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*)

Robert J. Owen and Jacqueline Xerry

Helicobacter Reference Unit, Laboratory of Enteric Pathogens, Health Protection Agency, 61 Colindale Avenue, London, NW9 5HT, UK

Correspondence

Robert J. Owen
rowen@phls.nhs.uk

Helicobacter pylori, a Gram-negative bacterium, is a causal agent of peptic ulcers and is estimated to infect the gastric mucosa of at least half of the world's population. As primary infections are acquired mainly by household contact, studies on family clusters provide a model for investigating transmission and the natural history of initial infection. Here, sequence typing exploiting genetic variation in core fragments of three key housekeeping loci (*ureI*, *atpA* and *ahpC*) was used to determine clonal descent amongst isolates of ten members of four families in Northern Ireland and a family with three generations in central England. Phylogenetic analysis of each locus for 73 strains of *H. pylori* from 11 countries indicated high background intraspecific diversity, apart from identical paired isolates from five unrelated patients and strains with identical sequence types (STs) detected in adult members of two families. In several families carrying strains with different STs, evidence of residual clonal descent was detected at one or two loci by comparison of nucleotide and amino acid sequences. Pathogenicity-associated genotypes were heterogeneous with respect to ST and amino acid type. Analysis of these three housekeeping genes provides unique evidence for precise tracing of clonal descent in isolates of *H. pylori* in family groups.

Received 11 June 2002

Accepted 6 January 2003

INTRODUCTION

Helicobacter pylori, a Gram-negative microaerobic bacterium, has a worldwide distribution in the human population, is a key causal factor in peptic ulcer disease and is an important risk factor in the initial stages of the development of gastric cancer (Dunn *et al.*, 1997; Calam & Baron, 2001). Seroepidemiological studies of *H. pylori* indicate infection rates of about 30 % in the UK and other developed countries, but rates in many developing countries are significantly higher (Pounder & Ng, 1995; Brown, 2000; Feldman, 2001). The precise routes of *H. pylori* infection remain largely unclear and baseline information on the epidemiology of individual strain types is limited. Infection occurs predominantly by household contact in childhood, probably through person-to-person transmission by an oral–oral or faecal–oral route (Mendall *et al.*, 1992; Rowland, 2000; Feldman, 2001). Acquisition of new infection seems to occur mainly within the first 10 years of life (Malaty *et al.*, 2002) although

in some high-risk groups, acquisition is most likely in the first 2 years and will persist for decades unless eradicated (Rowland, 2000).

An understanding of the natural history of initial infection is provided by investigation of intrafamilial transmission. Several immunological studies have demonstrated clustering of infection amongst family members (Drumm *et al.*, 1990; Malaty *et al.*, 1991; Oderda *et al.*, 1991; Dominici *et al.*, 1999), including antibody profiling (Ng *et al.*, 2001). The importance of adult–child transmission has been demonstrated, with parents, in particular infected mothers, playing a key role as a source of *H. pylori* (Rothenbacher *et al.*, 1999; Malaty *et al.*, 2000; Taneike *et al.*, 2001). Most of these epidemiological investigations have relied on serology and urea breath-tests to define intrafamilial clustering, but genotyping provides more precise information about transmission of specific strains between individuals. The genomic heterogeneity of *H. pylori* is well-documented (Owen *et al.*, 2001b), and isolates in families have been investigated using a variety of DNA-based methods that have included ribotyping (Nwokolo *et al.*, 1992; Bamford *et al.*, 1993; Taneike *et al.*, 2001), arbitrarily primed (RAPD) typing (van der Ende *et al.*, 1996), PCR-RFLP and PFGE analysis (Han *et al.*, 2000) and

Abbreviations: MLST, multilocus sequence typing; RAPD, random amplified polymorphic DNA; ST, sequence type.

sequencing of pathogenicity-associated genes such as *vacA*, *flaA* and *flaB* (Suerbaum *et al.*, 1998). Population genetic analysis by multilocus sequence typing (MLST), which is based on the concept of defining isolates according to allelic variation using seven internal fragment sequences (about 500 bp) of housekeeping loci (Enright & Spratt, 1999), has demonstrated that *H. pylori* has a non-clonal (panmictic) population structure, due to horizontal gene transfer and frequent recombination resulting in extensive genetic rearrangements (Achtman *et al.*, 1999; Maggi Solcà *et al.*, 2001). A similar conclusion was reached from sequence analysis of several virulence-associated loci (Achtman *et al.*, 1999). Even so, *H. pylori* appears to be clonal over short periods of time, as there is evidence that clonal descent is readily detected within some family groups in Germany (Suerbaum *et al.*, 1998; Han *et al.*, 2000) and in Japan (Malaty *et al.*, 2000). Subsequent development of diversity and loss of clonality may occur during mixed colonization by unrelated strains (Falush *et al.*, 2001; Israel *et al.*, 2001).

Our aim in this study was to explore the existence of clonality in *H. pylori* infecting family groups by investigation of sequence diversity within core regions of three housekeeping genes (*ureI*, *atpA* and *ahpC*). To assess the discriminatory power of sequence typing for *H. pylori* and to obtain an indication of the overall background species diversity at these loci, we examined isolates from two gastric sites, as well as strains from single sites, within unrelated individuals. We also analysed inferred amino acid sequences to gain an insight into the frequency of non-synonymous variation, and we determined heterogeneity in two key pathogenicity-associated markers (*cagA* status and *vacA* allelic type) to assess their contribution to strain diversity within the family clusters.

METHODS

Bacterial strains. The 73 isolates and reference strains of *H. pylori* studied were as follows. Sixteen family isolates were tested: one isolate was obtained from each of six members of a family (representing three generations) with a high incidence of duodenal ulcer disease in central England (Coventry set) (Nwokolo *et al.*, 1992); one isolate was also obtained from each of ten members of four Northern Ireland (Belfast) families, each comprising a child (index case) being investigated for abdominal pain and the respective healthy parent(s) at a paediatric clinic in Belfast (Bamford *et al.*, 1993). Ten isolates representing paired gastric (antrum and corpus) samples were from five unrelated individuals in London with dyspeptic symptoms. The genome-sequenced strain NCTC 12455 (= strain 26695) (Tomb *et al.*, 1997) was obtained from the National Collection of Type Cultures (NCTC) in lyophilized form, and 46 miscellaneous isolates were selected from our laboratory collection. Fifteen of these originated from antral biopsy cultures from unrelated adult dyspeptic patients in London, UK ($n = 13$ isolates) undergoing routine upper gastrointestinal endoscopic investigation for various presentations between 1982 and 1999, and Belfast, Northern Ireland ($n = 2$). Twenty-nine isolates originated from ten other countries, namely Nigeria ($n = 8$), South Africa ($n = 6$), Japan ($n = 3$), Canada (NCTC 13082, NCTC 13084 and NCTC 13086), France ($n = 2$), Turkey ($n = 2$), Australia (NCTC 11637^T, the nomenclatural type strain, and NCTC 11638), China ($n = 1$), USA (NCTC 13081 = Tx30a) and Peru (NCTC 13090). Two other isolates were the

genome-sequenced strain, *H. pylori* J99 (USA) (Alm *et al.*, 1999) provided by Professor Diane Taylor (University of Alberta, Canada), and strain SS1 (Australia) (Lee *et al.*, 1997) provided by Dr S. Rijpkema (National Institute for Biological Standards, Potters Bar, UK). Also obtained from the NCTC were the type strains of *Helicobacter nemestrinae* NCTC 12491^T, originating from the gastric mucosa of a captive primate but recently reported to be a strain of *H. pylori* (Suerbaum *et al.*, 2002), and *Helicobacter mustelae* NCTC 12198^T.

Culture of isolates and template DNA preparation. All strains of *H. pylori* and the type strains of *H. nemestrinae* and *H. mustelae* were cultured for 2–3 days on Columbia agar base (Oxoid) containing 10 % (v/v) defibrinated horse blood, and incubated at 37 °C under micro-aerophilic conditions (4 % O₂, 5 % H₂, 5 % CO₂ and 86 % N₂, by vol.) in a variable-atmosphere incubator (Don Whitley Scientific). Stock cultures were preserved (after minimum passages) on glass beads in Nutrient Broth (Oxoid) containing 10 % (v/v) glycerol over liquid N₂ or at –80 °C. Genomic DNA was extracted from sweep cultures of each isolate using the cetyltrimethylammonium bromide (CTAB) method (Wilson, 1987).

Selection of loci for sequence typing. The *ureI* and *atpA* loci, encoding enzymes for intermediary metabolism, were chosen according to the criteria defined by Achtman *et al.* (1999). The *ureI* locus (HP0071) is part of the urease gene cluster (Mobley, 2001), and its product is a urea accessory (transporter) protein that is thought to be an integral cytoplasmic membrane protein, forming a proton-gated urea channel regulating cytoplasmic urease (Weeks *et al.*, 2000). The *atpA* locus (HP1134) encodes a homologue of the ATP synthase F1 α subunit, which has a key function in the synthesis of ATP. The *ahpC* gene was included here as it is a key housekeeping gene, although it was not used in previous MLST studies (Achtman *et al.*, 1999; Maggi Solcà *et al.*, 2001). The AhpC subunit, partially involved in alkyl hydroperoxide reductase (Ahp) activity, is encoded by *ahpC* (HP1563/JHP1471), a cysteine-based peroxidase homologue, which may be involved in oxidative stress by reduction of hydrogen peroxide and a wide range of organic peroxides to the corresponding alcohols (Kelly *et al.*, 2001). AhpC was identified as a major 26 kDa antigen in *H. pylori* and was present in other species of *Helicobacter* (Lundstrom *et al.*, 2001).

PCR and sequencing. Diluted DNA (100 ng) was used as the target to amplify an internal fragment for each locus. For *ureI*, the following primers for amplification and sequencing were designed from the sequence (HP0071) of NCTC 12455 (Tomb *et al.*, 1997): forward primer ureIS4, 5'-GGAAGGAAAAGGCAATGC-3' and reverse primer ure-IAS2, 5'-CTAAACGCTCTATGATCA-3'. These were used as alternatives to the previously described primers (Achtman *et al.*, 1999) which in our hands did not amplify *ureI* from every strain. For *atpA* (HP1134) and *ahpC* (HP1563), previously described primers (Achtman *et al.*, 1999; Lundstrom *et al.*, 2001) were used for amplification and sequencing of fragments of 627 and 541 bp, respectively. PCR amplifications were performed in a total volume of 50 μ l containing 100 ng template DNA, 1.5 mM MgCl₂, 0.05 mM each dNTP, 0.4 mM each oligonucleotide primer, 0.2 μ l (1 U) *Taq* polymerase (Life Technologies) and 5 μ l 10 $\times$ buffer provided by the manufacturer. Automated sequencing from both strands of PCR products was then performed commercially (MWG Biotech) and on-site by standard protocols using a CEQ 2000XL instrument (Beckman Coulter). All new sequences were deposited in the EMBL database (see Table 1).

Analysis of sequence and allele and sequence type (ST) assignment. The sequences used in the phylogenetic analyses were as follows. For *ureI* and *atpA*, these each comprised the five *H. pylori* reference strain sequences obtained from the GenBank database (Table 1) and 68 new *H. pylori* sequences. For *ahpC*, these comprised two *H. pylori* reference strain sequences from GenBank (Table 1) and 71 new

Table 1. *Helicobacter* nucleotide sequences from GenBank and new sequences determined in the study

Reference strain	<i>ureI</i>	<i>atpA</i>	<i>ahpC</i>
<i>H. pylori</i> :			
NCTC 11637 ^T	AJ239707	AJ239602	See below‡
NCTC 11638	AJ239708	AJ239601	See below‡
NCTC 12455 (26695)	AE000528	AE000619	AE000654
SS1	AJ239706	AJ239603	See below‡
J99	AE001445	AE001533	AE001569
68 field strains	See below*	See below†	See below‡
<i>H. nemestrinae</i> NCTC 12491 ^T	AJ421789	AJ421784	–
<i>H. mustelae</i> NCTC 12198 ^T	–	AJ132983	–

**ureI* sequences deposited in GenBank/EMBL: AJ550812–AJ550822.

†*atpA* sequences deposited in GenBank/EMBL: AJ550823–AJ550836.

‡*ahpC* sequences deposited in GenBank/EMBL: AJ550837–AJ550850.

sequences. The *atpA* and *ureI* sequences for *H. nemestrinae* NCTC 12491^T and the *atpA* sequence for *H. mustelae* NCTC 12198^T were also obtained from GenBank. The *ahpC* sequences for *H. mustelae* and *H. nemestrinae* were determined. All sequences were stored and aligned using BioEdit (Hall, 1999). Sequences were converted in SeqVerter (GeneStudio) for use in TREECON version 1.3 (Van de Peer & De Wachter, 1994), implementing CLUSTAL W multiple alignment to construct and draw phylogenetic trees. Nucleotide sequences were analysed by pairwise comparison and clustering using the neighbour-joining method of Saitou & Nei (1987) with the Jukes–Cantor correction. Bootstrap analysis was performed with 100 resampled datasets from evolutionary distance, based on alignments of nucleotide sequences (Van de Peer & De Wachter, 1994). Each tree was rooted by using the *H. mustelae* sequence as the outgroup.

***cagA* and *vacA* genotyping.** Details of the primers and PCR conditions for the two genotyping assays based on the *H. pylori* pathogenicity-associated factors *cagA* and *vacA* have been described elsewhere. Briefly, PCR conditions for the two assays for *cagA* (a marker for the right-hand end of the *cag* pathogenicity island and for the *cagI* region) were as described for the F1/B1 primers and the D008/R008 primers (Slater *et al.*, 1999; Owen *et al.*, 2001a). Genotyping based on *vacA* signal (s1 and s2) and mid-region (m1 and m2) alleles was also performed as described previously (Atherton *et al.*, 1999).

RESULTS

Overall nucleotide sequence variation

Specific internal fragments of the *ureI* and *atpA* loci from 68 field strains of *H. pylori* and of the *ureI* locus of *H. mustelae* NCTC 12198^T were amplified, sequenced and aligned with *ureI* and *atpA* sequences from GenBank for five reference strains of *H. pylori* and *H. nemestrinae* NCTC 12491^T, and the *atpA* sequence of *H. mustelae*. Nucleotide sequences of *ahpC*

for 68 field strains and three reference strains (NCTC 11637, NCTC 11638 and SS1) of *H. pylori*, *H. nemestrinae* NCTC 12491^T and *H. mustelae* NCTC 12198^T were also determined, and were aligned with the *H. pylori* NCTC 12455 and J99 *ahpC* sequences from GenBank. Separate analyses were performed on each set of nucleotide sequences; the alignments showed that the three loci each exhibited a high level of polymorphism, with most variations involving single nucleotides. None of the sequences contained gaps or insertions. Phylogenetic tree configurations showed no major clusters of isolates, and none of the trees were completely congruent (Figs 1, 2 and 3). *H. nemestrinae* clustered within the range of *H. pylori* diversity, whereas *H. mustelae* clustered separately in each of the analyses.

Assignment of nucleotide STs

From analysis of the sequences for the 73 isolates of *H. pylori* and the reference strains of *H. nemestrinae* and *H. mustelae*, the *ureI*, *atpA* and *ahpC* regions were found to be highly polymorphic (giving a total of 62, 63 and 65 alleles, respectively) for *H. pylori* including *H. nemestrinae*, with the sequences differing at one or more nucleotide positions. Each gene-sequence variant within the three sets was designated an allele number (see Figs 1, 2 and 3), and these were combined for the 73 strains of *H. pylori* and for the strains of *H. nemestrinae* and *H. mustelae* to give a total of 66 nucleotide STs (numbered ST1–ST66). All isolates had different STs apart from those in paired (antrum/corpus) sets from five patients (Table 2), and in two family sets where there were common STs within families (Table 3).

Assignment of derived amino acid STs

The *ureI*, *atpA* and *ahpC* nucleotide sequences for *H. pylori*, *H. nemestrinae* and *H. mustelae* were translated into derived amino acid (gene product) sequences, and phylogenetic analyses were performed (data not shown). Clustering pattern differences predominantly reflected nucleotide substitutions at non-synonymous sites, and 26, 15 and 23 amino acid STs were identified for *UreI*, *AtpA* and *AhpC*, respectively. Each amino acid ST was assigned a number and the type frequencies are shown in Fig. 4. The designated amino acid profiles, representing the combined types for each of the six reference strains and the five identical paired isolate sets, are listed in Table 2. Although the combined profiles of strains from different individuals were unique, certain amino acid types were highly conserved within the dataset, notably *UreI* type 4 ($n = 12$), *AtpA* types 1 ($n = 42$) and 3 ($n = 14$), and *AhpC* type 9 ($n = 12$). *AtpA* profile 1 was the most highly conserved, despite the geographic diversity of those isolates, indicating selective constraints on amino acid replacements in that protein due to possible functional or structural requirements.

Analysis of family isolates

Isolates of *H. pylori* from five family clusters were examined (Table 3). Isolates with the same STs were present in two

Fig. 1. Phylogenetic tree constructed from analysis of *ureI* nucleotide sequences of *H. pylori* reference strains and clinical isolates. The isolate laboratory number is indicated, followed by the country of origin and the allelic type (where UI prefix indicates *ureI* type). In the Belfast sets, families are indicated in parentheses by M, H, K and B; F = father, M = mother, C = child. In the Coventry set, S = son (index case), GF = grandfather, U = uncle and F = father. GM* and U* indicate grandmother and uncle in the same family who were directly related to each other but not to the other family members studied. Site of isolation in paired strains is indicated by antrum and corpus. Bootstrap values (expressed as percentages of 100 replications) over 50 % are shown at the key branching points. The tree was rooted using *H. mustelae* NCTC 12198^T as the outgroup.

Fig. 2. Phylogenetic tree constructed from analysis of *atpA* sequences of *H. pylori* reference and field strains. See legend to Fig. 1 for other details. *atpA* strain types are prefixed by AA.

Fig. 3. Phylogenetic tree constructed from analysis of *ahpC* sequences of reference and field strains of *H. pylori*. See legend to Fig. 1 for other details. Strain types are prefixed by AC.

Table 2. Allelic and STs, amino acid profiles and pathogenicity-associated genotypes of *H. pylori* reference strains and paired (antrum/corpus) isolates

Strain	Allelic profile			ST	Amino acid profile			Genotype*
	<i>ureI</i>	<i>atpA</i>	<i>ahpC</i>		UreI	AtpA	AhpC	
Reference:								
NCTC 12455	1	1	1	1	1	1	1	<i>cagA</i> ⁺ /s1m1
J99	2	2	2	37	2	2	2	<i>cagA</i> ⁺ /s1m1
NCTC 11637 ^T	3	3	3	28	3	1	1	<i>cagA</i> ⁺ /s1m1
NCTC 11638	4	4	4	41	4	3	3	<i>cagA</i> ⁺ /s1m1
SS1	5	5	5	11	4	1	4	<i>cagA</i> ⁺ /s2m2
NCTC 13081	7	7	7	34	6	1	6	<i>cagA</i> ⁻ /s2m2
Paired isolates:								
H2966/H2967	28	26	28	5	1	1	11	<i>cagA</i> ⁺ /s1m1
H2993/H2994	29	27	29	57	8	8	15	<i>cagA</i> ⁻ /s2m2
H2320/H2321	30	28	30	24	5	1	9	<i>cagA</i> ⁺ /s1m2
H2682/H2683	31	29	31	3	1	1	15	<i>cagA</i> ⁺ /s1m2
H3030/H3031	32	30	32	43	4	3	6	<i>cagA</i> ⁺ /s2m2

*Pathogenicity-associated genotypes defined on presence of *cagA* and *vacA* allelic type (Dunn *et al.*, 1997).

Table 3. Analysis of *H. pylori* strain types infecting family members

Family and member	Strain no.	Genotype	ST	Allelic profile (<i>ureI</i> – <i>atpA</i> – <i>ahpC</i>)	Amino acid profile (UreI–AtpA–AhpC)
Belfast-K:					
Subject 1 – Child: boy (11 years)	H3019	<i>cagA</i> ⁺ /s1m1	64	36–36–38	1–10–9
Subject 2 – Father (40 years)	H3020	<i>cagA</i> ⁺ /s1m1	2	36–37–38	1–1–9
Subject 3 – Mother (35 years)	H3021	<i>cagA</i> ⁺ /s1m2	2	36–37–38	1–1–9
Belfast-M:					
Subject 1 – Child: boy (13 years)	H3027	<i>cagA</i> ⁺ /s1m1	33	42–43–45	18–1–2
Subject 2 – Father (43 years)	H3026	<i>cagA</i> ⁺ /s1m2	9	41–42–44	4–1–9
Subject 3 – Mother (43 years)	H356	<i>cagA</i> ⁺ /s1m2	9	41–42–44	4–1–9
Belfast-H:					
Subject 1 – Child: boy (10 years)	H3022	<i>cagA</i> ⁺ /s1m1	4	37–38–40	1–1–12
Subject 2 – Father (34 years)	H3023	<i>cagA</i> ⁺ /s1m1	5	37–38–39	1–1–12
Belfast-B:					
Subject 1 – Child: girl (10 years)	H357	<i>cagA</i> ⁻ /s1m2	49	39–40–42	17–3–17
Subject 2 – Mother (37 years)	H3024	<i>cagA</i> ⁺ /s1m2	13	38–39–41	13–1–18
Coventry:					
Subject 1 – Child: boy (18 years)	H3018	<i>cagA</i> ⁺ /s1m2	20	34–32–37	16–1–6
Subject 2 – Grandfather (GF) (71 years)	H3014	<i>cagA</i> ⁺ /s1m1	19	34–32–34	16–1–1
Subject 3 – Father (46 years)	H3016	<i>cagA</i> ⁺ /s1m2	21	34–34–36	16–1–12
Subject 4 – Uncle (50 years)	H3017	<i>cagA</i> ⁺ /s1m1	63	34–35–34	16–9–1
Subject 5 – Grandmother (GM*) (51 years)	H3013	<i>cagA</i> ⁺ /s1m2	10	33–31–33	4–1–17
Subject 6 – Uncle* (32 years)	H3015	<i>cagA</i> ⁺ /s1m1	25	35–33–35	5–1–11

*Individuals that belong to a separate branch of the family.

Fig. 4. Frequency distributions of inferred amino acid STs (gene products) derived from three *H. pylori* housekeeping gene loci: numbers of (a) UreI types, including single strain types ($n = 26$); (b) AtpA types, including single strain types ($n = 15$); (c) AhpC types, including single strain types ($n = 23$). Other type numbers are for strains not included in the present analyses.

families from Belfast (K and M) and in both of these, the strain from the child had a different ST from those of the parents. Nevertheless, members of family K all shared the *ureI* allele 36 and at the amino acid level also shared AhpC amino acid type 9. Likewise, in family M, all members shared AtpA amino acid type 1. Strains H3022 and H3023 from family H had identical amino acid profiles, although the mother in that family was not infected with *H. pylori*. Strains H357 (child) and H3024 (mother) from family B had different STs, with no shared alleles or shared amino acid types. For the Coventry family group, all members were infected by strains with unique STs. However, some alleles were conserved, as *ureI* allele 34 was present in isolates from four members of three generations. Furthermore, AtpA amino acid type 1 was present in five members and AhpC amino acid type 1 was present in two members (grandfather and uncle).

Associations with *cagA* status and *vacA* genotype

The *cagA* status and *vacA* genotype were determined for all 73 strains of *H. pylori* in the study. The genotypes of the

reference strains and multiple isolates are listed in Table 2 and those of the family isolates are listed in Table 3. In addition, *H. nemestrinae* NCTC 12491^T had the typical *cagA*⁺/*vacA* s1m1 genotype. In the paired isolates of *H. pylori*, members of each set had identical *cagA*/*vacA* types, as did isolates in the families with matching STs, except for family Belfast-K where the isolates from the mother (H3021) and father (H3020) differed in *vacA* m-type.

DISCUSSION

Our analyses of the core DNA sequences of three housekeeping loci (*atpA*, *ureI* and *ahpC*) indicate extensive diversity amongst the 73 strains of *H. pylori* from individuals in the UK and ten other countries and the type strain of *H. nemestrinae*, recently proposed as a synonym of *H. pylori* (Suerbaum *et al.*, 2002). The findings are fully consistent with the level of allelic diversity reported previously for the *ureI* and *atpA* loci in other geographically diverse strains of *H. pylori* (Achtman *et al.*, 1999; Falush *et al.*, 2001). We also examined sequence variation in *ahpC* as it encodes alkyl hydroperoxide reductase, which may be important in defence against oxidative stress in *H. pylori* and whose overexpression has been linked with the development of resistance to metronidazole (McAtee *et al.*, 2001). The *ahpC* locus, which exhibited similar levels of variation to those observed in the other two loci, has been identified in several *Helicobacter* species of human and animal origin (Lundstrom *et al.*, 2001). Homologues of *ureI* have also been identified in several species (Owen *et al.*, 2001c), so a combination of *ahpC* and *ureI* sequence data might provide the basis for a genus-wide sequence typing scheme. The interspecies conservation of *atpA* is not yet known, but in the present study a homologue was identified in *H. mustelae* NCTC 12198^T.

Members of the five families examined were colonized by more than one unrelated strain of *H. pylori*. The fact that strains with identical STs were present in two members of each of two of the five families gives support to the accepted model that intrafamilial transmission provides the main focus for *H. pylori* spread. However, at present there is no general agreement on the precise route: one possibility is spread between spouses, for which there is evidence from several serology-based studies (Brenner *et al.*, 1999; Singh *et al.*, 1999; Stone *et al.*, 2000). However, genotyping provides conflicting results, with detection of the same strain in both partners apparently being uncommon in spouses examined in Singapore (Kuo *et al.*, 1999) and Japan (Suzuki *et al.*, 1999), whereas others reported the presence of identical strains in 8 of 18 couples in Greece (Georgopoulos *et al.*, 1996), and in one of four couples in Germany (Suerbaum *et al.*, 1998). The fact that we found strains with identical STs in husband and wife in two Belfast families (K and M) confirms that interspouse transmission is also a possibility in UK households, although the family K isolates had different *vacA* m-types. Pathogenicity-associated genes such as *vacA*, where change in m-type could be attributed to the loss or gain of a

75 bp insert in the mid-region, may therefore provide less stable markers than housekeeping genes.

The second possibility for intrafamilial spread is by parent-to-child transmission. Several serology-based studies have shown that infected parents, in particular mothers, have a key role (Brenner *et al.*, 1998) and that there is a twofold increase in risk when both parents are infected (Dominici *et al.*, 1999). Transmission from mother to child has been demonstrated by strain genotyping in Japan (Malaty *et al.*, 2000; Taneike *et al.*, 2001) and in Germany (Han *et al.*, 2000). However, it was reported that even when both parents in one family in Germany had an identical strain, the child isolate was different (Suerbaum *et al.*, 1998). In the present study, one subject (a 14-year-old boy) was infected with a strain that had the same amino acid profile as the strain infecting the father (Belfast family H). However, the mother in that family was not infected with *H. pylori*, so either father-to-son transmission or the converse, son-to-father transmission, might have occurred. The latter would seem more likely, as it would be probable that the father had been colonized since early childhood (< 10 years of age). A Japanese family study (Taneike *et al.*, 2001) showed that a son was probably infected originally from the father, but became reinfected after eradication with a strain from the mother. In a German family study (Suerbaum *et al.*, 1998), DNA sequence data showed that in three of four families, mother and child had related strains, whereas only one family provided direct evidence of father-child strain relatedness. The extended Coventry family provided additional opportunities of possible parent-to-child transmission. However, the grandmother (subject 5) and her son (subject 6), both members of a separate branch of the family, were infected by strains that lacked matching alleles, so transmission was unlikely to have taken place. By contrast, the grandfather (subject 2) shared *ureI* allele 34 with two of his sons (subjects 3 and 4) and also *ahpC* allele 34 with one son, indicating transmission of a related strain. There was also evidence of transmission between father (subject 3) and son (subject 1) in generations 2 and 3, as these isolates had one shared allele. Interestingly, there was a close association between the grandfather and grandson (subject 1), indicating conservation of two alleles over three generations and a possible direct infection, as the father had a different *atpA* allele. These data indicate that the alleles could have remained stable for at least 70 years in the family. Comparisons of the *cagA* and *vacA* s1 genotypes for the family isolates indicated that they were generally conserved, but the m-type could vary from m1 to m2 in passing between some subjects, such as between Coventry subjects 2 and 3, and as noted above in Belfast family K.

Other possible *H. pylori* transmission routes for which there are limited data include sibling-to-sibling infections (Goodman & Correa, 2000; Rothenbacher *et al.*, 2000). Isolates from the two brothers in generation 2 of the Coventry family (subjects 3 and 4) were found to share *ureI* allele 34, which suggests that the isolates may have been derived from a common ancestral strain. DNA fingerprinting of isolates from two parents and six daughters in the

Netherlands showed that all were closely related, but not identical (van der Ende *et al.*, 1996), whereas identical strains were found in only two of six Lithuanian families (Chalkaukas *et al.*, 1998). It is possible that some family infections (particularly in developing countries) arise from common environmental sources, but there are no genotyping data available to confirm the involvement of water, food or domestic pets (Brown, 2000). However, as the evidence to date is limited to a relatively small number of families worldwide, it remains difficult to reach firm conclusions about any one predominant intrafamilial route of transmission for *H. pylori*.

Assessment of interfamilial transmission is further complicated by the possibility of mixed *H. pylori* infections in individual subjects. The present study of family subjects was based on cultures originating from antral biopsies, for which there was no evidence of simultaneous infections with more than one strain or strain subtype. The presence of different strains at other gastric sites cannot be excluded, but investigations of paired antrum and corpus biopsies in unrelated subjects in this study, and in previous investigations of colonization patterns before treatment, showed that infections of UK individuals were typically by identical or closely related subtypes (Owen, 1993; Owen *et al.*, 1999). Studies on individuals in other countries suggest a spectrum of diversity in different individuals; for instance, analysis of the parent strain J99 and its derivatives after 6 years indicated continuous microevolution, attributed to loss and acquisition of exogenous DNA over time (Israel *et al.*, 2001).

The introduction of DNA sequence typing to investigate clonal descent and microdiversity of *H. pylori* within individuals and in families (Suerbaum *et al.*, 1998; Falush *et al.*, 2001; Israel *et al.*, 2001) is a significant development because it provides an unambiguous approach, avoiding the problems of interpreting conventional genotyping data based on DNA fragment fingerprinting by ribotyping, RAPD, PCR-RFLP and PFGE. To date, direct sequencing has used both core fragments of pathogenicity-associated genes (Suerbaum *et al.*, 1998) and of housekeeping genes (Achtman *et al.*, 1999; Falush *et al.*, 2001; Maggi Solcà *et al.*, 2001). We have extended this approach to designate STs, albeit in a restricted form based on three housekeeping genes, to define clonal descent or evidence of residual clonality within families, as interpretation was equivocal using ribotyping and restriction digest analysis (Nwokolo *et al.*, 1992; Bamford *et al.*, 1993). Thus, in the Coventry family, the *ureI* allele 34 was identified as unique and conserved in isolates from three generations, and the *atpA* allele 32 was conserved in the child and the grandfather. These findings suggest that some alleles may have been present for decades in strains causing infections within some families. Strains within families may have different STs due to differences in single alleles; even so, it was evident from the gene trees that such strains were more similar to other strains of the same family than to strains from unrelated individuals – a finding supported by analyses including additional *ureI* and *atpA* sequence data on some 46 strains from GenBank (R. J. Owen,

unpublished results). We also found that analysis of amino acid sequences provides evidence of common descent in some loci, after removal of differences due to synonymous mutations. However, as certain amino acid sequences (such as AtpA type 1) were highly conserved amongst strains from unrelated individuals, their use as markers is less precise for defining unique family lineages. Overall, our study has shown that sequence typing provides unambiguous information about common descent, based on unique shared alleles of housekeeping genes that are not present in other members of the population. These alleles can be used to identify genetic drift, as *H. pylori* strains are transmitted between individuals in the same or different generations of a family.

In conclusion, related strains of *H. pylori* can be transmitted amongst different family members. Over time, the species genotype may change as populations continually adapt to different human hosts by the formation of successful recombinants. Some alleles of housekeeping loci are sufficiently stable to provide family-specific markers, so sequence typing provides a precise means of identifying and monitoring their distribution and evolution. Sequence databases using a combination of data from housekeeping and pathogenicity-associated loci will facilitate development of a global typing strategy for future studies on *H. pylori* transmission.

ACKNOWLEDGEMENTS

This work was supported in part by the PHLS Central R&D Small Scientific Initiative Fund 2000.

REFERENCES

- Achtman, M., Azuma, T., Berg, D. E. & 7 other authors (1999). Recombination and clonal groupings within *Helicobacter pylori* from different geographical regions. *Mol Microbiol* **32**, 459–470.
- Alm, R. A., Ling, L. S., Moir, D. T. & 20 other authors (1999). Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori*. *Nature* **397**, 176–180.
- Atherton, J. C., Cover, T. L., Twells, R. J., Morales, M. R., Hawkey, C. J. & Blaser, M. J. (1999). Simple and accurate PCR-based system for typing vacuolating cytotoxin alleles of *Helicobacter pylori*. *J Clin Microbiol* **37**, 2979–2982.
- Bamford, K. B., Bickley, J., Collins, J. S., Johnston, B. T., Potts, S., Boston, V., Owen, R. J. & Sloan, J. M. (1993). *Helicobacter pylori*: comparison of DNA fingerprints provides evidence for intrafamilial infection. *Gut* **34**, 1348–1350.
- Brenner, H., Rothenbacher, D., Bode, G. & 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.
- Brenner, H., Rothenbacher, D., Bode, G., Dieudonne, P. & Adler, G. (1999). Active infection with *Helicobacter pylori* in healthy couples. *Epidemiol Infect* **122**, 91–95.
- Brown, L. M. (2000). *Helicobacter pylori*: epidemiology and routes of transmission. *Epidemiol Rev* **22**, 283–297.
- Calam, J. & Baron, J. H. (2001). ABC of the upper gastrointestinal tract: pathophysiology of duodenal and gastric ulcer and gastric cancer. *BMJ* **323**, 980–982.
- Chalkaukas, H., Kersulyte, D., Cepulienė, I., Urbonas, V., Ruzeviciene, D., Barauskiene, A., Raudonikiene, A. & Berg, D. E. (1998). Genotypes of *Helicobacter pylori* in Lithuanian families. *Helicobacter* **3**, 296–302.
- Dominici, P., Bellentani, S., Di Biase, A. R. & 9 other authors (1999). Familial clustering of *Helicobacter pylori* infection: population based study. *BMJ* **319**, 537–540.
- Drumm, B., Perez-Perez, G. I., Blaser, M. J. & Sherman, P. M. (1990). Intrafamilial clustering of *Helicobacter pylori* infection. *N Engl J Med* **322**, 359–363.
- Dunn, B. E., Cohen, H. & Blaser, M. J. (1997). *Helicobacter pylori*. *Clin Microbiol Rev* **10**, 720–741.
- Enright, M. C. & Spratt, B. G. (1999). Multilocus sequence typing. *Trends Microbiol* **7**, 482–487.
- Falush, D., Kraft, C., Taylor, N. S., Correa, P., Fox, J. G., Achtman, M. & Suerbaum, S. (2001). Recombination and mutation during long-term gastric colonization by *Helicobacter pylori*: estimates of clock rates, recombination size, and minimal age. *Proc Natl Acad Sci U S A* **98**, 15056–15061.
- Feldman, R. A. (2001). Epidemiologic observations and open questions about disease and infection caused by *Helicobacter pylori*. In *Helicobacter pylori: Molecular and Cellular Biology*, pp. 29–51. Edited by M. Achtman & S. Suerbaum. Wymondham: Horizon.
- Georgopoulos, S. D., Mentis, A. F., Spiliadis, C. A., Tzouveleakis, L. S., Tzelepi, E., Moshopoulos, A. & Skandalis, N. (1996). *Helicobacter pylori* infection in spouses of patients with duodenal ulcers and comparison of ribosomal RNA gene patterns. *Gut* **39**, 634–638.
- Goodman, K. J. & Correa, P. (2000). Transmission of *Helicobacter pylori* among siblings. *Lancet* **355**, 358–362.
- Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser* **41**, 95–98.
- Han, S.-R., Zschausch, H.-C. E., Meyer, H.-G. W., Schneider, T., Loos, M., Bhakdi, S. & Maeurer, M. J. (2000). *Helicobacter pylori*: clonal population structure and restricted transmission within families revealed by molecular typing. *J Clin Microbiol* **38**, 3646–3651.
- Israel, D. A., Salama, N., Krishna, U., Rieger, U. M., Atherton, J. C., Falkow, S. & Peek, R. M., Jr (2001). *Helicobacter pylori* genetic diversity within the gastric niche of a single human host. *Proc Natl Acad Sci U S A* **98**, 14625–14630.
- Kelly, D. J., Hughs, N. J. & Poole, R. K. (2001). Microaerobic physiology: aerobic respiration, anaerobic respiration and carbon dioxide metabolism. In *Helicobacter pylori: Physiology and Genetics*, pp. 113–133. Edited by H. L. T. Mobley, G. L. Mendz & S. L. Hazell. Washington, DC: American Society for Microbiology.
- Kuo, C.-H., Poon, S.-K., Su, Y.-C., Su, R., Chang, C.-S. & Wang, W.-C. (1999). Heterogeneous *Helicobacter pylori* isolates from *H. pylori*-infected couples in Taiwan. *J Infect Dis* **180**, 2064–2068.
- Lee, A., O'Rourke, J., De Ungria, M. C., Robertson, B., Daskalopoulos, G. & Dixon, M. F. (1997). A standardized mouse model of *Helicobacter pylori* infection: introducing the Sydney strain. *Gastroenterology* **112**, 1386–1397.
- Lundstrom, A. M., Sundaeus, V. & Bolin, I. (2001). The 26-kilodalton, AhpC homologue, of *Helicobacter pylori* is also produced by other *Helicobacter* species. *Helicobacter* **6**, 44–54.
- Maggi Solcà, N., Bernasconi M. V., Valsangiacomo, C., Van Doorn, L.-J. & 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 IS605 genotyping. *Microbiology* **147**, 1693–1707.
- Malaty, H. M., Graham, D. Y. & Klein, P. D., Evans, D. G., Adam, E. &

- Evans, D. J. (1991). Transmission of *Helicobacter pylori* infection. Studies in families of healthy individuals. *Scand J Gastroenterol* **26**, 927–932.
- Malaty, H. M., Kumagai, T., Tanaka, E., Ota, H., Kiyosawa, K., Graham, D. Y. & 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.
- Malaty, H. M., El-Kasabany, A., Graham, D. Y., Miller, C. C., Reddy, S. G., Srinivasan, S. R., Yamaoka, Y. & Berenson, G. S. (2002). Age at acquisition of *Helicobacter pylori* infection: a follow-up study from infancy to adulthood. *Lancet* **359**, 931–935.
- McAttee, C. P., Hoffman, P. S. & Berg, D. E. (2001). Identification of differentially regulated proteins in metronidazole resistant *Helicobacter pylori* by proteome techniques. *Proteomics* **1**, 516–521.
- Mendall, M. A., Goggin, P. M., Molineaux, N., Levy, J., Toosy, T., Strachan, D. & Northfield, T. C. (1992). Childhood living conditions and *Helicobacter pylori* seropositivity in adult life. *Lancet* **339**, 896–897.
- Mobley, H. L. T. (2001). Urease. In *Helicobacter pylori: Physiology and Genetics*, pp. 179–191. Edited by H. L. T. Mobley, G. L. Mendz & S. L. Hazell. Washington, DC: American Society for Microbiology.
- Ng, B. L., Ng, H. C., Goh, K. T. & Ho, B. (2001). *Helicobacter pylori* in familial clusters based on antibody profile. *FEMS Immunol Med Microbiol* **30**, 139–142.
- Nwokolo, C. U., Bickley, J., Attard, A. R., Owen, R. J., Costas, M. & Fraser, I. A. (1992). Evidence of clonal variants of *Helicobacter pylori* in three generations of a duodenal ulcer disease family. *Gut* **33**, 1323–1327.
- Oderda, G., Vaira, D., Holton, J., Ainley, C., Altare, F., Boero, M., Smith, A. & Ansaldi, N. (1991). *Helicobacter pylori* in children with peptic ulcer and their families. *Dig Dis Sci* **36**, 572–576.
- Owen, R. J. (1993). Microbiological aspects of *Helicobacter pylori* infection. *Commun Dis Rep CDR Rev* **3**, R51–56.
- Owen, R. J., Slater, E. R., Gibson, J., Lorenz, E. & 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., Peters, T. M., Varea, R., Teare, E. L. & Saverymuttu, S. (2001a). Molecular epidemiology of *Helicobacter pylori* in England: prevalence of *cag* pathogenicity island markers and IS605 presence in relation to patient age and severity of gastric disease. *FEMS Immunol Med Microbiol* **30**, 65–71.
- Owen, R. J., Taylor, D. E., Wang, G. & Van Doorn, L.-J. (2001b). Heterogeneity and subtyping. In *Helicobacter pylori: Physiology and Genetics*, pp. 343–378. Edited by H. L. T. Mobley, G. L. Mendz & S. L. Hazell. Washington, DC: American Society for Microbiology.
- Owen, R. J., Xerry, J. & Chisholm, S. A. (2001c). Sequence diversity within the *Helicobacter pylori ureI* locus and identification of homologues in other species of *Helicobacter* of human and animal origin. *Gut* **49** (Suppl. 11), A9.
- Pounder, R. E. & Ng, D. (1995). The prevalence of *Helicobacter pylori* infection in different countries. *Aliment Pharmacol Ther* **9** (Suppl. 2), 33–39.
- Rothenbacher, D., Bode, G., Berg, G., Knayer, U., Gonser, T., Adler, G. & Brenner, H. (1999). *Helicobacter pylori* among preschool children and their parents: evidence of parent-child transmission. *J Infect Dis* **179**, 398–402.
- Rothenbacher, D., Bode, G. & Brenner, H. (2000). *Helicobacter pylori* among siblings. *Lancet* **355**, 1998.
- Rowland, M. (2000). Transmission of *Helicobacter pylori*: is it all child's play? *Lancet* **355**, 332–333.
- Saitou, N. & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**, 406–425.
- Singh, V., Trikha, B., Vaiphei, K. Nain, C. K., Thennarasu, K. & Singh, K. (1999). *Helicobacter pylori*: evidence for spouse-to-spouse transmission. *J Gastroenterol Hepatol* **14**, 519–522.
- Slater, E., Owen, R. J., Williams, M. & Pounder, R. E. (1999). Conservation of the *cag* pathogenicity island of *Helicobacter pylori*: associations with vacuolating cytotoxin allele and IS605 diversity. *Gastroenterology* **117**, 1308–1315.
- Stone, M. A., Taub, N., Barnett, D. B. & Mayberry, J. F. (2000). Increased risk of infection with *Helicobacter pylori* in spouses of infected subjects: observations in a general population sample from the UK. *Hepatogastroenterology* **47**, 433–436.
- Suerbaum, S., Smith, J. M., Bapumia, K., Morelli, G., Smith, N. H., Kunstmann, E., Dyrek, I. & Achtman, M. (1998). Free recombination within *Helicobacter pylori*. *Proc Natl Acad Sci U S A* **95**, 12619–12624.
- Suerbaum, S., Kraft, C., Dewhirst, F. E. & Fox, J. G. (2002). *Helicobacter nemestrinae* ATCC 49396^T is a strain of *Helicobacter pylori* (Marshall *et al.* 1985) Goodwin *et al.* 1989, and *Helicobacter nemestrinae* Bronsdon *et al.* 1991 is therefore a junior heterotypic synonym of *Helicobacter pylori*. *Int J Syst Evol Microbiol* **52**, 437–439.
- Suzuki, J., Muraoka, H., Kobayashi, I., Fujita, T. & Mine, T. (1999). Rare incidence of interspousal transmission of *Helicobacter pylori* in asymptomatic individuals in Japan. *J Clin Microbiol* **37**, 4174–4176.
- 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.
- Tomb, J. F., White, O., Kerlavage, A. R. & 39 other authors (1997). The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature* **388**, 539–547.
- Van de Peer, Y. & De Wachter, R. (1994). TREECON for Windows: a software package for the construction and drawing of evolutionary trees for the Microsoft Windows environment. *Comput Appl Biosci* **10**, 569–570.
- van der Ende, A., Rauws, E. A. J., Feller, M., Mulder, C. J., Tytgat, G. N. & Dankert, J. (1996). Heterogeneous *Helicobacter pylori* isolates from members of a family with a history of peptic ulcer disease. *Gastroenterology* **111**, 638–647.
- 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.
- Wilson, K. (1987). Preparation of genomic DNA from bacteria. In *Current Protocols in Molecular Biology*, pp. 2.4.1–2.4.5. Edited by F. M. Ausubel, R. Brent, R. E. Kingston, D. D. Moore, J. A. Smith, J. G. Seidman & K. Struhl. New York: Wiley.