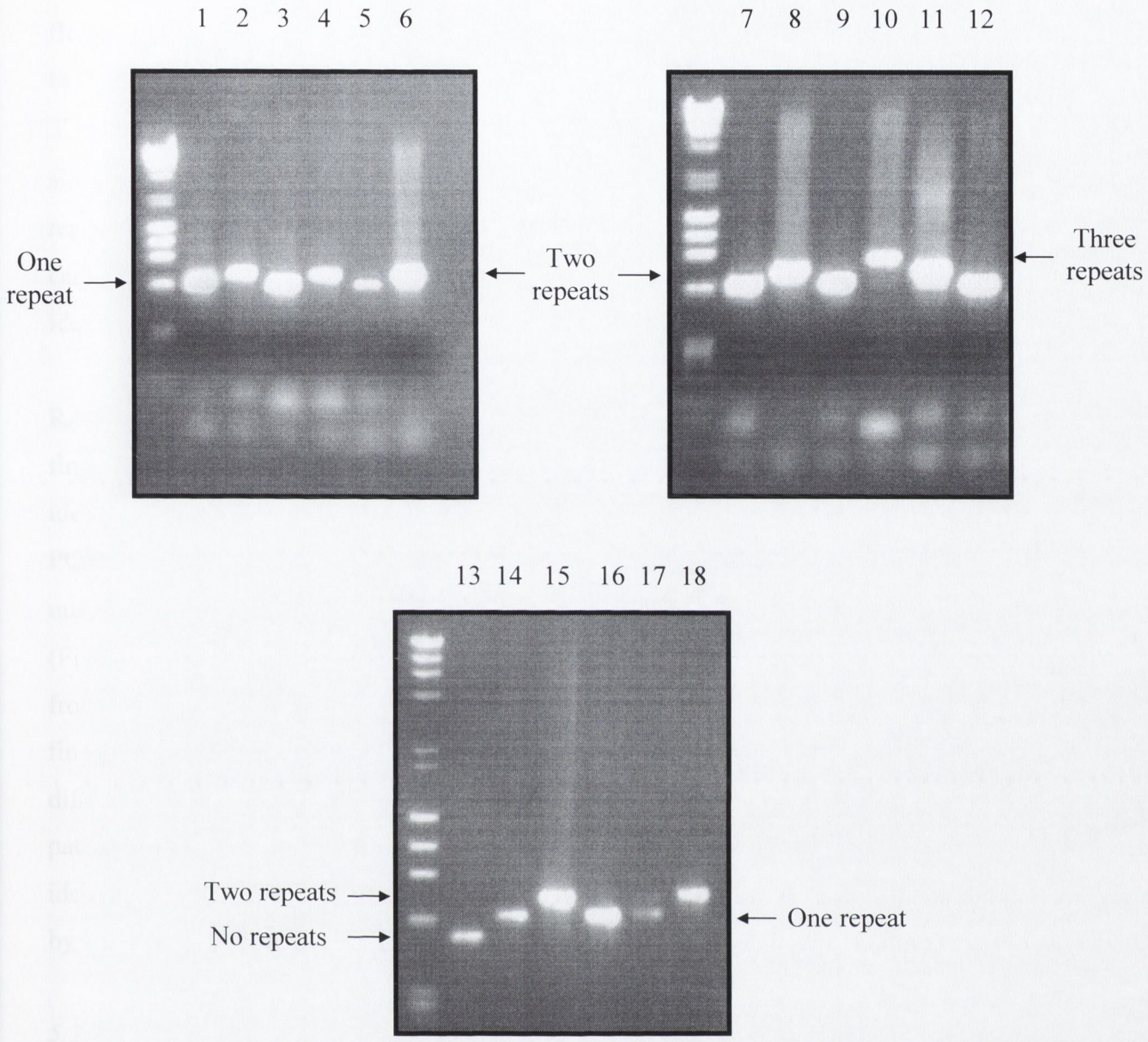


Fig. 5.1. Amplification of the 3′-end of the *cagA* gene in nine paired *H. pylori* isolates showing differences in the number of repeat sequences.

PCR was carried out using primers CAG1 and CAG2 (Table 3.1). The number of repeat sequences differs within each pair: 1 = 79862A, 2 = 79862C; 3 = MI 571, 4 = MI 575; 5 = FR 150, 6 = FR 151; 7 = 45688A, 8 = 45688C; 9 = 15500A, 10 = 15500C; 11 = 87212A, 12 = 87212C; 13 = MI 512, 14 = MI 541; 15 = MI 355, 16 = MI 356; 17 = 34683A, 18 = 34683C. A = antrum, C = corpus

5.2.5. RAPD-PCR analysis.

Based on genotyping all pairs of isolates displayed identical allelic combinations with respect to 5 genetic loci, while 9 of these pairs displayed differences in the number of repeats at the 3′-end of the *cagA* gene. These initial observations suggested that 22 pairs might represent single strains and that the 9 pairs showing variations at the 3′-end of the *cagA* gene might represent two different strains inhabiting the same patient. RAPD-PCR fingerprinting was carried out on all 31 paired isolates to determine whether each pair represented isolates with identical (related) or different (mixed infection) profiles.

RAPD-PCR revealed that 22 (71.0%) of the pairs displayed identical or nearly identical fingerprint profiles while 9 pairs (29.0%) displayed different profiles. Of the 22 putative identical pairs on the basis of genotyping, 5 pairs constituted isolates with different RAPD-PCR fingerprints (e.g., see Fig. 5.2), only 4 of the 9 pairs of isolates showing variations in the number of repeats at the 3′-end of the *cagA* gene displayed different RAPD-PCR profiles (Fig. 5.3). Interestingly, RAPD-PCR revealed that isolates 577A / 577C and 9613A / 9613C from siblings in Family 1 displayed identical fingerprints, that were dissimilar to the fingerprints of the pair of isolates F610 / F611 from the father, even although they exhibited different *vacA* alleles and sequence variants preceding the *vacA* gene. The RAPD-PCR patterns also revealed that none of the pairs of isolates within Family 2 displayed similar or identical fingerprints between pairs indicating that each member of the family was colonised by a different strain.

5.2.6. Restriction fragment length polymorphism (RFLP) and sequence analysis of paired H. pylori *isolates.*

To provide additional confirmation that paired *H. pylori* isolates that exhibited identical RAPD-PCR profiles were indeed identical a number of different analyses were performed on isolate pairs with representative s m / VacAP / in-del types. It was not the intention that such tests should be applied comprehensively to all 22 pairs of isolates that displayed identical or nearly identical fingerprint profiles by RAPD-PCR.

5.2.6.1. RFLP analysis of paired *H. pylori* isolates at the *napA*, *omp* (JHP0429), *hsdR*, *galE*, *gmd*, and *flnF* loci.

To determine whether paired *H. pylori* isolates that exhibited identical RAPD-PCR profiles also exhibited identical restriction fragment-length polymorphism (RFLP) patterns, six loci

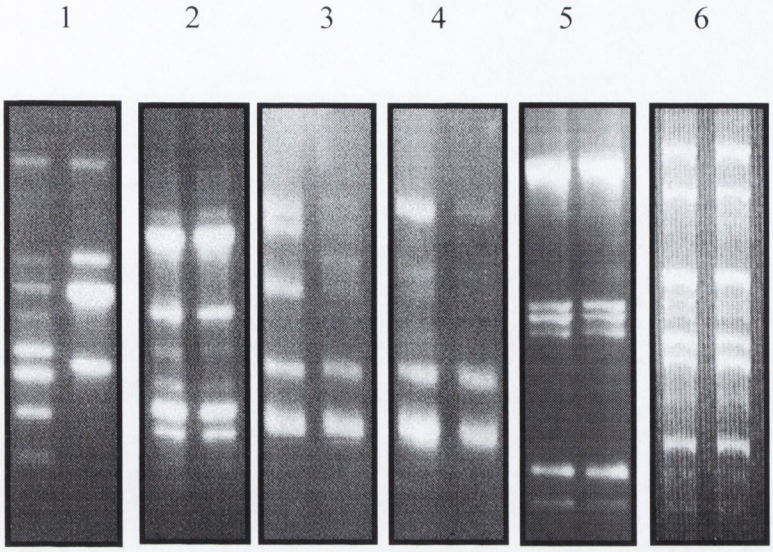


FIG. 5.2. RAPD-PCR fingerprint profiles of six putatively related pairs of *H. pylori* isolates based on specific loci genotyping.

RAPD-PCR profiles were generated using primer PI (Table 1) for all pairs except pairs from 5 and 6 where RAPD primer combinations PII / PV and PI / PV (Table 1) were used, respectively. Pairs from left to right; 1 = 222A / 2 22C, 2 = FR205 / FR206, 3 577A / 577C, 4 = 9613A / 9613C, 5 = F610 / F611, 6 = MI 458 / MI 459. Pairs 1–3 represent two different *H. pylori* strains inhabiting the same host, whereas pairs 4–6 are two isolates representing one *H. pylori* strain. Pairs 3 and 4 are from members of the same family (Family 1); their fingerprint profiles are highly similar and suggest that both members of this family are colonised by the same strain. A = antrum, C = corpus.

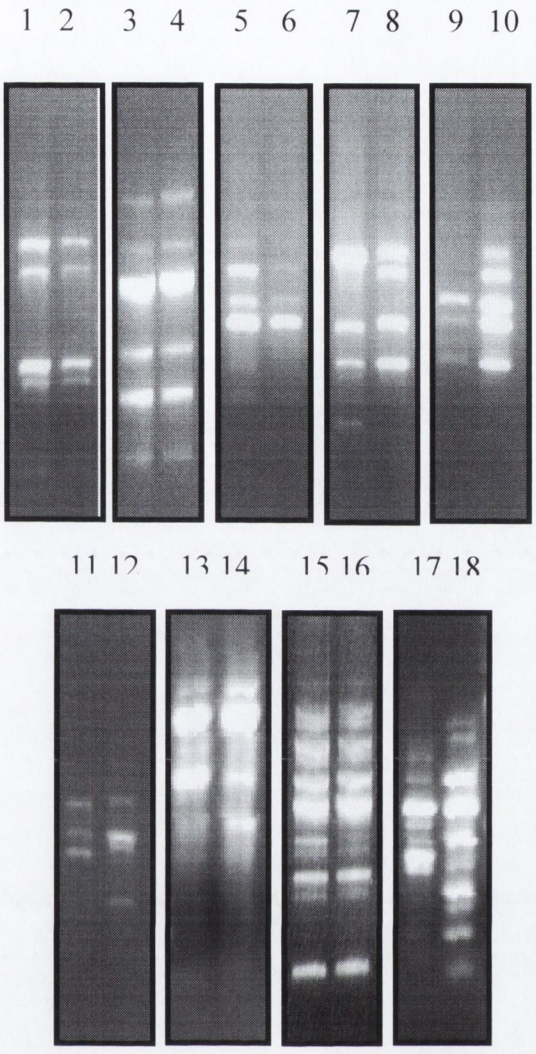


FIG. 5.3. RAPD-PCR profiles of the nine paired *H. pylori* isolates showing variations in the number of repeat sequences at the 3′-end of the *cagA* gene within each pair and putatively deemed to be unrelated.

Lanes: 1 = 45688A, 2 = 45688C, 3 = 15500A, 4 = 15500C, 5 = 87212A, 6 = 87212C, 7 = 79862A, 8 = 79862C, 9 = FR 150, 10 = FR 151, 11 = MI 571, 12 = MI 575, 13 = MI 512, 14 = MI 541, 15 = MI 355, 16 = MI 356, 17 = 34683A, 18 = 34683C. These fingerprint profiles demonstrate that each of the pairs 45688A / 45688C, 15500A / 15500C, 87212A / 87212C, MI 512 / MI541 and MI 355 / MI356 represent single strains inhabiting the same individuals as they display identical profiles. The other four pairs (79862A / 79862C, FR150 / FR151, MI571 / MI575 and 34683A / 34683C) represent two strains inhabiting the same individuals, as their fingerprint patterns are different to each other. A = antrum, C = corpus.

within *Hin*dIII-restricted genomic DNA from a pair of isolates (MI458 and MI459 that appeared to be identical on the basis of s m / VacAP / in-del typing and RAPD-PCR fingerprinting) were investigated. These six loci included housekeeping genes and genes that may be under selection pressure. These genes were *flhF* (a flagellar biosynthesis protein), *hsdR* (or *hsd_3*, a type I restriction enzyme), *omp* / JHP0429 (a putative outer membrane protein), *gmd* (a GDP-mannose dehydratase enzyme), *galE* (a UDP-glucose 4-epimerase enzyme) and *napA* (neutrophil activating protein A). The probes for each gene were amplified by PCR. From Fig. 5.4 it can be seen that the RFLP patterns for each isolate within this pair were identical. This confirms that the genomes of these isolates are highly similar. It has been shown previously that RFLP patterns from unrelated *H. pylori* isolates tend to differ (Akopyanz *et al.*, 1992).

5.2.6.2. RFLP analysis of the *vacA* locus.

To assess whether the RFLP patterns differ within pairs of unrelated isolates, one related pair of *H. pylori* isolates (167233A / 167233C) and one pair of unrelated isolates (79862A / 79862C, both on the basis of RAPD-PCR) were investigated with a *vacA* probe. Fig. 5.5 demonstrates that within this pair of related isolates the RFLP patterns are identical, whilst the unrelated pair of isolates displayed distinct RFLP patterns at this locus.

5.2.6.3. Sequence analysis of paired *H. pylori* isolates.

To determine whether RAPD-PCR analysis was sufficient to indicate whether a pair of isolates shared an ancestral relationship, the *napA* gene (encoding the Neutrophil-Activating Protein A) nucleotide sequence of a pair of isolates with identical RAPD-PCR profiles (MI355 / MI356) were compared (Fig. 5.6). These isolates differed with respect to the number of 3′-end repeats in the *cagA* gene. The partial DNA sequences of the *napA* genes of isolates MI355 and MI356 show 100% identity (Fig. 5.7). This indicates that these isolates originated from the same founder strain, as sequence divergence between orthologous genes is common whereas identical sequences are rare (Achtman *et al.*, 1999).

5.2.7. FAFLP-PCR genotyping.

FAFLP-PCR was applied (i) to confirm earlier observations produced by RAPD-PCR fingerprinting, and (ii) to determine if minor genetic differences could be detected between paired isolates displaying identical RAPD-PCR profiles, as FAFLP-PCR is a more sensitive fingerprinting technique. Due to the fact that *H. pylori* isolates are genetically diverse (Go *et*

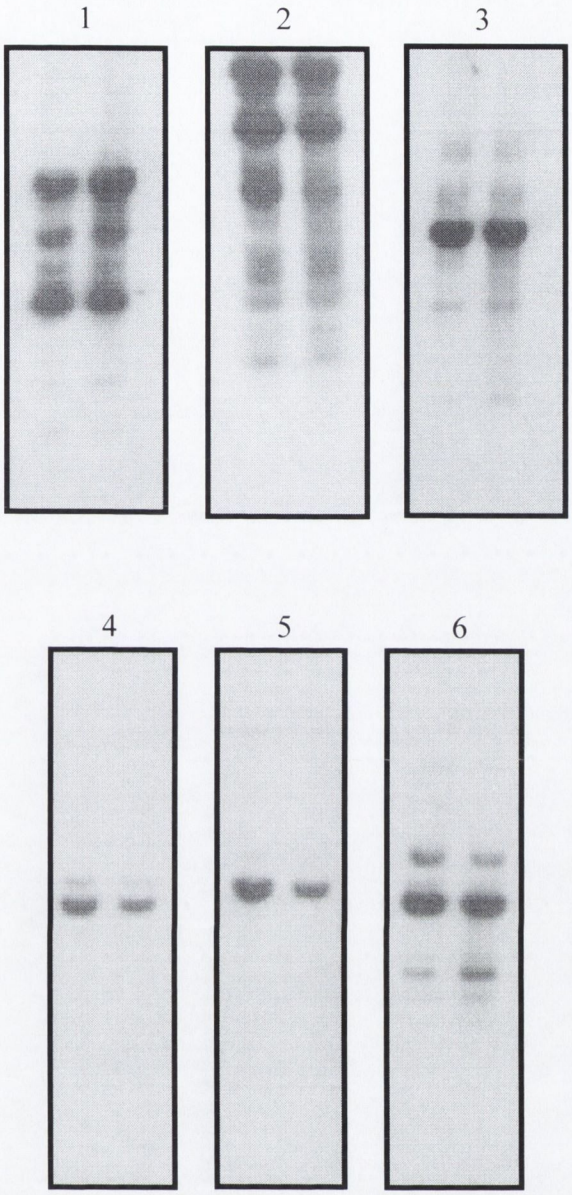


Fig. 5.4. Southern blot of *Hin*dIII-restricted genomic DNA from paired *H. pylori* isolates MI458 and MI459 using six different probes.

Blots 1–6 contain restricted genomic DNA from isolates MI458 and MI459 (s1a m2, VacAP3, in-del type III) in the left and right lanes, respectively. Each blot was hybridised with a gene probe for different genes (1 = *napA*, 2 = *omp*, 3 = *hsdR*, 4 = *galE*, 5 = *gmd*, 6 = *flhF*). The RFLP patterns for each isolate within this pair are identical for each probe used.

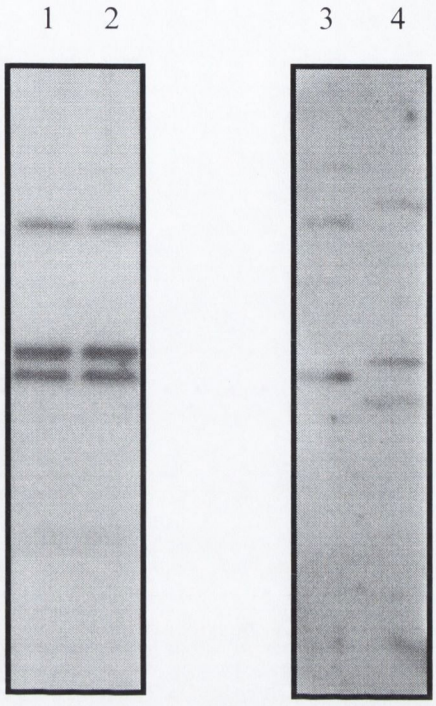


Fig. 5.5. Southern blot of *Hin*dIII-restricted genomic DNA from one ancestrally related pair of *H. pylori* isolates and one unrelated pair of isolates using a *vacA* probe.

Both blots contain *Hin*dIII-restricted genomic DNA from *H. pylori* isolates. Lanes 1-4 contain DNA from isolate pairs 167233A / 167233C (s1a m1, VacAP4, in-del type IV) and 79862A / 79862C (s1a m2, VacAP3, in-del type IV), respectively. From these blots (and the blots form Fig. 5.4) it can be seen that ancestrally related isolates share identical RFLP patterns, whereas unrelated isolates can display distinct RFLP patterns. A = antrum, C = corpus.

Isolate MI355

CTAAATGGGCTTCCAGCATCCAAATGGATTTTTGCAACTTGGCCAATTGATCGTCCGCATA
AGTTACGGTAACTTTATCGCCTTCTTTTTCAGCGGTGTTAGAGAGCTCTTTAAATTCTTTTT
CTAGGTGTTTGTAGTCCTCTAGAATTTCTTTAAAGATGTCTTTAGAGTGGAAGCTCGTTTTA
GTTTCTTCTTTAACACGAGTGAGTTTGATCGCTTCGGATAAAGTGACTAAGGGGTGGTGTC
CTAATTGAACGATCCTTTCAGCGAGATCGTCAAACATGTCCGCAAACTCTTCATAAATTTC
TTCAGTGGCTTTATGCACATTGAAAAAATCGGTGCCTTTCACATTCCAATGGAAGTTATGC
ACTTTCATAAATAACACGATCGCATCCGCTTGCAAATGTTTTAAAATTTCAAATGTTTTCA
TCAAAAGTCCTTTTTTTAAAATTGTTTTAGGCTTTTATTTTCATAAAAACCTTGATC

Isolate MI356

GCTAAATGGGCTTCCAGCATCCAAATGGATTTTTGCAACTTGGCCAATTGATCGTCCGCAT
AAGTTACGGTAACTTTATCGCCTTCTTTTTCAGCGGTGTTAGAGAGCTCTTTAAATTCTTTT
TCTAGGTGTTTGTAGTCCTCTAGAATTTCTTTAAAGATGTCTTTAGAGTGGAAGCTCGTTTT
AGTTTCTTCTTTAACACGAGTGAGTTTGATCGCTTCGGATAAAGTGACTAAGGGGTGGTGT
CCTAATTGAACGATCCTTTCAGCGAGATCGTCAAACATGTCCGCAAACTCTTCAAAATTTC
TTCAGTGGCTTTATGCACATTGAAAAAATCGGTGCCTTTCACATTCCAATGGAAGTTATGC
ACTTTCATAAATAACACGATCGCATCCGCTTGCAAATGTTTTAAAATTTCAAATGTTTTCA
TCAAAAGTCCTTTTTTTAAAATTGTTTTAGGCTTTTATTTTCATAAAAACCTTGA

Fig. 5.6. *napA* nucleotide sequences of paired *H. pylori* isolates MI355 and MI356 (s1a m2, VacAP3, in-del type IV).

Partial sequences of the *napA* gene were amplified using primers N1 and N3 (Table 3.1). Sequencing was performed by MWG-Biotech™.

```
MI355           -CTAAATGGGCTTCCAGCATCCAAATGGATTTTTGCAACTTGGCCAATTGATCGTCCGCA 59
MI356           GCTAAATGGGCTTCCAGCATCCAAATGGATTTTTGCAACTTGGCCAATTGATCGTCCGCA 60
                 ************************************************************

MI355           TAAGTTACGGTAACTTTATCGCCTTCTTTTTCAGCGGTGTTAGAGAGCTCTTTAAATTCT 119
MI356           TAAGTTACGGTAACTTTATCGCCTTCTTTTTCAGCGGTGTTAGAGAGCTCTTTAAATTCT 120
                ************************************************************

MI355           TTTTCTAGGTGTTTGTAGTCCTCTAGAATTTCTTTAAAGATGTCTTTAGAGTGGAAGCTC 179
MI356           TTTTCTAGGTGTTTGTAGTCCTCTAGAATTTCTTTAAAGATGTCTTTAGAGTGGAAGCTC 180
                ************************************************************

MI355           GTTTTAGTTTCTTCTTTAACACGAGTGAGTTTGATCGCTTCGGATAAAGTGACTAAGGGG 239
MI356           GTTTTAGTTTCTTCTTTAACACGAGTGAGTTTGATCGCTTCGGATAAAGTGACTAAGGGG 240
                ************************************************************

MI355           TGGTGTCCTAATTGAACGATCCTTTCAGCGAGATCGTCAAACATGTCCGCAAACTCTTCA 299
MI356           TGGTGTCCTAATTGAACGATCCTTTCAGCGAGATCGTCAAACATGTCCGCAAACTCTTCA 300
                ************************************************************

MI355           TAAATTTCTTCAGTGGCTTTATGCACATTGAAAAAATCGGTGCCTTTCACATTCCAATGG 359
MI356           TAAATTTCTTCAGTGGCTTTATGCACATTGAAAAAATCGGTGCCTTTCACATTCCAATGG 360
                ************************************************************

MI355           AAGTTATGCACTTTCATAAATAACACGATCGCATCCGCTTGCAAATGTTTTAAAATTTCA 419
MI356           AAGTTATGCACTTTCATAAATAACACGATCGCATCCGCTTGCAAATGTTTTAAAATTTCA 420
                ************************************************************

MI355           AATGTTTTCATCAAAAGTCCTTTTTTTAAAATTGTTTTAGGCTTTTATTTTCATAAAAAC 479
MI356           AATGTTTTCATCAAAAGTCCTTTTTTTAAAATTGTTTTAGGCTTTTATTTTCATAAAAAC 480
                ************************************************************

MI355           CTTGATC 486
MI356           CTTGA-- 485
                *****
```

Fig. 5.7. Alignment of *napA* sequences from paired *H. pylori* isolates MI355 and MI356.

Sequences were aligned using ClustalW™. These isolates show 100% nucleotide identity. These paired isolates are, therefore, ancestrally related.

al., 1996, Salaun *et al.*, 1998), if the FAFLP profiles of two isolates from the same individual displayed highly similar fingerprints they were accepted as being ancestrally related. Out of the 31 pairs of isolates examined, 9 (29.0%) pairs were not used for further FAFLP-PCR analysis in this study as these pairs either did not yield profiles for all three selective primers (+ A, + C and + G) (see Fig. 4.1), or they did not yield profiles at all (degraded DNA) for each isolate within a pair.

Visual analysis of the fingerprints run on a sequencing gel clearly showed that some pairs were nearly identical while others possessed different profiles (Fig. 5.8). However, visual inspection is not enough to determine minor genetic differences between pairs.

Three different electropherograms were generated for each of the 22 paired isolates, with each of the three selective primer pairs used (*Mse*I adaptor sequence + 0 / *Eco*RI adaptor sequence + A, + G or + C). The Genescan™ FAFLP profiles display the diversity found between paired isolates (Figs. 5.9 and 5.10). These electropherograms revealed three situations that can occur between paired isolates; (i) all peaks representing amplified DNA fragments were shared by both isolates in a pair (Fig. 5.9 A), (ii) the peaks in one isolate of a pair were not shared by its corresponding isolate (Fig. 5.9 B) and (iii) both isolates within a pair shared the majority of peaks, but there were a small number of peaks that occured in one isolate but not in the other (Fig. 5.10 A) In the latter case such minor differences can sometimes be picked up by one selective primer pair but not by another (c.f., Fig. 5.10 A and B). Situation (i) indicates that two isolates are representative of the same strain, situation (ii) indicates a mixed infection, and situation (iii) indicates a pair of ancestrally related isolates that exhibit microevolution.

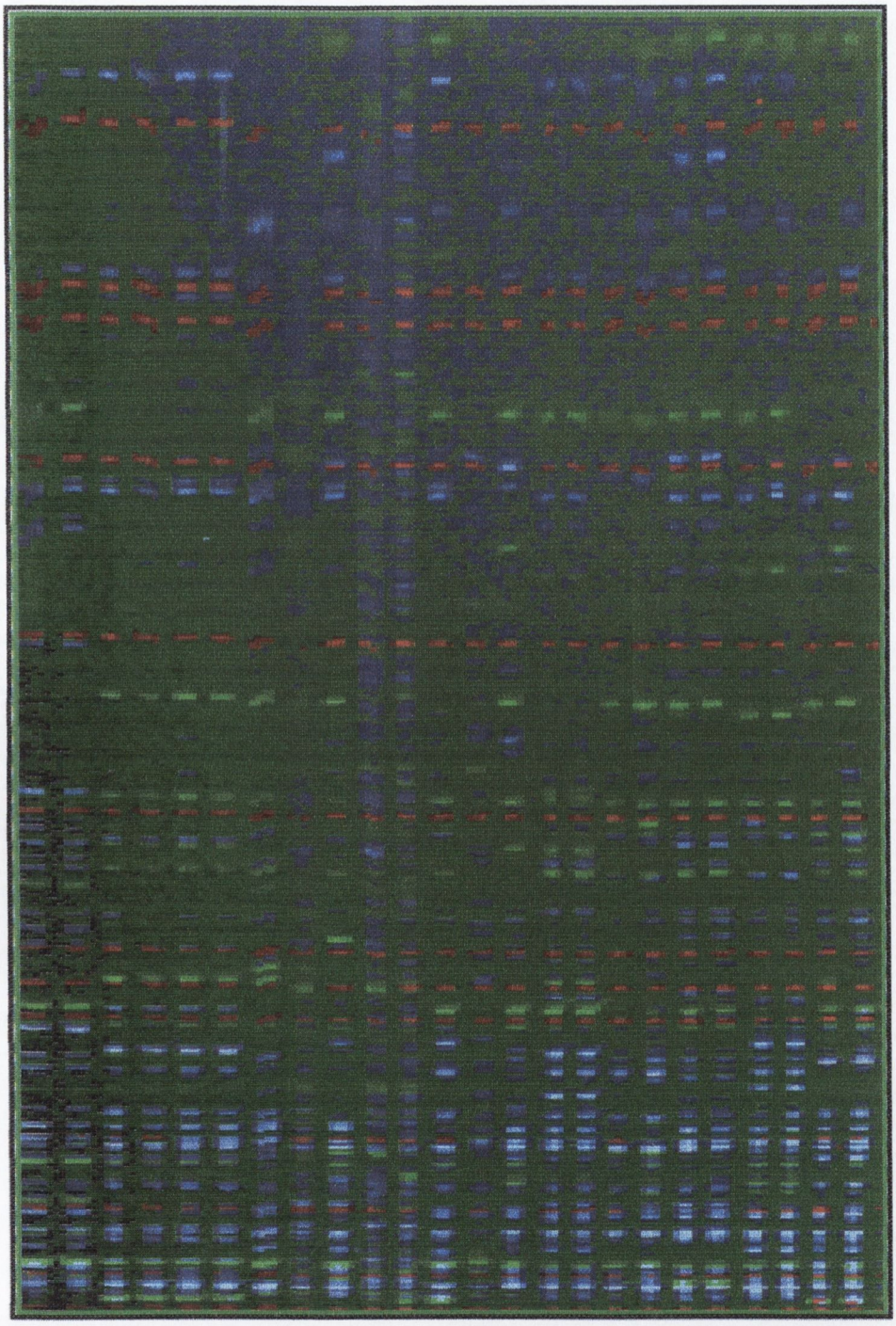
5.2.8. Phylogenetic analysis of FAFLP-PCR profiles.

The electropherogram peaks for all three selective primers were converted into binary format depending on whether a peak was present [1] or absent [0]. The binary data for each isolate were aligned and a neighbour-joining phylogenetic tree was generated (Fig. 5.11). This tree shows that out of the 22 pairs of isolates examined, 16 (72.7%) were found to have ancestral relationships and 6 (27.3%) were found to be unrelated (mixed infection). With the exception of isolate pairs 577A / 577C and 9613A / 9613C that grouped together, the other 14 pairs of isolates that had been previously shown to be related by RAPD-PCR grouped correspondingly, and displayed minor variations in their fingerprint profiles as indicated by the differing branch lengths between each isolate of each pair. These minor variations

FIG. 5.8. FAFLP -PCR fingerprints of 12 *H. pylori* paired isolates.

PCR products were run on an acrylamide sequencing gel. The primer pairs (*Mse*I + 0 / *Eco*RI + A - FAM [blue bands] and *Mse*I + 0 / *Eco*RI + G - JOE [green bands]) were used in a multiplex reaction. Red coloured bands indicate the internal lane standard, Genescan Rox 500. Isolates by lanes: 1 =167233A, 2 = 167233C, 3 =577A, 4 = 577C, 5 = 9613A, 6 = 9613C, 7 = MI571, 8 = MI575, 9 = 34683A, 10 = 34683C, 11 = 58745A, 12 = 58745C, 13 = 15500A, 14 =v 15500C, 15 = 59393A, 16 = 59393C, 17 = 68545A, 18 = 68545C, 19 = 87212A, 20 = 87212C, 21 = MI458, 22 = MI459, 23 = MI506, 24 = MI520. A = antrum, C = corpus.

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24



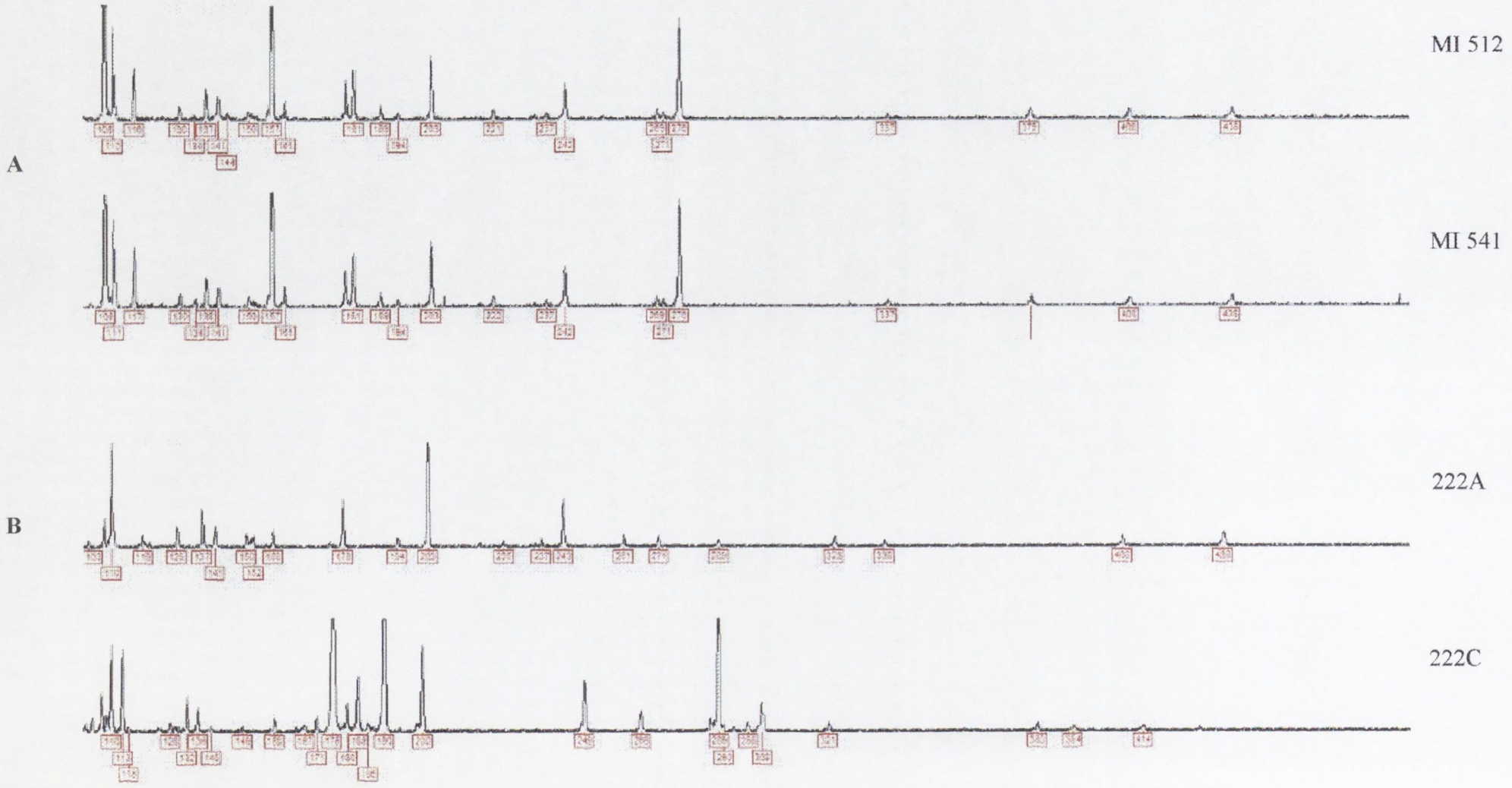

A
MI 512
MI 541
B
222A
222C

Fig. 5.9. Electropherogram of two paried *H. pylori* isolates with the selective primer + G.

(A) Electropherograms obtained for isolate pair MI512 and MI541 using selective primer pair *Mse*I + 0 / *Eco*RI + C are shown. The profiles for each isolate are identical indicating that they are probably the same strain.

(B) Electropherograms obtained for isolate pair 222A and 222C using selective primer pair *Mse*I + 0 / *Ec* RI + C are shown. The profiles of each isolate are dissimilar. These profiles indicate that this pair of isolates represent two different strains inhabiting the same individual. RAPD-PCR analysis supports these fingings (Fig. 5.3).

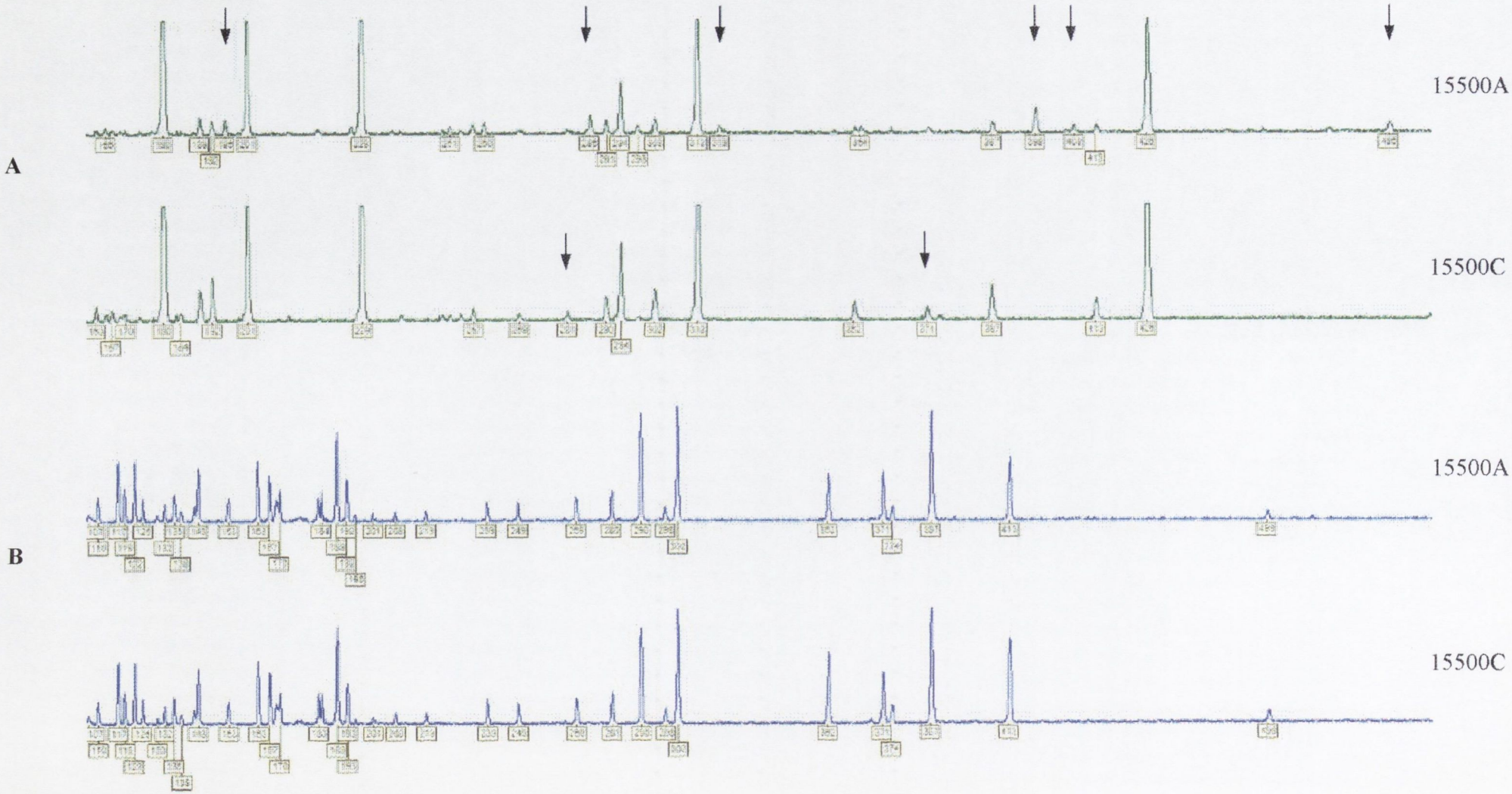

A
15500A
15500C
B
15500A
15500C

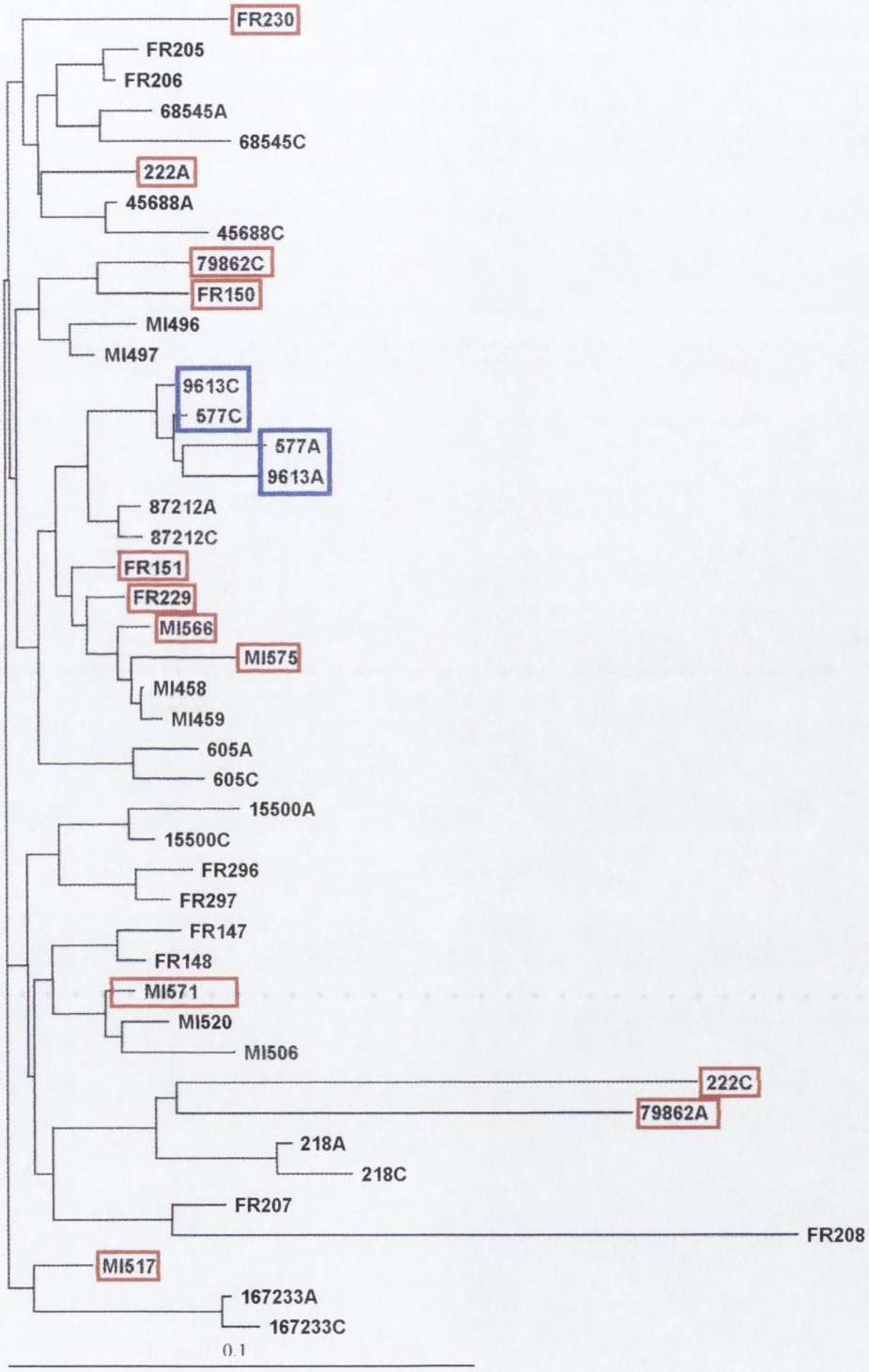
Fig. 5.10. Electropherogram of paried isolates 15500A / 15500C using two different selective primers (+ G and + A).

(A) Electropherograms obtained for isolate pair 15500A and 15500C using selective primer pair *Mse*I + 0 / *Eco*RI + G are shown. The fact that the majority of peaks are shared by each isolate indicates that they share an ancestral relationship. However, there are peaks in each isolate that is not contained in the other (Black arrows). This indicates that these isolates were once identical but minor genetic differences arose via microevolution.

(B) Electropherograms obtained for isolate pair 15500A and 15500C using selective primer pair *Mse*I + 0 / *Eco*RI + A are shown, minor genetic differences can be seen. This demonstrates that microevolution may be detected by one selective primer but not another. Therefore, it is necessary to analyse isolates with more than one selective primer pair to detect microevolution.

Fig. 5.11. Neighbour-joining tree developed from the binary data obtained by genotypic comparisons by the Genotyper™ macro.

This tree displays the genetic similarities of paired *H. pylori* isolates based on their FAFLP profiles. Red boxes indicate pairs of isolates that did not group with each other. Isolates from two members of the same family group (Family 1) together (blue boxes) indicating that both members of this family were colonised by the same strain. The bottom scale displays the number of fingerprint band substitutions per site.


FR230
FR205
FR206
68545A
68545C
222A
45688A
45688C
79862C
FR150
MI496
MI497
9613C
577C
577A
9613A
87212A
87212C
FR151
FR229
MI566
MI575
MI458
MI459
605A
605C
15500A
15500C
FR296
FR297
FR147
FR148
MI571
MI520
MI506
222C
79862A
218A
218C
FR207
FR208
MI517
167233A
167233C
0.1

indicated the presence of microevolution. There was no evidence to suggest that paired antral or paired antrum and corpus isolates taken at the same endoscopic session or isolates taken some time apart (sequentially) exhibited more or less microevolution.

5.2.9. VacA protein analysis of microevolved H. pylori *isolates.*

Microevolution was detected in 16 of the 22 pairs fully analysed by FAFLP-PCR (see Section 5.2.8). The loci affected by genomic rearrangements between paired isolates showing highly similar genomic profiles have been shown to vary (Bjorkholm *et al.*, 2001; Israel *et al.*, 2001b; Beesley, 2002) and seem to be random. Since the genomic sequences affected by microevolution herein were not known and because microevolution has been shown to have an effect at the protein level (Bjorkholm *et al.*, 2001), the presence of the VacA protein was examined in the paired *H. pylori* isolates.

Out of the 31 pairs of isolates, 19 pairs (61.3%) expressed the VacA cytotoxin, 8 pairs (25.8%) failed to express the toxin and 4 pairs (12.9%) had one isolate that expressed the toxin at a different level compared to its corresponding isolate (Fig. 5.12). The pairs of isolates that displayed varying levels of VacA expression were 577A / 577C, 45688A / 45688C, 34683A / 34683C and 79862A / 79862C. As indicated by FAFLP-PCR and / or RAPD-PCR analysis isolate pairs 577A / 577C and 45688A / 45688C represent ancestrally related isolates that have undergone microevolution (i.e., 2 out of the 16 microevolved paired isolates). The other two pairs of isolates represent mixed infections, which explains the variation in VacA expression in these latter instances. In the case of isolate pair 577A / 577C the VacA cytotoxin was expressed by isolate 577C but not by 577A, whereas with isolate pair 45688A / 45688C different levels of VacA expression were apparent based on equal loading of total protein (Fig. 5.12). Within both pairs isolates had an identical *vacA* genotype, namely, VacAP3 s1a m2.

5.2.10. vacA transcriptional analysis of microevolved H. pylori *isolates.*

The presence of *vacA* mRNA transcripts was investigated using RT-PCR within paired isolates 577A / 577C and 45688A / 45688C to determine whether variations in VacA expression were due variation in transcriptional levels at the *vacA* locus. Specific reverse primers for the mid-region and signal sequence of the *vacA* gene were used to generate cDNA for this locus in paired isolates 577A / 577C and 45688A / 45688C. Both mid-region and signal sequence primers were capable of detecting *vacA* mRNA transcripts (Fig. 5.13).

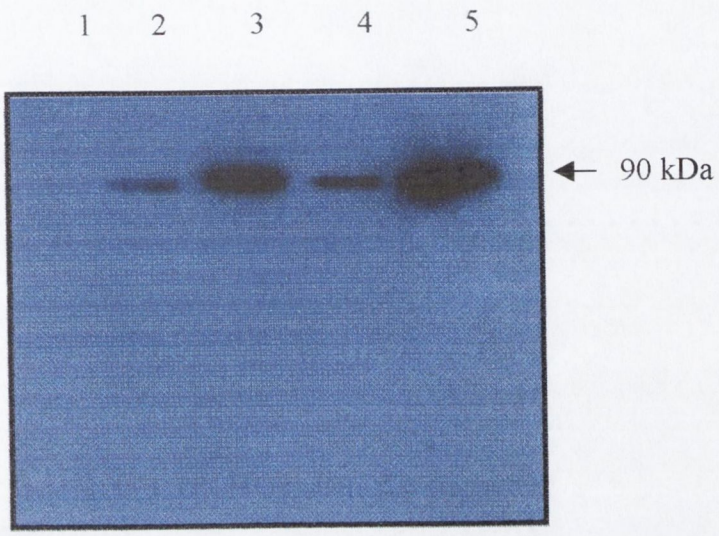


Fig. 5.12. Western immuno-blot analysis using anti-VacA antibodies.

Lanes: 1 = 577A, 2 = 577C, 3 = 45688A, 4 = 45688C, 5 = NCTC 11638 ($VacA^+$). Paired isolates 577A / 577C and 45688A / 45688C are ancestrally related *H. pylori* isolates. A band at ~ 90 kDa indicates the presence of VacA. Comparisons are based on equal loading of protein onto each lane.

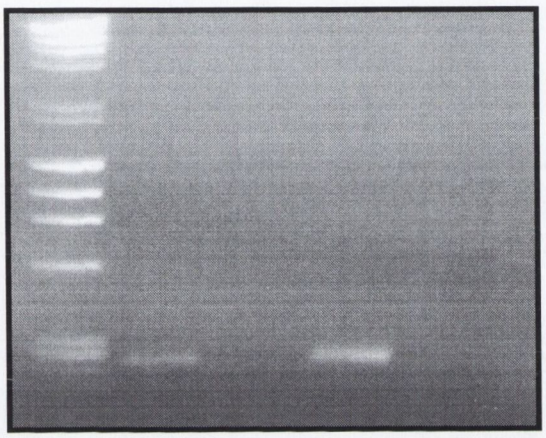
Fig. 5.13. Reverse transcription polymerase chain reaction (RT-PCR) analysis of *H. pylori vacA* mRNA transcripts.

(A). *vacA* transcripts analysed in paired *H. pylori* isolates 45688A (lane 1) and 45688C (lane 3). *vacA* mRNA transcripts were converted to cDNA using the reverse signal sequence primer VA1-R (Table 3.1) in the presence of reverse transcriptase. The signal sequence of the *vacA* cDNA was amplified using primers VA1-F and VA1-R (Table 3.1). A 190-bp PCR product indicates the presence of *vacA* mRNA transcripts. Lanes 2 and 4 represent negative controls (no reverse transcriptase) for isolates 45688A and 45688C, respectively.

(B). *vacA* transcripts analysed in paired *H. pylori* isolates 577A (lane 4) and 577C (lane 2). Analysis included a positive control of the *galE* gene. *vacA* and *galE* mRNA transcripts were converted to cDNA using the reverse primers VAG-R (*vacA* mid-region) and P2, respectively (Table 3.1) in the presence of reverse transcriptase. The *galE* gene and mid-region of the *vacA* gene were amplified using primer pairs P1 / P2 and VAG-F / VAG-R, respectively (Table 3.1). A 645-bp PCR product indicates the presence of *vacA* mRNA transcripts, and a 461-bp PCR product indicates the presence of *galE* transcripts. Lanes 1 and 3 represent the negative controls (no reverse transcriptase) for isolates 577C and 577A, respectively.

1 2 3 4

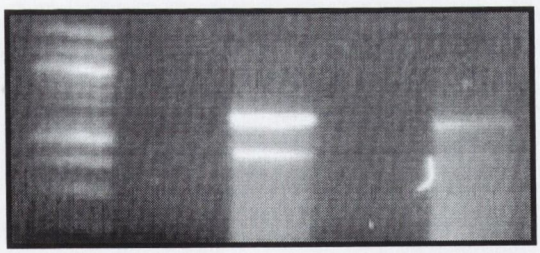
(A)



190 bp

1 2 3 4

(B)



645 bp *vacA*

461 bp *galE*

Negative controls where, RT-PCR reactions were carried out without reverse transcriptase, were used to confirm that the amplified products were generated from cDNA and not genomic DNA. These results demonstrated that all isolates investigated transcribed the *vacA* gene.

5.3 Discussion.

5.3.1. Genotyping of paired isolates at 5 different loci.

As mentioned in Chapter 3 the *vacA* locus presents as a mosaic structure (Atherton *et al.*, 1995). The frequencies of the *vacA* alleles that occurred in the paired isolates studied in this Chapter were almost identical to the frequencies reported for the single isolates (see Table 3.3). Interestingly, no allelic variation at the *vacA* locus was detected within each pair (Table 5.1). By itself this would suggest that each pair used in this study represented single strains as previous reports have shown that pairs of *H. pylori* isolates displaying *vacA* allelic variation represent genetically unrelated isolates colonising the same host (Kuipers *et al.*, 2000, Owen and Xerry, 2003).

Similarly, no differences were detected between the motifs or sequence variants occurring at the 3′-end of the *cag* PAI or the sequences preceding the *vacA* and *ribA* genes within each pair (Table 5.1). The fact that each pair displays conserved alleles, in-del motifs and sequence variants at five different loci strongly suggested that each pair represented a single strain colonising either both the antrum and the corpus or two different sites in the antrum.

5.3.2 cagA *genotyping of paired isolates.*

The *cagA* status was determined in all isolates. Similarly to the single isolates previously studied each isolate within each pair contained the *cagA* gene. However, analysis of the 3′-end of the *cagA* gene demonstrated that variation in the number of repeat sequences occured within nine pairs (Fig. 5.1). Variation in the repeat number has been shown to vary in single isolates in this (Chapter 3) and others studies (Evans *et* al., 1998, Yamaoka *et al.*, 1999a, 2000). For eight out of the nine isolate pairs (89%) investigated one isolate had a difference of one repeat to its corresponding isolate. In the case of isolate pair 15500A / 15500C the difference was two repeats between isolates (Table 5.1). Five out of these nine isolates pairs represented antrum and corpus isolates. For four of the latter pairs the corpus isolate had the extra repeat sequence. Subsequently, isolates within three of these four pairs (45688A / 45688C, 15500A / 15500C, 87212A / 87212C) has identical RAPD-PCR fingerprints and were ancestrally related by FAFLP-PCR (Table 5.1, Fig. 5.3). As genetic differences were discovered within nine of the 31 paired isolates it was reasonable to assume, based on a previous report (Kuipers *et al.*, 2000), that each pair represented two genetically unrelated strains inhabiting the same individual.

5.3.3. RAPD-PCR analysis of paired H. pylori *isolates.*

RAPD-PCR analysis is a reproducible fingerprinting technique that has been used to differentiate *H. pylori* strains (Marshall *et al.*, 1995), to detect the diversity within a *H. pylori* population from a certain geographical region (Chalkauskas *et al.*, 1998; Gzyl *et al.*, 1999), and to examine the development of a sub- or quasi-species in *H. pylori* isolates emerging during colonisation of the same host (Kersulyte *et al.*, 1999; Kuipers *et al.*, 1999; Kuo *et al.*, 1999; Israel *et al.*, 2001b). RAPD-PCR analysis was applied to all 31 paired isolates in this study. On the basis of their DNA fingerprints, 22 pairs were shown to be genetically related, while nine were genetically unrelated. Interestingly, the nine unrelated pairs did not encompass the nine pairs (Section 5.3.2) that displayed variations in the repeat number at the 3′-end of the *cagA* gene. Of the 22 pairs of isolates that appeared to be identical on the basis of genotyping at six different loci, five pairs displayed different RAPD-PCR fingerprints indicating that the isolates within these pairs were unrelated. Thus, genotyping at multiple loci alone does not necessarily prove relatedness or unrelatedness.

5.3.4. Southern blotting of related and unrelated H. pylori *isolates.*

Previous studies have shown that RFLP patterns differ between unrelated *H. pylori* isolates (Akopyanz *et al.*, 1992; Vos *et al*, 1999). The RFLP patterns of a representative ancestrally related pair of *H. pylori* isolates, as evidenced by RAPD-PCR, were investigated. As expected each isolate within this pair displayed identical RFLP patterns for each gene probe used. In addition, the RFLP profiles of an unrelated pair of *H. pylori* isolates (by RAPD-PCR) hybridised with a *vacA* probe displayed dissimilar patterns, confirming their non-relatedness. This result is interesting as both isolates within this unrelated pair exhibited identical *vacA* alleles, and yet there is some sequence diversity present, as demonstrated by Southern blotting. This finding further confirms the observation made by RAPD-PCR fingerprinting that genotyping paired isolates at multiple loci is not sufficient to differentiate unrelated *H. pylori* isolates.

5.3.4. Sequence analysis of paired H. pylori *isolates.*

As *H. pylori* is a genetically diverse organism it is very rare to find two isolates with genes with identical nucleotide sequences (Achtman *et al.*, 1999). In the case of paired isolates MI355 and MI356 the nucleotide sequences of the *napA* gene in each isolate were identical demonstrating that they represent one strain. As the paired isolates MI355 and MI356 display differences in repeat number at the 3′-end of the *cagA* gene and yet exhibit identical RAPD-

PCR fingerprints, it is safe to extrapolate that the other four pairs of isolates that share this characteristic represent a single strain even although they display 3′-end repeat number differences in the *cagA* gene.

5.3.5. FAFLP-PCR analysis of paired H. pylori *isolates.*

Since five of the 22 paired *H. pylori* isolates exhibited genetic differences within each pair and yet RAPD-PCR analysis demonstrated identical or almost identical fingerprints, a more sensitive fingerprinting technique (FAFLP-PCR) was applied to all 31 paired isolates to determine the genetic relatedness of each pair.

FAFLP-PCR generated a unique fingerprint for each isolate. However, upon visual inspection of the fingerprints run on the ABI 377 sequencer, the differences in profiles were difficult to see. Therefore, it was easier to analyse the differences in fingerprint profiles by examining the electropheorograms generated by both sequencing machines (ABI 377 and ABI 100; Chapter 3). Due to the fact that *H. pylori* is a diverse organism (Go *et al.*, 1996, Salaun *et al.*, 1998) it is unlikely that two unrelated isolates would display similar profiles. Therefore, although 16 of the 22 paired isolates investigated by FAFLP-PCR in this study displayed minor differences between isolates, the nearly identical profiles indicate that each isolate within a pair is derived from the same progenitor strain. The profiles of the paired isolates demonstrated the resolving capability of the FAFLP-PCR technique as with some pairs the profiles of each isolate only showed minor differences when using one selective primer (Fig. 5.10 A). Herein lies the benefit of the FAFLP-PCR technique as it usually generates four different fingerprints for each isolate. With four fingerprints the likelihood of detecting minor genetic differences between paired *H. pylori* isolates is increased. Unfortunately, the + T selective primer did not yield fingerprints for any of the isolates analysed. This could be because there were no fragments that could be amplified with this primer. However, this seems unlikely as *H. pylori* has an AT-rich genome (61%, Alm *et al.*, 1999). The fact that the + T selective primer did not work is more likely to be because each reaction was carried out in a multiplex PCR and the other primer (+ C) used in the same reaction overwhelmed the + T primer which was unable to compete and amplify any fragments. Regardless, the three selective primers used were sufficient to detect the minor genetic differences between the 16 ancestrally related *H. pylori* isolates that were difficult to detect by RAPD-PCR analysis.

5.3.6. Phylogenetic analysis of paired H. pylori *isolates.*

Although minor genetic differences were detected between paired *H. pylori* isolates, it was not possible to quantify the amount of divergence between ancestrally related isolates based on visual inspection of FAFLP-PCR profiles. Thus, it was necessary to construct a phylogenetic tree based on the fingerprints of each isolate, an approach previously used to assess sub-species development between *H. pylori* isolates emerging over time (Kuipers *et al.*, 1999). The Genotyper^{TM} software is capable of converting each peak on the electropherogram of each isolate into a binary format (peak present = 1, no peak present = 0). The result is that a unique binary sequence is obtained for each isolate that is representative of that isolates fingerprint. The binary sequences for all three selective primers were combined. This sequence was then converted into a nucleotide format (1 = G, 0 = C) in order to enable a tree to be created (Fig. 5.11), a system previously used to investigate the diversity between strains of *Pseudomonas aeriginosa* (Ahmed *et al.*, 2002) and *Mycobacterium tuberculosis* (Ahmed *et al.*, 2003a, 2003b).

Although all isolates displayed measurable differences to each other, 14 pairs of isolates from individuals grouped together, while the isolates 577A / 577C and 9613A / 9613C, from two members of the same family, clustered together in one group (Fig. 5.11). Thus, strains that originated from the same patient were more similar to each other than strains from different patients in the majority of cases. The differences between related isolates varied from pair to pair (Fig. 5.11). The node at each branch between a pair of isolates is the point at which each isolate in a pair share 100% identity in their fingerprint profiles. The further away an isolate is from this point, the more differences in fingerprint profiles an isolate has to its corresponding paired isolate. Some ancestrally related paired isolates display more differences than others. However, there was no association between the degree of microevolution and whether the isolates within pairs were antrum and corpus or both antrum isolates. There was also no association between the degrees of microevolution observed for pairs of isolates and whether they were taken at the same endoscopic session or obtained some time apart.

It is difficult to determine the cause of these genetic differences between ancestrally paired *H. pylori* isolates. There are some possible reasons to why fragments may vary between the fingerprints of paired isolates. Primarily the causes of these differences are probably due to primers not annealing to the template DNA. Primers may not anneal due to point mutations, insertions and deletions [a common occurrence in *H. pylori* (Alm *et* al., 1999)] at the primer

binding sites. Secondly, if excision or insertion of DNA sequences occurs between the annealing sites for two primers in one isolate and not the other, the same region of DNA will be amplified from both isolates but each isolate will generate a different size fragment.

It has been possible to determine the exact nature of the differences in FAFLP-PCR profiles within *Mycobacterium tuberculosis* isolates *in silico* (Ahmed *et al.*, 2002). This requires the artificial digestion of the entire genome of *M. tuberculosis* with *Mse*I and *Eco*RI and determining what size and how many fragments will be generated in a fingerprint. However, this *in silico* strategy of determining the differences in FAFLP-PCR profiles will only work with clonal bacteria and thus is not possible with such a genetically diverse organism as *H. pylori*.

5.3.7. Analysis of familial isolates.

With the exception of the in-del motifs, the genotypes of the father's isolates (F610 / F611) were identical to one of his siblings (577A / 577C) within Family 1, whereas the other sibling within this family displayed differences in the *vacA* signal sequence and the DNA sequence preceding the *vacA* gene. Interestingly, RAPD-PCR and FAFLP-PCR fingerprinting demonstrated that this father and his two children were infected by different *H. pylori* strains, and that the two siblings were infected by the same strain, indicating transmission of isolates between siblings. This result again demonstrates the inability of multiple locus typing of genes to differentiate strains. Likewise in Family 2 isolate pairs 2187A / 218C (mother) and 222A / 222C (daughter's spouse) shared similar multiple locus types (with the exception of in-del type). However, these individuals were shown subsequently by RAPD-PCR analysis to be infected not only by two different strains, but also the daughter's spouse was infected with two different strains.

An interesting finding is that isolate pairs 577A / 577C and 9613A / 9613C from Family 1 both display the same RAPD-PCR and FAFLP-PCR fingerprints and yet display different signal sequence and upstream variants of the *vacA* gene. This strongly suggests that a recombination event occurred upstream of the *vacA* gene and within its 5′-end in one of these strains leading to the conversion of a VacAP3 / s1a genotype to a VacAP4 / s2 genotype or *vice versa*. Previously evidence for the emergence of *cag* recombinant strains during human infection has been presented by Kersulyte *et al.* (1999). Moreover, Owen and Xerry (2003) demonstrated that clonal variants within two members of the same family can exhibit different

mid-region genotypes. This finding represents the first report of a change in 5'-*vacA* genotype characteristics during infection. Currently it is not possible to be certain during at which stage of transmission this event might have occurred, i.e., parent to child, parent to parent, or sibling to sibling. These observations are, however, consistent with the recent data of Akade *et al.* (2003) on how genetic divergence may affect tissue tropism. The clustering of these paired isolates possibly indicates transmission of an antral *H. pylori* infection between siblings (Fig. 5.11). Perhaps after antral colonisation the 577A and 9613A populations adapted independently to the corpus region of their individual hosts. The transmission of isolates within both families will be analysed further in Chapter 7.

5.3.8. Phenotypic analysis of paired H. pylori *isolates.*

As all ancestrally related paired *H. pylori* isolates that were analysed by FAFLP-PCR were shown to exhibit microevolution, the presence or absence of the VacA toxin was examined in each pair to determine if paired isolates could exhibit phenotypic as well as genetic variation. Variation in expression of the VacA toxin was observed within four pairs of isolates. However, only two of these pairs were found to be ancestrally related pairs. VacA variation within a pair of *H. pylori* isolates that have identical RAPD-PCR profiles demonstrates a modification event that has led to the regulation of an important virulence factor. The presence or absence of VacA in *H. pylori* isolates was examined in whole cell lysates of cells grown on solid media, and not on whole cell lysates or supernatants of cells grown in broth. As VacA is a secreted toxin (Forsyth *et al.*, 1998) it is important to bear in mind that the relative absence of VacA within isolates 577A and 45688C (Fig. 5.12) may be due to the toxin being secreted into the medium, as VacA would not be a component of the whole cell lysates and thus not be detected by Western blotting. Also, the epitope that is recognised by the anti-VacA antibody may have been changed in isolate 577A due to mutations affecting the conformation of the toxin resulting in a toxin that could not be detected in the immunoblot. Lastly, these variations in VacA levels could possibly be due to a lack of transcription of the *vacA* gene because of an upstream mutation that affects transcriptional levels or a mutation within an as yet unknown repressor protein that regulates the transcription of the *vacA* gene (Forsyth *et al.*, 1998).

5.3.9. Transcriptional analysis of paired H. pylori *isolates that display phenotypic differences.*

To investigate whether the *vacA* gene was being transcribed within paired isolates 577A / 577C and 45688A / 45688C RT-PCR was applied to total cellular RNA from each isolate. The signal sequences (577A / 577C) and mid-regions (45688A / 45688C) of *vacA* transcripts were amplified to detect the presence of *vacA* mRNA. The VacA toxin was detected in isolates 45688A / 45688C at different levels. Therefore transcription should be taking place in order to produce this toxin. Just as would be expected both isolates were found to transcribe the *vacA* gene.

In the case of isolates 577A / 577C, it was found that the VacA toxin was detected in one isolate and not the other. To determine whether the *vacA* gene was transcriptionally regulated, total RNA samples from these isolates were investigated to see if they contained *vacA* mRNA. Both isolates were found to contain *vacA* mRNA indicating that this gene was being transcribed. Thus the cause of the lack of VacA toxin in isolate 577A was not transcriptional. Hence, it is likely that the VacA toxin in isolate 577A was not being translated, or was being secreted more efficiently compared to the toxin in isolate 577C.

5.3.9. Model for microevolution.

Initial colonisation of the gastric mucosa by a founder *H. pylori* strain leads to persistent infection (Marshall *et al.*, 1998). As a founder strain spreads throughout the stomach it encounters different selection pressures. Thus to survive and to meet these challenges *H. pylori* needs to able generate genetic diversity within a bacterial population (Taddei *et al.*, 1997) and to adapt to the new environment it encounters on entering a new host.

Factors that may impact on adaptation are (i) pH levels within different gastric niches and their influence on the survival of *H. pylori* and its ability to grow under acidic conditions (Karita & Blaser, 1998; Bijlsma *et al.*, 2000), (ii) differences in the distribution of mucin glycoforms in the gastric mucosa (Nordman *et al.*, 2002), (iii) differences in the viscosities of mucins and their influences on H^+, HCO_3^- and CO_2 diffusion (Tanaka *et al.*, 2002), (iv) differences in the biosynthesis of prostaglandins and hydroxyeicosatetraenoic acids in the regions of the stomach (Park *et al.*, 2003), (v) differences in the binding of *H. pylori* to gastric mucins in a pH-dependent manner (Nordman *et al.*, 1999), and (vi) differences in host genotype and physiology affecting niche characteristics (Akada *et al.*, 2003). It has also been

demonstrated that generation of microevolved or clonal variant strains can result in selection of antibiotic-resistant (clarithromycin) strains (Owen *et al.*, 1999). Alterations within the founder strain's DNA could potentially be selected for by these varying environmental factors found in different regions of the stomach. This would result in two populations colonising different regions of the stomach but with highly similar genomes yet displaying minor genetic differences.

Chapter 6

Intrafamilial spread of H. pylori: *A study of two families exhibiting clonal transmission of* H. pylori.

6.1. Introduction 110

6.2. Results 113

6.1.1. Analysis of the vacA *mosaic structure and the number of 3'-end* cagA *repeats within* H. pylori *isolates from two separate families* 113

6.2.1.2. Family 1 113

6.2.1.3. Family 2 113

6.2.2. RAPD-PCR analysis of H. pylori *isolates from two separate families* 113

6.2.2.1. Family 1 114

6.2.2.2. Family 2 114

6.2.3. FAFLP-PCR analysis of H. pylori *isolates from two separate families* 115

6.2.3.1. Family 1: FAFLP-PCR fingerprinting and phylogenetic analysis 115

6.2.3.2. Family 2: FAFLP-PCR fingerprinting and phylogenetic analysis 116

6.3. Discussion 117

6.3.1. Genotyping of familial H. pylori *isolates at three different loci* 117

6.3.2. Genomic fingerprinting of H. pylori *isolates from two separate families* 117

6.3.3. Clonal transmission of H. pylori *within families* 118

6.3.4. Clonal transmission and clinical outcome within a family unit infected with H. pylori 119

6.1. Introduction.

As previously mentioned, *Helicobacter pylori* is a panmictic organism. However, recent reports (and the report herein) indicate that this organism may be weakly clonal. In spite of *H. pylori's* controversial population structure the fact remains certain that *H. pylori* is an extremely genetically diverse organism. This trait allows different strains infecting unrelated individuals to be discriminated using DNA fingerprinting techniques (Clayton *et al.*, 1993; Owen *et al.*, 1994; Atherton *et al.*, 1995; Marshall *et al.*, 1995, 1996; Gibson *et al.*, 1998; Achtman *et al.*, 1999; Han *et al.*, 2000; Owen *et al.*, 2001; Li *et al.*, 2002b). This characteristic has also been useful when attempting to determine the pathway of transmission of *H. pylori* isolates.

In an effort to uncover how *H. pylori* transmission occurs researchers have analysed isolates from members of the same family, as transmission of other organisms has previously been shown to occur within the family unit (Genco and Loos 1991). Primarily analysis of familial *H. pylori* infections has been carried out using sero-positivity and urea breath tests. Using sero-positivity testing a population survey showed that *H. pylori* infection clusters within families belonging to the same population (Dominici *et al.*, 1999). This implies that either person-to-person transmission or a common source of exposure is responsible for *H. pylori* infection. In addition, Tindberg *et al.* (2001) demonstrated a lack of evidence for extraneous child-to-child transmission outside the family and suggested the importance of mother to child transmission of *H. pylori* infection. It has also been reported that index patients and their mothers had the same antibody profile in six different families, again supporting the idea of a mother-to-child transmission route (Elistur *et al.*, 1999). In contrast, Goodman *et al.* (2000) showed that transmission occurred mainly between siblings and not via parents. A study investigating common antibodies between a mother and her child demonstrated that anti-*H. pylori* antibodies can be passed on from the mother to the child (Blecker *et al.*, 1994). This phenomenon may complicate diagnosis and transmission analysis, as most of the previous reports demonstrating mother-to-child transmission were based on sero-positivity tests. Thus, more reliable techniques were needed for familial analysis of *H. pylori* strains.

DNA fingerprinting of *H. pylori* isolates from family members can be more useful as it provides precise information about transmission of specific strains between individuals, due to different *H. pylori* strains exhibiting distinct fingerprints. Although, molecular typing of *H. pylori* isolates obtained from members of the same family is an accurate and reliable

technique for determining the route of transmission within a family, the pathway of transmission has not yet been defined. However, DNA fingerprinting has yielded more informative results as father-to-child transmission has been shown to be just as important as mother-to-child transmission (Bamford *et al.*, 1993; Wang *et al.*, 1993), whereas epidemiological and seropositivity studies emphasise the importance of mother-to-child transmission alone. This finding has been confirmed by more recent report (Shimizu *et al.*, 2000). Taneike *et al.* (2001) demonstrated, via ribotyping, that the fingerprint of an *H. pylori* isolate taken from a child displayed the same fingerprint profile as that of his father. Moreover, after eradication of the child's infection he was re-infected with the strain that his mother was carrying. In accordance with this finding studies in Germany report the importance of father and mother to child transmission (Brenner *et al.*, 1998; Rothenbacher *et al.*, 1999). This indicates that both mother and father to child transmission can occur and that both are equally important routes. However, Goodman *et al.* (2000) reported the importance of transmission via siblings from the same family.

Oral carriage of *H. pylori* has been reported (Dowsett *et al.*, 1999). With the mouth acting as a reservoir for *H. pylori*, it is possible to see how infection can be spread within families. However, there are also studies that provide evidence of hosts acquiring infections from water (Karita *et al.*, 2003) and animal sources (Papiez *et al.*, 2003). Water and animal sources of transmission were reported in Japanese and Polish studies, respectively (Karita *et al.*, 2003; Papiez *et al.*, 2003). Therefore, it seems quite plausible that transmission of *H. pylori* within different parts of the world occurs via different routes. As the manners and customs associated with each region change so too might the pathway of *H. pylori* transmission.

Although population genetic analysis by multi-locus sequence typing (MLST) has demonstrated that *H. pylori* has a non-clonal structure (Enright and Spratt 1999), analysis of familial isolates has demonstrated that this organism can be clonal over a short period of time in Germany (Suerbaum *et al.*, 1998; Han *et al.*, 2000) and Japan (Malaty *et al.*, 2000). The lack of diversity found between the two sequenced genomes of *H. pylori* strains J99 and 26995 (Alm *et al.*, 1999) and its clonal behaviour over short periods of time suggest that a large array of clonal *H. pylori* lineages may be coexisting and may be evolving in isolation from one another. As previously shown in Chapter 5, two paired *H. pylori* isolates from children of the same family displayed unusual clustering that implied sibling transmission, as revealed by FAFLP-PCR fingerprinting. FAFLP-PCR fingerprinting was applied to two sets

of familial *H. pylori* isolates to determine the route of transmission of this organism within Irish families.

6.2. Results.

Nineteen *H. pylori* isolates obtained from two unrelated families were analysed in this study. Twelve of these isolates belonged to Family 1 and seven belonged to Family 2 (Table 6.1). In Family 1 three isolates were obtained from the mother, two from the father, and seven from four of their children. Isolates from Family 2 comprised two from the mother, two from her son, one from her daughter and two from the daughter's spouse.

6.2.1. Analysis of the vacA *mosaic structure and the number of 3'-end* cagA *repeats within* H. pylori *isolates from two separate families.*

6.2.1.2. Family 1.

As carried out in previous chapters the 's' and 'm' types and number of 3'-end *cagA* repeats were determined in the same manner for the *H. pylori* isolates within this family. Ten (83.3%) of the 12 family isolates exhibited an s1a m2 *vacA* type and 2 (16.7%) exhibited an s2 m2 *vacA* type (Table 6.2). Ten (83.3%) of these isolates exhibited 2 repeats at the 3'-end of the *cagA* gene, whilst 2 (16.7%) displayed 6 repeats (Table 6.2). All the mother's isolates shared the same *vacA* type (F561, F562 and F563; s1a m2), but isolates F562 and F563 had six repeat sequences at the 3'-end of the *cagA* gene, whereas isolate F561 had two. All the other paired isolates within this family displayed no variation in *vacA* alleles or *cagA* 3'-repeat number within. The *vacA* type of the isolates from one of the siblings differed from that of the rest (9613A / 9613C; s2 m2).

6.2.1.3. Family 2.

The *vacA* type and number of 3'-end *cagA* repeats were determined within all isolates of Family 2. Five (71.4%) of the seven family isolates were of *vacA* type s1a m2 and two (28.6%) were of *vacA* type s2 m2 (Table 6.2). Five (71.4%) of the seven family isolates contained two repeats sequences at the 3'-end of the *cagA* gene, whilst the other two contained one repeat sequence. All paired isolates within Family 2 displayed no variation in *cagA* 3'-repeat sequence number or *vacA* genotype within each pair. The *vacA* type of the son's isolates was different from the other family members (605A / 605C; s2 m2).

6.2.2. RAPD-PCR analysis of H. pylori *isolates from two separate families.*

RAPD-PCR fingerprinting was carried out on all 19 isolates within both families using two different primers individually (PIII and PIV for Family 1, with the exception of isolates F610

Table 6.1. Familial *H. pylori* isolates used in this study.

	Isolate	Date of isolation	Area of isolation	Clinical symptom	Family member
Family 1:					
	F561	Nov 1997	A		Mother
	F562	Nov 1997	C	ASY[a]	
	F563	Nov 1997	A		
	F610	Jan 1998	A	ASY[a]	Father
	F611	Jan 1998	C		
	577A	May 1998	A	DU[b]	Sibling
	577C	May 1998	C		
	9613A	May 1998	A	ASY[a]	Sibling
	9613C	May 1998	C		
	F431	April 1997	C	DU[a]	Sibling
	98-018	Jan 1998	A	ASY[a]	Sibling
	F613	Jan 1998	C		

Family 2:					
	218A	Aug 1998	A	ASY[a]	Mother
	218C	Aug 1998	C		
	605A	May 1998	A	ASY[a]	Son
	605C	May 1998	C		
	221C	Aug 1998	C	ASY[a]	Daughter
	222A	Aug 1998	A	ASY[a]	Daughter's
	222C	Aug 1998	C		spouse

[a]ASY-asymptomatic.
[b]DU-duodenal ulcer.

Table 6.2. The *vacA* types and the numbers of *cagA* 3′-end repeat sequences indicated of familial isolates used in this study.

	Isolate	*vacA* type	Number of *cagA* 3′ repeats
Family 1:			
	F561	s1a m2	2
	F562	s1a m2	6
	F563	s1a m2	6
	F610	s1a m2	2
	F611	s1a m2	2
	577A	s1a m2	2
	577C	s1a m2	2
	9613A	s2 m2	2
	9613C	s2 m2	2
	F431	s1a m2	2
	98-018	s1a m2	2
	F613	s1a m2	2
Family 2:			
	218A	s1a m2	2
	218C	s1a m2	2
	605A	s2 m2	2
	605C	s2 m2	2
	221C	s1a m2	2
	222A	s1a m2	1
	222C	s1a m2	1

/ F611 where the combination of RAPD primers PII and PV were used; PI and PIII for Family 2, see Table 2.3, Chapter 2). RAPD primers PIV and PI were used exclusively on families 1 and 2, respectively, as each primer yielded more informative fingerprints for the familial isolates with which they were used.

6.2.2.1. Family 1.

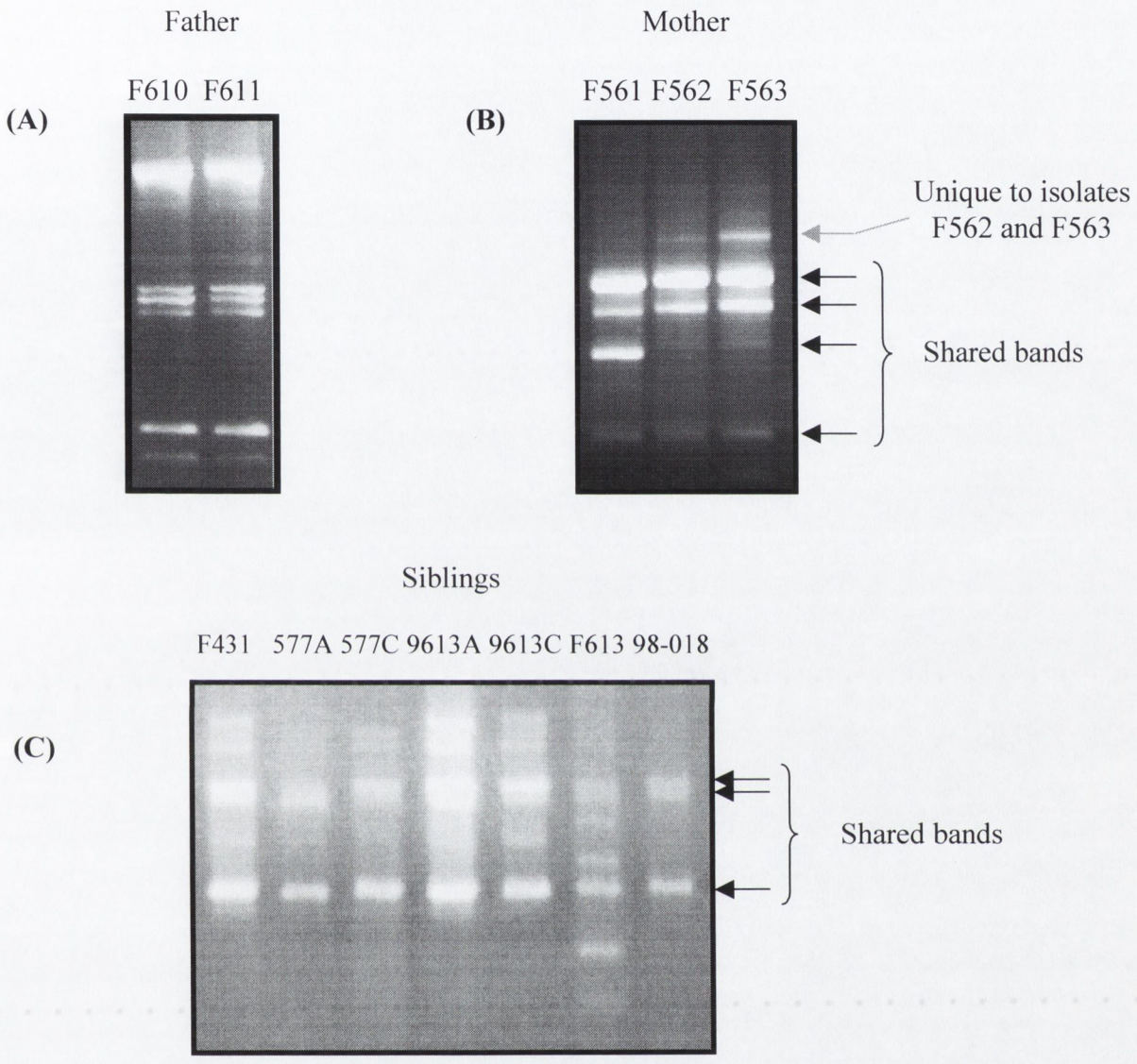
RAPD-PCR fingerprinting revealed that three different strains were infecting the members of Family 1. Figs. 6.1 and 6.2 displays the RAPD fingerprints using primers PIII and PIV, respectively, for each isolate within this family. The father's paired antral isolates (F610 / F611) exhibit identical profiles and are distinct to the mother's and siblings isolates. The mother's three antral isolates (F561 / F562 / F563) exhibited similar profiles and were distinct from the father's and siblings' isolates. Finally, the isolates form each sibling displayed very similar profiles that were distinct from the mother's and father's profiles. Each of the four siblings is therefore infected with a related strain, indicating transmission within siblings. Although isolate 577C displayed highly similar profiles to all the other sibling isolate profiles, this isolate lacked an amplified fragment within its RAPD-PCR fingerprint (using RAPD primer IV, Fig. 6.2) that was conserved within the fingerprints of the other sibling isolates. The absence of this band was reproducible upon repeated fingerprinting attempts. This demonstrates a genomic variation between paired isolates displaying highly similar fingerprint profiles. This is in accordance with the finding that these isolates (577A/ 577C and 9613A / 9613C) also exhibited sequence diversity at the start of and preceding the *vacA* gene dictating that these ancestrally related isolates were different (see Chapter 5, Sections 5.2.1 and 5.2.2).

6.2.2.2. Family 2.

RAPD-PCR fingerprinting revealed that four different strains were infecting the members of this family. Figs. 6.3 and 6.4 displays the RAPD fingerprints using primers PI and PIII, respectively, for each isolate within this family. The mother's paired antrum and corpus isolates (218A / 218C) and daughter's corpus isolate (221C) show identical fingerprints, demonstrating transmission of *H. pylori* between mother and daughter. Interestingly, the fingerprint of isolate 218A (using RAPD primer PI) had an extra band that was not present in isolates 218C and 221C, which makes these two isolates more alike each other than 218A. The son's paired antrum and corpus isolates (605A / 605C) possessed identical fingerprints that were distinct from any other isolate's fingerprint within this family (data not shown). The

Fig. 6.1. RAPD-PCR fingerprints of *H. pylori* isolates from Family 1 using primer PIII fractionated on a 1.8% agarose gel.

It can be seen from these fingerprints that three different *H. pylori* strains are colonising this family. The father harboured one strain **(A)**, the mother also harboured one strain **(B)**, and the siblings harboured strains that were very similar **(C)**. The mother's three isolates shared many amplified fragments in common, although some of these were weak. However, there is a band absent in the profile of isolate F561 that is present in both isolates F563 and F563. All the sibling isolates display RAPD-PCR patterns quite similar to each other, indicating that each sibling is infected with the same strain.


Father
F610 F611
(A)
Mother
F561 F562 F563
(B)
Unique to isolates
F562 and F563
Shared bands
Siblings
F431 577A 577C 9613A 9613C F613 98-018
(C)
Shared bands

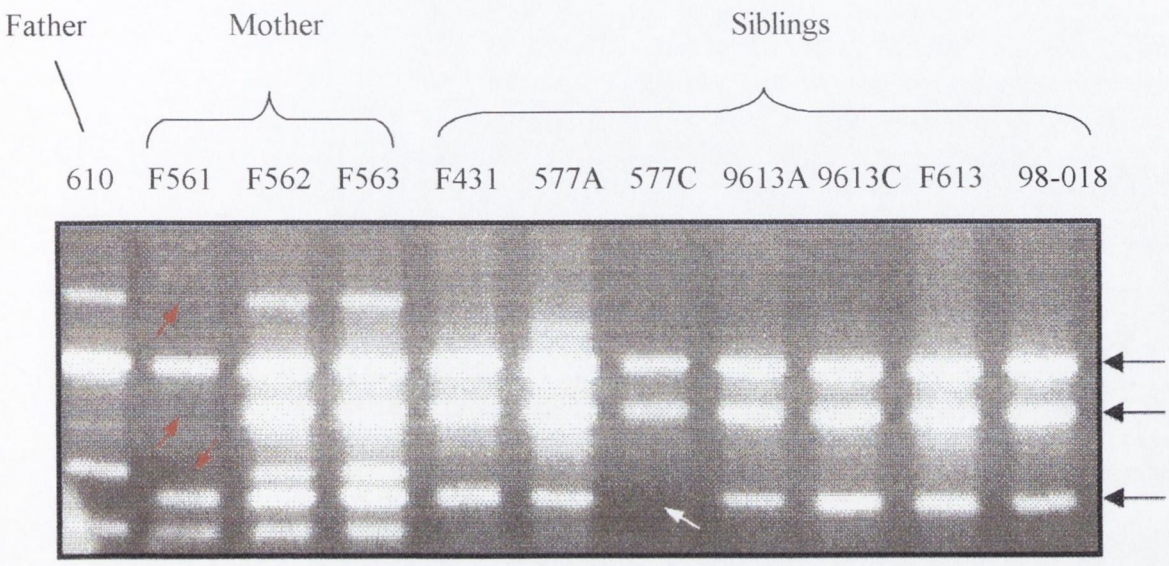


Fig. 6.2. RAPD-PCR fingerprints of *H. pylori* isolates from Family 1 using primer PIV fractionated on a 1.8% agarose gel.

One of the father's isolates (F611) was not included. The fingerprints of the siblings isolates show conserved bands (black arrows) demonstrating that a similar strain has infected all the children. However, isolate 577C lacks a band that is conserved throughout the fingerprints of the other siblings (white arrow). The fingerprints from the mother's three isolates indicate that a she is infected with three similar isolates. However, there are some bands missing in isolate F561 that are conserved between isolates F562 and F563 (red arrows).



Fig. 6.3. RAPD-PCR fingerprints of *H. pylori* isolates from Family 2 using primer PI fractionated on a 1.8% agarose gel.

The above fingerprints demonstrate that the mother and daughter in this family share a similar strain. The daughter's spouse harbours isolates with two different fingerprints (mixed infection). The fingerprint of isolate 605A is distinct from that of the other isolates (isolate 605C not shown). An extra band present in isolate 218A can be seen (white arrow). The red arrow indicates an amplified band that is present in isolates 218A / 218C but not present in isolate 221C.



Fig. 6.4. RAPD-PCR fingerprints of *H. pylori* isolates from Family 2 using primer PIII fractionated on a 1.8% agarose gel.

These fingerprints of the mother's and daughter's isolates are similar. The son's isolate is distinct for the other isolates. The isolates from the daughter's spouse display a common pattern.

daughter's spouse was infected with two different strains from the antrum and corpus (222A / 222C, mixed infection). Only one band was capable of being amplified within these isolates using RAPD primer PIII, indicating the necessity when applying RAPD fingerprinting to use more than one primer. These data indicate that the mother infected the daughter, or *vice versa*.

6.2.3. FAFLP-PCR analysis of H. pylori *isolates from two separate families.*

FAFLP-PCR fingerprinting was carried out on all 19 isolates from the two families using three selective primers (+ A, + G, + C). However, isolates F610 and F611 (father), F562 (mother) and 98-018 (sibling) from Family 1 did not yield FAFLP-PCR fingerprints for any of the three selective primers used in this study, despite repeated attempts, and were not included in the analysis. A Neighbour-joining phylogenetic tree was constructed for each family. Rectangular cladogram and radial phylogenetic trees were constructed in order to display different perspectives of the relationships of each isolate within each family.

6.2.3.1. Family 1: FAFLP-PCR fingerprinting and phylogenetic analysis.

The rectangular cladogram and radial phylogenetic trees of eight isolates from Family 1 are shown in Figs. 6.5 and 6.6, respectively. These two trees confirm the findings from RAPD-PCR fingerprinting, that two different *H. pylori* strains infected the mother and her children. However, as FAFLP-PCR fingerprinting is a more sensitive technique than RAPD-PCR and can detect microevolution between related *H. pylori* strains, a more precise depiction of the relatedness of each familial isolate is presented. As mentioned in Chapter 5, paired isolates 577A / 577C and 9613A / 9613C clustered together when phylogenetically compared to other paired isolates (Fig. 5.11). In this case the antral isolates clustered together and the corpus isolates clustered together yet remained separate from the antral isolates. This implied that the two antral isolates from these siblings (577A and 9613A) were more like each other than their corresponding paired corpus isolates. Likewise, the siblings' corpus isolates (577C and 9613C) were more like each other than their corresponding antral isolates. However, phylogenetic analysis of these isolates in relation to other isolates from the same family showed that they clustered differently (Figs. 6.5 and 6.6). In this case the two isolates from each sibling (antrum and corpus) clustered together in relation to the sibling they were obtained from (i.e., 577A clustered with 577C, and 9613A clustered with 9613C, Figs. 6.5 and 6.6). As the same fingerprints and phylogenetic program were used, the difference in clustering between the isolates from these two siblings must be related to the isolates they were being compared to. Nevertheless, it can be concluded that these isolates are related and

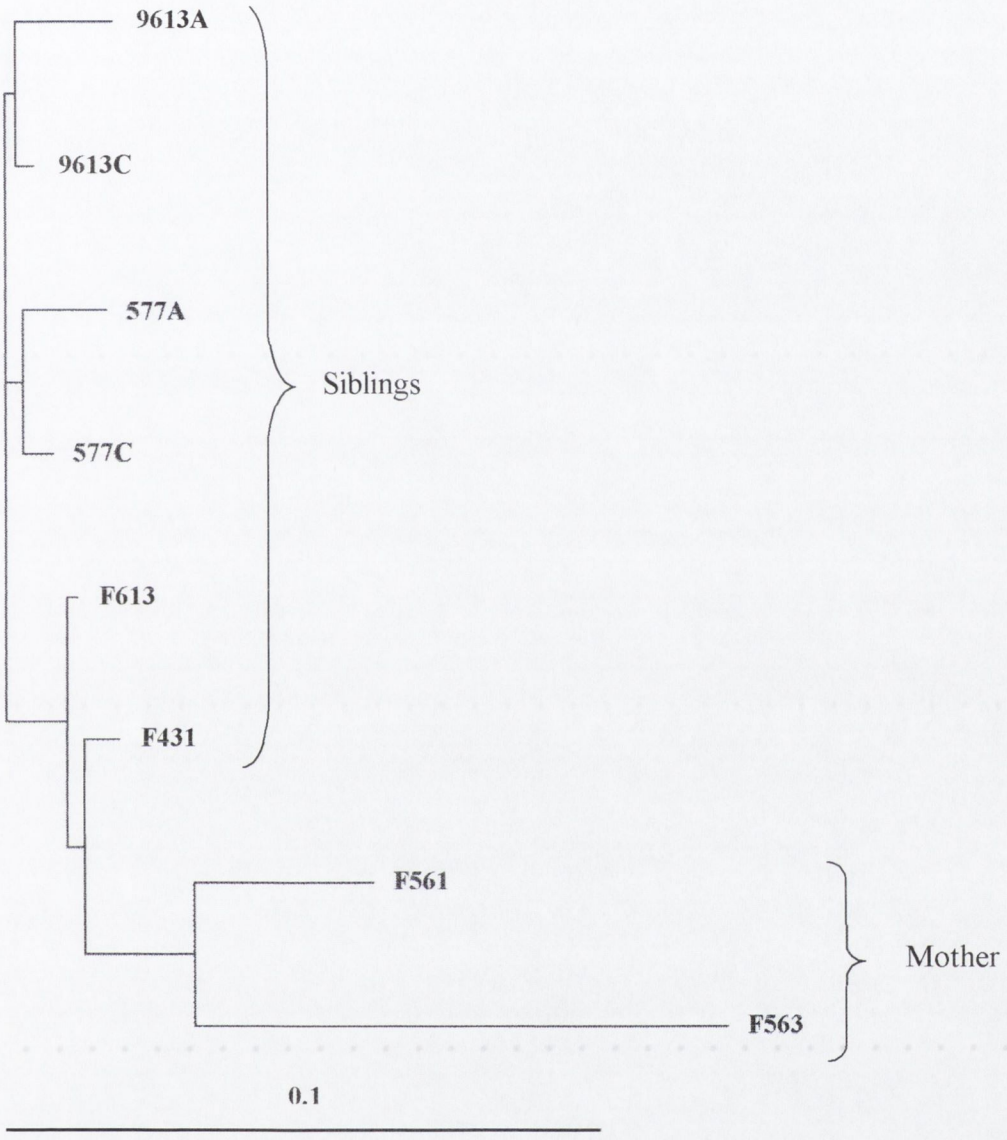


Fig. 6.5. Rectangular cladogram neighbour-joining phylogenetic tree of *H. pylori* isolates from Family 1 based on similarities of FAFLP-PCR fingerprints.

This tree displays the relatedness of each isolate within this family based on FAFLP-PCR fingerprints. The father's isolates (F610 / F611) are not included. The bottom scale displays the number of fingerprint band substitutions per site.

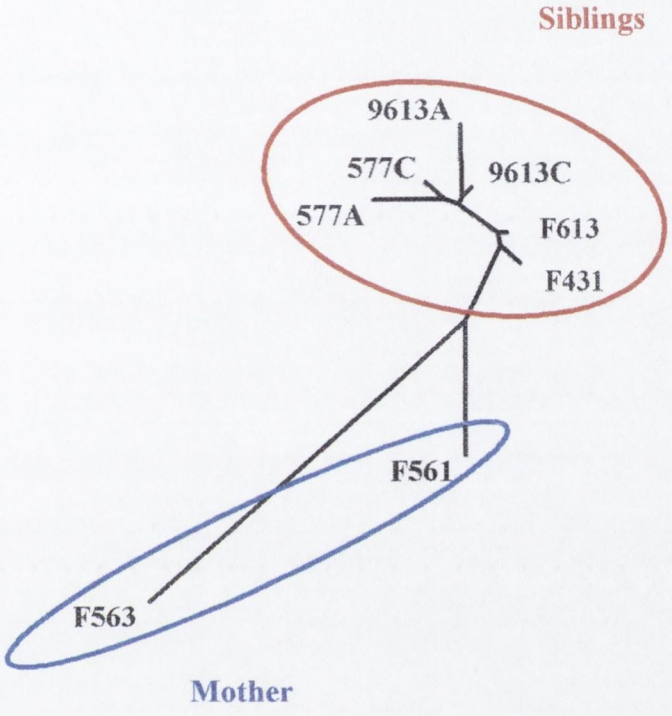


Fig. 6.6. Radial neighbour-joining phylogenetic tree of *H. pylori* isolates from Family 1 based on similarities of FAFLP-PCR fingerprints.

This style of tree depicts the relationship between the isolates from the mother and her children in a clearer form. The siblings' isolates (red oval) are distinct from the mother's isolates (blue oval). The bottom scale displays the number of fingerprint band substitutions per site.

that both siblings are infected with the a similar strain, an observation also made by RAPD-PCR.

Figs. 6.5 and 6.6 demonstrate that each isolate within a pair (577A / 577C, 9613A / 9613C and F561 / F563) shares more identity with its corresponding isolate than with any other isolate, indicating that microevolution has occurred. In addition, all sibling isolates show more identity to each other than to the mother's isolates, an finding illustrated better within the radial neighbour-joining tree (Fig. 6.6). This result, along with RAPD-PCR analysis, demonstrates that sibling transmission has occurred within this family, and that parental infection had no bearing on sibling infection.

6.2.3.2. Family 2: FAFLP-PCR fingerprinting and phylogenetic analysis.

The rectangular cladogram and radial phylogenetic trees of eight isolates from Family 2 are shown in Figs. 6.7 and 6.8, respectively. FAFLP-PCR analysis of the Family 2 isolates confirmed that four different *H. pylori* strains infected this family. The mother's and daughter's isolates shared more identity with each other than with any other isolates, indicating that mother to daughter transmission or *visa versa* had occurred. The son's isolates (605A / 605C) show more identity to each other than to any other isolate. The daughter's spouse harboured two unrelated isolates (222A / 222C, mixed infection). These FAFLP-PCR findings were in accordance with the findings of the RAPD-PCR analysis. However, FAFLP-PCR has revealed more information relating to the strains from mother and daughter. Although both of the mother's isolates and the daughter's isolate are related, the corpus strains from both mother and daughter show more identity to each other than to the mother's antral isolate, a result that supports RAPD-PCR analysis. Thus, it appears that whoever passed the *H. pylori* infection between mother and daughter, it was the isolate from the corpus that was transmitted between these two members of the same family.

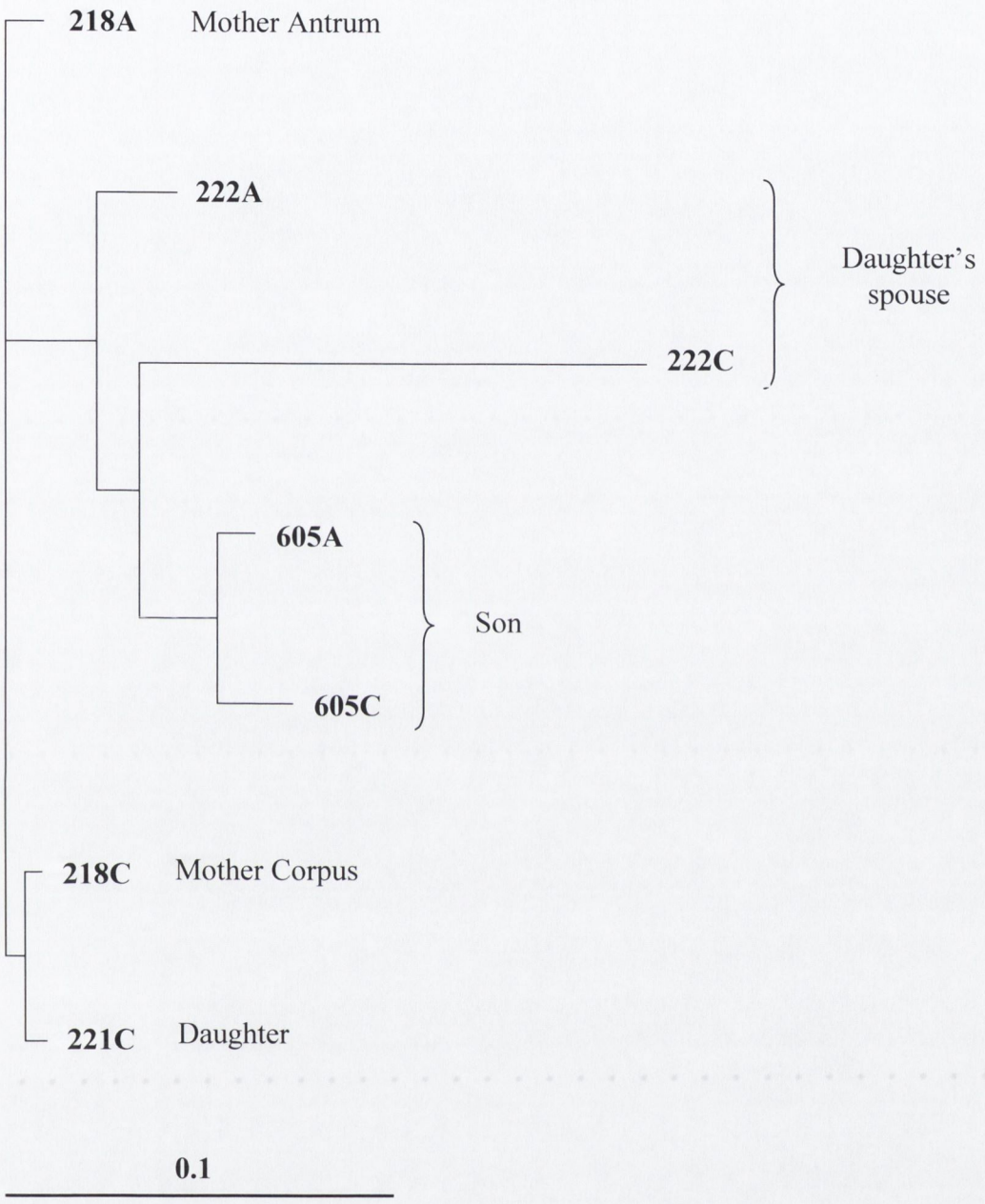


Fig. 6.7. Rectangular cladogram neighbour-joining phylogenetic tree of *H. pylori* isolates from Family 2 based on similarities of FAFLP-PCR fingerprints.

The mother's corpus isolate and the daughter's isolate (corpus) share more identity to each other than to the mother's antral isolate (218A), even though the isolate 218A is highly similar to these isolates. Clustering of the isolates from the son and the daughter's spouse show that these isolates are distinct from each other and the isolates from other family members. The bottom scale displays the number of fingerprint band substitutions per site.

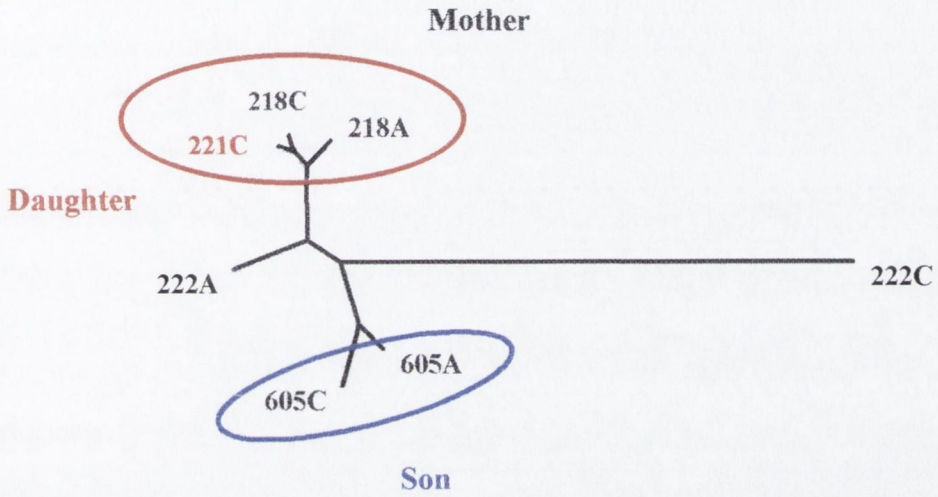


Fig. 6.8. Radial neighbour-joining phylogenetic tree of *H. pylori* isolates from Family 2 based on similarities of FAFLP-PCR fingerprints.

This depiction allows a clearer view of the relationships of these familial isolates. The mother's and daughter's isolates are related to each other, although the corpus isolates are more alike. The son harbours two isolates that have undergone microevolution. The daughter's spouse harbours two unrelated strains. The bottom scale displays the number of fingerprint band substitutions per site.

6.3. Discussion.

Although uncovering the route of transmission or source of an infection does not clarify the molecular mechanisms of an organism's virulence, understanding epidemiological information is necessary to prevent infection or re-infection after anti-microbial treatment. Moreover, knowledge of how *H. pylori* is passed throughout a population can be used to gain insights into the enigma for this organism's extensive genetic diversity.

6.3.1. Genotyping of familial H. pylori *isolates at three different loci.*

Characterisation of familial isolates by genotyping of different loci (*vacA*, *cagA* and *iceA*) has been used to characterise *H. pylori* strains within families (Yamaoka *et al.*, 2000a). Yamaoka *et al.* (2000a) showed that *H. pylori* genotypes were conserved between mothers and their children and distinct from their father's genotypes within two Hispanic families. As observed in Chapter 5, using paired isolates, genotyping of multiple loci cannot differentiate between single strain and mixed strain infections of *H. pylori* in the same individual. Therefore, it is wise to assume that analysis of familial *H. pylori* isolates using this method will not unambiguously reveal clonal transmission of *H. pylori* isolates or the number of strains infecting a family. The multiple locus genotyping data implied that there were three different strains within Family 1 (s1a m2, 2 *cagA* repeats; s1a m2, 6 *cagA* repeats; s2 m2, 2 *cagA* repeats) and three different strains within Family 2 (s1a m2 2 repeats, s1a m2 1 repeat, s2 m2 2 repeats). This conclusion proved to be inaccurate and demonstrates the superior sensitivity of RAPD-PCR and FAFLP-PCR analysis over multiple locus genotyping. Moreover, Owen and Xerry (2003) demonstrated that two isolates from a husband and wife in Belfast exhibited the same MLST type indicating clonal transmission of a *H. pylori* strain between spouses. However, the *vacA* types of these isolates were s1 m1 and s1 m2 in husband and wife, respectively. As both these isolates were shown to be clonal variants, this finding indicates that the *vacA* locus is not a stable marker and therefore does not give an accurate depiction of transmission of *H. pylori* within a family.

6.3.2. Genomic fingerprinting of H. pylori *isolates from two separate families.*

RAPD-PCR and FAFLP-PCR analyses provide a more accurate depiction of the spread of *H. pylori* infection between familial members than multiple-locus genotyping. An interesting observation is that both techniques complement each other when used in concert. For instance, RAPD and FAFLP analyses demonstrated that one strain resided within the mother and daughter in Family 2. However, FAFLP analysis can give a more accurate depiction of

the relatedness than RAPD analysis. For example, RAPD analysis demonstrated that isolates 218A and 218C (mother) contained an amplified fragment that was not present in isolate 221C (Fig. 6.3 red arrow), and isolate 218A also contained a high molecular weight fragment that was absent in isolates 218C and 221C (white arrow). [The absence of an amplified fragment between ancestrally related isolates could also be seen in Family 1 (Fig. 6.2, 577C)]. This suggests that isolates 218A, 218C and 221C are highly similar yet exhibit some diversity from each other. FAFLP analysis revealed that isolates 218C and 221C exhibited more relatedness to each other than to isolate 218A (Figs. 6.7 and 6.8), although all three isolates are ancestrally related. This relationship between these two corpus isolates that was not detected by RAPD analysis. This observation implies that the infection from mother to daughter, or *vice versa*, was passed from corpus to corpus. Unfortunately there was no antral isolate from the daughter available for comparative analysis. Akada *et al.* (2003) demonstrated that different *H. pylori* strains can exhibit dissimilar tissue tropism targeting different gastric niches, implying that infection does not always start in the antrum and then spread to the corpus.

Another benefit of using FAFLP and RAPD analysis in conjunction is that sometimes RAPD fingerprinting alone can yield ambiguous results. For example, isolates 222A and 222C from Family 2 had identical RAPD profiles using primer PIII, yet dissimilar profiles using primer PI. Use of primer PIII alone may have lead to the conclusion that these two isolates were in some way related. However, FAFLP analysis revealed that these two isolates exhibit a high degree of diversity. Thus, to obtain meaningful results using RAPD analysis more than one primer may be needed to provide indicative data.

6.3.3. *Clonal transmission of* H. pylori *within families.*

In the present study it was found that clonal transmission of *H. pylori* occurs within families, either between siblings or between mother and daughter. These routes of transmission have previously been described (Shimizu *et al.*, 1999; Shimizu *et al.*, 2000). In addition, FAFLP analysis revealed that clonal transmission of *H. pylori* also resulted in minor differences arising in genomic fingerprints between clonally transmitted strains.

Although *H. pylori* is a genetically diverse organism, a recent study has suggested the presence of many clonal lineages of *H. pylori* evolving in isolation from one another within families (Han *et al.*, 2000). This theory was confirmed in this study where isolates with

identical or nearly identical DNA fingerprints were found in two or more members of the same family.

6.3.4. Clonal transmission and clinical outcome within a family unit infected with H. pylori.

A study that monitored the spread of a *H. pylori* infection within a family demonstrated that a mother and her two sons were infected with the same strain, as revealed by repetitive extragenic palindromic (REP)-PCR fingerprinting, whilst the daughter was infected with an unrelated strain (Miehlke *et al.*, 1999). The mother and two sons all exhibited different forms of gastric cancer and the daughter suffered from a duodenal ulcer. This observation led to the conclusion that clonal transmission of a *H. pylori* strain within a family unit is a significant factor in determining clinical disease. However, a more recent study of gastric cancer patients found that specific virulence factors or similar *H. pylori* strains did not correlate with a specific histological pattern or disease even among those sharing the same environment in childhood (Li *et al.*, 2001).

The present study demonstrated that all siblings within Family 1 were infected with a highly similar strain. Two members of this family exhibit duodenal ulcers while the other two were asymptomatic. This result seems to confirm the conclusions made by Li *et al.* (2002a). However, the clinical symptoms may not have yet occurred within two of these siblings, and duodenal ulceration could potentially arise within these latter family members at a later stage. As the age of these siblings and their clinical history post initial diagnosis is not known, a conclusion cannot be drawn on this point.

The fact that isolates with highly similar RAPD-PCR and FAFLP-PCR fingerprints were present in two members of each family studies gives support to the accepted model that intrafamilial transmission provides the main focus for *H. pylori* spread. However, the precise route of transmission is as yet unknown. One route of interest has been between spouses. Studies in Greece and Germany have shown that some spouses can be infected with the same strain (Gerogopoulos *et al.*, 1996; Suerbaum *et al.*, 1998), whereas in Singapore and Japan this mode of transmission is uncommon (Kuo *et al.*, 1999; Suzuki *et al.*, 1999). No evidence of transmission between spouses was detected in both families studied. This is in agreement with the findings of Kuo *et al.* (1999) and Suzuki *et al.* (1999). However, Owen and Xerry (2003) detected isolates exhibiting the same sequence type in husband and wife in two

different Belfast families demonstrating that this mode of transmission can occur within Northern Irish isolates.

Another possible mode of transmission within families is via parent to child, with contrasting arguments as to whether mother-to-child or father-to-child transmission is more important (Bamford *et al.*, 1993; Wang *et al.*, 1993; Shimizu *et al.*, 1999, 2000; Taneike *et al.*, 2001). However, it has been reported in Germany that a mother and father were both infected with the same strain, while their child was infected with an unrelated strain (Suerbaum *et al.*, 1998). The present analysis supports both the mode of transmission from mother to child (daughter) and from sibling to sibling. From Family 2 it is apparent that the mother and daughter share clonal variants of the same strain (Figs. 6.7 and 6.8). Interestingly, the daughter's isolate (corpus) is more like the mother's corpus isolate than her mother's antral isolate. As infection usually occurs at an early age (< 10 years of age), it is likely that the mother infected the daughter and that the daughter was infected with a *H. pylori* residing in the corpus of the mother's stomach.

The other mode of transmission found in this study was transmission via siblings (Family 1). From Fig. 6.6 is appears that, although all siblings were infected with clonal variants, the strains infecting these children group separately indicating a possible route of transmission within these siblings. Isolates F431 and F613 are more similar to each other than the isolates from the other siblings. Likewise isolates 577A / 577C and 9613A / 9613C show more identity to each other than to isolates F431 and F613. Thus, it is possible that either sibling F431 or F613 was first infected with an *H. pylori* strain and then one of these children passed the infection onto the other siblings, or *vice versa*.

This analysis demonstrates intrafamial transmission of *H. pylori* infections. Passage of a strain from one host to another was associated with slight changes in the genetic fingerprints of this strain. This clonal variation implies that over time the species genotype may change as populations continually adapt to different human hosts by formation of successful variants.

Chapter 7

Microevolution at the 3′-end of the cagA *gene between paired* Helicobacter pylori *isolates alters the number of Tyrosine Phosphorylation Motifs (TPMs) located at this locus.*

7.1 Introduction 123

7.2. Results 127

7.2.1. Amplification of cagA *3'-repeat sequences from two paired* H. pylori *isolates* 127

7.2.2. Cloning of cagA *repeat sequences into the pSFV1 vector* 128

7.2.2.1. Digestion and ligation of modified *cagA* repeats and pSFV1 vector 128

7.2.2.2. Screening for inserts 128

7.2.2.3. Checking orientation of *cagA* repeat sequences 129

7.2.2.3. Sequence analysis of pSFV1 clones 129

7.2.3. in vitro *transcription of the pSFV1 clones* 131

7.2.4. Growth and transfection of AGS cells 132

7.2.5. Analysis of cellular proteins within transfected AGS cells 132

7.2.5.1. Analysis of protein fractions from transfected AGS cells 132

7.2.5.2. Tyrosine phosphorylation analysis of transfected AGS cells 133

7.3. Discussion 135

7.3.1. Tyrosine phosphorylation motifs within H. pylori cagA *3'- repeat sequences* 135

7.3.2. Sequence analysis of pSFV1 clones 136

7.3.3. AGS cell transfection 138

7.3.4. Analysis of polypeptides expressed in AGS cells 138

7.3.5. Analysis of tyrosine phosphorylation pattern of polypeptides expressed in AGS cells transfected with recombinant pSFV1 bearing 3'-end cagA *sequences* 139

7.1 Introduction.

Helicobacter pylori has been known for some time to be the cause of a variety of gastric illnesses from peptic ulcer disease to gastric adenocarcinoma (Lee *et al.*, 1993). The exact mechanisms by which *H. pylori* causes these diseases are as yet unclear. The vacuolating cytotoxin (VacA) has attracted much interest as it causes acidic vacuoles to form and swell in gastric epithelial cells. It was believed that the morphological changes caused by VacA were the cause of thinning of the gastric epithelium and the formation of gastric lesions (Telford *et al.*, 1994). Although VacA is an important virulence factor, another locus within the *H. pylori* genome was found to be associated with particular *vacA* alleles and disease outcome (Tummuru *et al.*, 1993). This locus was termed the cytotoxin-associated gene A or the *cagA* gene. Subsequent epidemiological analysis of the *cagA* gene showed that the presence of this locus was associated with increased severity of disease outcome (Covacci *et al.*, 1993; Blaser, 1994). Although there have been reports showing that the presence of the *cagA* gene is associated with some clinical diseases, it has also been suggested that the *cagA* gene does not predict clinical outcome due to its high prevalence (75–95%) in strains in several geographical regions with a relatively low incidence of clinical disease (Hacker and Kaper, 2000). Studies in Japan, Korea and China showed that more than 90% of *H. pylori* strains contained the *cagA* gene irrespective of clinical outcome (Miehlke *et al.*, 1996; Pan *et al.*, 1997; Shimoyama *et al.*, 1997). As no other sequenced gene in databanks shared sequence similarity with the *cagA* gene the function of the CagA protein remained a mystery for many years.

Subsequent to the discovery of the *cagA* gene it was found that this locus resided within a 40-kb pathogenicity island [the *cag* PAI (Censini *et al.*, 1996)] that encodes a type IV secretory apparatus. The *cag* PAI in the fully sequenced *H. pylori* strain 26695 consists of 27 putative genes, six of which have been identified as homologues of the components of the prototype type IV secretion system of *Agrobacterium tumefaciens virB* operon (Akopyants *et al.*, 1998). More recently it was demonstrated that *H. pylori* strains with intact *cag* PAI's could induce strong inflammation and ulceration in the stomach of Mongolian gerbils, whereas an isogenic strain with a defect in the *cag* PAI could not provoke these symptoms (Ogura *et al.*, 2000). It is now known that *H. pylori's cag*-associated type IV secretory apparatus is a sheated needle-like structure reminiscent of type III secretion apparatuses (Rohde *et al.*, 2003).

Many bacterial species contain secretory systems that encode syringe-like structures, which are used to transfer bacterial proteins into host cells (Nagai and Roy, 2003; Aizawa, 2001; Cornelis and van Gijsegem, 2000). This knowledge led to the discovery that the CagA protein is transferred into epithelial host cells where it becomes phosphorylated and triggers profound host cytoskeleton rearrangements (Segal *et al.*, 1999, Aski *et al.*, 2000, Odenbreit *et al.*, 2000, Stein *et al.*, 2000). The scattering phenotype induced by translocated and phosphorylated CagA has been termed the 'hummingbird' phenotype (Segal *et al.*, 1996, 1999). This phenotype alters the regular epithelial structure in such a way that a cup shaped pedestal forms beneath the bacterium. The epithelial cells used in these studies have been AGS cells. These are cells derived an adenocarcinoma of the stomach of a 54-year-old Caucasian female without prior anti-cancer treatment. As AGS cells are gastric cells, they enable *in vitro* infection studies to be carried out that closely mimic what occurs *in vivo*.

Delivery of virulence proteins via type III secretion mechanisms has been well established (Heuck 1998; Cornelis and van Gijsegem, 2000). Although various pathogens have been shown to use type IV secretion systems for conjugational DNA transfer (*Escherichia coli*), for transfer of virulence plasmids into plant cells (*Agrobacterium tumefaciens*), or for toxin secretion *Bordetella pertussis*) (Covacci *et al.*, 1999), bacterial proteins had never before been documented to be transferred by type IV systems. Thus, CagA is the first bacterial virulence protein found to be translocated by a type IV secretion system (Nagai and Roy, 2003). Although the discovery that CagA is translocated into epithelial cells and becomes phosphorylated was a major advance in the understanding of the function of this protein, the mode by which it induces cytoskeletal rearrangements and its role in pathogenesis is still unknown.

Subsequent studies were directed at uncovering the kinase that phosphorylates CagA to understand the pathway by which CagA induces cytoskeletal rearrangements in epithelial cells. CagA becomes phosphorylated on a tyrosine residue, a process in eukaryotic cells that is a critical element in signal tranduction pathways involved in the regulation of biological responses, including cell growth and differentiation. Protein kinase substrates differ due to the different amino acid sequence specificity for each kinase family. Based on previous investigations the amino acid motif EPIYA was demonstrated to be an essential tyrosine phosphorylation motif (TPM) for CagA-induced rearrangements (Backert *et al.*, 2001). Interestingly, Selbach *et al.* (2002) suggested that this motif shares a relationship with the Src

phosphorylation consensus motif [EEIY(G/E)]. Based on observations like this, many studies have demonstrated that Src kinases are responsible for the phosphorylation of CagA (Ashi *et al.*, 2000; Backert *et al.*, 2000; Christie *et al.*, 2000; Odenbreit *et al.*, 2000; Stein *et al.*, 2000, 2002; Selbach *et al.*, 2002).

The treatment of mammalian cells with growth factors, such as hepatocyte growth factor (HGF) induces morphological changes similar to the hummingbird phenotype induced by *cagA*$^{+}$ *H. pylori* strains (Hatakeyama, 2003). This led to the theory that the CagA protein may trigger a growth factor-like response within gastric epithelial cells and that CagA signalling may be similar to an HGF-triggered signalling pathway. The scattering phenotype induced by HGF has been reported to require an Src homology 2 (SH2) domain-containing protein tyrosine phosphatase, SHP-2 (Kodama *et al.*, 2000). SH2 domains are known to be protein modules that specifically interact with phosphotyrosine-containing peptides (Songyang *et al.*, 1993). Indeed, recent reports have demonstrated that CagA not only becomes phosphorylated but also that it binds to and stimulates SHP-2 in a tyrosine phosphorylation-dependent manner (Tsutsumi *et al.*, 2003; Yamazaki *et al.*, 2003). The role of SHP-2 remains to be defined. However, it is thought to be primarily a positive regulator of cell growth and development. Thus, since SHP-2 is involved in cell growth and differentiation and CagA influences the activity of this phosphatase, it is highly likely that CagA–SHP-2 interactions are responsible for *H.* pylori-induced cytoskeletal rearrangements in epithelial cells. In addition, the C-terminal Src kinase (Csk) binds to phosphorylated CagA and inhibits the further phosphorylation of other CagA molecules (Tsutsumi *et al.*, 2002). Tsutsumi *et al.* (2002) suggested that this may be a way of regulating CagA in order to prevent excess CagA toxicity and to enable *cagA*-positive *H. pylori* to persistently infect the human stomach for decades. This would also provoke chronic gastric damage by occasional but transient deregulation of SHP-2 upon injection of cellular CagA. Regulation is brought about by SHP-2 and Csk competing to bind with phosphorylated CagA, and phosphorylated CagA–Csk interactions reducing the amount of phosphorylated CagA present in the cell and hence CagA available to bind to SHP-2. Figure 7.1 outlines the interaction of CagA with SHP-2 and Csk.

Analysis of paired *H. pylori* isolates has demonstrated that two isolates from the same individual can have identical fingerprint profiles, yet have a different number of repeat sequences located at the 3′-end of the *cagA* gene (Chapter 5). Within a repeat sequence lies an EPIYA motif, which, as mentioned previously, has been shown to be a critical motif involved

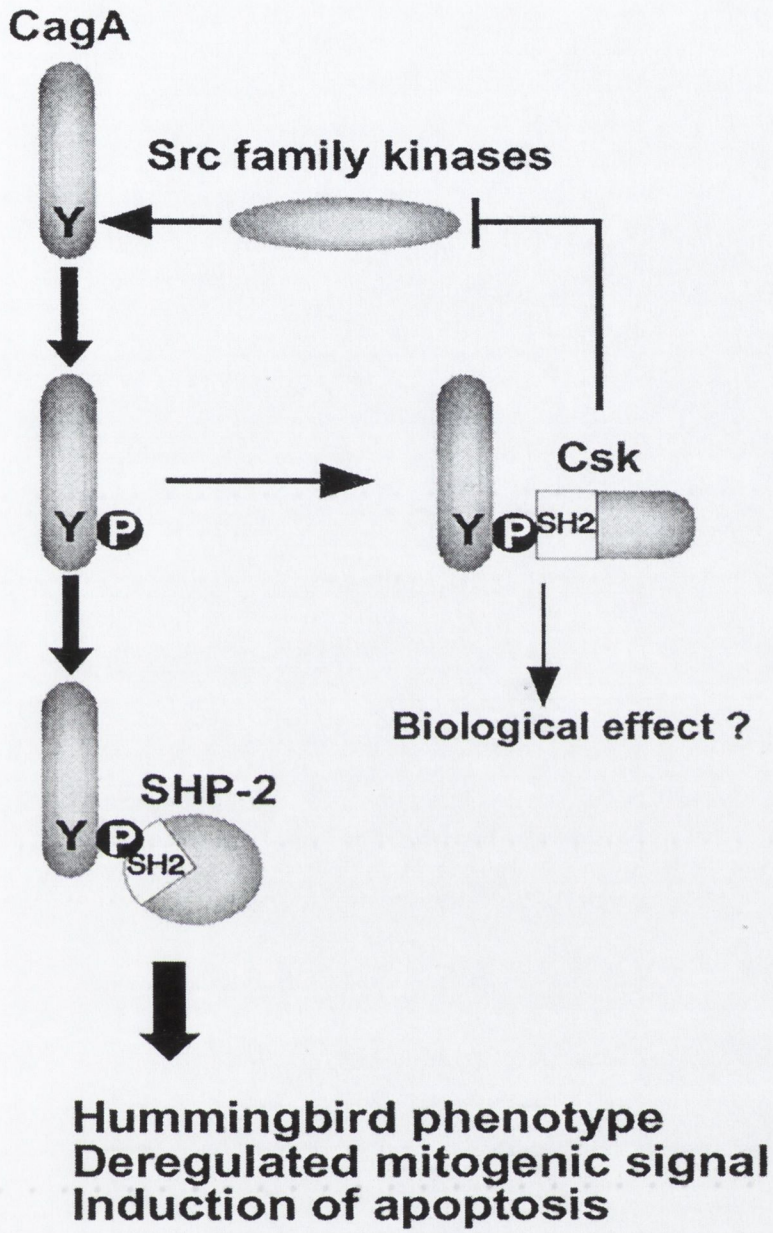


Fig. 7.1. Model for the regulation of CagA–SHP-2 signalling by Csk (adapted from Tsutsumi *et al.*, 2003).

CagA-Csk complex attenuates the CagA-SHP-2 signalling by down-regulating complex formation between CagA and SHP-2. This process involves (i) competitive inhibition of SHP-2 binding to CagA by Csk, and (ii) reduced CagA phosphorylation by inactivation of Src family kinases, which is caused by the activation of Csk upon CagA binding.

in the phosphorylation of the CagA protein. The benefit or otherwise to an isolate of increasing or decreasing the number of repeat sequences within its *cagA* gene remains unclear. However, because varying the number of these repeat sequences also means that the number of TPMs is increased or decreased as well suggests that this variation can alter the phosphorylation potential of the CagA protein, and perhaps affect its properties, e.g., binding to Csk or binding to SHP-2.

The study herein describes the cloning of the *cagA* repeat sequences from two pairs of ancestrally related *H. pylori* isolates recovered from individual patients that displayed variations in the number of repeat sequences between isolates. Each repeat was cloned into a Semliki Forest Virus vector (pSFV1), a viral vector based on the replicative cycle of this virus (Fig. 7.2) that is used to manipulate mammalian cells into generating genetically engineered viral particles (Liljeström and Garoff, 1993; Strauss and Strauss, 1994; Fig. 7.3). This study employs an adapted strategy to express heterologous proteins in AGS cells via the pSFV1 vector (Fig. 7.4). The aims of this strategy were to confirm whether the EPIYA motif was tyrosine phosphorylated and whether the variation in number of *cagA* repeats might alter the degree of phosphorylation of the CagA protein.

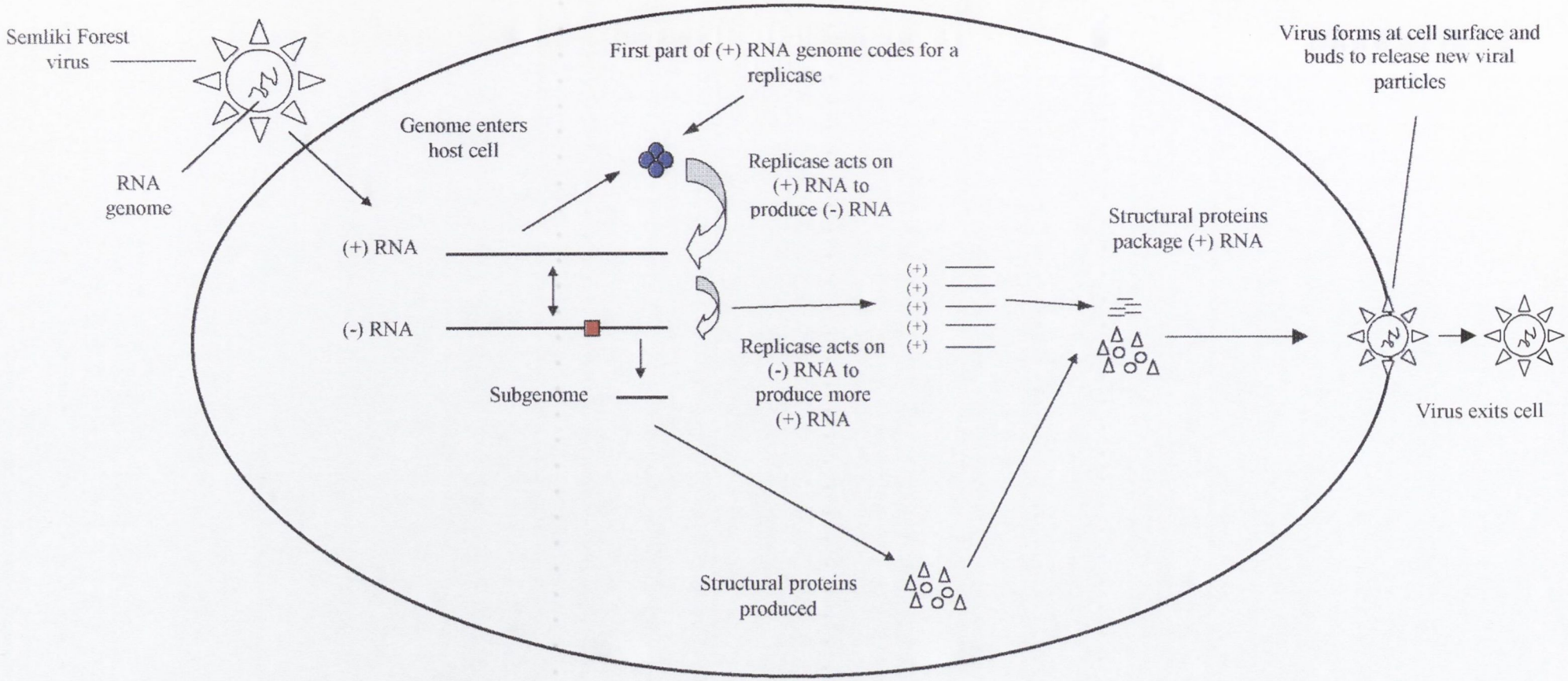


Fig. 7.2. Schematic representation of the Semliki Forest virus replicative cycle.

Semliki Forest virus is endocytosed into the host cell and its genome is released. The (+) RNA is translated by the host cell. The first two thirds of the (+) RNA encodes a replicase. This replicase acts on the (+) RNA to produce (-) RNA. The replicase can now act on the (-) RNA to generate more (+) RNA. This results in a large amount of (+) RNA and (-) RNA being produced. The (-) RNA contains a 26S promoter (red box) that allows for the last one third of the molecule to be transcribed and translated into structural proteins. The structural proteins package the (+) RNA strands and the virus is now ready to exit the cell.

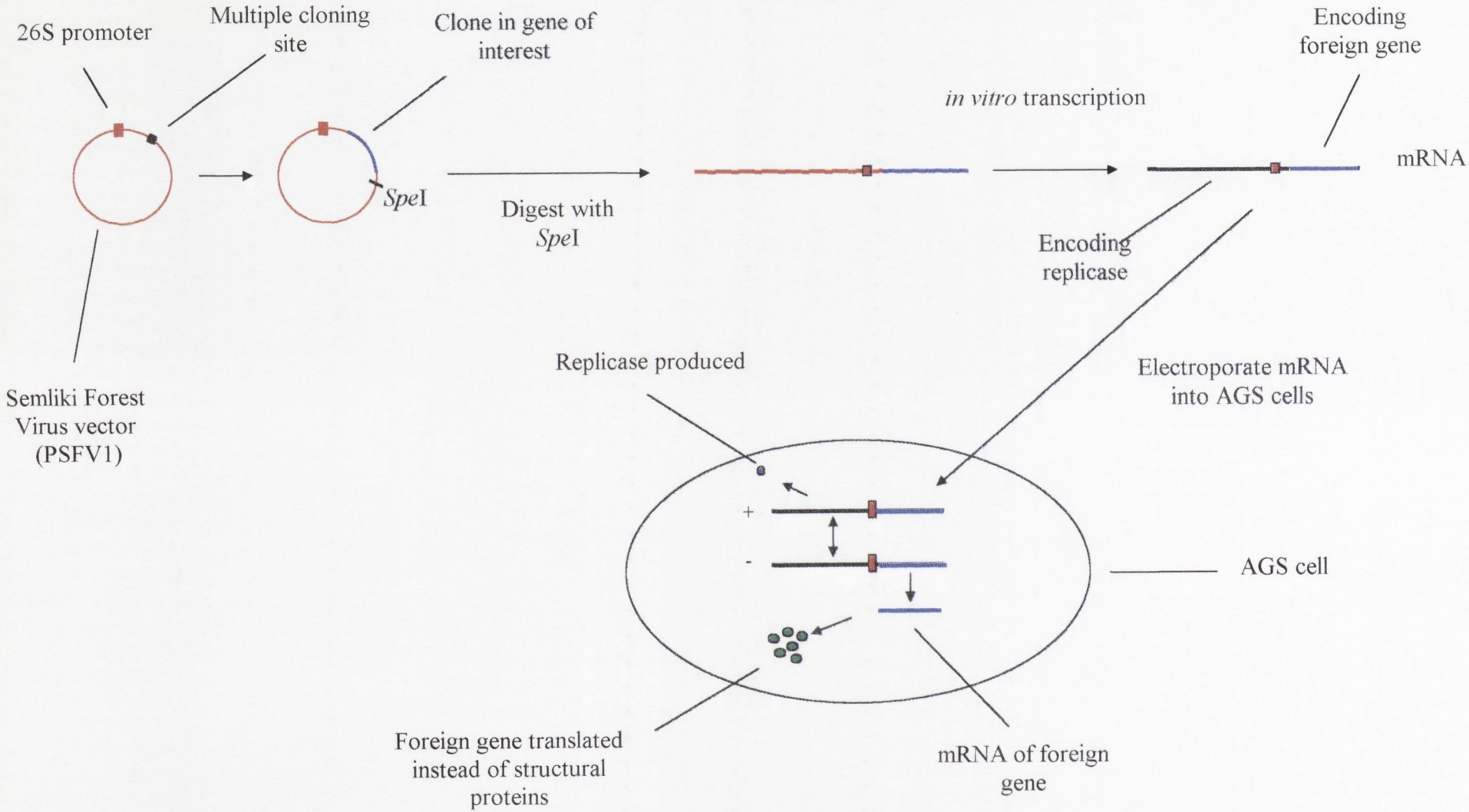
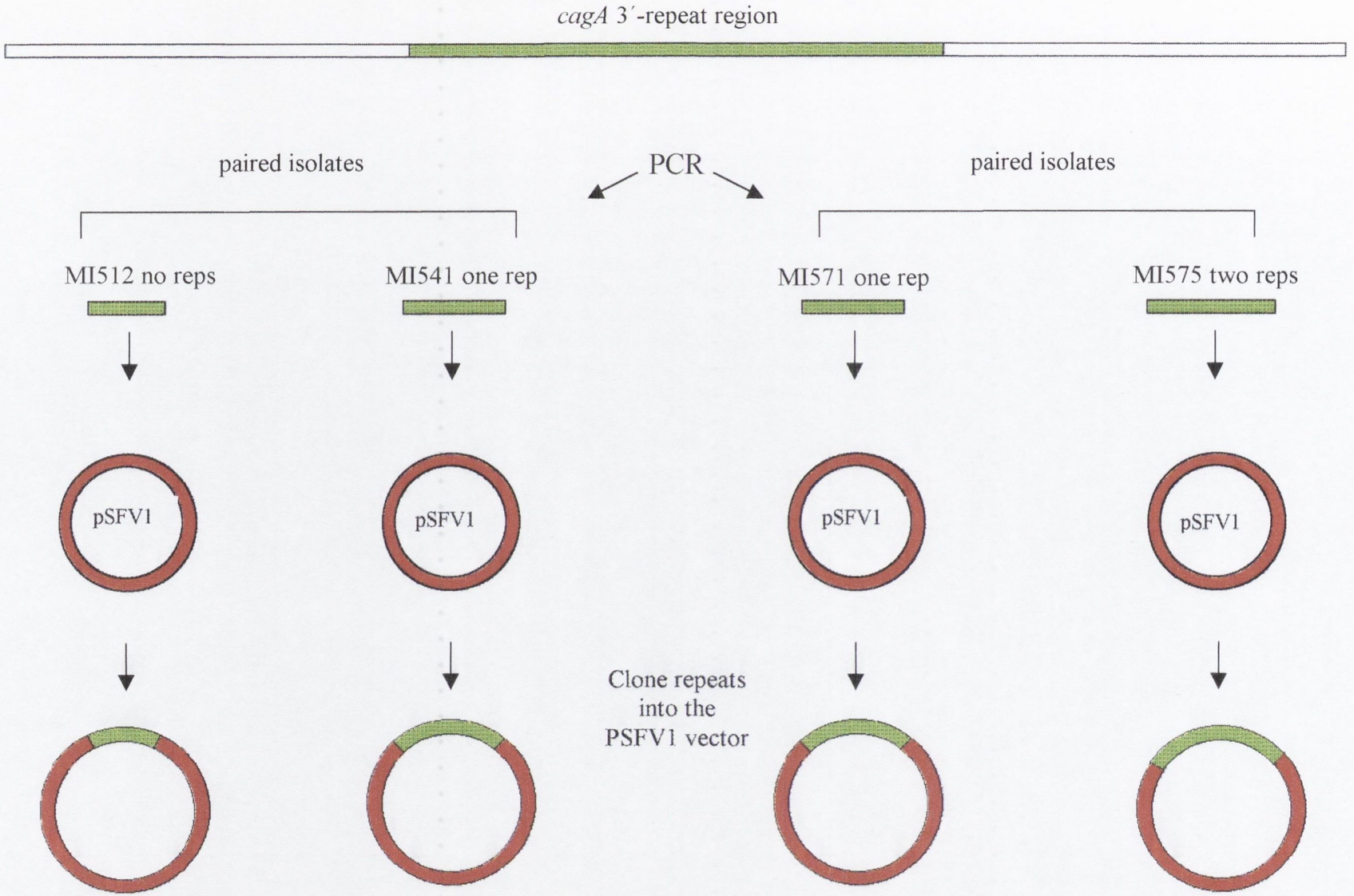

26S promoter
Multiple cloning site
Clone in gene of interest
Encoding foreign gene
in vitro transcription
mRNA
*Spe*I
Digest with *Spe*I
Encoding replicase
Semliki Forest Virus vector (PSFV1)
Replicase produced
Electroporate mRNA into AGS cells
+
-
AGS cell
Foreign gene translated instead of structural proteins
mRNA of foreign gene

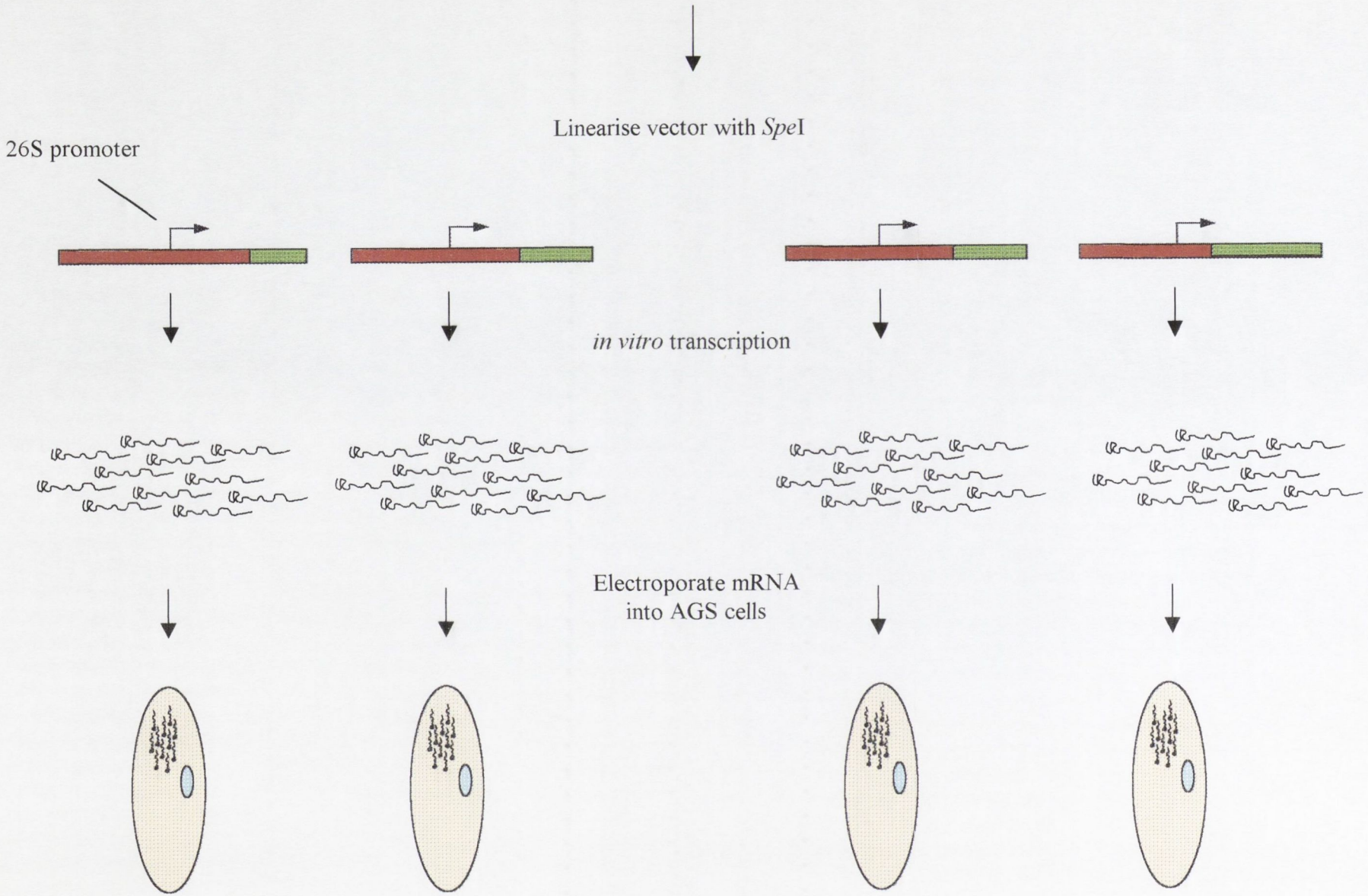
Fig. 7.3. Schematic diagram of how the pSFV1 vector can be used to produce heterologous proteins in mammalian cells.

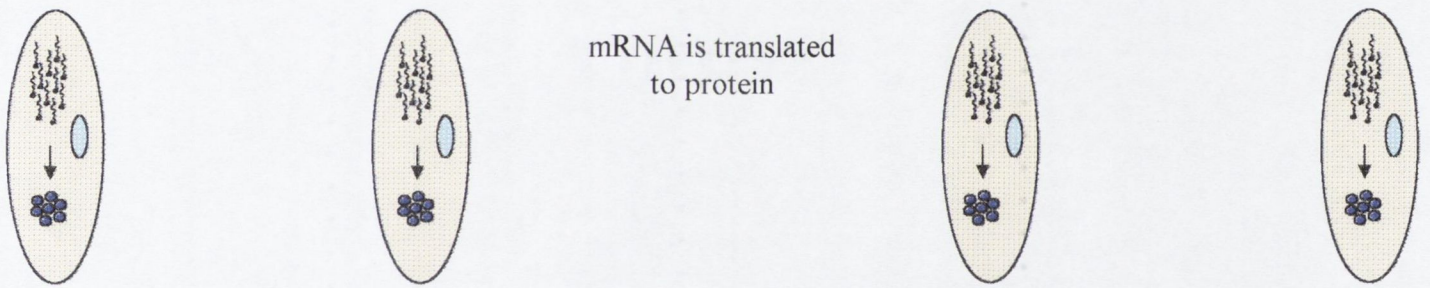
The Semliki Forest Virus vector (pSFV1) is 11,033 bp in size. It contains a region that codes for a replicase enzyme, a 26S promoter, a multiple cloning site (MCS) and a unique *Spe*I restriction site. For heterologous protein expression the gene of interest must be first cloned into the MCS. The whole vector is linearised by digestion with *Spe*I (the insert must not contain *Spe*I sites). The whole linearised vector can now be transcribed *in vitro* using an RNA polymerase. This mRNA now codes for a replicase enzyme and the foreign protein. The mRNA generated can now be introduced into the mammalian cells (AGS cells) via electroporation. Once inside the host cell the mRNA behaves in the same fashion as the RNA genome in the life cycle of the Semliki Forest Virus (Fig. 6.1). Briefly, the replicase encoded within the introduced (+) RNA is translated, this enzyme generates a (-) RNA copy, and the replicase generates many copies of (+) RNA and (-) RNA (200,000). The replicase now acts on the 26S promoter on the (-) RNA strand and transcribes the foreign gene instead of viral structural genes. The RNA transcript for the foreign gene will now be translated by host cellular organelles, and because there are so many copies of this transcript host protein synthesis will shut down and only foreign proteins will be generated. As this is a modification of the SFV system, a helper mRNA is usually introduced into the host cell as well. This helper mRNA codes for the structural proteins of the virus. These proteins will only package the (+) RNA resulting in virus particles with recombinant genomes. These particles can be used to infect mammalian cells and again only the foreign gene will be expressed.

7.4. Strategy used to determine the phosphorylation potential of *cagA* repeats.

The strategy used in this study is outlined in Fig. 7.3. Firstly two pair of isolates were chosen that displayed identical RAPD-PCR fingerprints within each pair. These pairs also displayed variations in the number of repeats at the 3′-end of the *cagA* gene within each pair. The cagA 3′-repeats were amplified by PCR. *Bam*HI sites, which flanked each fragment, were introduced into the primers, and an ATG start codon was introduced before each repeat as the whole gene was not being expressed, just the *cagA* 3′-repeat sequences. The pSFV1 vector and the fragments were digested with *Bam*HI. Each fragment was ligated into a vector. Once it was certain that each clone was present and in the right orientation, each vector was linearised by digestion with *Spe*I. The linear vectors were *in vitro* transcribed using RNA polymerase. Each recombinant mRNA was electroporated into an AGS cell culture. Once inside the foreign gene would be expressed at a high level. Total protein can then be extracted from each cell culture that has been transfected with the recombinant RNA. When each sample is run on an SDS-PAGE gel, there should be extra or more intense bands in these samples with respect to samples of total cellular protein from AGS cells electroporated with PBS and cells electroporated with an empty pSFV1 vector. When these samples and controls are transferred to a nitrocellulose membrane and probed with an anti-phosphotyrosine antibody, it is possible to determine if these heterologous proteins are tyrosine phosphorylated.


cagA 3′-repeat region
paired isolates
PCR
paired isolates
MI512 no reps
MI541 one rep
MI571 one rep
MI575 two reps
pSFV1
pSFV1
pSFV1
pSFV1
Clone repeats
into the
PSFV1 vector


Linearise vector with *Spe*I
26S promoter
in vitro transcription
Electroporate mRNA
into AGS cells



mRNA is translated
to protein

Extract total cellular protein

Run protein on SDS gel to determine if there are over-expressed proteins compared to cells with an empty vector and cells without a vector.

Transfer protein samples to nitrocellulose filter.

Probe with anti-tyrosine antibody.

Determine whether heterologous proteins are tyrosine phosphorylated.

7.2. Results.

7.2.1. Amplification of cagA *3′-repeat sequences from two paired* H. pylori *isolates.*

In designing primers used to amplify the *cagA* 3′-repeat sequences several factors needed to be considered. Firstly, the primer sequences had to flank the repeat region. However, they must only amplify the repeat region as there are other putative EPIYA motifs outside the repeat sequences and the purpose of this study was to determine whether the tyrosine residue(s) within the repeats is (are) phosphorylated. To enable expression of only part of a gene an artificial start codon (ATG) needed to be inserted upstream of the repeats (on the forward primer). Moreover, this start condon must be inserted so that the sequence of interest is in frame and directionally expressed as natural. As *Bam*HI was the restriction enzyme of choice for cloning, a *Bam*HI restriction site needed to be inserted into each primer so that these sites flanked the amplified repeat. A 'GCG' clamp was also needed on either side of the *Bam*HI sites in order to allow proper digestion. Finally, a ribosome-binding site (RBS, GCACC) was needed upstream of the start codon. The final primer sequences used were:

Forward: CagcloneF - GCG**ggatcc**GCG**GCACCATG**GGGCAAGCAGCGGGC

Clamp — *Bam* HI — Clamp — RBS — Start codon — *cagA* repeat sequence

Reverse: CagcloneR - CGC**ggatcc**GCG**GTAATTGTCTAGTTTCGC**

Clamp — *Bam* HI — Clamp — *cagA* repeat sequence

Fig. 7.5 demonstrates how these primers anneal to *cagA* sequences. These primers are specific for *cagA* sequences. However, their melting temperatures (Tm) vary considerably (Forward Tm = 82 °C, Reverse Tm = 70.9 °C). Due to this variation in Tm's these primers don't work very well together when trying to amplify *cagA* repeats from genomic DNA, as they amplify non-specific PCR products. Therefore, to overcome this, it was necessary to amplify the *cagA* repeat sequences from the chosen isolates using primers CAG1 and CAG2, purify the PCR products and then use these as templates for the amplification the *cagA* repeats using CagcloneF and CagcloneR. Using this strategy the repeats from each isolate (MI512 [no

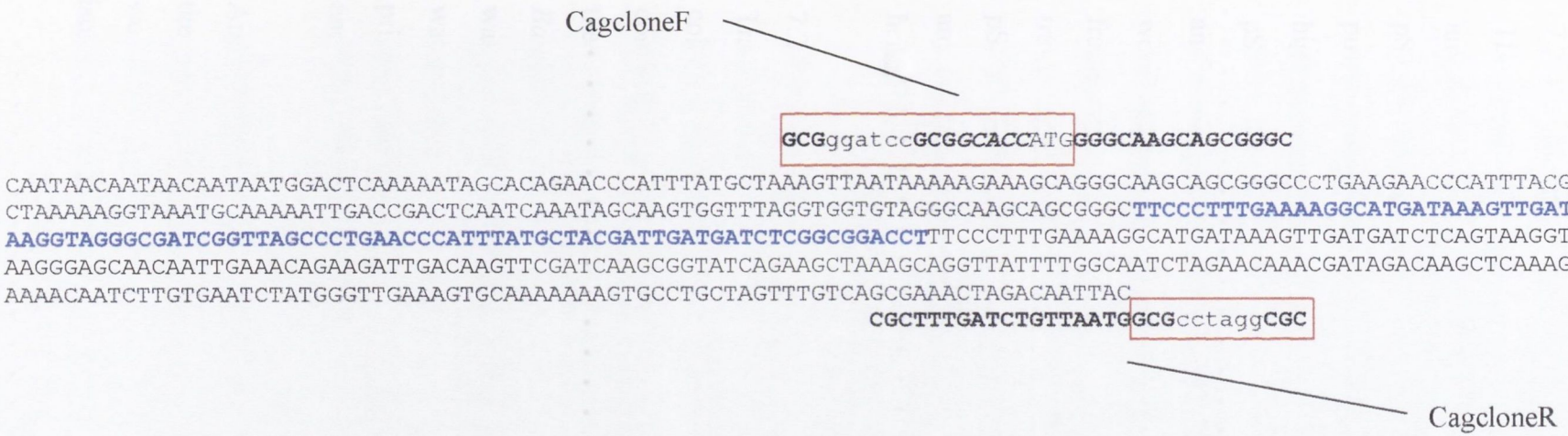


Fig. 7.5. Alignment of cloning primers with the 3′-end of the *cagA* gene of a European *H. pylori* isolate.

The 3′-end of the *cagA* gene from an Italian *H. pylori* isolate is shown (GenBank, AB057094). The repeat sequence of this strain is indicated in blue. The forward and reverse primers used for cloning are indicated. The non-complementary elements of these primers are indicated (red box). The non-complementary sequences within CagcloneF contain GCG clamps, a *Bam*HI sequence (ggatcc), a ribosome-binding site (GCACC) and a start codon (ATG). CagcloneR contains GCG clamps and a *Bam*HI site within its non-complementary sequences. It can be seen that these primers will amplify the 102-bp repeat sequences plus sequences downstream of the repeats (~ 250 bp). However, they will not amplify the regions upstream of the repeat sequences that also contain putative Tyrosine Phosphorylation Motifs.

repeats] / MI541 [one repeat] and 45688A [one repeat] / 45688C [two repeats]) with the various modifications were generated (Fig. 7.6).

7.2.2. Cloning of cagA *repeat sequences into the pSFV1 vector.*

7.2.2.1. Digestion and ligation of modified *cagA* repeats and pSFV1 vector.

The amplified *cagA* repeat sequences from each isolate were purified by gel extraction. Each amplicon was digested with *Bam*HI, and the digested fragments were again purified. The pSFV1 vector (Fig. 7.7) was also digested with *Bam*HI and purified. The digested and purified fragments were run on a gel to make certain that each fragment was present and at a high concentration, and to make certain that the pSFV1 vector was linear (Fig. 7.8). As the pSFV1 vector was digested with one restriction enzyme, there was a danger of it religating and not accepting any fragments. To overcome this problem the 5′-overhangs of the vector were dephosphorylated using Shrimp Alkaline Phosphatase (SAP). The digested vectors and fragments were ligated together overnight. The religated vector with its cloned insert was transformed into *Escherichia coli* strain DH5α and grown on plates containing ampicillin. The pSFV1 vector confers ampicillin resistance (strain DH5α is ampicillin sensitive). Therefore, any colonies that grew should contain a clone as the vector was incapable of religating and, hence had an inability to confer ampicillin resistance.

7.2.2.2. Screening for inserts.

Transformation of DH5α cells on average yielded approximately 10–20 c.f.u. per plate. Five colonies from each ligation reaction were harvested and used to inoculate 10 ml of L-broth containing ampicillin (100 μg/ml). These cultures were grown overnight. The bacteria were then pelleted, and the pSFV1 vectors were isolated. Each putative clone was digested with *Bam*HI and run on an 1.8% agarose gel. Unfortunately, this method of screening for clones was not useful as the fragments that were cloned were very small (278–482 bp) and the vector was so large (11,033 bp) that the small digested fragments could not be visualised. Therefore, primers that spanned the multiple cloning site of the pSFV1 vector were designed (SFV-1F and SFV-1R [Table. 7.1]).

Amplification of an empty pSFV1 vector with these primers (SFV-1F and SFV-1R) results in the generation of a 383-bp PCR product (Fig. 7.9 A). Any vector that contained an insert would produce a PCR product larger than this. Amplification of the putative clones for each isolate confirmed the presence of inserts in the pSFV1 vector (Fig. 7.9 B). If a 278-bp PCR

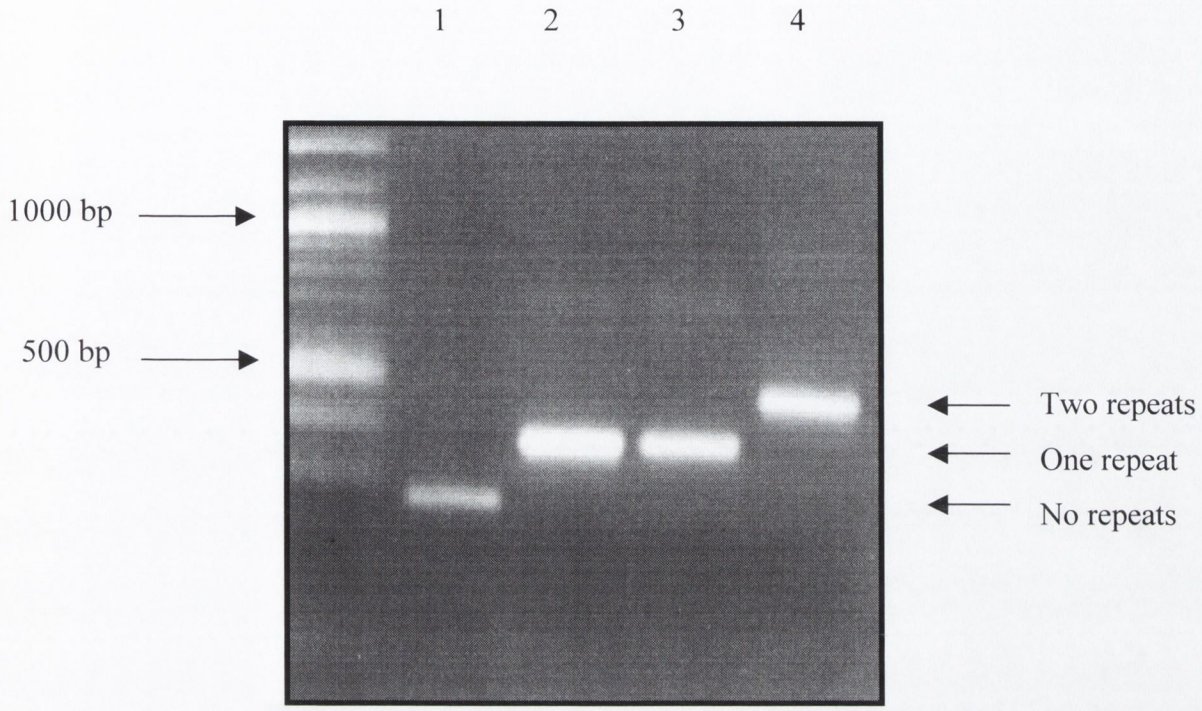


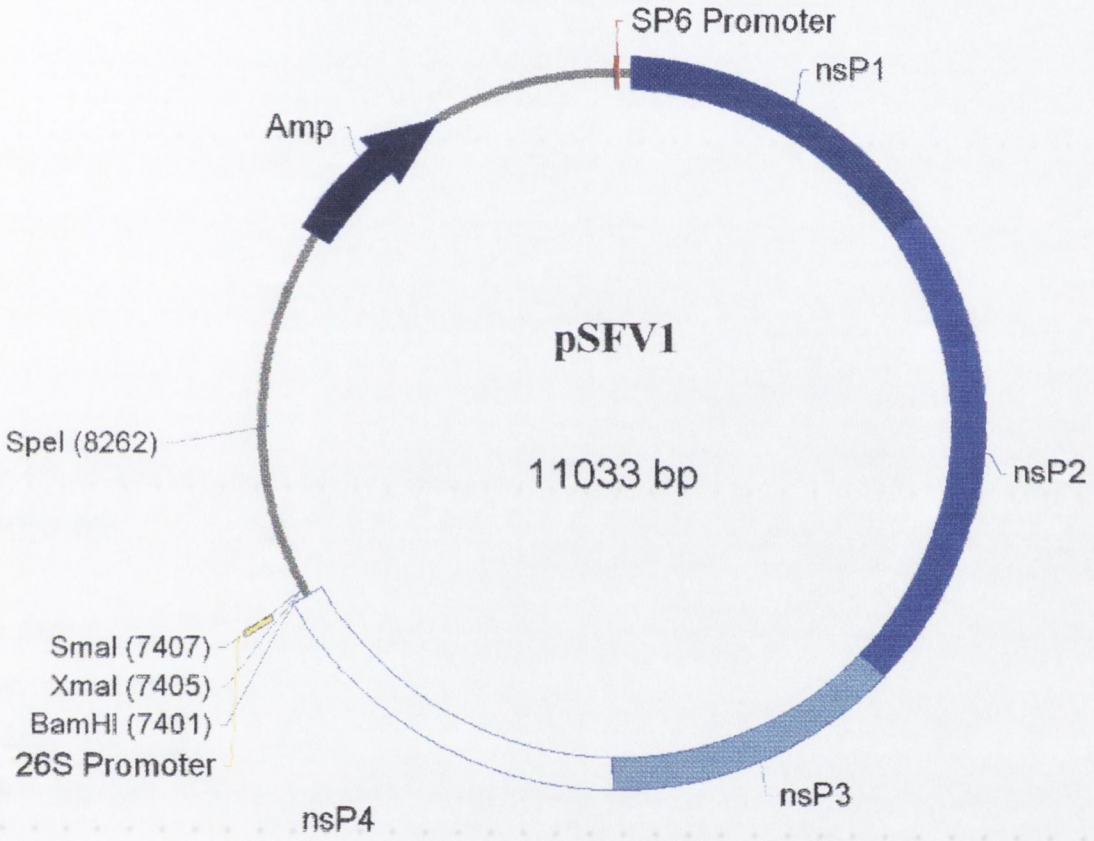
Fig. 7.6. Modified repeat sequences from two paired ancestrally related *H. pylori* isolates separated on a 1.8% agarose gel.

The repeat sequences from isolate pairs MI512 / MI541 (Lanes 1 and 2, respectively) and 45688A / 45688C (Lanes 3 and 4, respectively) were amplified using primers CagcloneF and CagcloneR. Isolate MI512 lacks a repeat sequence (278 bp), isolates MI541 and 45688A contain one repeat sequence (380 bp) and isolate 45688C contains two repeat sequences (482 bp). The primers modify the repeats by incorporating a *Bam*HI site, a start codon and a ribosome-binding site into their PCR products. A = antrum, C= corpus.

Fig. 7.7. Map of pSFV1 viral vector.

The nsP1-4 genes code for a polyprotein termed P1234, which is autocatalytically cleaved by the nsP2 protein, the C-terminal half of which is a papain-like cysteine endopeptidase. The N-terminal half of nsP2 is a viral helicase. The nsP2 protein may also be involved in regulation of the synthesis of the subgenomic 26S RNA. Non-structural protein nsP1 is a methyltransferase and guanylyltranferase involved in capping of viral (+) RNA. It is membrane bound and palmitoylated. Non-structural protein nsP4, the catalytic subunit of RNA polymerase, together with polyprotein P123 forms a short-lived RNA polymerase that copies the (+) RNA genome of SFV to the complementary (-) RNA. Cleavage of polyprotein P123 to non-structural proteins nsP1, nsP2 and nsP3 coverts P123 to a (+) RNA polymerase. The functions of nsP3, which is a phosphoprotein, in RNA replication are poorly characterised. The SP6 promoter drives the transcription and translation of the non-structural proteins.

The restriction enzymes *Sma*I, *Xma*I and *Bam*HI have cleavage sites located within the multiple cloning site (MCS), the nucleotide positions of which are indicated. The 26S promoter drives the transcription and translation of any cloned fragment within the MCS. The *Spe*I site for linearising the vector preceding *in vitro* transcription is indicated. (Liljeström and Garoff, 1993; Salonen *et al.*, 2003).


SP6 Promoter
nsP1
Amp
pSFV1
11033 bp
SpeI (8262)
nsP2
SmaI (7407)
XmaI (7405)
BamHI (7401)
26S Promoter
nsP3
nsP4

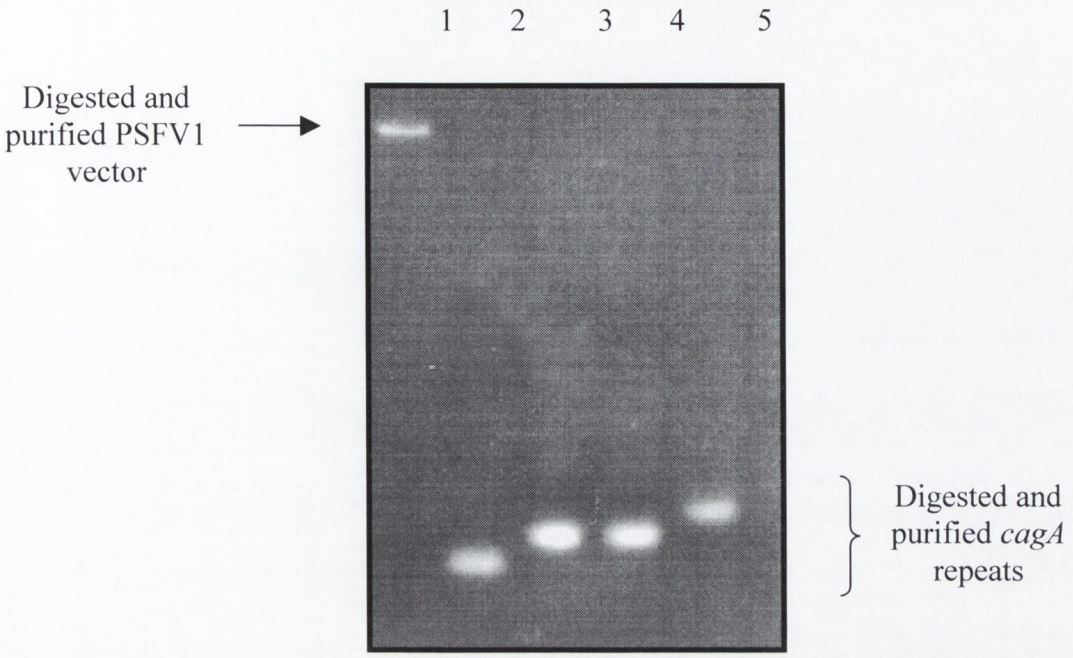
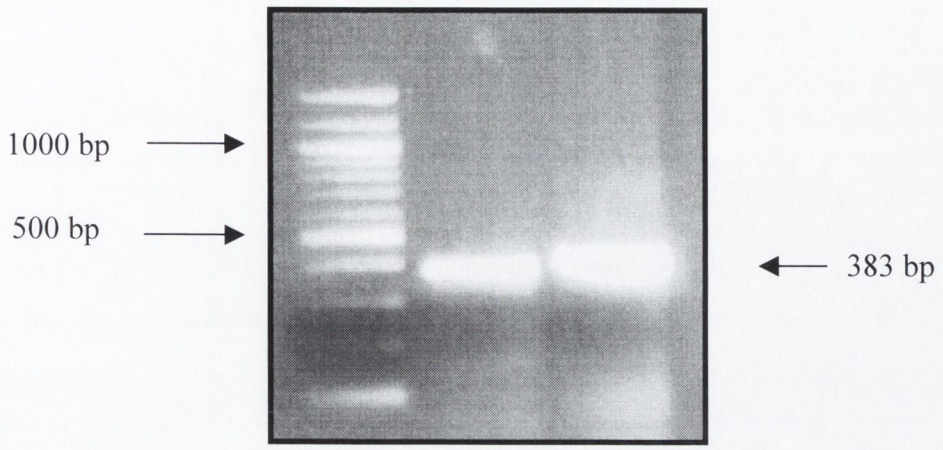


Fig. 7.8. Digested and purified pSFV1 vector and amplified *cagA* repeats run on a 1.8% agarose gel.

The digested and purified pSFV1 vector and the modified *cagA* PCR products are shown. Lane 1 = pSFV1 vector, lane 2 = MI512, lane 3 = MI541, lane 4 = 45688A and lane 5 = 45688C. The pSFV1 vector is much larger (11,033 bp) than the PCR products to be cloned (278 – 382 bp). The digested vector was treated with shrimp alkaline phoshpatase to prevent religation.

A.



B.

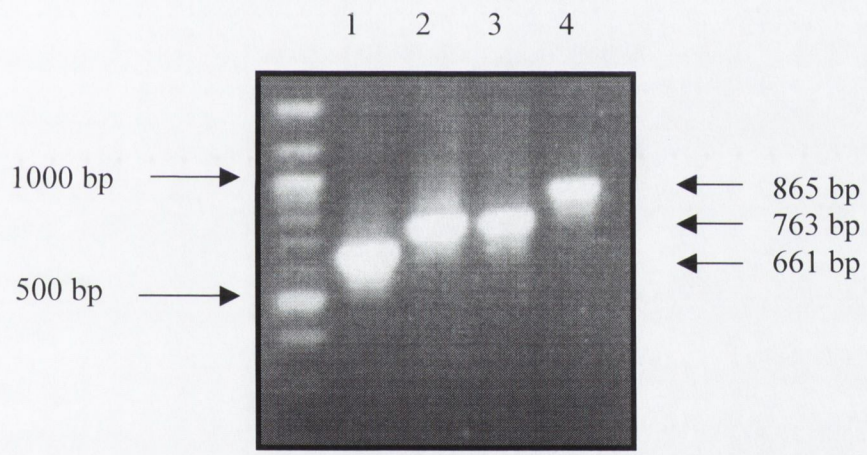


Fig. 7.9. A. and B. Amplification of inserts from cloned and empty vectors run on a 1.8% agarose gel.

As it was impossible to determine the presence of cloned sequences within the pSFV1 vector simply by digestion with *Bam*HI, primers that flanked the multiple cloning site of this vector were designed. An empty or uncloned vector yields a PCR product of 383 bp (**A**). If cloned properly, the *cagA* repeat sequences from isolates MI512 (no repeats), MI541 and 45688A (one repeat) and 45688C (two repeats) would yield PCR products of 661 bp, 763 bp and 865 bp, respectively (**B**).

product was generated using primers CagcloneF and CagcloneR then cloned *cagA* sequences did not contain a 102-bp repeat (e,g., isolate MI512). Therefore, using the SFV-1F and SFV-1R primers, if the vector contained the sequences from isolate MI512 the PCR product would be 661 bp (383 bp + 278 bp). Likewise, if a vector contained an insert from either isolate MI541 or 45688A (one repeat) the PCR product would be 763 bp (838 bp + 380 bp) and the PCR product for isolate 45688C (two repeats) would be 865 bp (383 bp + 482 bp). Upon screening, two of the three putative clones (obtained from the original five transformed colonies) for isolate MI512 were shown to contain an insert of the right size. One of the three clones for isolate MI541 contained an insert of the right size and four of the four clones for isolate 45688A contained an insert of the right size, but only clone obtained for isolate 45688C contained an insert of the right size (Fig. 7.9 B).

7.2.2.3. Checking orientation of *cagA* repeat sequences.

Although inserts were detected within the SFV1 vector, these inserts had to be in the correct orientation for expression purposes. As the cloning method used was not site-directed, the inserts could have been cloned in either orientation. PCR was used to determine whether the clones contained inserts in the correct orientation. The forward SFV-1F and the reverse primer CagcloneR were used in conjunction with each clone as template. As the distance from the annealing site for the forward primer (SFV-1F) to the cloning site and is 172 bp, if the clone was in the right orientation the PCR product would be the size of the insert plus 172 bp. For example, the insert for isolate MI512 is 278 bp. Therefore if a clone for isolate MI512 was in the right orientation, the PCR product generated using primers SFV-1F and CagcloneR (and a MI512 clone as template) would be 450 bp in size (172 bp + 278 bp). If a cloned DNA fragment was in the wrong orientation, no PCR product would be generated. This strategy is outlined in Fig. 7.10. Using this method at least one clone was found that was in the right orientation for each isolate (Fig. 7.11). Therefore, one clone was selected for each isolate and used for further analysis (designated SFV512 and SFV541, SFV45688A and SFV45688C)

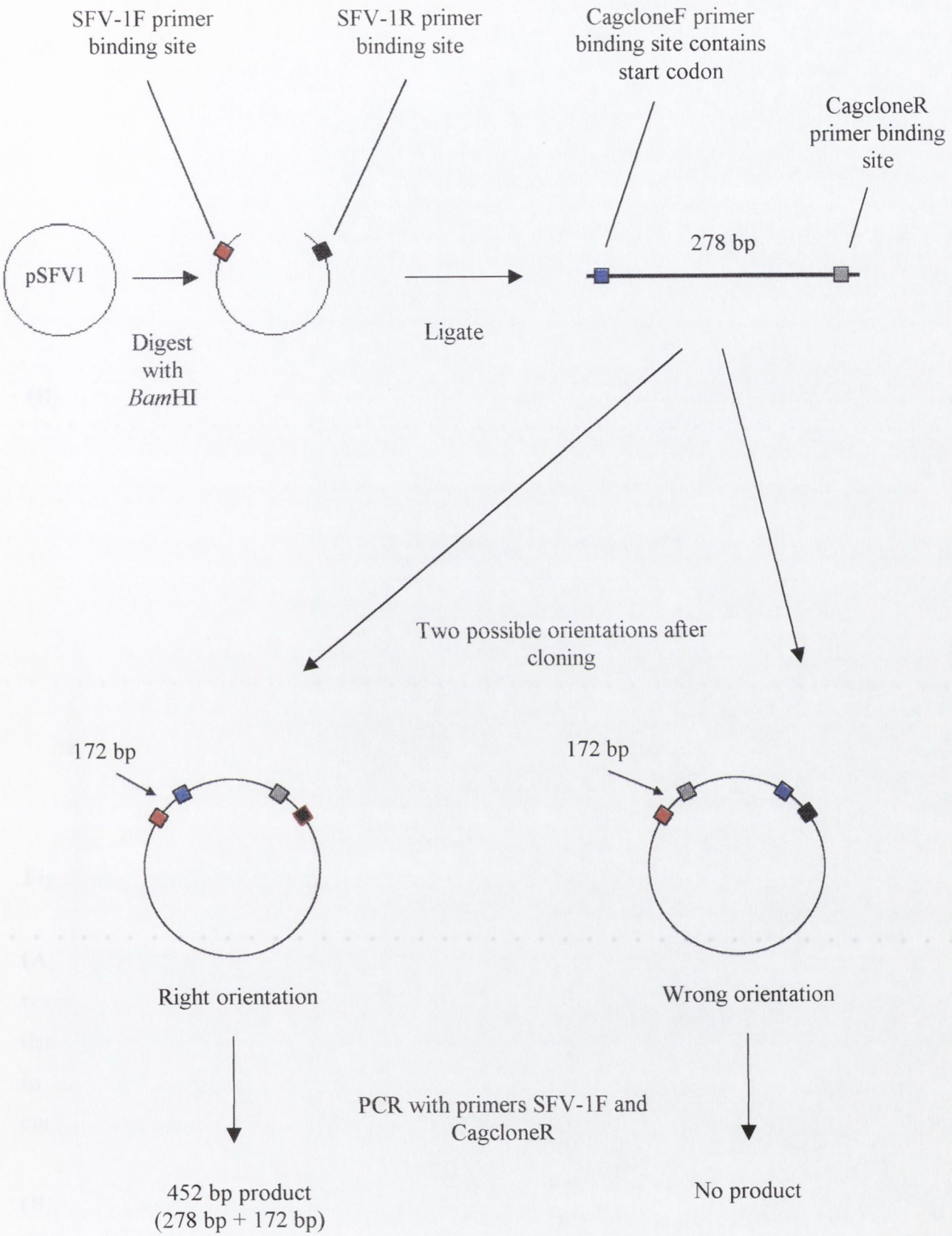
7.2.2.4. Sequence analysis of pSFV1 clones.

Although PCR can be informative as to whether a clone has been obtained and is in the correct orientation, it was necessary to sequence the clones to confirm (i) that the cloned DNA sequences did not contain any mutations as a result of the PCR, (ii) the presence or absence of EPIYA motifs, and (iii) that the sequences were in-frame with the start codon indicating that the polypeptide would be translated as predicted. Figs. 7.12 and 7.13 display the nucleotide

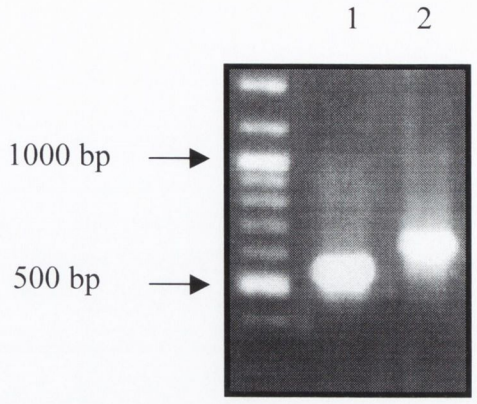
Fig. 7.10. PCR strategy to determine the right orientation of pSFV1 clones.

The pSFV1 vector and the fragment to be cloned are both cut with *Bam*HI. As both ends of vector and fragment have an affinity to combine the fragment can insert in one of two orientations. This is not desirable as one end of the fragment contains a start codon (ATG) and transcription and translation need to be driven in this direction. Thus, as fragments can insert in either orientation a screening method was needed to determine if the clones were in the right orientation.

From this diagram it can be seen that the primer binding sites change position depending on the orientation of the cloned fragments. This allows PCR to amplify the whole fragment and a section of the pSFV1 vector using primers SFV-1F and CagcloneR. If the fragment is in the correct orientation, a 452-bp PCR product will be obtained. If the fragment is in the incorrect orientation, no PCR product will be obtained, as each primer will act as a forward primer amplifying nothing.


SFV-1F primer binding site
SFV-1R primer binding site
CagcloneF primer binding site contains start codon
CagcloneR primer binding site
pSFV1
Digest with *Bam*HI
Ligate
278 bp
Two possible orientations after cloning
172 bp
172 bp
Right orientation
Wrong orientation
PCR with primers SFV-1F and CagcloneR
452 bp product (278 bp + 172 bp)
No product

(A)



(B)

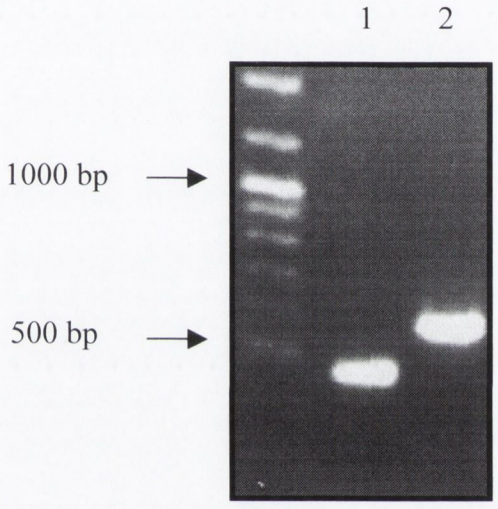


Fig. 7.11. PCR products run on a 1.8% agarose to confirm orientation of clones.

(A). Determining the orientation of clones SFV45688A and SFV45688C, lanes 1 and 2, respectively. Primers SFV-1F and CagcloneR (Table 6.1) were used in combination to determine the orientation of each clone. A 554-bp PCR product indicates that SFV45688A is in the correct orientation and a 656-bp PCR product indicates that SFV45688C is in the correct orientation.

(B). Determining the orientation of clones SFV512 and SFV541, lanes 1 and 2, respectively. Primers SFV-1F and CagcloneR (Table 6.1) were used in combination to determine the orientation of each clone. A 452-bp PCR product indicates that SFV512 is in the correct orientation and a 554-bp PCR product indicates that SFV541 is in the correct orientation.

Fig. 7.12. Nucleotide and translated amino acid sequences and alignment of amino acid sequences for clones SFV512 and SFV541.

(A). Nucleotide (278 bp) and translated amino acid (91 residues) sequences of cloned DNA inserted into clone SFV512 between primer sites CagcloneF and CagcloneR. Primer binding sites are in bold; the start codon contained within CagcloneF is in italics. The *cagA* repeat region within isolate MI512 does not contain a 102-bp repeat sequence. However, it contains a truncated form of this repeat sequence (blue) that is characteristic of Irish and non-Asian isolates. The amino acid sequence of this clone contains two tyrosine residues (Y). Both of these residues do not fit the consensus sequence for a tyrosine phosphorylation motif (TPM). The *Spe*I site is indicated (bold italics, A/CTAGT).

(B). Nucleotide (380 bp) and translated amino acid (125 residues) sequences of cloned DNA inserted into clone SFV541 between primer sites CagcloneF and CagcloneR. Primer binding sites are in bold; the start codon contained within CagcloneF is in italics. The *cagA* repeat region within isolate MI541 contains one 102-bp repeat sequence. It also contains the truncated form of the repeat sequence (blue). The amino acid sequence of this clone contains three tyrosine residues. Two of these residues, as in clone SFV512, do not fit the consensus sequence for a tyrosine phosphorylation motif (TPM). However, it does contain an EPIYA motif (black box) within its repeat region that can be phosphorylated. The *Spe*I site is indicated (bold italics, A/CTAGT).

(C). Amino acid alignment of translated polypeptides of clones SFV512 and SFV541. There is no difference in the amino acid sequence of each cloned fragment, except for the presence of a repeat sequence in SFV541.

(A).

GCACC*ATG*GGGCAAGCAGCGGGCTTCCCTTTGAAAAGGCATGATAAAGTTGATGATCTCAGTAAGGTAGG
GCTTTCAAGGAATCAAGAATTGGCTCAGAAAATTGACAATCTCAATCAGGCGGTATCAGAAGCTAAAACA
TGTCATTTTGGCAACCTAGATCAAATGATAGACAAGCTCAAAGATTCTACAAAAAAGAATGTTATGAATC
TATATGTTGAAAGTGCAAAAAAAGTGCCT***ACTAGT***TTGTCAGCGAAACTAGACAATTACTAATAATGA

M G Q A A G F P L K R H D K V D D L S K V G L S R N Q E L A Q K I D N
L N Q A V S E A K T C H F G N L D Q M I D K L K D S T K K N V M N L Y
V E S A K K V P T S L S **A K L D N Y S S S**

(B).

GCACC*ATG*GGGCAAGCAGCGGGCTTCCCTTTGAAAAGGCATGATAAAGTTGATGATCTCAGTAAGGTAGG
GCTTTCAGCTAGCCCTGAACCCATTTACGCTACGATTGATGATCTCGGCGGACCTTTCCCTTTGAAAAGG
CACGATAAAGTTGATGATCTCAGTAAGGTAGGGCTTTCAAGGAATCAAGAATTGGCTCAGAAAATTGACA
ATCTCAATCAGGCGGTATCAGAAGCTAAAACATGTCATTTTGGCAACCTAGATCAAATGATAGACAAGCT
CAAAGATTCTACAAAAAAGAATGTTATGAATCTATATGTTGAAAGTGCAAAAAAAGTGCCT***ACTAGT***TTG
TCAGCGAAACTAGACAATTACTAATAATGA

M G Q A A G F P L K R H D K V D D L S K V G L S A S P E P I Y A T I D
D L G G P F P L K R H D K V D D L S K V G L S R N Q E L A Q K I D N L
N Q A V S E A K T C H F G N L D Q M I D K L K D S T K K N V M N L Y V
E S A K K V P T S L S **A K L D N Y S S S**

(C).

```
SFV512          MGQAAG----------------------------------FPLKRHDKVDDLSKVGLSKN
SFV541          MGQAAGFPLKRHDKVDDLSKVGLSASPEPIYATIDDLGGPFPLKRHDKVDDLSKVGLSKN
                ******                                    ********************

SFV512          QELAQKIDNLNQAVSEAKTCHFGNLDQMIDKLKDSTKKNVMNLYVESAKKVPTSLSAKLD
SFV541          QELAQKIDNLNQAVSEAKTCHFGNLDQMIDKLKDSTKKNVMNLYVESAKKVPTSLSAKLD
                ************************************************************

SFV512          NYSSS
SFV541          NYSSS
                *****
```

Fig. 7.13. Nucleotide and translated amino acid sequences and alignment of amino acid sequences for clones SFV45688A and SFV45688C.

(A). Nucleotide (380 bp) and translated amino acid (125 residues) sequences of cloned DNA inserted into clone SFV45688A between primer sites CagcloneF and CagcloneR. Primer binding sites are in bold; the start codon contained within CagcloneF is in italics. The *cagA* repeat region within isolate 45688A contains one 102-bp repeat sequences and the truncated form of this repeat sequence (blue). The amino acid sequence of this clone contains two tyrosine residues (Y). One of these residues does not fit the consensus sequence for a tyrosine phosphorylation motif (TPM). However, the repeat sequence contains an EPIYA motif (black box), that can be phosphorylated. This sequence does not contain an *Spe*I site.

(B). Nucleotide (482 bp) and translated amino acid (159 residues) sequences of cloned DNA inserted into clone SFV45688C between primer sites CagcloneF and CagcloneR. Primer binding sites are in bold; the start codon contained within CagcloneF is in italics. The *cagA* repeat region within isolate 45688C contains two 102-bp repeat sequences (red and green). It also contains the truncated form of the repeat sequence (blue). The amino acid sequence of this clone contains five tyrosine residues. One of these residues (similarly to SFV45688A) does not fit the consensus sequence for a tyrosine phosphorylation motif (TPM). However, there are three tyrosine residues located in this sequence that are not in SFV45688A. One of these is expected (located within the repeat region), yet two are present that were not expected and have probably arisen by spontaneous mutation. Also, the two EPIYA motifs located within the two repeat sequences do not fit the consensus sequence completely (blue boxes), one or both of which might become phosphorylated. The *Spe*I site is indicated (bold italics, A/CTAGT).

(C). Amino acid alignment of clones SFV45688A and SFV45688C. These clones share 97% amino acid identity, excluding the presence of an extra repeat sequence in SFV45688C.

(A).

GCACC*ATG*GGGGCAAGCAGCGGGCTTCCCTTTGAAAAGGCATGATAAAGTTGATGATCTCAGTAAGGTAGG
GCGATCGGTTAGCCCTGAACCCATTTATGCTACGATTGATGATCTCGGCGGACCTTTCCCTTTGAAAAGG
CATGATAAAGTTGATGATCTCAGTAAGGTAGGGCTTTCAAGGAATCAAGAATTGACTCAGAAAATTGACA
ATCTCAATCAAGCGGTATCAGAAGCTAAAGCAGGTTTTTTTGGCAATCTAGAACAAACGATAGACAAGCT
CAAAGATTCTACAAAACACAATCCCATGAATCTATGGGTTGAAAGTGCAAAAAAAGTGCCTGCTAGTTTG
TCAGCGAAACTAGACAATTACTAATAATGA

M G Q A A G F P L K R H D K V D D L S K V G R S V S P E P I Y A T I D
D L G G P F P L K R H D K V D D L S K V G L S R N Q E L T Q K I D N L
N Q A V S E A K A G F F G N L E Q T I D K L K D S T K H N P M N L W V
E S A K K V P A S L S A K L D N Y S S S

(B).

GCACC*ATG*GGGGCAAGCAGCGGGCTTCCCTTTGAAAAGGCATGATAAAGTTGATGATCTCAGTAAGGTGGG
GCTTTCAGCTAGCCCTGAACCCATTTATGCTACGATTGATGATCTCGGCGGACCTTTCCCTTCGAAAAGG
TATGATAAAGTTGATGATCTCAGTAAGGTAGGGCTTTCAGCTAACCATGAACCCATTTACGTTACGATTG
ATGATCTCGGCGGACCTTTCCCTTTGAAAAGGTATGATAAAGTTGATGATCTCAGTAAGGTAGGGCTTTC
AAGGAATCAAGAATTGACTCAGAAAATTGACAATCTCAATCAAGCGGTATCAGAAGCTAAATCAGGTTTT
TTTGGCAATCTAGAGCAAACGATAGACAAGCTCAAAGATTCTACAAAACACAATCCCATGAATCTATGGG
TTGAAAGTGCAAAAAAAGTGCCT*ACTAGT*TTGTCAGCGAAACTAGACAATTACTAATAATGA

M G Q A A G F P L K R H D K V D D L S K V G L S A S P E P I Y A T I D
D L G G P F P S K R Y D K V D D L S K V G L S A N H E P I Y V T I D D
L G G P F P L K R Y D K V D D L S K V G L S R N Q E L T Q K I D N L N
Q A V S E A K S G F F G N L E Q T I D K L K D S T K H N P M N L W V E
S A K K V P T S L S A K L D N Y S S S

(C).

```
SFV45688A           MGQAAGFPLKRHDKVDDLSKVGRSVSPEPIYATIDDLGGP--------------------
SFV45688C           MGQAAGFPLKRHDKVDDLSKVGLSASPEPIYATIDDLGGPFPSKRYDKVDDLSKVGLSAN
                    ********************** *.***************

SFV45688A            --------------FPLKRHDKVDDLSKVGLSRNQELTQKIDNLNQAVSEAKAGFFGNLE
SFV45688C            EEPIYVTIDDLGGPFPLKRYDKVDDLSKVGLSRNQELTQKIDNLNQAVSEAKSGFFGNLE
                           *****:**********************************:*******

SFV45688A            QTIDKLKDSTKHNPMNLWVESAKKVPASLSAKLDNYSSS
SFV45688C            QTIDKLKDSTKHNPMNLWVESAKKVPTSLSAKLDNYSSS
                     ***************************:************
```

and translated amino acid sequences of each clone. The sequences shown are the nucleotides present between the primers used for the cloning reactions (CagcloneF and CagcloneR). These primers can be seen to flank each repeat sequence. With each clone the *cagA* repeats were amplified correctly and the number of repeats predicted to be in each isolate were also correct. All clones contain a tyrosine residue that is located within the reverse primer sequence. However, this residue is outside the *cagA* repeat sequences and is not part of the EPIYA. It also does not fit the TPM consensus sequence and therefore would be predicted not to become phosphorylated.

In clones SFV512 and SFV541 there is another tyrosine residue that is located outside the repeat sequences. This residue also does not conform to a TPM consensus sequence and therefore also would be predicted not to become phosphorylated. Thus, both clones (SFV512 and SFV541) contain two tyrosine residues that are located in the same position in both clones that potentially may not become phosphorylated. The fact that these residues are conserved between two isolates further confirms how these isolates are ancestrally related. Except for the presence of a *cagA* repeat sequence in SFV514, both clones share 100% homology at the amino acid level. Although both clones share two tyrosine residues at the same position, clone SFV512 lacks repeat sequences while clone SFV541 contains one *cagA* repeat sequence. As mentioned previously these *cagA* repeat sequences contain an EPIYA motif. The sequence of clone SFV541 shows that within its repeat sequence an EPIYA motif is present. Thus, clone SFV541 contains one putative TPM and clone SFV512 lacks a putative TPM.

Clones SFV45688A and SFV45688C are similar to clones SFV512 and SFV541 with respect to the tyrosine residue located within the reverse primer sequence. However, the second tyrosine residue located outside the repeat sequence in clones SFV512 and SFV514 are not present in clones SFV45688A and SFV45688C. Interestingly, with clones SFV512 and SFV541 the only difference between them on a sequence level was the presence of a 102-bp *cagA* repeat sequence in clone SFV514. However, although clone SFV45688C contains an extra repeat sequence compared to clone SFV45688A, it also displays a slightly different amino acid sequence and contains three more tyrosine residues than clone SFV45688A. In total clone SFV45688A contains five tyrosine residues. Two of these are EPIYA motifs located in its two repeat sequences, one is within the reverse primer sequence, one is outside the repeat sequences and the last is located within the second repeat sequence nineteen amino

acid residues before the tyrosine residue within the EPIYA motif. Only the tyrosine residues in the EPIYA motifs located in the *cagA* repeat sequences fit the consensus sequence for a TPM, and the other tyrosine residues should not be phosphorylated.

7.2.3. in vitro *transcription of the pSFV1 clones.*

During the SFV replicative cycle the initial stage involves the virus passing its RNA genome into a host cell (Fig. 7.2). To mimic this *in vitro* it was necessary to transcribe the pSFV1 clones. The first step in this process was to linearise the four clones with *Spe*I (the pSFV1 vector contains one *Spe*I site). Thus, it is favourable that cloned fragments do not contain *Spe*I sites, as these would interfere with linearising the pSFV1 clone. Unfortunately, the cloned sequences in SFV512, SFV541 and SFV45688C all contained one *Spe*I site (Fig. 7.12 A and B, Fig. 7.13 B). This meant that this enzyme would cut the cloned sequences resulting in truncated transcripts. The *Spe*I site has always been used for linearisation of the pSFV1 vector, as it is necessary to leave a poly-A tail at the end of the transcript that is important for processing and protecting the viral RNA (Strauss and Strauss, 1994). Linearisation with another enzyme would result in an RNA transcript without a poly-A tail and heterologous gene expression would not be carried out to full capacity.

To overcome this problem a partial digestion of the pSFV1 clones were carried out with restriction endonuclease *Spe*I (Chapter 2). This would generate linear clones cut at the pSFV1 *Spe*I site, linear clones cut at the *Spe*I site within the cloned fragments and clones cut at the two *Spe*I sties within the vector and the cloned sequence. Thus, *in vitro* transcription could be carried out resulting in three types of mRNA transcript. However, only one type of transcript would be processed once inside the host cell. Partial digestion of each clone and an unmodified pSFV1 vector was carried out using *Spe*I (Fig. 7.14.). As the vector is large and the inserts are small, the digested clones look approximately the same size. However, as this was a partial digestion, other bands (undigested vector) can be seen at low intensity. *In vitro* transcription using RNA polymerase was carried out on these digested clones and the mRNA generated was run on a gel (Fig. 7.15). The transcribed RNA was then run on a gel to check the integrity of the RNA and to determine the size of the transcripts (Fig. 7.15). From this gel it can be seen that the RNA transcripts did not vary much in size. Again the reason for this could be that there is very little size difference between each clone. If a situation arose where *in vitro* transcription only transcribed a linear vector that was cut within the insert (partial digestion), a much smaller RNA transcript would be produced as an extra ~ 860 bases

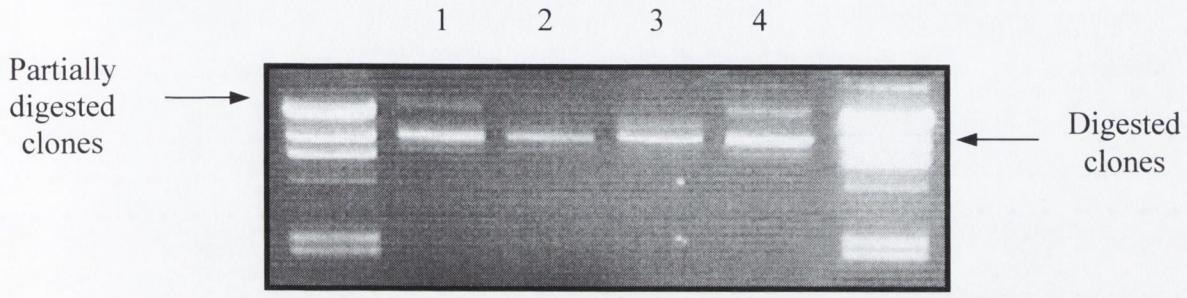


Fig. 7.14. Partial digestion of pSFV1 clones with *Spe*I run on a 1% agarose gel.

The partial digest of clones SFV512, SFV541, SFV45688A and 45688C using *Spe*I are in lanes 1, 2, 3 and 4, respectively. From lane 1 and 4 it can be seen that not all the vector is digested as other bands can be seen in these lanes. The ladder on the left contains 5 μg DNA and the ladder on the right contains 10 μg DNA, this was used as standards to estimate how much template to use for *in vitro* transcription. All clones are linear and are approximately the same size.

1 2 3 4 5

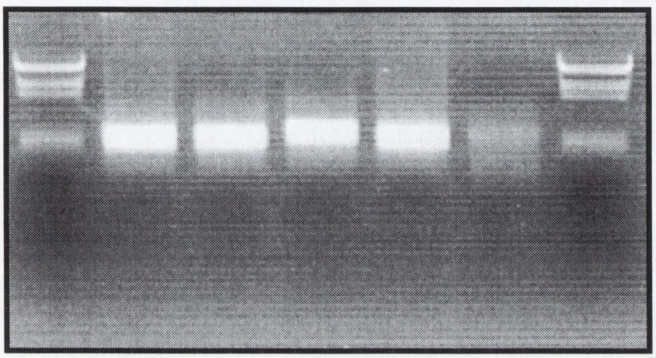


Fig. 7.15. *in vitro* transcripts of clones separated on a 1% agarose gel.

The clones SFV512, SFV541, SFV45688A and SFV45688C and the pSFV1 empty vector (lanes 1-5, respectively) were transcribed *in vitro* and run on an agarose gel to check the integrity of the transcripts. As the empty vector sample is considerably less concentrated than the clones, a larger volume of this sample was used for transfection.

between the multiple cloning site and the *Spe*I site. As all transcripts are relatively the same size as that of the empty vector control, it can be assumed that the *in vitro* transcription was successful.

7.2.4. Growth and transfection of AGS cells.

AGS cells were grown as previously described (Stein, *et al.*, 1999). As AGS cells are derived from a gastric carcinoma, they provide a system that closely mimics the interactions between *H. pylori* and epithelial cells *in vivo*. Initial growth was from a monolayer at passage 12. This monolayer was split into three different flasks and allowed to grow for three days until the cells in each flask were confluent. The cells were split three times again to achieve nine flasks of cells. Once the nine flasks of cells were confluent they were washed with DMSO and stored in liquid nitrogen. Thus, nine stocks of AGS cells at passage 14 were available for further experimentation.

Once the mRNA samples from the four pSFV1 clones and the empty vector had been prepared, frozen AGS cells were split and grown twice to yield six confluent monolayers at passage 14. Each monolayer was electroporated with an equal amount of either mRNA from one of the four clones or mRNA from an empty vector or PBS. After electroporation each transfected cell suspension was grown in a new flask. As pSFV1 mRNA causes cell death after 48 h, the cells were harvested after 24 h. Each transfected culture was approximately 70% confluent, indicating that most cells survived the electroporation process and the uptake of foreign mRNA.

7.2.5. Analysis of cellular proteins within transfected AGS cells.

7.2.5.1. Analysis of protein fractions from transfected AGS cells.

Total protein was isolated from the six transfected AGS cell cultures. Four of these cell cultures represented transfection with pSFV1 clones containing various *cagA* repeats from *H. pylori* isolates, one represented transfection with an empty pSFV1 vector and one was a negative PBS control. A lysis buffer that contained protease inhibitiors and phosphatase inhibitors was used to obtain total cellular protein from each culture. The concentration of each protein preparation was determined by a Bradford™ assay. The protein concentration for each preparation was equalised, and 20 µg of total cellular protein from each sample was analysed by SDS-polyacrylamide gel electrophoresis.

As 1 kb of DNA is equal to 37 kDa protein the potential sizes of the translated recombinant sequences could be predicted, namely SFV512 = 10.3 kDa, SFV541 and SFV45688A = 14 kDa and SFV45688C = 16.8 kDa. The fractionated total cellular protein for each pSFV1-transfected sample and controls can be seen in Fig. 7.16. From this figure it is evident that equal amounts of protein were fractionated for each sample. Within the molecular mass range of 7–17 kDa there are some polypeptide bands that are present at a higher intensity in cells electroporated with recombinant pSFV1 mRNA than in the negative control. However, the empty vector control also displayed these high intensity proteins that were not present in the negative control. This implies that, although the pSFV1 vector did not contain any *H. pylori* sequences, it had an effect on the protein profile of AGS cells.

However, when analysing the proteins that were present in the transfected cells but not in the negative control, there was one band that was of high intensity in all the cells electroporated mRNA from pSFV1 clones but of low intensity in cells electroporated with mRNA from the empty vector control. This implies that all cells electroporated with recombinant mRNA containing *H. pylori* sequences had a different effect on AGS cellular protein profiles than cells not transfected with mRNA and cells electroporated with empty vector pSFV1 mRNA. However, the predicted protein sizes encoded by the *cagA* 3′-repeat sequences ranged from approximately 10–17 kDa and yet there is no size variation between these high intensity 12-kDa bands. This suggested that these 12-kDa polypeptides may not be the recombinant polypeptides of interest. These bands were transferred to a nitrocellulose membrane and probed with an anti-CagA polyclonal antibody. This antibody did not bind to any polypeptides in the AGS protein samples suggesting that no CagA 3′-repeat polypeptides were present (data not shown). However, it is possible that putatively expressed recombinant polypeptides from C-terminal domains of the CagA protein do not possess the epitopes recognised by the anti-CagA antibody preparation used.

7.2.5.2. Tyrosine phosphorylation analysis of transfected AGS cells.

To define if the tyrosine phosphorylation patterns of AGS cells were affected by transfection with recombinant mRNA, a Western blot with an anti-tyrosine phosphorylation antibody was carried out on the six samples of total cellular protein from AGS cultures, four of which had been electroporated with recombinant mRNA (SFV512, SFV541, SFV45688A and SFV45688C), one of which had been electroporated with mRNA from an empty vector and one of which had been electroporated with PBS (Fig. 7.17). The three of these over-expressed

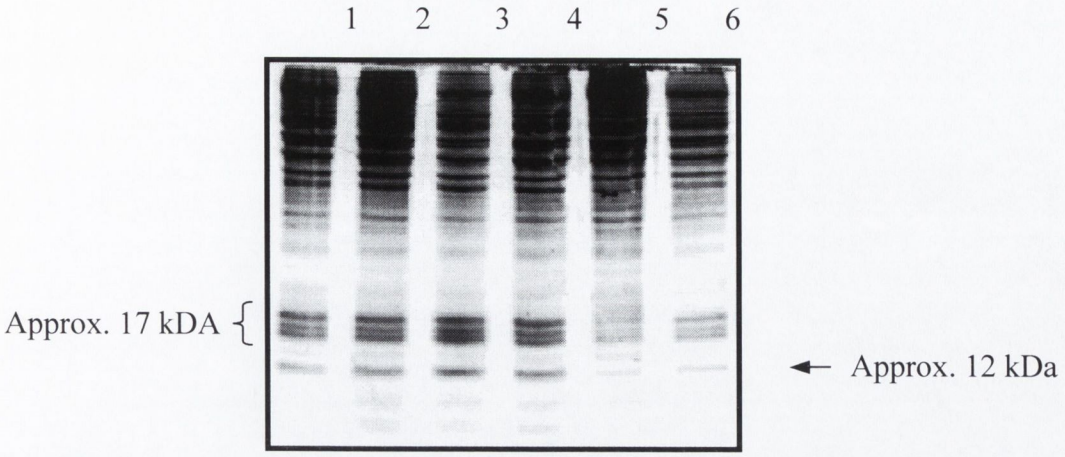


Fig. 7.16. Total cellular protein of AGS cells, electroporated with or without recombinant pSFV1 mRNA, fractionated on a 15% polyacrylamide gel by SDS-PAGE.

AGS cell cultures were electroporated with mRNA from the pSFV1 clones, mRNA from an empty vector or PBS (negative control, lane 5). Total protein from SFV512, SFV541, SFV45688C, SFV45688A and pSFV1 (empty vector) transfected AGS cells was run in lanes 1–4 and lane 6, respectively. All lanes contain equal amount of total cellular protein (20 μg). From analysis of the gel there are approximately three polypeptides that are over-expressed in the cultures treated with recombinant pSFV1 mRNA and also in that treated with the empty vector, although to a lesser extent (17 kDa range), compared to the non-transfected control cells. There is also a polypeptide of approximately 12 kDa that is over-expressed in the recombinant SFV-transfected cells (lanes 1–4) compared to the empty vector transfected cells (lane 6). This suggests that the expression of the proteins within the 17-kDa range is affected by pSFV1 mRNA and that the 12-kDa polypeptide is affected by the presence of *cagA* 3′-end polypeptide sequences.

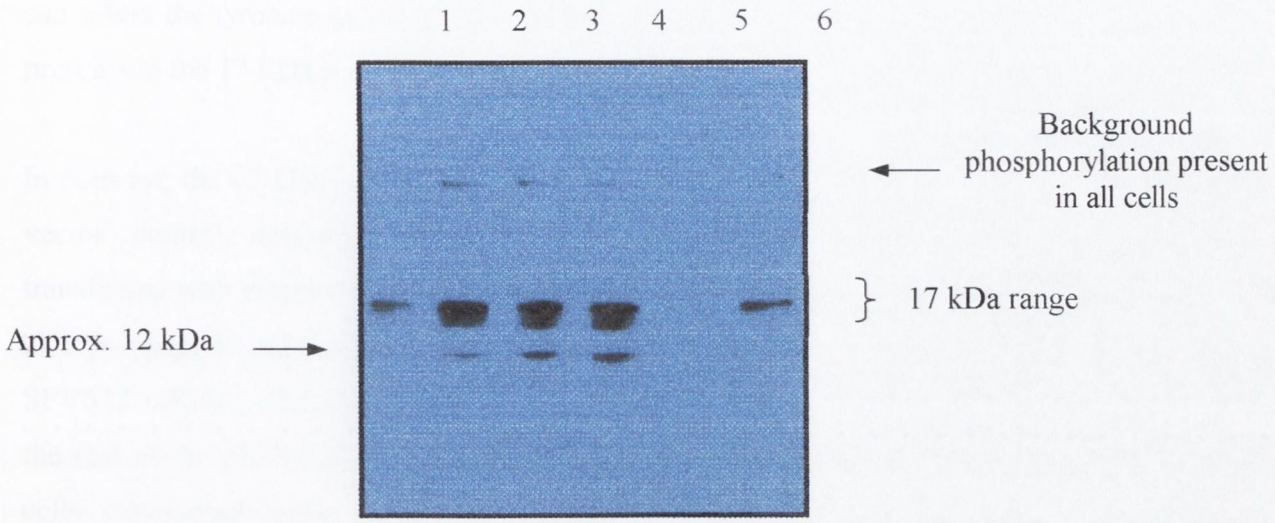


Fig. 7.17. Anti-PY Western blot of AGS cells transfected with and without recombinant pSFV1 mRNA.

The total cellular protein of AGS cells electroporated with recombinant pSFV1 mRNA was fractionated on a 15% polyacrylamide gel and then transferred to a PVDF membrane. The membrane was then treated with an anti-PY antibody. Lanes 1–4 contain total cellular protein from AGS cells electroporated with recombinant pSFV1 mRNA (SFV512, SFV541, SFV45688A and SFV45688C, respectively), lane 5 contains total cellular protein from AGS cells electroporated with PBS and lane 6 contains total cellular protein from AGS cells treated with pSFV1 RNA. The above phosphorylation pattern shows that the over-expressed polypeptides in the 17-kDa range were not only tyrosine phosphorylated in the cells treated with recombinant mRNA but also in the cells treated with 'empty vector' pSFV1 mRNA. This further suggests that the polypeptides within the 17-kDa range were pSFV1 associated proteins. However, the 12-kDa polypeptide that was present in the AGS cells tranfected with recombinant pSFV1 mRNA from SFV541, SFV45688A and SFV45688C is tyrosine phosphorylated whereas polypeptides of similar molecular mass (Fig. 7.17) present in cells transfected with 'empty vector' pSFV1 mRNA or SFV512 mRNA were not tyrosine phosphorylated.

polypeptides in the 17-kDa range were found to be tyrosine phosphorylated in all cells transfected with mRNA including the empty pSFV1 vector (Fig. 7.17). This indicates that when using pSFV1 as a mammalian heterologous expression system, pSFV1 mRNA alone can affect the tyrosine phoshporylation pattern of AGS cells. Thus, the three over-expressed proteins in the 17-kDa range were affected by pSFV1 alone.

In contrast, the 12-kDa polypeptide was not detected by the anti-PY antibodies in the 'empty vector' control, indicating that it did not become tyrosine phosphorylated in AGS cells transfected with empty vector pSFV1 mRNA. From Fig. 7.17 it can also be seen that this 12-kDa polypeptide did not become tyrosine phosphorylation within cells transfected with clone SFV512 mRNA, whereas this protein was tyrosine phosphorylated in cells transfected with the rest of the pSFV1 clones (SFV541, SFV45688A and SFV45688C). The 12-kDa band in cells transfected with the pSFV1 recombinants bearing one *cagA* 3′-repeat (SFV541, SFV45688A) showed approximately the same degree of tyrosine phosphorylation based on labelling with an anti-tyrosine antibody. In contrast the 12-kDa band in AGS cells transfected with the SFV45688C clone possessing two *cagA* repeats, was somewhat more intense than observed with the SFV541- and SFV45688A- transfected cells.

7.3. Discussion.

7.3.1. Tyrosine phosphorylation motifs within H. pylori cagA *3'- repeat sequences.*

As mentioned previously, variation in the number of *cagA* 3'-repeat sequences was found to occur between two *H. pylori* isolates recovered from the same individual that showed identical RAPD-PCR profiles (Chapter 5). The overall effect or advantage that this phenomenon has on *H. pylori* isolates remains unclear. However, with recent advances made regarding the functions of the CagA protein, the effect of changing the number of *cagA* 3'-repeat sequences within an *H. pylori* population may as yet be revealed. Sequence analysis of the 3'-end of the CagA protein reveals the presence of an EPIYA motif localised within each repeat region within Western and European *H. pylori* isolates (Covacci *et al.*, 1993; Tummuru *et al.*, 1993). Evans and Evans (2001) stated that a tyrosine phosphorylation motif (TPM) complies with the consensus sequence [RK]-X(2,3)-[DE]-X(2,3)-[Y], and that the CagA protein may have up to three TPM sequences, one near the N-terminus [QKFGDQRY, Y-122 in the 26695 genome (Tomb *et al.*, 1997)], one located after the poly-N motif [KNSTEPIY, Y-899 in the 26695 genome (Tomb *et al.*, 1997)] and a poorly conserved TPM located near the C-terminus [DKLKDSTKY, Y-1039 in the 26695 genome (Tomb *et al.*, 1997)].

The amino acid sequences preceding Y-899 fit the consensus sequence for a TPM and interestingly this TPM contains an EPIYA motif, a motif that is also found within CagA 3'-repeat sequences. As the sequences surrounding the EPIYA motif located within these repeat sequences did not fit the TPM consensus sequence [SASPEPIY, Y-972 in the 26695 genome (Tomb *et al.*, 1997)], it was assumed that the EPIYA motif was not a functional TPM. However, not all TPM sequences adhere to the consensus sequence stated by Evans and Evans (2001). In fact, Src kinases phosphorylate the consensus sequence EEIY[G/E] (Songyang *et al.*, 1995; Hunter 2000). Thus, although the EPIYA motif does not fit the consensus sequence completely, there is a striking similarity. Based on the idea that the EPIYA motif within the CagA repeat could be a sequence that is phosphorylated, it was found that indeed this motif (Y-972) was not only tyrosine phosphorylated (Backert *et al.*, 2001) but was also an essential motif for causing the CagA-induced cytoskeletal rearrangements previously observed by Segal *et al.* (1996).

Therefore, knowing that the EPIYA motif located within CagA repeats is an essential TPM and that ancestrally related *H. pylori* isolates with identical RAPD-PCR profiles display variation in the number of 3'-repeats they possess, the question can be posed as to whether a

microevolved population alters its tyrosine phosphorylation potential by gaining or loosing a *cagA* repeat sequence. With a view to investigating this, analysis of the *cagA* 3′-repeat sequences that encode these motifs was carried out on two pairs of ancestrally related *H. pylori* isolates that displayed variation in the number of repeat sequences they possessed, to determine whether variation in the number of repeat sequences within these pairs varied their tyrosine phosphorylation capability.

The approach used to determine the phosphorylation patterns of these repeat sequences was to employ a well-characterised animal cell vector (pSFV1). The pSFV1 vector is a tool used for producing heterologous proteins in mammalian cells, whereby foreign proteins consist of up to 25% of total cellular protein after transfection (Liljeström and Garoff, 1993). As the protein sequences under investigation were defined regions of the total CagA protein, it was desirable to produce a large quantity to facilitate analysis, hence the use of the pSFV1 vector. Thus, *cagA* repeats from each of two pairs of *H. pylori* isolates were cloned in to the pSFV1 vector. As only the sequences that encoded the EPIYA motifs were to be analysed, the sequences preceding the 102-bp repeats were not cloned as they also contain intact TPM sequences. As a result the sequences cloned encoded small putative polypeptides (MI512 - 10.3 kDa, MI541 and 45688A - 14 kDa and 45688C - 16.8 kDa).

Cloning the various *cagA* 3′-repeat fragments posed problems due to the large size of the pSFV1 vector and the small size of the DNA sequences to be cloned. A ratio of 10:1 *cagA* fragment to pSFV1 vector DNA was needed to ensure efficient cloning, but this was in itself a problem as the concentrations of the DNA fragments obtained after digestion with restriction endonuclease *Bam*HI were low. As the cloning procedure was not site-directed and problems relating to vector religation and incorrect orientation of inserts exist, the cloning frequency obtained was low. However, all the target DNA sequences encoding *cagA* repeats were cloned and confirmed by sequence analysis.

7.3.2. Sequence analysis of pSFV1 clones.

The cloned *cagA* 3′-repeat sequences encompassed the forward primer sequence, the 102-bp repeats, the 54-bp partial repeat sequence (characteristic of Western *H. pylori* isolates) and a 201-bp sequence downstream of the 102-bp repeat sequences that also contained the reverse primer sequence. The downstream sequences were included so as to make the final heterologous polypeptide larger, as analysis of the *cagA* sequences of genome-sequenced

strains J99 and 26995 had indicated that no functional TPM sequences were present within this downstream sequence.

The two tyrosine residues present in the amino acid sequence of the SFV512 clone did not fit the consensus sequence of a functional TPM and showed no resemblance to an EPIYA or EEIY[G/E] (SLSAKLDNYS) motif. It was, therefore, unlikely that either tyrosine residue would be phosphorylated, and as no other tyrosine residues were present, no tyrosine phosphorylation would be expected to be detected within this recombinant polypeptide. On the other hand, the *cagA* 3′ -repeat sequences of its paired isolate showed three tyrosine residues, two identical to the two residues in clone SFV512 and one being the EPIYA motif located within the repeat sequence. Therefore, as the EPIYA motif had been identified as an important TPM, phosphorylation of this motif was expected to occur within this domain.

The amino acid sequences of clones SFV45688A and SFV45688C showed a different pattern of tyrosine variation. Amino acid sequence analysis of these two clones displayed 96% sequence similarity within conserved regions (excluding the extra repeat), unlike the other clones (SFV521 and SFV541) that displayed 100% identity within conserved regions. One of these five amino acid changes within the conserved regions of clones SFV45688A and SFV45688C (Fig. 7.13) was a change of a histidine residue (45688A) to a tyrosine residue (45688C). In addition, within the second repeat sequence in isolate 45688C (not found in 45688A) there was a tyrosine residue where a histidine residue was present in the first repeat sequence. So, instead of having just one extra tyrosine residue (EPIYA motif) associated with the extra repeat not found in isolate 45688A, isolate 45688C has three more tyrosine residues within this region. However, the two extra unexpected tyrosines do not fit the consensus sequence for a TPM, and therefore were not expected to affect the phosphorylation pattern for these repeats.

The changes of these two histidine residues in isolate 45688A to tyrosine residues in isolate 45688C were brought about by single base changes, which is a common occurrence in *H. pylori*. Both of these single base mutations were a C→T change in the codon CAT (histidine) to TAT (tyrosine). These changes represent a change of an amino acid with a basic side chain (histidine) to an amino acid with an uncharged side chain (tyrosine). Both residues are relatively the same size and as the tyrosine residue is not part of a functional TPM these changes are relatively minor. The EPIYA motif within the second repeat sequence that was

only present in isolate 45688C does not completely fit the consensus sequence as the last residue is a valine (EPIYV) rather than an alanine (EPIYA). As both of these are non-polar amino acids, it was unknown whether this alteration would affect the phosphorylation potential of this motif.

7.3.3. AGS cell transfection.

The pSFV1 viral vector is 11,033 bp in size and contains bacterial and viral nucleotide sequences. The bacterial sequences include an ampicillin resistance cassette and an origin of replication, while the viral sequences encode non-structural proteins nsP1–4 (Fig. 7.7), and include the multiple cloning site. When a sequence of choice has been inserted into the pSFV1 vector, *in vitro* transcription of the *Spe*I-cleaved linearised pSFV1 is carried out. However, not all the pSFV1 sequence is transcribed, as the bacterial sequences are not needed for production of heterologous proteins. The bacterial sequences are located after the *Spe*I site and before the SP6 promoter. Therefore, when transcribing an empty vector the sequence of the transcript should be approximately 8,262 bases long.

The transcript sizes of the clones SFV512, SFV541, SFV45688A and SFV45688C should be 8,540 bp, 8,642 bp, 8,642 bp and 8,744 bp, respectively. The size difference between the largest clone transcript (8,744 bp) and the empty vector (8,262 bp) is 482 bp. This size difference is small when compared to the size of the whole transcript, and was not expected to be detected on an agarose gel (Fig. 7.15). If the transcription of the *cagA* inserts had been unsuccessful leading to the production of an mRNA truncate then a transcript 860 bp shorter than the empty vector mRNA would arise due to a deletion between the MCS and the *Spe*I restriction endonuclease site located wihtin the cloned *cagA* sequences (Section 7.23). Since no size differences were detected between the recombinant mRNAs and the empty vector mRNA it was assumed that transcription was successful.

7.3.4. Analysis of polypeptides expressed in AGS cells.

This study demonstrated that there were four over-expressed polypeptides in the total cellular protein of AGS cells electroporated with recombinant pSFV1 mRNA (Fig. 7.16). However, only one of these polypeptides was found in the total cellular protein of AGS cells electroporated with recombinant pSFV1 mRNA that was weakly expressed in the total cellular protein of AGS cells electroporated with pSFV1 mRNA lacking a *cagA* repeat sequence. Thus, three of the over-expressed polypeptides (of apparent size ~ 17 kDa) were

pSFV1 vector-associated proteins and one polypeptide (apparent size ~ 12kDa) appeared to be influenced solely by the presence of CagA sequences. The latter polypeptide did not possess the predicted size for each recombinant CagA 3′-repeat polypeptide based on sequence analysis. Thus, the 12-kDa polypeptides could either be a degraded form of the recombinant CagA polypeptides or a host-cell protein whose expression is affected by the presence of recombinant CagA polypeptides. With respect to the former suggestion, indicated in Fig. 7.7, the alphavirus nsP2 proteins is multifunctional, possessing helicase and proteinase activities. The cleavage sites for alphavirus non-structural proteins in the processing of P1234 polyproteins have been characterised (Fig. 7.18; ten Dam *et al.*, 1999). Analysis of the amino acid sequences of the cloned CagA 3′-repeat regions (Figs. 7.11 and 7.12) did not reveal any potential cleavage sites resembling those for the pSFV1 nsP2 proteinase or those of other alphaviruses. It thus seems unlikely that the 12-kDa polypeptide represent nsP2-cleaved truncates of CagA 3′-repeat polypeptides. However, the possibility of proteolysis by AGS cell associated proteinases cannot be excluded. As a 12-kDa protein is present in the control samples at low levels, it would appear more likely that this polypeptide is a host-cell protein. Nonetheless, from analysis of AGS total cellular protein it can be concluded that there are proteins that are affected by electroporation of recombinant pSFV1 mRNA.

7.3.5. Analysis of tyrosine phosphorylation pattern of polypeptides expressed in AGS cells transfected with recombinant pSFV1 bearing 3′-end cagA *sequences.*

The phosphorylation pattern found after immunoblotting the total cellular proteins from transfected AGS cells with anti-phosphotyrosine antibodies appear to follow the predicted phosphorylation pattern for CagA repeat sequences (Fig. 7.17) based on the presence or absence of EPIYA motifs. However, it seemed unlikely that the 12-kDa polypeptides within the transfected AGS cells were encoded by the recombinant *H. pylori* mRNA sequences (Section 7.3.5). The three over-expressed proteins, that are assumed to be proteins expressed under the influence of pSFV1 sequences (either host proteins affected by pSFV1 proteins or pSFV1 translated proteins), are tyrosine phosphorylated in all cells transfected with SFV mRNA. The paired isolates MI512 and MI541 lacked and contained one EPIYA motif, respectively. The 12-kDa protein was not tyrosine phosphorylated within cells electroporated with MI512 *cagA* 3′-end sequences. However, this polypeptide was tyrosine phosphorylated in cells transfected with MI541 *cagA* sequences. As the CagA protein has also been shown to interfere with epithelial cell signal transduction pathways in a phosphorylation-dependent

nsP1			nsP2				Alphaviruses
P2	P1		P1'	P2'	P3'	P4'	No.
G	A	↓	A	L	V	E	4
G	A	↓	G	S	V	E	3
G	A	↓	G	V	V	E	2[a]
G	A	↓	G	I	V	E	2
G	A	↓	G	V	V	N	1
G	E	↓	G	V	V	E	1

nsP2			nsP3				Alphaviruses
P2	P1		P1'	P2'	P3'	P4'	No.
G	A	↓	A	P	S	Y	4
G	C	↓	A	P	S	Y	6[a]
G	R	↓	A	P	A	Y	2
G	S	↓	A	P	A	Y	1

nsP3			nsP4				Alphaviruses
P2	P1		P1'	P2'	P3'	P4'	No.
G	G	↓	Y	I	F	S	7
G	A	↓	Y	I	F	S	6[a]

[a] These represent the cleavage sites for the nsP2 proteinase of pSFV1.

Fig. 7.18. Polyprotein 1234 cleavage sites of nsP2 cysteine endopeptidases (proteinases) of alphaviruses.

The cleavage sites for nsP1–nsP2, nsP2–nsP3, and nsP3–nsP4 of thirteen alphaviruses are shown - Sindbis, O'Nyong-nyong, Ross River, Semliki Forest, Barmah Forest, Eastern Equine Encephalitis, Rubella, AURA, Chikungunya, Ockelbo, Igbo Ora, Middelburg and Whataroa - according to their frequencies of sequence occurrence. The cheavage sites atre characyteriused by a conserved glycxine residue in position P2 and the cleavage specificity is commonly between sjhort aliphatic amino acids. Conserved or highly comnserved amino acids are coloured. The cleavage sites are indicated by vertical arrows. Data adapted from ten Dam *et al.* (1999).

manner (Backert *et al.*, 2001), it is tempting to speculate that the presence and absence of EPIYA motifs may be affecting the phosphorylation of this 12-kDa polypeptide.

Isolates 45688A and 45688C contain one and two CagA repeat sequences, respectively, indicating the presence of one and two TPM sequences, respectively. Analysis of the phosphorylation pattern of the 12-kDa polypeptides in AGS cells transfected with *cagA* sequences from isolates 45688A and 45688C showed that both polypeptides were phosphorylated. However, the 12-kDa polypeptide in AGS cells tranfected with *cagA* sequences from isolate 45688C displayed a somewhat higher level of phosphorylation. This could have been due to the presence of the extra TPM sequence within the *cagA* sequence of isolate 45688C. The extra TPM located within the second repeat sequence in isolate 45688C does not, however, contain a full EPIYA motif (EPIYV). This suggests that the change from an alanine residue to a valine residue may not have an effect on the phosphorylation characteristic of EPIYA motifs.

Higashi *et al.* (2002) described the tyrosine phosphorylation potential of CagA as fluctuating depending on the number of *cagA* 3′-repeat sequences (each repeat encodes one EPIYA motif) present within unrelated *H. pylori* strains. These authors also demonstrated that strains with CagA proteins with more EPIYA motifs can more severely perturb the SHP-2-dependent signalling pathway. The report herein describes analysis of microevolved *H. pylori* isolates that contain different numbers of *cagA* 3′-repeat sequences and, although the phosphorylation status of microevolved *cagA* 3′-repeat sequences remains to be determined, it can be concluded that these microevolved *cagA* repeats exhibit differing effects on the phosphorylation pattern of AGS cell proteins. If *H. pylori* isolates can increase or decrease the level of phosphorylation of its CagA antigen via microevolution, this may also alter the potency of the CagA antigen. Thus, it is tempting to speculate that *H. pylori* may reduce the number of *cagA* 3′-repeat sequences, thereby lowering its virulence to behave more as a commensal. This is not inconsistent with the idea that *H. pylori* may be beneficial to humans. *H. pylori* has been shown to have antimicrobial properties (Pütsep *et al.*, 1999) and to prevent diarrhoea in children (Rothenbacher *et al.*, 2000).

Chapter 8

General Discussion

8.1. Discussion 143
8.1.1. Population structure of Irish* H. pylori *isolates 143
8.1.2. Microevolution of H. pylori *isolates* in vivo 147
8.2. Future prospects 150
8.3. Concluding remarks 151

8.1. Discussion.

8.1.1. Population structure of Irish H. pylori *isolates.*

One of the most distinctive features of *H. pylori* is the genetic diversity it displays between clinical isolates. This characteristic of *H. pylori* is due to its natural competence (Israel *et al.*, 2000) and its freely recombining population structure (Go *et al.*, 1996). The diversity displayed between clinical *H. pylori* isolates has provoked attempts to associate certain variants of virulence genes with diseases linked to this organism. However, attempts to associate genotypic traits with disease caused by *H. pylori* isolates have had little success and have been largely restricted to the *vacA* and *cagA* loci (Atherton *et al.*, 1995; Blaser *et al.*, 1995, Graham and Yamaoka, 2000).

Although cataloguing the genetic diversity of *H. pylori* provides little information predicting clinical outcome, it has contributed to the differentiation of *H. pylori* isolates from separate geographical locations. On the basis of the nucleotide sequences of various genes Asian strains can be distinguished from those of white populations in Europe, North America, and South Africa (Suerbaum *et al.*, 1998, van der Ende *et al.*, 1998, Achtman *et al.*, 1999, Kersulyte *et al.*, 1999). Northern and Southern Asian strains can be further differentiated based on *vacA* mid-region variants [m1 and m2, respectively] (Ito *et al.*, 1997, Yang *et al.*, 1998, Pan *et al.*, 1998). Furthermore diversity within the *vacA* gene has also been used to segregate *H. pylori* isolates within Europe (van Doorn *et al.*, 1999a) and within certain ethnic groups in the United States (Yamaoka *et al.*, 2000). Analysis of the diversity within the urease gene was used to separate *H. pylori* strains from Polynesians and European settlers in New Zealand (Campbell *et al.*, 1997). The examination of the geographical partitioning of *H. pylori* led to the theory that this organism's genetic geography coincides with that for man (Covacci *et al.*, 1999).

With this in mind the primary aim of the present study was to characterise a collection of *H. pylori* isolates representing a bacterial population of this organism in Ireland. The *vacA* status of this collection of Irish *H. pylori* isolates was determined, as this method of genotyping has previously been shown to differentiate *H. pylori* isolates from different geographical regions (van Doorn *et al.*, 1999a). As mentioned previously the frequency and allelic mosaic of the *vacA* gene in Irish *H. pylori* isolates resembled those from studies carried out on Taiwanese (Lin *et al.*, 2000), Chinese (Pan *et al.*, 1998) and Northern and Eastern European isolates (van Doorn *et al.*, 1999a), although van Doorn *et al.* (1999a) reported the s1c allele to be

prominent in these East Asian countries. In addition, all three reports detected a high frequency of the *cagA* gene in isolates from Taiwan and China, a finding similar to that herein for Irish isolates.

Mosaic structures have been found to occur within the penicillin-binding protein (PBP) genes in *Streptococcus* (Dowson *et al.*, 1989) and *Neisseria* (Spratt *et al.*, 1989) species. As both these organisms, like *H. pylori*, are naturally competent and these genes are under strong selection pressures (due to the clinical use of penicillin derivatives), it is assumed that these mosaic structures were generated by the incorporation of foreign sequences into the *pbp* gene. The *vacA* gene of *H. pylori* encodes an important virulence factor and therefore is exposed to selection pressures. This implies, as is now accepted (Atherton *et al.*, 1995), that horizontal transfer of foreign DNA sequences brought about the mosaic structure within the vacA gene. As the two main signal sequence variants (s1 and s2) display quite disparate nucleotide sequences, as do the mid-region variants (m1 and m2, see Chapter 3, Figs 3.11 and 3.12), it is highly unlikely that one allelic form could have converted to the other by natural mutation. Indeed, the difference between the m1 and m2 mid-region alleles is the presence of a 75-bp inserted DNA sequence. It is also unlikely that the same recombinational event could have occurred within two separate bacterial populations incorporating identical foreign sequences within the *vacA* gene. Thus, it likely that two separate *H. pylori* populations exhibiting the same form of *vacA* mosaic structure were originally derived from a common ancestral strain. It is, therefore, interesting to find that Irish *H. pylori* isolates share a genetic relationship with isolates from certain parts of Asia, as there is no historical relationship between Ireland and these countries and a large distance separates these regions. It has, however, been argued that the distribution of different *vacA* subtypes reflects particular adaptations of *H. pylori* to specific host populations (van Doorn *et al.*, 1999a). This implies that the predominant *vacA* types in specific geographical regions do not reflect genetic relationships between different locations. This may, perhaps, explain the unusual relationship between Irish *H. pylori* isolates and specific regions within East Asia indicated by this method of typing.

Insertion-deletion (in-del) analysis of the *cag* PAI has been used as a reliable method to differentiate *H. pylori* populations from different geographical regions (Kersulyte *et al.*, 2000). In-del motifs represent several ancient insertion, substitution and deletion events at the 3′-end of the *cag* PAI. Irish *H. pylori* isolates exhibited the type II in-del motif as its predominant genotype. Differing gene sequences and the absence of insertion elements

(miniIS*605* and IS*606*) differentiate type II motifs from type I motifs. Type I motifs are common in *H. pylori* isolates from Southern Europe (Spain), South America, Africa, South Asia and North America. This suggests that there is little relationship between Irish *H. pylori* isolates and isolates from these regions. It does, however, suggest that Irish isolates have a relationship with isolates from East Asia (South China, Hong Kong, North China and Japan) and Northeast Europe (Sweden and Lithuania), based on the predominant in-del motifs found in these regions, a finding that supports *vacA* genotype associations. Kersulyte *et al.* (2000) reported the frequencies of the type II motifs within East Asia and Northeast Europe to be 95% and 42%, respectively. As the frequency of the type II motif found in Irish isolates is 49% it might be concluded that Irish isolates resemble North-eastern Europe and not East Asian isolates at this locus. However, although the predominant in-del motif within Irish isolates was type II, the next most common was type IV (35%). In contrast, the second most common motif within North-eastern European isolates was type I, and the only type IV motif previously described was found within one English isolate studied (Kersulyte *et al.*, 2000). A more extensive study of English strains by in-del analysis is required to clarify the frequency of occurrence of type IV motifs among isolates from different ethnic groups, including the Irish diaspora. Thus, the present collection of Irish *H. pylori* isolates displays a rare in-del genotype that may be representative of an insular European population.

An interesting finding arising from characterisation of various loci within this population of Irish *H. pylori* isolates was the relative lack of allelic and sequence variant diversity found. For example, two *vacA* allelic combinations (s1a m1 and s1a m2) were represented in ~96% of these isolates, all isolates contained the *cagA* gene, ~90% of isolates possessed one or two repeat sequences at the 3′-end of the *cagA* gene, ~84% had either type II or IV in-del motifs and ~90% of isolates contained the RibAP1 and VacAP3 sequence variants upstream of their *ribA* and *vacA* genes, respectively. It has been suggested that a lack of genetic diversity within a *H. pylori* population is associated with a traditionally isolated or insular human population (Campbell *et al.*, 1997; Evans *et al.*, 1999).

In order to determine the heterogeneity of Irish *H. pylori* isolates and to further characterise the associations of these isolates with those from different geographical regions, a sensitive fingerprinting technique (FAFLP-PCR) was applied to these strains. The overall outcome of FAFLP-PCR fingerprinting combined with phylogenetic analysis demonstrated that this population of Irish *H. pylori* isolates exhibited very little heterogeneity between fingerprint

profiles (~88% of isolates displayed ~85% homogeneity). Evans *et al.* (1999) compared the heterogeneity of *H. pylori* isolates within human populations from Göteborg (Sweden), Houston (Texas) and Minas Gerais (Brazil). Göteborg represents a relatively small genetically homogeneous human population, whilst Houston and Minas Gerais both represent large genetically heterogeneous human populations. The result was that a genetically homogeneous *H. pylori* population is associated with a relatively small, stable, genetically homogeneous human population. Thus, it is possible that the genetically homogeneous Irish *H. pylori* isolates examined in this study has resulted from spread within a relatively small insular human population.

As *H. pylori* is a panmitic organism and is prone to recombination, the genotype of a resident *H. pylori* isolate can accumulate small recombinational replacements from genetically distinct co-infecting isolates, thereby diversifying its genome. High rates of recombination would result in a *H. pylori* population with many genotypes. However, Spratt (2003) states that if all DNA replacements are from within the same population as the resident *H. pylori* population then recombination becomes a homogenising force. Thus, the lack of genetic diversity found within this Irish *H. pylori* collection could be due to exchange of nucleotide sequences, via recombination, that is restricted to or only occurs between *H. pylori* isolates from a resident population. With this in mind, a low amount of diversity within Irish *H. pylori* isolates could indicate a lack of polymorphisms introduced by other populations, again indicative of an insular population.

The relative lack of heterogeneity between Irish *H. pylori* isolates could also be due to a lack new polymorphic *H. pylori* DNA sequences being introduced into Irish isolates from strains from different geographical regions. The Irish isolates examined in this study were obtained before 1998. The principal demographical results of the 2002 Census (Central Statistics Office, 2002) report that immigration into Ireland has risen over the last 12 years (Fig. 8.1). Fig. 8.2 shows that before 1998 the influx of people from non-Irish (European and non-European) countries remained relatively constant. Thus, the majority *H. pylori* isolates obtained at and before 1998 would be indicative of isolates from patients of Irish nationality. The lack of immigration into Ireland at this time suggests that very few new polymorphic *H. pylori* DNA sequences were being introduced into the Irish *H. pylori* population from different geographical regions. Immigration into Ireland over the last 5 years has increased, and as approximately 7.2% of people now living in Ireland are of non-Irish origin (Fig. 8.3),

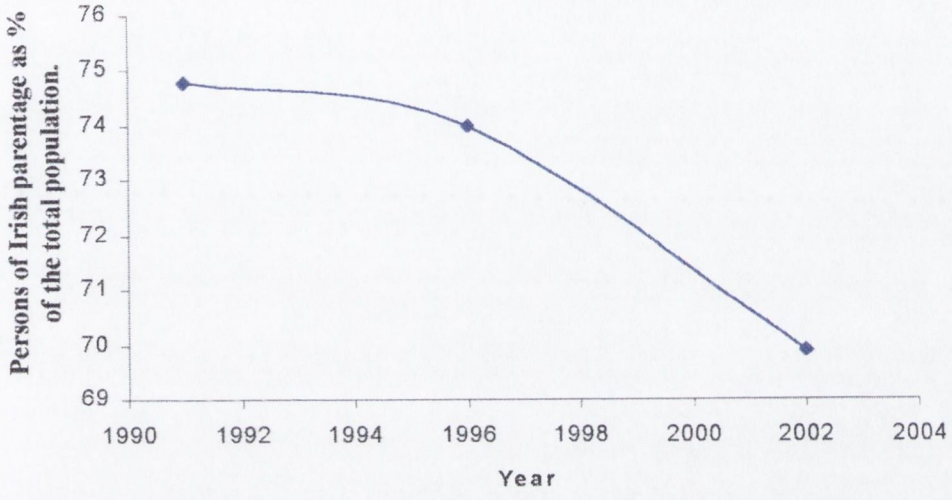


Fig. 8.1. Graph indicating the % of Irish nationals within the total population of the Republic of Ireland from 1991–2002.

The y-axis displays the % of the total Irish population that had Irish parents and the x-axis displays the year. The % of the population with a single Irish parent remained constant at approximately 20% (not shown). These data were obtained from the Census 2002 report (Central Statistics office, 2002).

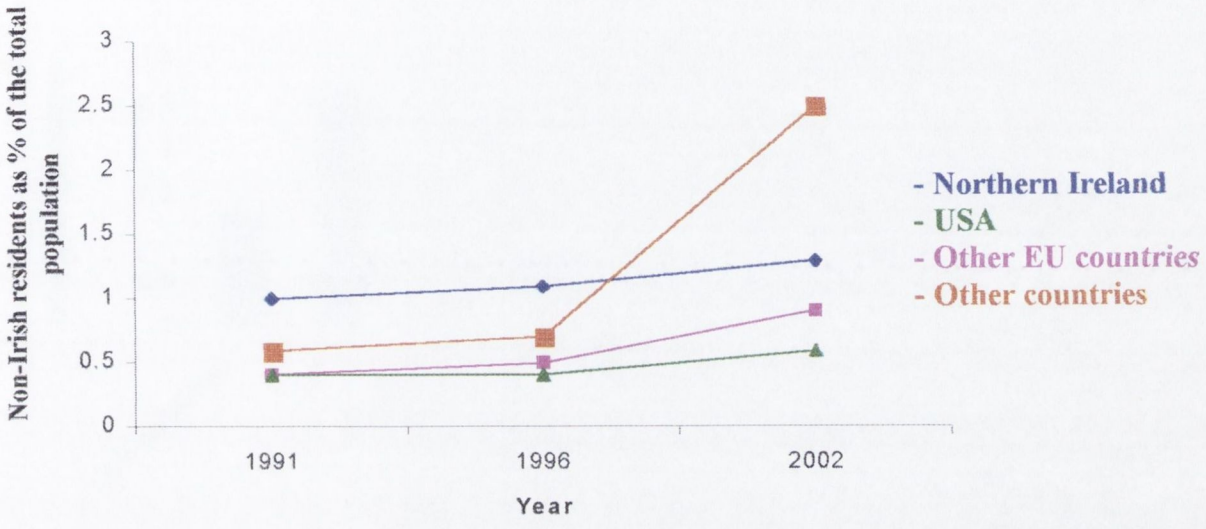


Fig. 8.2. Graph of the % non-Irish residents within the total population of the Republic of Ireland from 1991–2002.

The y-axis displays non-Irish residents as a % of the total Irish population and the x-axis displays the year. Immigration of people from the UK rose slightly from 3.8%–5.1% between 1991 and 2002 (not shown). Immigration of individuals from Northern Ireland, other EU countries (outside UK and Ireland) and the USA have risen slightly, whereas individuals from the rest of the world have increased dramatically. These data were obtained from the Census 2002 report (Central Statistics office, 2002).

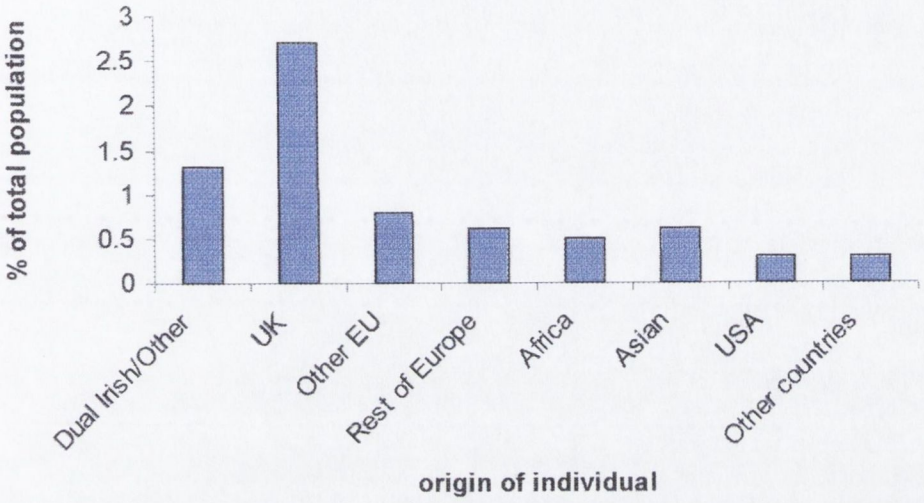


Fig. 8.3. Histograms indicating the % of immigrants from different countries within the total population of the Republic of Ireland in 2002.

The % of Irish nationals resident in Ireland in 2002 was approximately 91.6%. These data were obtained from the Census 2002 report (Central Statistics office, 2002).

thus the homogeneity of Irish *H. pylori* isolates may decline over the coming years. The increase of immigration of people into Ireland from non-EU and non-American countries (Fig. 8.2) suggests an influx of human populations from developing countries. As the incidence of *H. pylori* infection in the developing world can be over 90%, it is possible that a large number of current immigrants into Ireland are bringing *H. pylori* infections with them. Thus, if genetic exchange between newly introduced *H. pylori* strains and the current Irish *H. pylori* population occurs the genetic status of *H. pylori* isolates in Ireland may be altered. Fig. 8.4 outlines the factors associated with the genetic status of the Irish *H. pylori* isolates used in this study.

FAFLP-PCR fingerprinting was carried out not only on Irish *H. pylori* isolates but also on isolates from different geographical regions (India, UK, Peru, Spain, Japan and Africa). Although this group represented a small number of isolates from each country, it still enabled a picture of how Irish *H. pylori* isolates cluster with isolates from other geographical regions. The analysis once again demonstrated that Irish isolates are relatively homogeneous. In addition, this examination also demonstrated that Irish isolates are distinct, yet they share some ancestral relationship with European (UK) isolates and a more distant relationship with Asian (Indian) isolates. Falush *et al.* (2003) also demonstrated that European and Indian *H. pylori* isolates cluster within the same category. Thus, FAFLP-PCR while confirming what was previously observed herein with *vacA* and in-del typing provided a definitive result.

8.1.2. Microevolution of H. pylori *isolates* in vivo.

Microorganisms require a certain degree of flexibility to adapt to changes in their environment, for example, during times of nutrient limitation or, in relation to pathogens, during exposure to host cell defences. Adaptation to environmental changes can be brought about by activation of regulatory networks or through genome variability (Claverys *et al.*, 2000). The genomes of bacterial pathogens vary in size from *Mycoplasma genitalium* at 580 kbp to *Pseudomonas aeruginosa* at 6,300 kbp. Most pathogens have a conserved core set of genes that includes the 'minimal set' (Mushegian and Koonin, 1996), plus flexible genes consisting of strain-specific or species-specific genetic information that provide additional properties enabling a species to adapt to a specialised environment (Dobrindt and Hacker, 2001). Genome variability within bacterial populations is an interesting process that contributes to bacterial adaptation.

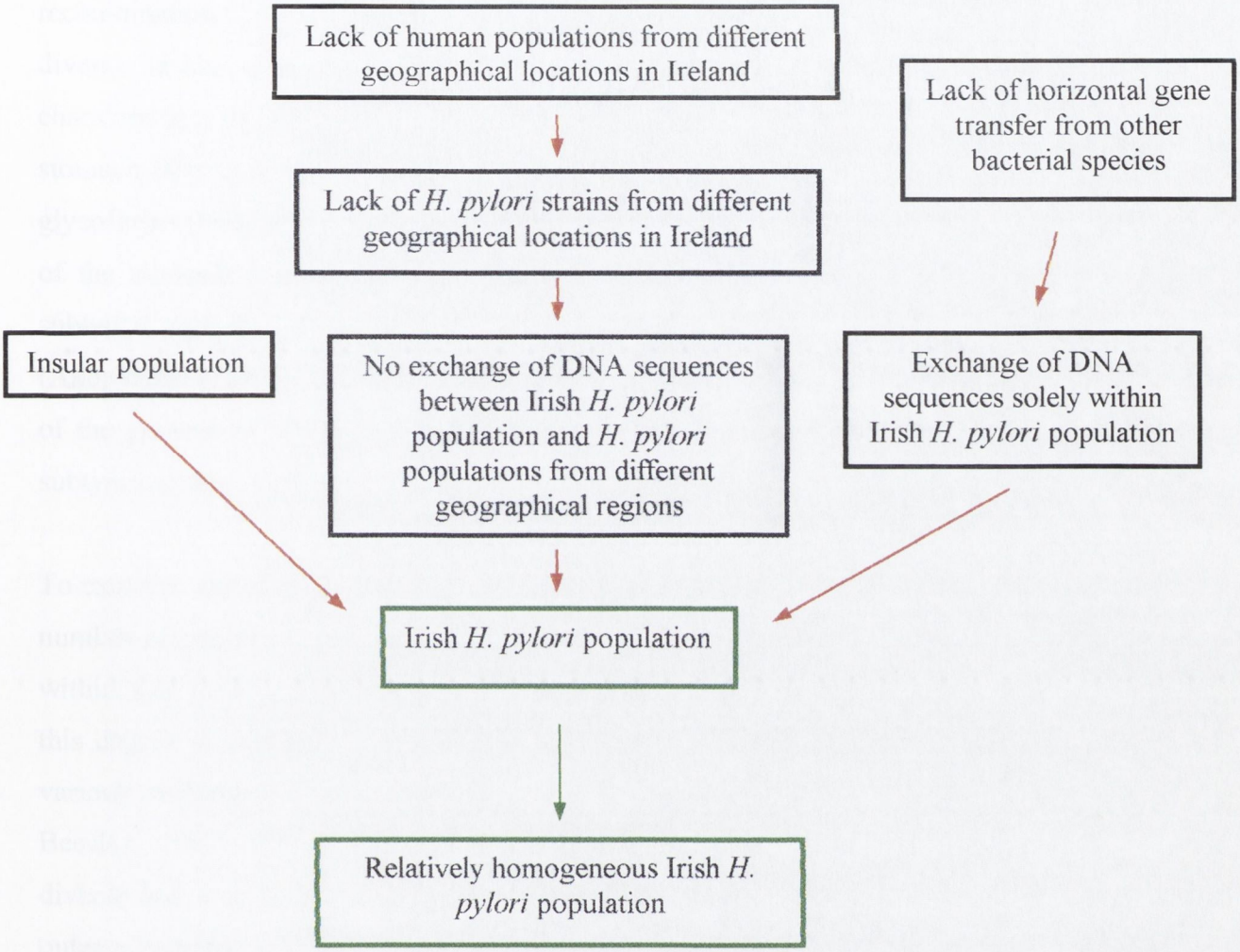


Fig. 8.4. Schematic representation of the probable causes of a relatively homogenous Irish *H. pylori* population.

Mutation, genetic rearrangements and recombination are mechanisms that result in the plasticity of bacterial genomes (Dobrindt and Hacker, 2001). Naturally competent organisms such as *Strep. pneumoniae* and *H. pylori* can readily generate genetic diversity (subtypic variants) within a population by integration of foreign DNA sequences into their genomes via recombination (Go *et al.*, 1996, Claverys *et al.*, 2000). As *H. pylori* resides within a harsh and diverse niche, selection for fitter subtypic variants would be expected to occur. Some characteristics of this diverse niche include differing pH levels within different regions of the stomach (Karita & Blaser, 1998; Bijlsma *et al.*, 2000), variation in the distribution of mucin glycoforms (Nordman *et al.*, 2002) and different viscosities of mucins within different regions of the stomach (Tanaka *et al.*, 2002). In addition, it has been documented that *H. pylori* subtypes can be selected for when passaged *in vivo*, demonstrating host adaptation (Akopyants, *et al.*, 1995, Janvier *et al.*, 1999, Kersulyte *et al.*, 1999). Thus, the plastic nature of the genome of *H. pylori* and the diverse environment in which it resides allow for fitter subtypes to arise within a bacterial population.

To examine the phenomenon of microevolution, FAFLP-PCR fingerprinting was applied to a number of paired *H. pylori* isolates from individuals. It was revealed that subtypic variation within a *H. pylori* population within the same gastric niche is a common occurrence, but that this degree of variation is not consistent. Indeed, microevolution has been described using various techniques (Marshall *et al.*, 1996; Bjorkholm *et al.*, 2001; Israel *et al.*, 2001b; Beesley, 2002). However, the genetic sequences affected by this adaptation process are diverse and varying from genes in the *cag* PAI (Bjorkholm *et al.*, 2001), to genes encoding outer membrane proteins (Beesley, 2002) and genes involved in antibiotic resistance (Israel *et al.*, 2001b).

The present study revealed that the number of repeat sequences at the 3′-end of the *cagA* gene can differ between ancestrally related isolates. Variation in the number of repeat sequences within an organism may alter its genetic repertoire and may result in the evolution of a quasi-species. In addition, it is as yet unknown whether variation in the number of repeat sequences within prokaryotic genomes occurs by chance or whether targeted variation exists (Belkum *et al.*, 1998). Thus, it is difficult to speculate whether the variation in repeat sequence number at the 3′-end of the *cagA* gene observed between paired microevolved *H. pylori* isolates results from selection of intentional or unintentional variation. However, as *H. pylori* lacks mismatch repair systems and as no gene products that specifically drive the variation in repeat number

have been identified, the *cagA* 3′-repeat variation within microevolved *H. pylori* isolates is probably unintentional.

Variability in repeat sequence number within bacterial populations has clear implications for virulence. For example, repeat sequences within genes responsible for lipopolysaccharide (LPS) biosynthesis within the *Haemophilus influenzae* genome are believed to be responsible for phase variation of the LPS (Weiser *et al.*, 1990). Similarly within *Neisseria gonorrhoeae*, phase variation within the LPS is believed to be a result of repeat sequences (van Putten *et al.*, 1993). Variation of the LPS aids in immune evasion and translocation across various anatomical and physiological barriers. The phase variation in the LPS of *N. gonorrhoeae* closely resembles the phase variation of the *H. pylori* LPS (also controlled by repetitive sequences), where antigenic mimicry occurs in order to evade the host's immune system (Applemelk *et al.*, 1996, 1998). Due to the well documented fact that CagA is an important virulence factor and the fact that repeat sequence variation occurs via sub-species generation, it is tempting to speculate that this variation may in some way affect the virulence of *H. pylori*. Moreover, recent reports describe the presence of important tyrosine phosphorylation motifs (TPM's) within these 3′-repeat sequences. Thus, variation in the number of *cagA* repeat sequences in turn varies the number of CagA TPM's that potentially affect the potency of CagA as a virulence molecule. Indeed, Higashi *et al.* (2002) demonstrated that CagA molecules with more sequence repeats undergo greater tyrosine phosphorylation and can induce greater morphological change in epithelial cells. This suggests that variation in the number of *cagA* 3′-repeat sequences between ancestrally related *H. pylori* isolates obtained from one individual signifies an evolutionary event in which the activity of an important virulence molecule is increased or decreased.

Analysis of the tyrosine phosphorylation potential of the CagA protein within paired ancestrally related *H. pylori* isolates yielded inconclusive results. However, it was observed that epithelial cells transfected with CagA sequences displayed different effects on tyrosine phosphorylation of host proteins depending on the number of CagA repeat sequences present. It is important to understand that it cannot be deduced as to whether adaptation of *H. pylori* isolates within the gastric mucosa increases or decreases (or both) the number of *cagA* repeat sequences. In the case of paired isolates MI512 and MI541 the number of repeats differs from none to one, respectively. If, in this case, a *H. pylori* isolate evolved via selective pressures to

lose its *cagA* repeat sequences (MI541 → MI512) an evolutionary change would have occurred that reduced this strain's virulence.

8.2. Future prospects.

A recent and comprehensive study by Falush *et al.* (2003) demonstrated that sequence analysis of *H. pylori* isolates recovered from twenty-seven countries displayed geographical partitioning of this bacterial species. This study also revealed that not only were particular polymorphisms associated with specific geographical regions, but also that mosaics from one population appear to have been introduced into another. The introduction of such mosaics was found to correlate with historical migration events, such as the prehistoric colonisation of Polynesia and the Americas, the Neolithic introduction of farming to Europe, the Bantu expansion within Africa, and the slave trade. Moreover, Falush *et al.* (2003) demonstrated a link between Asian and Amerindian strains supporting earlier observations (Ghose *et al.*, 2002). Since it is believed that Ireland has been inhabited for 10,000 years (Hill *et al.*, 2000) and *H. pylori* has been carried by human hosts for much longer (Ghose *et al.*, 2002, Falush *et al.*, 2003), perhaps comparative analysis of Irish *H. pylori* isolates with those from other geographical regions by FAFLP-PCR may reveal historical accounts of human migrations into Ireland. In addition, it would also be interesting to characterise *H. pylori* isolates from individuals living in Ireland who originated from various different countries and to compare such isolates with characterised Irish isolates. Such a study may provide insights into the genetic exchange between *H. pylori* populations from different geographical regions.

Further analysis of microevolution between paired *H. pylori* isolates should be carried out to uncover the exact DNA sequences affected by quasi-species development. Application of DNA arrays would provide accurate information as to which sequences are different within a pair of microevolved *H. pylori* isolates. It would also be interesting to uncover the mechanism(s) [if any] behind the generation of microevolution.

As the effect that *cagA* 3′-end repeat variation might have on the tyrosine phosphorylation status of the CagA protein between ancestrally related *H. pylori* isolates was not determined herein, perhaps this analysis should be attempted using a different approach. The drawback of the approach described herein was that the polypeptides to be analysed were small. The whole *cagA* gene from each isolate within a pair could be cloned and the TPM sequences abolished outside the 3′-repeat region by mutagenesis. The degree of tyrosine phosphorylation of the

two cloned CagA proteins with different numbers of 3′-repeats could then be determined in a cell line. Anti-CagA antibodies could be used using this approach as the whole CagA protein would be used.

8.3. Concluding remarks.

The estimated high incidence of *H. pylori* infection worldwide contrasts with the much lower incidence of *H. pylori*-associated disease. This suggests co-evolution of this bacterium with its human host and that disease is caused by strains with specific properties that increase virulence. There have been many attempts to associate *H. pylori* disease with specific virulence factors (Graham and Yamaoka, 2000). So far no specific factor has been identified that either predicts symptomatic disease or clinical outcome with epidemiological consistency. The application of microarray technology and multi-locus sequence typing may generate data to resolve this frustrating lack of correlation with pathogenicity or asymptomatic infection. It is essential that such future studies be carried out with isolates from different geographical regions because of the inconsistent linkages between putative virulence factors and disease from studies involving national populations of *H. pylori* that may have biased associations in the past.

Just as strains of *H. pylori* with supposed virulence attributes may be responsible for disease, some types of *H. pylori* may be protective where they have coevolved with their human host and thus have lost virulence through microevolutionary changes associated with life-long colonisation of the gastric mucosa. The idea that *H. pylori* may not be a "bad bug" (Richter *et al.*, 1998) has been controversial and indeed is alleged to "have caused unnecessary confusion among practicing physicians" (Graham, 2003). While this debate remains an open question, the diversity of *H. pylori* and of the human population give room for speculation that among the myriad of host-pathogen interactions that could potentially occur with microevolution of this bacterium, some could prove beneficial where the co-evolved bacterium and host reach an almost 'symbiotic' relationship.

Bibliography

Achtman, M. and Suerbaum, S. (2000). Sequence variation in *Helicobacter pylori*. *Trends Microbiol* **8**, 57-58.

Achtman, M., Azuma, T., Berg, D. E. and 10 other authors (1999). Recombination and clonal groupings within *Helicobacter pylori* from different geographical regions. *Mol Microbiol* **32**, 459-470.

Agostini, H. T., Yanagihara, R., Davis, V., Ryschkewitsch, C. F. and Stoner, G. L. (1997). Asian genotypes of JC virus in Native Americans and in a Pacific Island population: markers of viral evolution and human migration. *Proc Natl Acad Sci U S A* **94**, 14542-14546.

Ahmad, A., Govil, Y. and Frank, B. B. (2003). Gastric mucosa-associated lymphoid tissue lymphoma. *Am J Gastroenterol* **98**, 975-986.

Ahmed, N., A. A. Khan, D. E. Berg, and C. M. Habibullah. (2001). Genomic analysis reveals distinct biogeographic partitioning of *H. pylori* from South Indians and other world populations. Int. J. Med. Microbiol. **291, Suppl. 31:**31 (Abstract D-11).

Ahmed, N., Bal, A., Khan, A. A. and other authors (2002). Whole genome fingerprinting and genotyping of multiple drug resistant (MDR) isolates of *Pseudomonas aeruginosa* from endophthalmitis patients in India. *Infect Genet Evol* **1**, 237-242.

Ahmed, N., Alam, M., Abdul Majeed, A., Asad Rahman, S., Cataldi, A., Cousins, D. and Hasnain, S. E. (2003a). Genome sequence based, comparative analysis of the fluorescent amplified fragment length polymorphisms (FAFLP) of tubercle bacilli from seals provides molecular evidence for a new species within the *Mycobacterium tuberculosis* complex. *Infect Genet Evol* **2**, 193-199.

Ahmed, N., Caviedes, L., Alam, M., Rao, K. R., Sangal, V., Sheen, P., Gilman, R. H. and Hasnain, S. E. (2003b). Distinctiveness of *Mycobacterium tuberculosis* genotypes from human immunodeficiency virus type 1-seropositive and -seronegative patients in Lima, Peru. *J Clin Microbiol* **41**, 1712-1716.

Aizawa, S. I. (2001). Bacterial flagella and type III secretion systems. *FEMS Microbiol Lett* **202**, 157-164.

Akada, J. K., Ogura, K., Dailidiene, D., Dailide, G., Cheverud, J. M. & Berg, D. E. (2003). *Helicobacter pylori* tissue tropism: mouse-colonizing strains can target different gastric niches. *Microbiology* **149**, 1901-1909.

Akopyants, N. S., Clifton, S. W., Kersulyte, D. and 7 other authors (1998). Analyses of the *cag* pathogenicity island of *Helicobacter pylori*. *Mol Microbiol* **28**, 37-53.

Akopyants, N. S., Eaton, K. A. and Berg, D. E. (1995). Adaptive mutation and cocolonization during *Helicobacter pylori* infection of gnotobiotic piglets. *Infect Immun* **63**, 116-121.

Akopyanz, N., Bukanov, N. O., Westblom, T. U. and Berg, D. E. (1992a). PCR-based RFLP analysis of DNA sequence diversity in the gastric pathogen *Helicobacter pylori*. *Nucleic Acids Res* **20**, 6221-6225.

Akopyanz, N., Bukanov, N. O., Westblom, T. U., Kresovich, S. and Berg, D. E. (1992b). DNA diversity among clinical isolates of *Helicobacter pylori* detected by PCR-based RAPD fingerprinting. *Nucleic Acids Res* **20**, 5137-5142.

Alm, R. A. and Trust, T. J. (1999). Analysis of the genetic diversity of *Helicobacter pylori*: the tale of two genomes. *J Mol Med* **77**, 834-846.

Alm, R. A., Ling, L. S., 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. and Trust, T. J. (2000). Comparative genomics of *Helicobacter pylori*: analysis of the outer membrane protein families. *Infect Immun* **68**, 4155-4168.

Andrutis, K. A., Fox, J. G., Schauer, D. B., Marini, R. P., Li, X., Yan, L., Josenhans, C. and Suerbaum, S. (1997). Infection of the ferret stomach by isogenic flagellar mutant strains of *Helicobacter mustelae*. *Infect Immun* **65**, 1962-1966.

Ang, S., Lee, C. Z., Peck, K., Sindici, M., Matrubutham, U., Gleeson, M. A. and Wang, J. T. (2001). Acid-induced gene expression in *Helicobacter pylori*: study in genomic scale by microarray. *Infect Immun* **69**, 1679-1686.

Angelakopoulos, H. and Hohmann, E. L. (2000). Pilot study of *phoP/phoQ*-deleted *Salmonella enterica* serovar *typhimurium* expressing *Helicobacter pylori* urease in adult volunteers. *Infect Immun* **68**, 2135-2141.

Anonymous (1985). Validation of the publication of new names and new combinations previously effectively published outside the IJSB: list no. 17. *Int J Syst Bacteriol* **35,** 223-225.

Appelmelk, B. J., Simoons-Smit, I., Negrini, R. and other authors (1996). Potential role of molecular mimicry between *Helicobacter pylori* lipopolysaccharide and host Lewis blood group antigens in autoimmunity. *Infect Immun* **64**, 2031-2040.

Appelmelk, B. J., Shiberu, B., Trinks, C. and 10 other authors (1998). Phase variation in *Helicobacter pylori* lipopolysaccharide. *Infect Immun* **66**, 70-76.

Asahi, M., Azuma, T., Ito, S. and other authors (2000). *Helicobacter pylori* CagA protein can be tyrosine phosphorylated in gastric epithelial cells. *J Exp Med* **191**, 593-602.

Atherton, J. C. (1997). The clinical relevance of strain types of *Helicobacter pylori*. *Gut* **40**, 701-703.

Atherton, J. C., Cao, P., Peek, R. M., Jr., Tummuru, M. K., Blaser, M. J. and Cover, T. L. (1995). Mosaicism in vacuolating cytotoxin alleles of *Helicobacter pylori*. Association of specific *vacA* types with cytotoxin production and peptic ulceration. *J Biol Chem* **270**, 17771-17777.

Atherton, J. C., Peek, R. M., Jr., Tham, K. T., Cover, T. L. and Blaser, M. J. (1997). Clinical and pathological importance of heterogeneity in *vacA*, the vacuolating cytotoxin gene of *Helicobacter pylori*. *Gastroenterology* **112**, 92-99.

Ausubel, R. M., Brent, R., Kingston, R. E., Moore, D. E., 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.

Avenaud, P., Marais, A., Monteiro, L., Le Bail, B., Bioulac Sage, P., Balabaud, C. and Mégraud, F. (2000). Detection of *Helicobacter* species in the liver of patients with and without primary liver carcinoma. *Cancer* **89**, 1431-1439.

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. *Cell Microbiol* **2**, 155-164.

Backert, S., Moese, S., Selbach, M., Brinkmann, V. and Meyer, T. F. (2001). Phosphorylation of tyrosine 972 of the *Helicobacter pylori* CagA protein is essential for induction of a scattering phenotype in gastric epithelial cells. *Mol Microbiol* **42**, 631-644.

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.

Banatvala, N., Mayo, K., Megraud, F., Jennings, R., Deeks, J. J. and Feldman, R. A. (1993). The cohort effect and *Helicobacter pylori*. *J Infect Dis* **168**, 219-221.

Barabino, A. (2002). *Helicobacter pylori*-related iron deficiency anemia: a review. *Helicobacter* 7, 71-75.

Barry, M. A., Lai, W. C. and Johnston, S. A. (1995). Protection against mycoplasma infection using expression-library immunization. *Nature* **377**, 632-635.

Beesley, S. M. (2002). Microevolution in *Helicobacter pylori*. Ph.D. Thesis, Trinity College, University of Dublin.

Bereswill, S., Schonenberger, R., Thies, C., Stahler, F., Strobel, S., Pfefferle, P., Wille, L. and Kist, M. (2000). New approaches for genotyping of *Helicobacter pylori* based on amplification of polymorphisms in intergenic DNA regions and at the insertion site of the *cag* pathogenicity island. *Med Microbiol Immunol (Berl)* **189**, 105-113.

Bijlsma, J. J., Lie, A. L. M., Nootenboom, I. C., Vandenbroucke-Grauls, C. M. & Kusters, J. G. (2000). Identification of loci essential for the growth of *Helicobacter pylori* under acidic conditions. *J Infect Dis* **182**, 1566-1569.

Birnboim, H. C. and Doly, J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Research* 7, 1513-1523.

Bizzozero, G. (1893). Ueber die Schlauchformigen drusen des Magendarmkanals und die Bezienhungen ihres Epithels zu dem Oberflachenepithel der Schleimhaut. *Arch Mikr Anat* **42,** 82.

Björkholm, B., Lundin, A., Sillén, A., Guillemin, K., Salama, N., Rubio, C., Gordon, J. I., Falk, P. and Engstrand, L. (2001). Comparison of genetic divergence and fitness between two subclones of *Helicobacter pylori*. *Infect Immun* **69**, 7832-7838.

Blanchard, T. G., McGovern, K. J., Youngman, P., Gutierez, J. and Czinn S. J. (1999). Is gamma-glutamyl transpeptidase essential for infection of the gastric mucosa by *H. pylori*? *Gut* **45,** A26.

Blaser, M. J. (1990). *Helicobacter pylori* and the pathogenesis of gastroduodenal inflammation. *J Infect Dis* **161**, 626-633.

Blaser, M. J. (1993). *Helicobacter pylori*: microbiology of a 'slow' bacterial infection. *Trends Microbiol* **1**, 255-260.

Blaser, M. J. (1994). *Helicobacter pylori* phenotypes associated with peptic ulceration. *Scand J Gastroenterol* **205**, (Suppl) 1-5.

Blaser, M. J. and Parsonnet J. (1994). Parasitism by the "slow" bacterium *Helicobacter pylori* leads to altered gastric homeostasis and neoplasia. *J Clin Invest* **94**, 4-8.

Blaser, M. J. and Berg, D. E. (2001). *Helicobacter pylori* genetic diversity and risk of human disease. *J Clin Invest* **107**, 767-773.

Blaser, M. J., Perez-Perez, G. I., Kleanthous, H., Cover, T. L., Peek, R. M., Chyou, P. H., Stemmermann, G. N. and Nomura, A. (1995). Infection with *Helicobacter pylori* strains possessing *cagA* is associated with an increased risk of developing adenocarcinoma of the stomach. *Cancer Res* **55**, 2111-2115.

Blecker, U., Lanciers, S., Keppens, E. and Vandenplas, Y. (1994). Evolution of *Helicobacter pylori* positivity in infants born from positive mothers. *J Pediatr Gastroenterol Nutr* **19**, 87-90.

Bode, G., Mauch, F. and Malfertheiner, P. (1993). The coccoid forms of *Helicobacter pylori*. Criteria for their viability. *Epidemiol Infect* **111**, 483-490.

Bodger, K. and Crabtree, J. E. (1998). *Helicobacter pylori* and gastric inflammation. *Br Med Bull* **54**, 139-150.

Boneca, I. G., de Reuse, H., Epinat, J. C., Pupin, M., Labigne, A. and Moszer, I. (2003). A revised annotation and comparative analysis of *Helicobacter pylori* genomes. *Nucleic Acids Res* **31**, 1704-1714.

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.

Brenner, H., Rothenbacher, D., Bode, G. and Adler, G. (1998). Parental history of gastric or duodenal ulcer and prevalence of *Helicobacter pylori* infection in preschool children: population based study. *Bmj* **316**, 665.

Byrne, M. F., Kerrigan, S. W., Corcoran, P. A., Atherton, J. C., Murray, F. E., Fitzgerald, D. J. and Cox, D. M. (2003). *Helicobacter pylori* binds von Willebrand factor and interacts with GPIb to induce platelet aggregation. *Gastroenterology* **124**, 1846-1854.

Campbell, S., Fraser, A., Holliss, B., Schmid, J. and O'Toole, P. W. (1997). Evidence for ethnic tropism of *Helicobacter pylori*. *Infect Immun* **65**, 3708-3712.

Cantorna, M. T. and Balish, E. (1990). Inability of human clinical strains of *Helicobacter pylori* to colonize the alimentary tract of germfree rodents. *Can J Microbiol* **36**, 237-241.

Caugant, D. A., Mocca, L. F., Frasch, C. E., Froholm, L. O., Zollinger, W. D. and Selander, R. K. (1987). Genetic structure of *Neisseria meningitidis* populations in relation to serogroup, serotype, and outer membrane protein pattern. *J Bacteriol* **169**, 2781-2792.

Cellini, L., Di Campli, E., Di Candia, M. and Marzio, L. (2003). Molecular fingerprinting of *Helicobacter pylori* strains from duodenal ulcer patients. *Lett Appl Microbiol* **36**, 222-226.

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. *Proc Natl Acad Sci U S A* **93**, 14648-14653.

Central Statistics Office. (2002). Census 2002 – Preliminary Report, pp. 84. Stationary office, Dublin, Ireland.

Chalkauskas, H., Kersulyte, D., Cepuliene, I., Urbonas, V., Ruzeviciene, D., Barakauskiene, A., Raudonikiene, A. and Berg, D. E. (1998). Genotypes of *Helicobacter pylori* in Lithuanian families. *Helicobacter* **3**, 296-302.

Chevalier, C., Thiberge, J.-M., Ferrero, R. L. and Labigne, A. (1999). Essential role of *Helicobacter pylori* gamma-glutamyltranspeptidase for the colonization of the gastric mucosa of mice. *Mol Microbiol* **31**, 1359-1372.

Chmiela, M., Kowalewicz-Kulbat, M., Miszczak, A., Wisniewska, M., Rechcinski, T., Kolodziej, K., Kasprzak, J., Wadström, T. and Rudnicka, W. (2003). A link between *Helicobacter pylori* and/or *Chlamydia* spp. infections and atherosclerosis. *FEMS Immunol Med Microbiol* **36**, 187-192.

Chow, W. H., Blaser, M. J., Blot, W. J., Gammon, M. D., Vaughan, T. L., Risch, H. A., Perez-Perez, G. I., Schoenberg, J. B., Stanford, J. L., Rotterdam, H., West, A. B. and Fraumeni, J. F. Jr. (1998). An inverse relation between *cagA*+ strains of *Helicobacter pylori* infection and risk of esophageal and gastric cardia adenocarcinoma. *Cancer Res* **58**, 588-590.

Christie, P. J. and Vogel, J. P. (2000). Bacterial type IV secretion: conjugation systems adapted to deliver effector molecules to host cells. *Trends Microbiol* **8**, 354-360.

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? *Mol Microbiol* **35**, 251-259.

Clayton, C. L., Pallen, M. J., Kleanthous, H., Wren, B. W. and Tabaqchali, S. (1990). Nucleotide sequence of two genes from *Helicobacter pylori* encoding for urease subunits. *Nucleic Acids Res* **18**, 362.

Clayton, C. L., Kleanthous, H., Morgan, D. D., Puckey, L. & Tabaqchali, S. (1993). Rapid fingerprinting of *Helicobacter pylori* by polymerase chain reaction and restriction fragment length polymorphism analysis. *J Clin Microbiol* **31**, 1420-1425.

Clyne, M., Ocroinin, 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. *Infect Immun* **68**, 4335-4339.

Colding, H., Hartzen, S. H., Roshanisefat, H., Andersen, L. P. and Krogfelt, K. A. (1999). Molecular methods for typing of *Helicobacter pylori* and their applications. *FEMS Immunol Med Microbiol* **24**, 193-199.

Cornelis, G. R. and Van Gijsegem, F. (2000). Assembly and function of type III secretory systems. *Annu Rev Microbiol* **54**, 735-774.

Correa, P. (1992). Human gastric carcinogenesis: a multistep and multifactorial process. *Cancer Res* **52**, 6735-6740.

Corthésy-Theulaz, I., Porta, N., Glauser, M., Saraga, E., Vaney, A. C., Haas, R., Kraehenbuhl, J. P., Blum, A. L. and Michetti, P. (1995). Oral immunization with *Helicobacter pylori* urease B subunit as a treatment against *Helicobacter* infection in mice. *Gastroenterology* **109**, 115-121.

Covacci, A. and Rappuoli, R. (1998). *Helicobacter pylori*: molecular evolution of a bacterial quasi-species. *Curr Opin Microbiol* **1**, 96-102.

Covacci, A., Censini, S., Bugnoli, M. and 7 other authors (1993). Molecular characterization of the 128-kDa immunodominant antigen of *Helicobacter pylori* associated with cytotoxicity and duodenal ulcer. *Proc Natl Acad Sci U S A* **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. and Blaser, M. J. (1992). Purification and characterization of the vacuolating toxin from *Helicobacter pylori*. *J Biol Chem* **267**, 10570-10575.

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.

Cullen, D. J., Collins, B. J., Christiansen, K. J., Epis, J., Warren, J. R., Surveyor, I. and Cullen, K. J. (1993). When is *Helicobacter pylori* infection acquired? *Gut* **34**, 1681-1682.

Cussac, V., Ferrero, R. L. and Labigne, A. (1992). Expression of *Helicobacter pylori* urease genes in *Escherichia coli* grown under nitrogen-limiting conditions. *J Bacteriol* **174**, 2466-2473.

Danesh, J., Wong, Y., Ward, M. and Muir, J. (1999). Chronic infection with *Helicobacter pylori*, *Chlamydia pneumoniae*, or *cytomegalovirus*: population based study of coronary heart disease. *Heart* **81**, 245-247.

Desai, M., Tanna, A., Wall, R., Efstratiou, A., George, R. & Stanley, J. (1998). Fluorescent amplified-fragment length polymorphism analysis of an outbreak of group A *streptococcal* invasive disease. *J Clin Microbiol* **36**, 3133-3137.

Dhaenens, L., Szczebara, F. and Husson, M. O. (1997). Identification, characterization, and immunogenicity of the lactoferrin-binding protein from *Helicobacter pylori*. *Infect Immun* **65**, 514-518.

DiPetrillo, M. D., Tibbetts, T., Kleanthous, H., Killeen, K. P. and Hohmann, E. L. (1999). Safety and immunogenicity of *phoP/phoQ*-deleted *Salmonella typhi* expressing *Helicobacter pylori* urease in adult volunteers. *Vaccine* **18**, 449-459.

Dixon, M. F. (2001). Pathology of gastritis and peptic ulceration. In Helicobacter pylori *physiology and genetics*, pp. 459-469. Edited by H. L. T. Mobley, G. L. Mendz and S. L. Hazell: ASM press.

Dobrindt, U. and Hacker, J. (2001). Whole genome plasticity in pathogenic bacteria. *Curr Opin Microbiol* **4**, 550-557.

Doenges, J. L. (1939). Spirochetes in the gastric glands of Macacus *rhesus* and of man without related diseases. *Arch. Pathol.* **27,** 469-477.

Dominici, P., Bellentani, S., Di Biase, A. R., Saccoccio, G., Le Rose, A., Masutti, F., Viola, L., Balli, F., Tiribelli, C., Grilli, R., Fusillo, M. and Grossi, E. (1999). Familial clustering of *Helicobacter pylori* infection: population based study. *Brit Med J* **319**, 537-540.

Dore, M. P., Bilotta, M., Vaira, D., Manca, A., Massarelli, G., Leandro, G., Atzei, A., Pisanu, G., Graham, D. Y. and Realdi, G. (1999a). High prevalence of *Helicobacter pylori* infection in shepherds. *Dig Dis Sci* **44**, 1161-1164.

Dore, M. P., Sepulveda, A. R., Osato, M. S., Realdi, G. and Graham, D. Y. (1999b). *Helicobacter pylori* in sheep milk. *Lancet* **354**, 132.

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. *J Clin Microbiol* **37**, 2456-2460.

Dowson, C. G., Hutchison, A., Brannigan, J. A., George, R. C., Hansman, D., Linares, J., Tomasz, A., Smith, J. M. and Spratt, B. G. (1989). Horizontal transfer of penicillin-binding protein genes in penicillin-resistant clinical isolates of *Streptococcus pneumoniae*. *Proc Natl Acad Sci U S A* **86**, 8842-8846.

Drumm, B., Pérez-Pérez, G. I., Blaser, M. J. and Sherman, P. M. (1990). Intrafamilial clustering of *Helicobacter pylori* infection. *N Engl J Med* **322**, 359-363.

Dubois, A., Fiala, N., Heman-Ackah, L. M., Drazek, E. S., Tarnawski, A., Fishbein, W. N., Pérez-Pérez, G. I. and Blaser, M. J. (1994). Natural gastric infection with *Helicobacter pylori* in monkeys: a model for spiral bacteria infection in humans. *Gastroenterology* **106**, 1405-1417.

Dundon, W. G., Marshall, D. G., O Moráin, C. A. and Smyth, C. J. (2000). Population characteristics of Irish *Helicobacter pylori* isolates: a tRNA-associated locus. *Ir J Med Sci* **169**, 137-140.

Dunn, B. E., Cohen, H. and Blaser, M. J. (1997). *Helicobacter pylori*. *Clin Microbiol Rev* **10**, 720-741.

Eaton, K. A. and Krakowka, S. (1994). Effect of gastric pH on urease-dependent colonization of gnotobiotic piglets by *Helicobacter pylori*. *Infect Immun* **62**, 3604-3607.

Eaton, K. A., Suerbaum, S., Josenhans, C. and Krakowka, S. (1996). Colonization of gnotobiotic piglets by *Helicobacter pylori* deficient in two flagellin genes. *Infect Immun* **64**, 2445-2448.

Elitsur, Y., Adkins, L., Saeed, D. and Neace, C. (1999). *Helicobacter pylori* antibody profile in household members of children with H. *pylori* infection. *J Clin Gastroenterol* **29**, 178-182.

Enright, M. C. and Spratt, B. G. (1999). Multilocus sequence typing. *Trends Microbiol* 7, 482-487.

Enroth, H., Bjorkhölm, B. and Engstrand, L. (1999). Occurence of resistance mutation and clonal expansion in *Helicobacter pylori multiple*-strain infection: a potential risk in clarithromycin-based therapy. *Clin Infect Dis* **28**, 1305-1307.

Ernst, P. B. and Gold, B. D. (2000). The disease spectrum of *Helicobacter pylori*: the immunopathogenesis of gastroduodenal ulcer and gastric cancer. *Annu Rev Microbiol* **54**, 615-640.

Eun, C. S., Han, D. S., Park, J. Y., Jeon, Y. C., Hahm, J. S., Kim, K. S. and Kang, J. O. (2003). Changing pattern of antimicrobial resistance of *Helicobacter pylori* in Korean patients with peptic ulcer diseases. *J Gastroenterol* **38**, 436-441.

Evans, D. J., Jr. 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.

Evans, D. J., Jr., Evans, D. G., Kirkpatrick, S. S. and Graham, D. Y. (1991). Characterization of the *Helicobacter pylori* urease and purification of its subunits. *Microb Pathog* **10**, 15-26.

Evans, D. J., Jr., Evans, D. G., Engstrand, L. and Graham, D. Y. (1992). Urease-associated heat shock protein of *Helicobacter pylori*. *Infect Immun* **60**, 2125-2127.

Evans, D. J., Jr., Evans, D. G., Takemura, T., Nakano, H., Lampert, H. C., Graham, D. Y., Granger, D. N. and Kvietys, P. R. (1995). Characterization of a *Helicobacter pylori* neutrophil-activating protein. *Infect Immun* **63**, 2213-2220.

Evans, D. J., Jr., Queiroz, D. M., Mendes, E. N. and Evans, D. G. (1998). Diversity in the variable region of *Helicobacter pylori cagA* gene involves more than simple repetition of a 102-nucleotide sequence. *Biochem Biophys Res Commun* **245**, 780-784.

Evans, D. G., Queiroz, D. M., Mendes, E. N., Svennerholm, A. M. and Evans, D. J., Jr. (1999). Differences among *Helicobacter pylori* strains isolated from three different populations and demonstrated by restriction enzyme analysis of an internal fragment of the conserved gene hpaA. *Helicobacter* **4**, 82-88.

Falk, P. G., Roth, K. A., Borén, T., Westblom, T. U., Gordon, J. I. and Normark, S. (1993). An *in vitro* adherence assay reveals that *Helicobacter pylori* exhibits cell lineage-specific tropism in the human gastric epithelium. *Proc Natl Acad Sci U S A* **90**, 2035-2039.

Falk, P. G., Bry, L., Holgersson, J. and Gordon, J. I. (1995). Expression of a human alpha-1,3/4-fucosyltransferase in the pit cell lineage of FVB/N mouse stomach results in production of Leb-containing glycoconjugates: a potential transgenic mouse model for studying *Helicobacter pylori* infection. *Proc Natl Acad Sci U S A* **92**, 1515-1519.

Falush, D., Wirth, T., Linz, B. and 15 other authors (2003). Traces of human migrations in *Helicobacter pylori* populations. *Science* **299**, 1582-1585.

Fantry, G. T., Zheng, Q. X., Darwin, P. E., Rosenstein, A. H. and James, S. P. (1996). Mixed infection with *cagA*-positive and *cagA*-negative strains of *Helicobacter pylori*. *Helicobacter* **1**, 98-106.

Farsak, B., Yildirir, A., Akyon, Y., Pinar, A., Oc, M., Boke, E., Kes, S. and Tokgozoglu, L. (2000). Detection of *Chlamydia pneumoniae* and *Helicobacter pylori* DNA in human atherosclerotic plaques by PCR. *J Clin Microbiol* **38**, 4408-4411.

Feil, E. J. and Spratt, B. G. (2001). Recombination and the population structures of bacterial pathogens. *Annu Rev Microbiol* **55**, 561-590.

Ferea, T. L. and Brown, P. O. (1999). Observing the living genome. *Curr Opin Genet Dev* **9**, 715-722.

Fernando, N., Holton, J., Vaira, D., DeSilva, M. and Fernando, D. (2002). Prevalence of *Helicobacter pylori* in Sri Lanka as determined by PCR. *J Clin Microbiol* **40**, 2675-2676.

Figueiredo, C., Quint, W. G., Sanna, R. and other authors (2000). Genetic organization and heterogeneity of the *iceA* locus of *Helicobacter pylori*. *Gene* **246**, 59-68.

Forsyth, M. H., Atherton, J. C., Blaser, M. J. and Cover, T. L. (1998). Heterogeneity in levels of vacuolating cytotoxin gene (*vacA*) transcription among *Helicobacter pylori* strains. *Infect Immun* **66**, 3088-3094.

Fox, J. G., Correa, P., 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, P. A., Hu, L. T. and Mobley, H. L. (1992). Use of polymerase chain reaction-amplified *Helicobacter pylori* urease structural genes for differentiation of isolates. *J Clin Microbiol* **30**, 739-741.

Freedberg, A. S. and Baron, L. E. (1940). The presence of spirochetes in human gastric mucosa. *Am. J. Med. Sci.* **7**, 443-445.

Frenck, R. W., Jr. and Clemens, J. (2003). *Helicobacter* in the developing world. *Microbes Infect* **5**, 705-713.

Funk, D. J., Helbling, L., Wernegreen, J. J. and Moran, N. A. (2000). Intraspecific phylogenetic congruence among multiple symbiont genomes. *Proc R Soc Lond B Biol Sci* **267**, 2517-2521.

Genco, R. J. and Loos, B. G. (1991). The use of genomic DNA fingerprinting in studies of the epidemiology of bacteria in periodontitis. *J Clin Periodontol* **18**, 396-405.

Genta, R. M., Hamner, H. W. and Graham, D. Y. (1993). Gastric lymphoid follicles in *Helicobacter pylori* infection: frequency, distribution, and response to triple therapy. *Hum Pathol* **24**, 577-583.

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.

Ghiara, P., Marchetti, M., Blaser, M. J., Tummuru, M. K., 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. *Infect Immun* **63**, 4154-4160.

Ghose, C., Pérez-Pérez, G. I., Dominguez-Bello, M. G., Pride, D. T., Bravi, C. M. and Blaser, M. J. (2002). East Asian genotypes of *Helicobacter pylori* strains in Amerindians provide evidence for its ancient human carriage. *Proc Natl Acad Sci U S A* **99**, 15107-15111.

Gibson, J. R., Slater, E., Xerry, J., Tompkins, D. S. and Owen, R. J. (1998). Use of an amplified-fragment length polymorphism technique to fingerprint and differentiate isolates of *Helicobacter pylori*. *J Clin Microbiol* **36**, 2580-2585.

Go, M. F. (1996). *Helicobacter pylori*: its role in ulcer disease and gastric cancer and how to detect the infection. *Acta Gastroenterol Latinoam* **26**, 45-49.

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. *J Bacteriol* **178**, 3934-3938.

Goodman, K. J. and Correa, P. (2000). Transmission of *Helicobacter pylori* among siblings. *Lancet* **355**, 358-362.

Goodwin, C. S. and Armstrong, J. A. (1990). Microbiological aspects of *Helicobacter pylori* (Campylobacter *pylori*). *Eur J Clin Microbiol Infect Dis* **9**, 1-13.

Goodwin, C. S., McCulloch, R. K., Armstrong, J. A. and Wee, S. H. (1985). Unusual cellular fatty acids and distinctive ultrastructure in a new spiral bacterium (*Campylobacter pyloridis*) from the human gastric mucosa. *J Med Microbiol* **19**, 257-267.

Goulding, J. N., Hookey, J. V., Stanley, J., Olver, W., Neal, K. R., Ala'Aldeen, D. A. & Arnold, C. (2000). Fluorescent amplified-fragment length polymorphism genotyping of *Neisseria meningitidis* identifies clones associated with invasive disease. *J Clin Microbiol* **38**, 4580-4585.

Grady, R., O'Neill, G., Cookson, B. & Stanley, J. (2000). Fluorescent amplified-fragment length polymorphism analysis of the MRSA epidemic. *FEMS Microbiol Lett* **187**, 27-30.

Graham, D. Y. (1999). *Helicobacter pylori*, GERD, NSAIDs, and cancer: where we really stand. *Am J Gastroenterol* **94**, 1420-1421.

Graham, D. Y. (2003). The changing epidemiology of GERD: geography and *Helicobacter pylori*. *Am J Gastroenterol* **98**, 1462-1470.

Graham, D. Y. and Yamaoka, Y. (2000). Disease-specific *Helicobacter pylori* virulence factors: the unfulfilled promise. *Helicobacter* **5 Suppl 1**, S3-9; discussion S27-31.

Grove, D. I. and Koutsouridis, G. (2002). Increasing resistance of *Helicobacter pylori* to clarithromycin: is the horse bolting? *Pathology* **34**, 71-73.

Gzyl, A., Augustynowicz, E., Dzierzanowska, D., Rozynek, E., Dura, W., Celinska-Cedro, D. and Berg, D. E. (1999). Genotypes of *Helicobacter pylori* in Polish population. *Acta Microbiol Pol* **48**, 261-275.

Haas, R., Meyer, T. F. and van Putten, J. P. (1993). Aflagellated mutants of *Helicobacter pylori* generated by genetic transformation of naturally competent strains using transposon shuttle mutagenesis. *Mol Microbiol* **8**, 753-760.

Hacker, J. and Kaper, J. B. (2000). Pathogenicity islands and the evolution of microbes. *Annu Rev Microbiol* **54**, 641-679.

Han, S. R., Zschausch, H. C., Meyer, H. G., 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. *J Clin Microbiol* **38**, 3646-3651.

Handt, L. K., Fox, J. G., Dewhirst, F. E., Fraser, G. J., Paster, B. J., Yan, L. L., Rozmiarek, H., Rufo, R. and Stalis, I. H. (1994). *Helicobacter pylori* isolated from the domestic cat: public health implications. *Infect Immun* **62**, 2367-2374.

Handt, L. K., Fox, J. G., Yan, L. L., Shen, Z., Pouch, W. J., Ngai, D., Motzel, S. L., Nolan, T. E. and Klein, H. J. (1997). Diagnosis of *Helicobacter pylori* infection in a colony of rhesus monkeys (*Macaca mulatta*). *J Clin Microbiol* **35**, 165-168.

Harrington, C. A., Rosenow, C. and Retief, J. (2000). Monitoring gene expression using DNA microarrays. *Curr Opin Microbiol* **3**, 285-291.

Hatakeyama, M. (2003). *Helicobacter pylori* CagA--a potential bacterial oncoprotein that functionally mimics the mammalian Gab family of adaptor proteins. *Microbes Infect* **5**, 143-150.

Hazell, S. L., Lee, A., Brady, L. and Hennessy, W. (1986). *Campylobacter pyloridis* and gastritis: association with intercellular spaces and adaptation to an environment of mucus as important factors in colonization of the gastric epithelium. *J Infect Dis* **153**, 658-663.

Higashi, H., Tsutsumi, R., Fujita, A., Yamazaki, S., Asaka, M., Azuma, T. & Hatakeyama, M. (2002). Biological activity of the *Helicobacter pylori* virulence factor CagA is determined by variation in the tyrosine phosphorylation sites. *Proc Natl Acad Sci U S A* **99**, 14428-14433.

Hill, E. W., Jobling, M. A. & Bradley, D. G. (2000). Y-chromosome variation and Irish origins. *Nature* **404**, 351-352.

Hofreuter, D., Odenbreit, S., Puls, J., Schwan, D. and Haas, R. (2000). Genetic competence in *Helicobacter pylori*: mechanisms and biological implications. *Res Microbiol* **151**, 487-491.

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. *Mol Microbiol* **41**, 379-391.

Homuth, G., Domm, S., Kleiner, D. and Schumann, W. (2000). Transcriptional analysis of major heat shock genes of *Helicobacter pylori*. *J Bacteriol* **182**, 4257-4263.

Höök-Nikanne, J., Berg, D. E., Peek, R. M., Jr., Kersulyte, D., Tummuru, M. K. and Blaser, M. J. (1998). DNA sequence conservation and diversity in transposable element IS605 of *Helicobacter pylori*. *Helicobacter* **3**, 79-85.

Hoshino, F. B., Katayama, K., Watanabe, K., Takahashi, S., Uchimura, H. and Ando, T. (2000). Heterogeneity found in the *cagA* gene of *Helicobacter pylori* from Japanese and non-Japanese isolates. *J Gastroenterol* **35**, 890-897.

Hueck, C. J. (1998). Type III protein secretion systems in bacterial pathogens of animals and plants. *Microbiol Mol Biol Rev* **62**, 379-433.

Huesca, M., Goodwin, A., Bhagwansingh, A., Hoffman, P. and Lingwood, C. A. (1998). Characterization of an acidic-pH-inducible stress protein (hsp70), a putative sulfatide binding adhesin, from *Helicobacter pylori*. *Infect Immun* **66**, 4061-4067.

Hunter, T. (2000). Signaling-2000 and beyond. *Cell* **100**, 113-127.

Hurtado, A. and Owen, R. J. (1994). Identification of mixed genotypes in *Helicobacter pylori* from gastric biopsy tissue by analysis of urease gene polymorphisms. *FEMS Immunol Med Microbiol* **8**, 307-313.

Husson, M. O., Legrand, D., Spik, G. and Leclerc, H. (1993). Iron acquisition by *Helicobacter pylori*: importance of human lactoferrin. *Infect Immun* **61**, 2694-2697.

Illingworth, D. S., Walter, K. S., Griffiths, P. L. and Barclay, R. (1993). Siderophore production and iron-regulated envelope proteins of *Helicobacter pylori*. *Zentralbl Bakteriol* **280**, 113-119.

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: ASM press.

Israel, D. A., Lou, A. S. and Blaser, M. J. (2000). Characteristics of *Helicobacter pylori* natural transformation. *FEMS Microbiol Lett* **186**, 275-280.

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. *J Clin Invest* **107**, 611-620.

Israel, D. A., Salama, N., Krishna, U., Rieger, U. M., Atherton, J. C., Falkow, S. and Peek, R. M., Jr. (2001b). *Helicobacter pylori* genetic diversity within the gastric niche of a single human host. *Proc Natl Acad Sci U S A* **98**, 14625-14630.

Ito, Y., Azuma, T., Ito, S. and 7 other authors (1997). Analysis and typing of the *vacA* gene from *cagA*-positive strains of *Helicobacter pylori* isolated in Japan. *J Clin Microbiol* **35**, 1710-1714.

Janssen, P. J., Audit, B. and Ouzounis, C. A. (2001). Strain-specific genes of *Helicobacter pylori*: distribution, function and dynamics. *Nucleic Acids Res* **29**, 4395-4404.

Janvier, B., Grignon, B., Audibert, C., Pezennec, L. and Fauchère, J. L. (1999). Phenotypic changes of *Helicobacter pylori* components during an experimental infection in mice. *FEMS Immunol Med Microbiol* **24**, 27-33.

Jenks, P. J. (2001). *Helicobacter pylori*: towards new therapeutic targets. *J Med Microbiol* **50**, 485-488.

Jiang, Q., Hiratsuka, K. and Taylor, D. E. (1996). Variability of gene order in different *Helicobacter pylori* strains contributes to genome diversity. *Mol Microbiol* **20**, 833-842.

Josenhans, C. and Suerbaum, S. (2001). *Helicobacter* motility and chemotaxis. In Helicobacter pylori: *molecular and cellular biology*, pp. 171-184. Edited by M. Achtman, and S. Suerbaum: Horizon scientific press.

Joyce, E. A., Chan, K., Salama, N. R. and Falkow, S. (2002). Redefining bacterial populations: a post-genomic reformation. *Nat Rev Genet* **3**, 462-473.

Kalantar, J., Eslick, G. D. and Talley, N. J. (1999). Chronic gastritis and nonulcer dyspepsia. *Curr Top Microbiol Immunol* **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. *Res Microbiol* **147**, 661-669.

Karita, M. & Blaser, M. J. (1998). Acid-tolerance response in *Helicobacter pylori* and differences between *cagA+* and *cagA-* strains. *J Infect Dis* **178**, 213-219.

Karita, M., Teramukai, S. and Matsumoto, S. (2003). Risk of *Helicobacter pylori* transmission from drinking well water is higher than that from infected intrafamilial members in Japan. *Dig Dis Sci* **48**, 1062-1067.

Kersulyte, D., Chalkauskas, H. and Berg, D. E. (1999). Emergence of recombinant strains of *Helicobacter pylori* during human infection. *Mol Microbiol* **31**, 31-43.

Kersulyte, D., Mukhopadhyay, A. K., Velapatino, B. and 27 other authors (2000). Differences in genotypes of *Helicobacter pylori* from different human populations. *J Bacteriol* **182**, 3210-3218.

Kidd, M., Atherton, J. C., Lastovica, A. J. and Louw, J. A. (2001). Clustering of South African *Helicobacter pylori* isolates from peptic ulcer disease patients is demonstrated by repetitive extragenic palindromic-PCR fingerprinting. *J Clin Microbiol* **39**, 1833-1839.

Kim, K. S., Kang, J. O., Eun, C. S., Han, D. S. and Choi, T. Y. (2002). Mutations in the 23S rRNA gene of *Helicobacter pylori* associated with clarithromycin resistance. *J Korean Med Sci* **17**, 599-603.

Kodama, A., Matozaki, T., Fukuhara, A., Kikyo, M., Ichihashi, M. and Takai, Y. (2000). Involvement of an SHP-2-Rho small G protein pathway in hepatocyte growth factor/scatter factor-induced cell scattering. *Mol Biol Cell* **11**, 2565-2575.

Koike, T., Ohara, S., Sekine, H., Iijima, K., Abe, Y., Kato, K., Toyota, T. and Shimosegawa, T. (2001). *Helicobacter pylori* infection prevents erosive reflux oesophagitis by decreasing gastric acid secretion. *Gut* **49**, 330-334.

Krakowka, S., Morgan, D. R., Kraft, W. G. and Leunk, R. D. (1987). Establishment of gastric *Campylobacter pylori* infection in the neonatal gnotobiotic piglet. *Infect Immun* **55**, 2789-2796.

Krienitz, W. (1906). Ueber das auftreten von spirochaeten verschiedener form im mageninhalt bei carcinoma ventriculi. *Dtsch. Med. Wochenschr.* **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., Klinkenberg-Knol, E. C. & Meuwissen, S. G. (1999). *Helicobacter pylori*, proton pump inhibitors and gastroesophageal reflux disease. *Yale J Biol Med* **72**, 211-218.

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. *J Infect Dis* **181**, 273-282.

Kuo, C. H., Poon, S. K., Su, Y. C., Su, R., Chang, C. S. and Wang, W. C. (1999). Heterogeneous *Helicobacter pylori* isolates from *H. pylori*-infected couples in Taiwan. *J Infect Dis* **180**, 2064-2068.

Kusters, J. G., Gerrits, M. M., Van Strijp, J. A. and Vandenbroucke-Grauls, C. M. (1997). Coccoid forms of *Helicobacter pylori* are the morphologic manifestation of cell death. *Infect Immun* **65**, 3672-3679.

Kwon, D. H., El-Zaatari, F. A., Woo, J. S., Perng, C. L., Graham, D. Y. and Go, M. F. (1998). REP-PCR fragments as biomarkers for differentiating gastroduodenal disease-specific *Helicobacter pylori* strains. *Dig Dis Sci* **43**, 980-987.

Labenz, J., Blum, A. L., Bayerdorffer, E., Meining, A., Stolte, M. and Borsch G. (1997). Curing *Helicobacter pylori* infection in patients with duodenal ulcer may provoke reflux esophagitis. *Gastroenterology* **112**, 1442-1447.

Labigne, A., Cussac, V. and Courcoux, P. (1991). Shuttle cloning and nucleotide sequences of *Helicobacter pylori* genes responsible for urease activity. *J Bacteriol* **173**, 1920-1931.

Lacy, B. E. and Rosemore, J. (2001). *Helicobacter pylori*: ulcers and more: the beginning of an era. *J Nutr* **131**, 2789S-2793S.

Lai, Y. P., Yang, J. C., Lin, T. Z., Wang, J. T. and Lin, J. T. (2003). CagA tyrosine phosphorylation in gastric epithelial cells caused by *Helicobacter pylori* in patients with gastric adenocarcinoma. *Helicobacter* **8**, 235-243.

Langenberg, W., Rauws, E. A., Widjojokusumo, A., Tytgat, G. N. and Zanen, H. C. (1986). Identification of *Campylobacter pyloridis* isolates by restriction endonuclease DNA analysis. *J Clin Microbiol* **24**, 414-417.

Lee, A., Fox, J. G., Otto, G. and Murphy, J. (1990). A small animal model of human *Helicobacter pylori* active chronic gastritis. *Gastroenterology* **99**, 1315-1323.

Lee, A., Fox, J. G. and Hazell, S. (1993). Pathogenicity of *Helicobacter pylori*: a perspective. *Infect Immun* **61**, 1601-1610.

Lee, A., O'Rourke, J., De Ungria, M. C., Robertson, B., Daskalopoulos, G. and Dixon, M. F. (1997). A standardized mouse model of *Helicobacter pylori* infection: introducing the Sydney strain. *Gastroenterology* **112**, 1386-1397.

Letley, D. P., Rhead, J. L., Twells, R. J., Dove, B. and Atherton, J. C. (2003). Determinants of non-toxicity in the gastric pathogen *Helicobacter pylori*. *J Biol Chem* **278**, 26734-26741.

Leunk, R. D., Johnson, P. T., David, B. C., Kraft, W. G. and Morgan, D. R. (1988). Cytotoxic activity in broth-culture filtrates of *Campylobacter pylori*. *J Med Microbiol* **26**, 93-99.

Li, L., Genta, R. M., Go, M. F., Gutierrez, O., Kim, J. G. and Graham, D. Y. (2002a). *Helicobacter pylori* strain and the pattern of gastritis among first-degree relatives of patients with gastric carcinoma. *Helicobacter* **7**, 349-355.

Li, L., Graham, D. Y., Gutierrez, O., Kim, J. G., Genta, R. M., El-Zimaity, H. M. and Go, M. F. (2002b). Genomic fingerprinting and genotyping of *Helicobacter pylori* strains from patients with duodenal ulcer or gastric cancer from different geographic regions. *Dig Dis Sci* **47**, 2512-2518.

Liljeström, P. and Garoff, H. (1993). *SFV Expression Systems: Technical Manual,* 2nd edn. Karolinska Institute, Huddinge: Novum Research Centre.

Lin, C. W., Wu, S. C., Lee, S. C. and Cheng, K. S. (2000). Genetic analysis and clinical evaluation of vacuolating cytotoxin gene A and cytotoxin-associated gene A in Taiwanese *Helicobacter pylori* isolates from peptic ulcer patients. Scand J Infect Dis **32,** 51-57.

Madinier, I. M., Fosse, T. M. and Monteil, R. A. (1997). Oral carriage of *Helicobacter pylori*: a review. *J Periodontol* **68**, 2-6.

Maeda, S., Otsuka, M., Hirata, Y. and other authors (2001). cDNA microarray analysis of *Helicobacter pylori*-mediated alteration of gene expression in gastric cancer cells. *Biochem Biophys Res Commun* **284**, 443-449.

Maggi Solca, N., Bernasconi, M. V., Valsangiacomo, C., Van Doorn, L. J. and Piffaretti, J. C. (2001). Population genetics of *Helicobacter pylori* in the southern part of Switzerland analysed by sequencing of four housekeeping genes (*atpD*, *glnA*, *scoB* and *recA*), and by *vacA*, *cagA*, *iceA* and IS*605* genotyping. *Microbiology* **147**, 1693-1707.

Mai, U. E., Pérez-Pérez, G. I., Allen, J. B., Wahl, S. M., Blaser, M. J. and Smith, P. D. (1992). Surface proteins from *Helicobacter pylori* exhibit chemotactic activity for human leukocytes and are present in gastric mucosa. *J Exp Med* **175**, 517-525.

Maiden, M. C. (1998). Horizontal genetic exchange, evolution, and spread of antibiotic resistance in bacteria. *Clin Infect Dis* **27 Suppl 1**, S12-20.

Maiden, M. C., Bygraves, J. A., Feil, E. and 10 other authors (1998). Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc Natl Acad Sci U S A* **95**, 3140-3145.

Majewski, S. I. and Goodwin, C. S. (1988). Restriction endonuclease analysis of the genome of *Campylobacter pylori* with a rapid extraction method: evidence for considerable genomic variation. *J Infect Dis* **157**, 465-471.

Malaty, H. M. and Graham, D. Y. (1994). Importance of childhood socioeconomic status on the current prevalence of *Helicobacter pylori* infection. *Gut* **35**, 742-745.

Malaty, H. M., Kumagai, T., Tanaka, E., Ota, H., Kiyosawa, K., Graham, D. Y. and Katsuyama, T. (2000). Evidence from a nine-year birth cohort study in Japan of transmission pathways of *Helicobacter pylori* infection. *J Clin Microbiol* **38**, 1971-1973.

Malfertheiner, P., Leodolter, A. and Peitz, U. (2000). Cure of *Helicobacter pylori*-associated ulcer disease through eradication. *Baillieres Best Pract Res Clin Gastroenterol* **14**, 119-132.

Marais, A., Mendz, G. L., Hazell, S. L. and Mégraud, F. (1999). Metabolism and genetics of *Helicobacter pylori*: the genome era. *Microbiol Mol Biol Rev* **63**, 642-674.

Marchetti, M., Arico, 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 Goodwin, C. S. (1987). Revised nomenclature of *C. pyloridis*. *Int. J. Syst. Bacteriol.* **37**, 68.

Marshall, B. J. and Warren, J. R. (1984). Unidentified curved bacilli in the stomach of patients with gastritis and peptic ulceration. *Lancet* **1**, 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*. *Med J Aust* **142**, 436-439.

Marshall, D. G., Chua, A., Keeling, P. W., Sullivan, D. J., Coleman, D. C. and Smyth, C. J. (1995). Molecular analysis of *Helicobacter pylori* populations in antral biopsies from individual patients using randomly amplified polymorphic DNA (RAPD) fingerprinting. *FEMS Immunol Med Microbiol* **10**, 317-323.

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. *J Appl Bacteriol* **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 (Pt 11)**, 2925-2939.

Marshall, D. G., Hynes, S. O., Coleman, D. C., O'Morain, C. A., Smyth, C. J. and Moran, A. P. (1999). Lack of a relationship between Lewis antigen expression and *cagA*, CagA, *vacA* and VacA status of Irish *Helicobacter pylori* isolates. *FEMS Immunol Med Microbiol* **24**, 79-90.

Martino, M. C., Stabler, R. A., Zhang, Z. W., Farthing, M. J., 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. *Infect Immun* **69**, 1697-1703.

McClain, M. S., Cao, P., Iwamoto, H., Vinion-Dubiel, A. D., Szabo, G., Shao, Z. and Cover, T. L. (2001). A 12-amino-acid segment, present in type s2 but not type s1 *Helicobacter pylori* VacA proteins, abolishes cytotoxin activity and alters membrane channel formation. *J Bacteriol* **183**, 6499-6508.

McGovern, K. J., Blanchard, T. G., Gutierrez, J. A., Czinn, S. J., Krakowka, S. and Youngman, P. (2001). gamma-Glutamyltransferase is a *Helicobacter pylori* virulence factor but is not essential for colonization. *Infect Immun* **69**, 4168-4173.

Meyer, J. M., Silliman, N. P., Wang, W. and other authors (2002). Risk factors for *Helicobacter pylori* resistance in the United States: the surveillance of *H. pylori* antimicrobial resistance partnership (SHARP) study, 1993-1999. *Ann Intern Med* **136**, 13-24.

Michetti, P., Corthesy-Theulaz, I., Davin, C. and other authors (1994). Immunization of BALB/c mice against *Helicobacter felis* infection with *Helicobacter pylori* urease. *Gastroenterology* **107**, 1002-1011.

Michetti, P., Kreiss, C., Kotloff, K. L. and 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., Lehn, N., Meining, A., Bastlein, E., Mannes, G. A., Stolte, M. and Bayerdorffer, E. (1996). *Helicobacter pylori* reinfection is rare in peptic ulcer patients cured by antimicrobial therapy. *Eur J Gastroenterol Hepatol* **8**, 1161-1163.

Miehlke, S., Thomas, R., Guiterrez, O., Graham, D. Y. and Go, M. F. (1999). DNA fingerprinting of single colonies of *Helicobacter pylori* from gastric cancer patients suggests infection with a single predominant strain. *J Clin Microbiol* **37**, 245-247.

Miehlke, S., Kirsch, C., Agha-Amiri, K., Gunther, T., Lehn, N., Malfertheiner, P., Stolte, M., Ehninger, G. & Bayerdorffer, E. (2000). The *Helicobacter pylori* vacA s1, m1 genotype and *cagA* is associated with gastric carcinoma in Germany. *Int J Cancer* **87**, 322-327.

Minnis, J. A., Taylor, T. E., Knesek, J. E., Peterson, W. L. and McIntire, S. A. (1995). Characterization of a 3.5-kbp plasmid from *Helicobacter pylori*. *Plasmid* **34**, 22-36.

Mitchell, H. M., Li, Y. Y., Hu, P. J., Liu, Q., Chen, M., Du, G. G., Wang, Z. J., Lee, A. and Hazell, S. L. (1992). Epidemiology of *Helicobacter pylori* in southern China: identification of early childhood as the critical period for acquisition. *J Infect Dis* **166**, 149-153.

Miyaji, H., Azuma, T., Ito, S. and 7 other authors (1999). The effect of *Helicobacter pylori* eradication therapy on gastric antral myoelectrical activity and gastric emptying in patients with non-ulcer dyspepsia. *Aliment Pharmacol Ther* **13**, 1473-1480.

Mobley, H. L., Island, M. D. and Hausinger, R. P. (1995). Molecular biology of microbial ureases. *Microbiol Rev* **59**, 451-480.

Monstein, H. J. and Ellnebo-Svedlund, K. (2002). Molecular typing of *Helicobacter pylori* by virulence-gene based multiplex PCR and RT-PCR analysis. *Helicobacter* 7, 287-296.

Montecucco, C., Papini, E., de Bernard, M. and Zoratti, M. (1999). Molecular and cellular activities of *Helicobacter pylori* pathogenic factors. *FEBS Lett* **452**, 16-21.

Mukhopadhyay, A. K., Kersulyte, D., Jeong, J. Y. and 9 other authors (2000). Distinctiveness of genotypes of *Helicobacter pylori* in Calcutta, India. *J Bacteriol* **182**, 3219-3227.

Murakami, K., Fujioka, T., Shiota, K., Ito, A., Fujiyama, K., Kodama, R., Kawasaki, Y., Kubota, T. and Nasu, M. (1995). Influence of *Helicobacter pylori* infection and the effects of its eradication on gastric emptying in non-ulcerative dyspepsia. *Eur J Gastroenterol Hepatol* 7 **Suppl 1**, S93-97.

Mushegian, A. R. and Koonin, E. V. (1996). A minimal gene set for cellular life derived by comparison of complete bacterial genomes. *Proc Natl Acad Sci U S A* **93**, 10268-10273.

Musser, J. M. (1996). Molecular population genetic analysis of emerged bacterial pathogens: selected insights. *Emerg Infect Dis* **2**, 1-17.

Musser, J. M., Kroll, J. S., Granoff, D. M. and 7 other authors (1990). Global genetic structure and molecular epidemiology of encapsulated *Haemophilus influenzae*. *Rev Infect Dis* **12**, 75-111.

Nagai, H. and Roy, C. R. (2003). Show me the substrates: modulation of host cell function by type IV secretion systems. *Cell Microbiol* **5**, 373-383.

Ng, E. K., Thompson, S. A., Perez-Perez, G. I., Kansau, I., van der Ende, A., Labigne, A., Sung, J. J., Chung, S. C. and Blaser, M. J. (1999). *Helicobacter pylori* heat shock protein A: serologic responses and genetic diversity. *Clin Diagn Lab Immunol* **6**, 377-382.

Nordman, H., Boren, T., Davies, J. R., Engstrand, L. and Carlstedt, I. (1999). pH-dependent binding of *Helicobacter pylori* to pig gastric mucins. *FEMS Immunol Med Microbiol* **24**, 175-181.

Nordman, H., Davies, J. R., Lindell, G., de Bolos, C., Real, F. and Carlstedt, I. (2002). Gastric MUC5AC and MUC6 are large oligomeric mucins that differ in size, glycosylation and tissue distribution. *Biochem J* **364**, 191-200.

Ochman, H. and Jones, I. B. (2000). Evolutionary dynamics of full genome content in *Escherichia coli*. *Embo J* **19**, 6637-6643.

Ochman, H. and Moran, N. A. (2001). Genes lost and genes found: evolution of bacterial pathogenesis and symbiosis. *Science* **292**, 1096-1099.

Odenbreit, S. and Haas, R. (2002). *Helicobacter pylori*: impact of gene transfer and the role of the *cag* pathogenicity island for host adaptation and virulence. *Curr Top Microbiol Immunol* **264**, 1-22.

Odenbreit, S., Till, M. and Haas, R. (1996a). Optimized BlaM-transposon shuttle mutagenesis of *Helicobacter pylori* allows the identification of novel genetic loci involved in bacterial virulence. *Mol Microbiol* **20**, 361-373.

Odenbreit, S., Wieland, B. and Haas, R. (1996b). Cloning and genetic characterization of *Helicobacter pylori* catalase and construction of a catalase-deficient mutant strain. *J Bacteriol* **178**, 6960-6967.

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. *Mol Microbiol* **31**, 1537-1548.

Odenbreit, S., Puls, 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.

Ogura, K., Maeda, S., Nakao, M., Watanabe, T., Tada, M., Kyutoku, T., Yoshida, H., Shiratori, Y. and Omata, M. (2000). Virulence factors of *Helicobacter pylori* responsible for gastric diseases in Mongolian gerbil. *J Exp Med* **192**, 1601-1610.

O'Toole, P. W., Lane, M. C. and Porwollik, S. (2000). *Helicobacter pylori* motility. *Microbes Infect* **2**, 1207-1214.

Oudbier, J. H., Langenberg, W., Rauws, E. A. and Bruin-Mosch, C. (1990). Genotypical variation of *Campylobacter pylori* from gastric mucosa. *J Clin Microbiol* **28**, 559-565.

Owen, R. J. (1998). *Helicobacter*--species classification and identification. *Br Med Bull* **54**, 17-30.

Owen, R. J. and Xerry, J. (2003). Tracing clonality of *Helicobacter pylori* infecting family members from analysis of DNA sequences of three housekeeping genes (*ureI*, *atpA* and *ahpC*), deduced amino acid sequences, and pathogenicity-associated markers (*cagA* and *vacA*). *J Med Microbiol* **52**, 515-524.

Owen, R. J., Bickley, J., Hurtado, A., Fraser, A. and Pounder, R. E. (1994). Comparison of PCR-based restriction length polymorphism analysis of urease genes with rRNA gene profiling for monitoring *Helicobacter pylori* infections in patients on triple therapy. *J Clin Microbiol* **32**, 1203-1210.

Owen, R. J., Slater, E. R., Gibson, J., Lorenz, E. and Tompkins, D. S. (1999). Effect of clarithromycin and omeprazole therapy on the diversity and stability of genotypes of *Helicobacter pylori* from duodenal ulcer patients. *Microb Drug Resist* **5**, 141-146.

Owen, R. J., Ferrus, M. & Gibson, J. (2001). Amplified fragment length polymorphism genotyping of metronidazole-resistant *Helicobacter pylori* infecting dyspeptics in England. *Clin Microbiol Infect* **7**, 244-253.

Palmer, E. D. (1954). Investigation of the gastric mucosa spirochetes of the human. *Gastroenterol.* **27,** 218-220.

Pan, Z. J., van der Hulst, R. W., Feller, M., Xiao, S. D., Tytgat, G. N., 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. *J Clin Microbiol* **35**, 1344-1347.

Pan, Z. J., Berg, D. E., van der Hulst, R. W., Su, W. W., Raudonikiene, A., Xiao, S. D., Dankert, J., Tytgat, G. N. and van der Ende, A. (1998). Prevalence of vacuolating cytotoxin production and distribution of distinct *vacA* alleles in *Helicobacter pylori* from China. *J Infect Dis* **178**, 220-226.

Papiez, D., Konturek, P. C., Bielanski, W., Plonka, M., Dobrzanska, M., Kaminska, A., Szczyrk, U., Bochenek, A. and Wierzchos, E. (2003). Prevalence of *Helicobacter pylori* infection in Polish shepherds and their families. *Dig Liver Dis* **35**, 10-15.

Park, D. K., Jeong, S. K., Chung, M. G. and other authors (2003). [Glutathione levels in *Helicobacter pylori*-infected gastric mucosa]. *Korean J Gastroenterol* **42**, 267-273.

Parkin, D. M., Laara, E. and Muir, C. S. (1988). Estimates of the worldwide frequency of sixteen major cancers in 1980. *Int J Cancer* **41**, 184-197.

Parsonnet, J., Friedman, G. D., Vandersteen, D. P., Chang, Y., Vogelman, J. H., Orentreich, N. and Sibley, R. K. (1991). *Helicobacter pylori* infection and the risk of gastric carcinoma. *N Engl J Med* **325**, 1127-1131.

Pearson, A. D., Ireland, A., Bamforth, J., Walker, C., Booth, L., Hawtin, P., Holdstock, G. and Millward-Sadler, H. (1984). Polyacrylamide gel electrophoresis of spiral bacteria from the gastric antrum. *Lancet* **1**, 1349-1350.

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 Res* **27**, 3325-3333.

Peek, R. M., Jr., Thompson, S. A., Donahue, J. P., Tham, K. T., 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. *Proc Assoc Am Physicians* **110**, 531-544.

Peterson, W. L. (1991). Randomized controlled trials of peptic ulcer therapy in the *Helicobacter pylori* era. In *Helicobacter pylori in peptic ulceration and gastritis*, pp. 122-128. Edited by B. J. Marshall, R. W. McCallum and R. L. Guerrant: Blackwell Scientific Publications.

Piana, A., Are, B. M., Maida, I., Dore, M. P., Sotgiu, G., Realdi, G. and Mura, I. (2002). Genotypic characterization of clarithromycin-resistant *Helicobacter pylori* strains. *New Microbiol* **25**, 123-130.

Pütsep, K., Branden, C. I., Boman, H. G. and Normark, S. (1999). Antibacterial peptide from *H. pylori*. *Nature* **398**, 671-672.

Qasim, A., O'Morain, C. A. (2002). Review article: treatment of *Helicobacter pylori* infection and factors influencing eradication. *Aliment Pharmacol Ther* **16 Suppl 1**, 24-30.

Radin, M. J., Eaton, K. A., Krakowka, S., Morgan, D. R., Lee, A., Otto, G. and Fox, J. (1990). *Helicobacter pylori* gastric infection in gnotobiotic beagle dogs. *Infect Immun* **58**, 2606-2612.

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. *Mol Microbiol* **38**, 103-113.

Rasmussen, S. R., Aarestrup, F. M., Jensen, N. E. and Jorsal, S. E. (1999). Associations of *Streptococcus suis* serotype 2 ribotype profiles with clinical disease and antimicrobial resistance. *J Clin Microbiol* **37**, 404-408.

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 UreI membrane protein and an inner membrane proton barrier. *Mol Microbiol* **36**, 141-152.

Ricci, V., Ciacci, C., Zarrilli, R. and 7 other authors (1996). Effect of *Helicobacter pylori* on gastric epithelial cell migration and proliferation in vitro: role of VacA and CagA. *Infect Immun* **64**, 2829-2833.

Richter, J. E., Falk, G. W. and Vaezi, M. F. (1998). *Helicobacter pylori* and gastroesophageal reflux disease: the bug may not be all bad. *Am J Gastroenterol* **93**, 1800-1802.

Rohde, M., Puls, J., Buhrdorf, R., Fischer, W. and Haas, R. (2003). A novel sheathed surface organelle of the *Helicobacter pylori cag* type IV secretion system. *Mol Microbiol* **49**, 219-234.

Roma-Giannikou, E., Karameris, A., Balatsos, B., Panayiotou, J., Manika, Z., Van-Vliet, C., Rokkas, T., Skandalis, N. and Kattamis, C. (2003). Intrafamilial spread of *Helicobacter pylori*: a genetic analysis. *Helicobacter* **8**, 15-20.

Romaniuk, P. J., Zoltowska, B., Trust, T. J., Lane, D. J., Olsen, G. J., Pace, N. R. and Stahl, D. A. (1987). *Campylobacter pylori*, the spiral bacterium associated with human gastritis, is not a true Campylobacter sp. *J Bacteriol* **169**, 2137-2141.

Rothenbacher, D., Bode, G., Berg, G., Knayer, U., Gonser, T., Adler, G. and Brenner, H. (1999). *Helicobacter pylori* among preschool children and their parents: evidence of parent-child transmission. *J Infect Dis* **179**, 398-402.

Rothenbacher, D., Blaser, M. J., Bode, G. and Brenner, H. (2000). Inverse relationship between gastric colonization of *Helicobacter pylori* and diarrheal illnesses in children: results of a population-based cross-sectional study. *J Infect Dis* **182**, 1446-1449.

Ryan, K. A., Moran, A. P., Hynes, S. O., Smith, T., Hyde, D., O'Morain, C. A. and Maher, M. (2000). Genotyping of *cagA* and *vacA*, Lewis antigen status, and analysis of the poly-(C) tract in the alpha(1,3)-fucosyltransferase gene of Irish *Helicobacter pylori* isolates. *FEMS Immunol Med Microbiol* **28**, 113-120.

Sackmann, M., Morgner, A., Rudolph, B., Neubauer, A., Thiede, C., Schulz, H., Kraemer, W., Boersch, G., Rohde, P., Seifert, E., Stolte, M. and Bayerdoerffer, E. (1997). Regression of gastric MALT lymphoma after eradication of *Helicobacter pylori* is predicted by endosonographic staging. MALT Lymphoma Study Group. *Gastroenterology* **113**, 1087-1090.

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. *Proc Natl Acad Sci U S A* **97**, 14668-14673.

Salaun, L., Audibert, C., Le Lay, G., Burucoa, C., Fauchere, J. L. & Picard, B. (1998). Panmictic structure of *Helicobacter pylori* demonstrated by the comparative study of six genetic markers. *FEMS Microbiol Lett* **161**, 231-239.

Salonen, A., Vasiljeva, L., Merits, A., Magden, J., Jokitalo, E., and Kääriäinen. (2003). Properly folden nonstructural polyprotein directs the Semliki Forest virus replication complex to the endosomal compartment. *Journal of virology*. **77,** 1691-1702.

Salomon H. (1896). Ueber das spirillum saugetiermangens und sien verhalten zu den belegzellen (abstract 1). *Zentralbl Bakteriol* **19**, 433-442.

Salyers, A. A. and Amabile-Cuevas, C. F. (1997). Why are antibiotic resistance genes so resistant to elimination? *Antimicrob Agents Chemother* **41**, 2321-2325.

Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989). *Molecular Cloning: a Laboratory Manual,* 2nd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Salaun, L., Audibert, C., Le Lay, G., Burucoa, C., Fauchere, J. L. & Picard, B. (1998). Panmictic structure of *Helicobacter pylori* demonstrated by the comparative study of six genetic markers. *FEMS Microbiol Lett* **161**, 231-239.

Santos, A., Queiroz, D. M. M., Ménard, A., Marais, A., Rocha, G. A., Oliveira, C. A., Nogueira, A. M. M. F., Uzeda, M. and Mégraud, F. (2003). New pathogenicity marker found in the plasticity region of the *Helicobacter pylori* genome. *J Clin Microbiol* **41**, 1651-1655.

Satin, B., Del Giudice, G., Della Bianca, V. and 7 other authors (2000). The neutrophil-activating protein (HP-NAP) of *Helicobacter pylori* is a protective antigen and a major virulence factor. *J Exp Med* **191**, 1467-1476.

Schmitt, W. and Haas, R. (1994). Genetic analysis of the *Helicobacter pylori* vacuolating cytotoxin: structural similarities with the IgA protease type of exported protein. *Mol Microbiol* **12**, 307-319.

Scott, D. R., Weeks, D., Hong, C., Postius, S., Melchers, K. and Sachs, G. (1998). The role of internal urease in acid resistance of *Helicobacter pylori*. *Gastroenterology* **114**, 58-70.

Segal, E. D., Falkow, S. and Tompkins, L. S. (1996). *Helicobacter pylori* attachment to gastric cells induces cytoskeletal rearrangements and tyrosine phosphorylation of host cell proteins. *Proc Natl Acad Sci U S A* **93**, 1259-1264.

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*. *Proc Natl Acad Sci U S A* **96**, 14559-14564.

Selbach, M., Moese, S., Hauck, C. R., Meyer, T. F. and Backert, S. (2002). Src is the kinase of the *Helicobacter pylori* CagA protein *in vitro* and *in vivo*. *J Biol Chem* **277**, 6775-6778.

Shibayama, K., Kamachi, K., Nagata, N. and 9 other authors (2003). A novel apoptosis-inducing protein from *Helicobacter pylori*. *Mol Microbiol* **47**, 443-451.

Shimizu, T., Oguchi, S., Yamashiro, Y., Segawa, O., Ohkura, R., Wakisaka, N. & Yamamoto, T. (1999). *Helicobacter pylori* transmission between a boy with duodenal ulcer and his father. *Pediatr Infect Dis J* **18**, 655-656.

Shimizu, T., Yarita, Y., Kaneko, K., Yamashiro, Y., Segawa, O., Ohkura, R., Taneike, I. & Yamamoto, T. (2000). Case of intrafamilial *Helicobacter pylori* reinfection after successful eradication therapy. *Pediatr Infect Dis J* **19**, 901-903.

Shimoyama, T., Fukuda, S., Tanaka, M., Mikami, T., Saito, Y. and Munakata, A. (1997). High prevalence of the CagA-positive *Helicobacter pylori* strains in Japanese asymptomatic patients and gastric cancer patients. *Scand J Gastroenterol* **32**, 465-468.

Skouloubris, S., Thiberge, J. M., 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. *Infect Immun* **66**, 4517-4521.

Smeets, L. C. and Kusters, J. G. (2002). Natural transformation in *Helicobacter pylori*: DNA transport in an unexpected way. *Trends Microbiol* **10**, 159-162; discussion 162.

Smith, J. M. (1995). Do bacteria have population genetics? In *Population genetics of bacteria*, pp. 1-12. Edited by S. Baumberg, J. P. W. Young, E. M. H. Wellington and J. R. Saunders: Cambridge University Press.

Smith, J. M., Smith, N. H., O'Rourke, M. & Spratt, B. G. (1993). How clonal are bacteria? *Proc Natl Acad Sci U S A* **90**, 4384-4388.

Smith, S. I., Cantet, F., Angelini, F., Marais, A., Mégraud, F., Bayerdoffer, E. and Miehlke, S. (2002a). Discriminatory power of RAPD, PCR-RFLP and southern blot analyses of *ureCD* or *ureA* gene probes on *Helicobacter pylori* isolates. *Z Naturforsch [C]* **57**, 516-521.

Smith, S. I., Kirsch, C., Oyedeji, K. S., Arigbabu, A. O., Coker, A. O., Bayerdoffer, E. and Miehlke, S. (2002b). Prevalence of *Helicobacter pylori vacA*, *cagA* and *iceA* genotypes in Nigerian patients with duodenal ulcer disease. *J Med Microbiol* **51**, 851-854.

Solnick, J. V. and Vandamme, P. (2001). Taxonomy of the *Helicobacter* genus. In Helicobacter pylori *physiology and genetics*, pp. 39-51. Edited by H. L. T. Mobley, G. L. Mendz and S. L. Hazell: ASM press.

Solnick, J. V., Hansen, L. M. and Syvanen, M. (1997). The major sigma factor (RpoD) from *Helicobacter pylori* and other gram-negative bacteria shows an enhanced rate of divergence. *J Bacteriol* **179**, 6196-6200.

Songyang, Z., Shoelson, S. E., Chaudhuri, M. and 7 other authors (1993). SH2 domains recognize specific phosphopeptide sequences. *Cell* **72**, 767-778.

Songyang, Z., Carraway, K. L., 3rd, Eck, M. J. and 7 other authors (1995). Catalytic specificity of protein-tyrosine kinases is critical for selective signalling. *Nature* **373**, 536-539.

Spohn, G. & Scarlato, V. (1999). Motility of *Helicobacter pylori* is coordinately regulated by the transcriptional activator FlgR, an NtrC homolog. *J Bacteriol* **181**, 593-599.

Spratt, B. G. (2003). Microbiology. Stomachs out of Africa. *Science* **299**, 1528-1529.

Spratt, B. G. and Maiden, M. C. (1999). Bacterial population genetics, evolution and epidemiology. *Philos Trans R Soc Lond B Biol Sci* **354**, 701-710.

Spratt, B. G., Zhang, Q. Y., Jones, D. M., Hutchison, A., Brannigan, J. A. and Dowson, C. G. (1989). Recruitment of a penicillin-binding protein gene from *Neisseria flavescens* during the emergence of penicillin resistance in *Neisseria meningitidis*. *Proc Natl Acad Sci U S A* **86**, 8988-8992.

Stein, M., Rappuoli, R. and Covacci, A. (2000). Tyrosine phosphorylation of the *Helicobacter pylori* CagA antigen after *cag*-driven host cell translocation. *Proc Natl Acad Sci U S A* **97**, 1263-1268.

Stein, M., Bagnoli, F., Halenbeck, R., Rappuoli, R., Fantl, W. J. and Covacci, A. (2002). c-Src/Lyn kinases activate *Helicobacter pylori* CagA through tyrosine phosphorylation of the EPIYA motifs. *Mol Microbiol* **43**, 971-980.

Stingl, K., Altendorf, K. and Bakker, E. P. (2002). Acid survival of *Helicobacter pylori*: how does urease activity trigger cytoplasmic pH homeostasis? *Trends Microbiol* **10**, 70-74.

Stolte, M. (1992). *Helicobacter pylori* gastritis and gastric MALT-lymphoma. *Lancet* **339**, 745-746.

Strauss, J. H. and Strauss, E. G. (1994). The alphaviruses: gene expression, replication, and evolution. *Microbiol Rev* **58**, 491-562.

Strobel, S., Bereswill, S., Balig, P., Allgaier, P., Sonntag, H. G. and Kist, M. (1998). Identification and analysis of a new *vacA* genotype variant of *Helicobacter pylori* in different patient groups in Germany. *J Clin Microbiol* **36**, 1285-1289.

Suerbaum, S. (1995). The complex flagella of gastric *Helicobacter* species. *Trends Microbiol* **3**, 168-170; discussion 170-161.

Suerbaum, S. and Achtman M. (2001). Population genetics. In Helicobacter pylori *physiology and genetics*, pp. 355-361. Edited by H. L. T. Mobley, G. L. Mendz and S. L. Hazell: ASM press.

Suerbaum, S., Smith, J. M., Bapumia, K., Morelli, G., Smith, N. H., Kunstmann, E., Dyrek, I. and Achtman, M. (1998). Free recombination within *Helicobacter pylori*. *Proc Natl Acad Sci U S A* **95**, 12619-12624.

Suzuki, J., Muraoka, H., Kobayasi, I., Fujita, T. and Mine, T. (1999). Rare incidence of interspousal transmission of *Helicobacter pylori* in asymptomatic individuals in Japan. *J Clin Microbiol* **37**, 4174-4176.

Taddei, F., Radman, M., Maynard-Smith, J., Toupance, B., Gouyon, P. H. and Godelle, B. (1997). Role of mutator alleles in adaptive evolution. *Nature* **387**, 700-702.

Tanaka, S., Meiselman, H. H., Engel, E., Guth, P. H., Furukawa, O., Wenby, R. B., Lee, J. & Kaunitz, J. D. (2002). Regional differences of H+, HCO3-, and CO2 diffusion through native porcine gastroduodenal mucus. *Dig Dis Sci* **47**, 967-973.

Taneike, I., Tamura, Y., Shimizu, T., Yamashiro, Y. & Yamamoto, T. (2001). *Helicobacter pylori* intrafamilial infections: change in source of infection of a child from father to mother after eradication therapy. *Clin Diagn Lab Immunol* **8**, 731-739.

Tee, W., Lambert, J., Smallwood, R., Schembri, M., Ross, B. C. and Dwyer, B. (1992). Ribotyping of *Helicobacter pylori* from clinical specimens. *J Clin Microbiol* **30**, 1562-1567.

Telford, J. L., Ghiara, P., Dell'Orco, M. and other authors (1994). Gene structure of the *Helicobacter pylori* cytotoxin and evidence of its key role in gastric disease. *J Exp Med* **179**, 1653-1658.

ten Dam, E., Flint, M. and Ryan M. (1999). Virus-encoded proteinases of the *Togaviridae*. *Journal of General Virology*. **80,** 1879-1888.

Testerman, T. L., McGee, D. J. and Mobley, L. T. (2001). Adherence and colonization. In Helicobacter pylori *physiology and genetics*, pp.381-417. Edited by H. L. T. Mobley, G. L. Mendz and S. L. Hazell: ASM press.

Thoreson, A. C., Hosseini, N., Svennerholm, A. M. and Bolin, I. (2000). Different *Helicobacter pylori* strains colonize the antral and duodenal mucosa of duodenal ulcer patients. *Helicobacter* **5**, 69-78.

Tindberg, Y., Bengtsson, C., Granath, F., Blennow, M., Nyren, O. and Granstrom, M. (2001). *Helicobacter pylori* infection in Swedish school children: lack of evidence of child-to-child transmission outside the family. *Gastroenterology* **121**, 310-316.

Tomb, J. F., White, O., Kerlavage, A. R. and 21 other authors (1997). The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature* **388**, 539-547.

Tsuda, M., Karita, M. and Nakazawa, T. (1993). Genetic transformation in *Helicobacter pylori*. *Microbiol Immunol* **37**, 85-89.

Tsutsumi, R., Higashi, H., Higuchi, M., Okada, M. and Hatakeyama, M. (2003). Attenuation of *Helicobacter pylori* CagA x SHP-2 signaling by interaction between CagA and C-terminal Src kinase. *J Biol Chem* **278**, 3664-3670.

Tummuru, M. K., 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. *Infect Immun* **61**, 1799-1809.

Tummuru, M. K., Cover, T. L. and Blaser, M. J. (1994). Mutation of the cytotoxin-associated *cagA* gene does not affect the vacuolating cytotoxin activity of *Helicobacter pylori*. *Infect Immun* **62**, 2609-2613.

Uemura, N., Okamoto, S., Yamamoto, S., Matsumura, N., Yamaguchi, S., Yamakido, M., Taniyama, K., Sasaki, N. and Schlemper, R. J. (2001). *Helicobacter pylori* infection and the development of gastric cancer. *N Engl J Med* **345**, 784-789.

Utas, S., Ozbakir, O., Turasan, A. and Utas, C. (1999). *Helicobacter pylori* eradication treatment reduces the severity of rosacea. *J Am Acad Dermatol* **40**, 433-435.

van Belkum, A. (2003). High-throughput epidemiologic typing in clinical microbiology. *Clin Microbiol Infect* **9**, 86-100.

van der Ende, A., Rauws, E. A., Feller, M., Mulder, C. J., Tytgat, G. N. and Dankert, J. (1996). Heterogeneous *Helicobacter pylori* isolates from members of a family with a history of peptic ulcer disease. *Gastroenterology* **111**, 638-647.

van der Ende, A., Pan, Z. J., Bart, A., van der Hulst, R. W., Feller, M., Xiao, S. D., Tytgat, G. N. and Dankert, J. (1998). cagA-positive *Helicobacter pylori* populations in China and The Netherlands are distinct. *Infect Immun* **66**, 1822-1826.

van der Ende, A., van Doorn, L. J., Rooijakkers, S., Feller, M., Tytgat, G. N. and Dankert, J. (2001). Clarithromycin-susceptible and -resistant *Helicobacter pylori* isolates with identical randomly amplified polymorphic DNA-PCR genotypes cultured from single gastric biopsy specimens prior to antibiotic therapy. *J Clin Microbiol* **39**, 2648-2651.

van Doorn, L. J., Figueiredo, C., Sanna, R., Pena, S., Midolo, P., Ng, E. K., Atherton, J. C., Blaser, M. J. and Quint, W. G. (1998a). Expanding allelic diversity of *Helicobacter pylori vacA*. *J Clin Microbiol* **36**, 2597-2603.

van Doorn, L. J., Figueiredo, C., Sanna, R., Plaisier, A., Schneeberger, P., de Boer, W. and Quint, W. (1998b). Clinical relevance of the *cagA*, *vacA*, and *iceA* status of *Helicobacter pylori*. *Gastroenterology* **115**, 58-66.

van Doorn, N. E., Namavar, F., Kusters, J. G., van Rees, E. P., Kuipers, E. J. and de Graaff, J. (1998c). Genomic DNA fingerprinting of clinical isolates of *Helicobacter pylori* by REP-PCR and restriction fragment end-labelling. *FEMS Microbiol Lett* **160**, 145-150.

van Doorn, L. J., Figueiredo, C., Mégraud, F. and 14 other authors (1999a). Geographic distribution of *vacA* allelic types of *Helicobacter pylori*. *Gastroenterology* **116**, 823-830.

van Doorn, L. J., Figueiredo, C., Sanna, R., Blaser, M. J. and Quint, W. G. (1999b). Distinct variants of *Helicobacter pylori cagA* are associated with *vacA* subtypes. *J Clin Microbiol* **37**, 2306-2311.

van Putten, J. P. (1993). Phase variation of lipopolysaccharide directs interconversion of invasive and immuno-resistant phenotypes of *Neisseria gonorrhoeae*. *Embo J* **12**, 4043-4051.

Versalovic, J. and Fox, J. G. (2001). Taxonomy and phylogeny of *Helicobacter*. In Helicobacter pylori: *molecular and cellular biology*, pp. 15-28. Edited by M. Achtman, and S. Suerbaum. Horizon scientific press.

Vicari, J. J., Peek, R. M., Falk, G. W., Goldblum, J. R., Easley, K. A., Schnell, J., Perez-Perez, G. I., Halter, S. A., Rice, T. W., Blaser, M. J. and Richter, J. E. (1998). The seroprevalence of *cagA*-positive *Helicobacter pylori* strains in the spectrum of gastroesophageal reflux disease. *Gastroenterology***115**, 50-57.

Vieth, M. and Stolte, M. (1999). Re: Richter et al.: possibly protective properties of *Helicobacter pylori* in connection with GERD. *Am J Gastroenterol* **94**, 3068-3069.

Vos, P., Hogers, R., Bleeker, M. and other authors (1995). AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res* **23**, 4407-4414.

Walker, M. M. and Dixon, M. F. (1996). Gastric metaplasia: its role in duodenal ulceration. *Aliment Pharmacol Ther* **10 Suppl 1**, 119-128.

Wang, G., Humayun, M. Z. and Taylor, D. E. (1999). Mutation as an origin of genetic variability in *Helicobacter pylori*. *Trends Microbiol* 7, 488-493.

Wang, J. T., Sheu, J. C., Lin, J. T., Wang, T. H. and Wu, M. S. (1993). Direct DNA amplification and restriction pattern analysis of *Helicobacter pylori* in patients with duodenal ulcer and their families. *J Infect Dis* **168**, 1544-1548.

Wang, X., Sturegard, E., Rupar, R., Nilsson, H. O., Aleljung, P. A., Carlen, B., Willen, R. and Wadstrom, T. (1997). Infection of BALB/c A mice by spiral and coccoid forms of *Helicobacter pylori*. *J Med Microbiol* **46**, 657-663.

Warburton-Timms, V. J., Charlett, A., Valori, R. M., Uff, J. S., Shepherd, N. A., Barr, H. and McNulty, C. A. (2001). The significance of *cagA*(+) *Helicobacter pylori* in reflux oesophagitis. *Gut* **49**, 341-346.

Watanabe, T., Tada, M., Nagai, H., Sasaki, S. and Nakao, M. (1998). *Helicobacter pylori* infection induces gastric cancer in mongolian gerbils. *Gastroenterology* **115**, 642-648.

Weeks, D. L., Eskandari, S., Scott, D. R. & Sachs, G. (2000). A H+-gated urea channel: the link between *Helicobacter pylori* urease and gastric colonization. *Science* **287**, 482-485.

Weiser, J. N., Maskell, D. J., Butler, P. D., Lindberg, A. A. and Moxon, E. R. (1990). Characterization of repetitive sequences controlling phase variation of *Haemophilus influenzae* lipopolysaccharide. *J Bacteriol* **172**, 3304-3309.

Wong, B. C., Lam, S. K., Ching, C. K. & other authors (1999). Seroprevalence of cytotoxin-associated gene A positive *Helicobacter pylori* strains in Changle, an area with very high prevalence of gastric cancer in south China. *Aliment Pharmacol Ther* **13**, 1295-1302.

Wotherspoon, A. C., Ortiz-Hidalgo, C., Falzon, M. R. and Isaacson, P. G. (1991). *Helicobacter pylori*-associated gastritis and primary B-cell gastric lymphoma. *Lancet* **338**, 1175-1176.

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. *Infect Immun* **63**, 94-98.

Xu, Q. and Blaser, M. J. (2001). Promoters of the CATG-specific methyltransferase gene hpyIM differ between iceA1 and iceA2 *Helicobacter pylori* strains. *J Bacteriol* **183**, 3875-3884.

Xu, Q., Peek, R. M., Jr., Miller, G. G. and Blaser, M. J. (1997). The *Helicobacter pylori* genome is modified at CATG by the product of *hpyIM*. *J Bacteriol* **179**, 6807-6815.

Yakoob, J., Hu, G., Fan, X. and Zhang, Z. (2001). *Helicobacter pylori* detection in Chinese subjects: a comparison of two common DNA fingerprinting methods. *Br J Biomed Sci* **58**, 239-243.

Yamaji, Y., Mitsushima, T., Ikuma, H. and other authors (2001). Inverse background of *Helicobacter pylori* antibody and pepsinogen in reflux oesophagitis compared with gastric cancer: analysis of 5732 Japanese subjects. *Gut* **49**, 335-340.

Yamaoka, Y., Kodama, T., Kashima, K., Graham, D. Y. and Sepulveda, A. R. (1998). Variants of the 3' region of the *cagA* gene in *Helicobacter pylori* isolates from patients with different *H. pylori*-associated diseases. *J Clin Microbiol* **36**, 2258-2263.

Yamaoka, Y., El-Zimaity, H. M., Gutierrez, O., Figura, N., Kim, J. G., Kodama, T., Kashima, K., Graham, D. Y. and Kim, J. K. (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. *J Clin Microbiol* **37**, 2274-2279.

Yamaoka, Y., Malaty, H. M., Osato, M. S. and Graham, D. Y. (2000a). Conservation of *Helicobacter pylori* genotypes in different ethnic groups in Houston, Texas. *J Infect Dis* **181**, 2083-2086.

Yamaoka, Y., Osato, M. S., Sepulveda, A. R., Gutierrez, O., Figura, N., Kim, J. G., Kodama, T., Kashima, K. and Graham, D. Y. (2000b). Molecular epidemiology of *Helicobacter pylori*: separation of *H. pylori* from East Asian and non-Asian countries. *Epidemiol Infect* **124**, 91-96.

Yamaoka, Y., Orito, E., Mizokami, M. and other authors (2002). *Helicobacter pylori* in North and South America before Columbus. *FEBS Lett* **517**, 180-184.

Yamazaki, S., Yamakawa, A., Ito, Y., Ohtani, M., Higashi, H., Hatakeyama, M. and Azuma, T. (2003). The CagA protein of *Helicobacter pylori* is translocated into epithelial cells and binds to SHP-2 in human gastric mucosa. *J Infect Dis* **187**, 334-337.

Yang, J. C., Kuo, C. H., Wang, H. J., Wang, T. C., Chang, C. S. and Wang, W. C. (1998). Vacuolating toxin gene polymorphism among *Helicobacter pylori* clinical isolates and its association with m1, m2, or chimeric *vacA* middle types. *Scand J Gastroenterol* **33**, 1152-1157.

Yokota, K., Hirai, Y., Haque, M. and other authors (1994). Heat shock protein produced by *Helicobacter pylori*. *Microbiol Immunol* **38**, 403-405.

Yoshida, N., Granger, D. N., Evans, D. J., Jr., Evans, D. G., Graham, D. Y., Anderson, D. C., Wolf, R. E. and Kvietys, P. R. (1993). Mechanisms involved in *Helicobacter pylori*-induced inflammation. *Gastroenterology* **105**, 1431-1440.

Yoshiyama, H., Nakamura, H., Kimoto, M., Okita, K. and Nakazawa, T. (1999). Chemotaxis and motility of *Helicobacter pylori* in a viscous environment. *J Gastroenterol* **34 Suppl 11**, 18-23.