Epidemiol. Infect. (1995). **114**. 257–266

Allelic variation in the *Helicobacter pylori* flagellin genes *fla*A and *fla*B: its consequences for strain typing schemes and population structure

K. J. FORBES, Z. FANG AND T. H. PENNINGTON

Medical Microbiology, Aberdeen University, Forresterhill, Aberdeen AB9 2ZD

(*Accepted 21 November 1994*)

SUMMARY

Extensive DNA sequence diversity was noted in *Helicobacter pylori* flagellin genes *fla*A and *fla*B. PCR amplified sequences from 49 isolates were digested with *Alu*I, *Hin*dIII, *Mbo*I or *Msp*I, the resultant patterns were compared between the different isolates and these used to differentiate the isolates from each other. Evidence that the extensive diversity that was found in these genes is the result of reassortment of sequences between strains in the bacterial population is presented, such that a comparatively small number of individual sequence mutations can recombine together in random combinations to form a greater number of distinct alleles. Geographical differences in the predominant patterns in the *fla*A alleles were also observed and could reflect regional differences either in the human host population or in the bacterial population. In view of the genetic complexity of this species, molecular typing schemes designed to identify related strains may falsely associate strains if the methods do not characterize sufficient genetic sites to exclude chance associations of genetic markers in strains which are actually not closely related to each other.

INTRODUCTION

Helicobacter pylori has been strongly implicated as a causative agent of gastric ulcers in humans [1], whilst its presence has also been correlated with gastric cancer [2]. Carriage rates vary world-wide, generally being higher in less developed countries and in areas of poorer sanitation. Acquisition can occur in childhood, with lifelong carriage being normal and spontaneous eradication rare [3]. Against this background of stability is the genetic structure of the population of *H. pylori*. Numerous studies have noted the extensive genetic diversity of this organism, with the identification of indistinguishable strains in different hosts being generally uncommon. This diversity is observable by many different methods: pulsed field gel electrophoretic (PFGE) separation of large genomic DNA fragments [5]. PCR amplification of random chromosomal DNA sequences (RAPD–PCR) [6]. PCR amplification of specific genes such as urease or flagellin [7, 8].

The polar flagellae of *H. pylori* are a major virulence factor. They are required for motility and their presence has been shown to correlate with gastritis in a piglet model [9]. The *fla*A gene codes for flagellin, the main protein component of

the flagella, and this gene has been cloned and sequenced [10]. The function of the FlaB protein is not known but its gene has also been sequenced [11]. 58% amino acid sequence homology has been noted between *fla*A and *fla*B of one strain of *H. pylori*, with much of the variability being located in the central half of the gene, and indeed the conserved flanking regions show extensive similarity to *fla*A of *H. mustelae* and to *Campylobacter coli* [11]. We have used PCR amplification of the *H. pylori* flagellin genes *fla*A and *fla*B to study the genetic diversity of this species. Restriction digestion of the *fla*A PCR products allowed the easy differentiation of the 49 isolates studied here into 37 groups. This degree of discrimination would permit the identification of individual strains in clinical situations such as the differentiation of recrudescence from reinfection in patients who have been treated for the eradication of *H. pylori*. However, a lack of correlation between groupings of strains based on *fla*A was noted when compared to the unlinked *fla*B and horizontal gene transfer is proposed as a contributory mechanism for the extensive genetic diversity of this species. The implications of this for typing *H. pylori* strains using only a single locus test are discussed along with geographical differences in the distribution of strain types.

MATERIALS AND METHODS

H. pylori type strains were obtained from National Collection of Type Cultures: strains NCTC11637, NCTC11638 (both isolated in Australia), NCTC11916 (isolated in Manchester, UK), NCTC12384 (isolated in France) and NCTC12385 (isolated in USA). Other *H. pylori* strains were obtained from biopsies from patients with gastric ulcers: from Aberdeen Royal Hospitals' Trust Microbiology Diagnostic Laboratory, Grampian, UK there were 18 isolates collected in 1991 and 8 collected in 1993, from Dublin, Ireland there were 5 isolates and from London, UK 13 isolates. All isolates were from independent sources with no known epidemiological links. Strains were cultured on blood agar in a microaerophilic atmosphere and stored at −70 °C.

Genomic DNA was prepared by resuspending the bacteria from a 5-day-old plate culture into 1·5 ml TE8 (10 mM Tris-HCl (pH 8·0), 1 mM EDTA), centrifuging to pellet the cells, resuspending them in lysis buffer (500 μl SET (150 mM NaCl, 15 mM EDTA, 60 mM Tris-HCl (pH 8·3)), 60 μl 10% (w/v) SDS, 60 μl proteinase K (20 μg/ml)) and incubating at 50 °C for 2 h with gentle agitation. Proteins were removed by the addition of 500 μl of TE8 saturated phenol, and following vigorous agitation and centrifugation the supernatant was removed, phenol extracted a second time, vigorously mixed with an equal volume of chloroform/isoamyl alcohol (24:1, v/v) and centrifuged. From the aqueous supernatant the DNA was precipitated after the addition of one-tenth volume of 7·5 M ammonium acetate and two volumes of ethanol at −20 °C. The alcohol was removed following centrifugation, the DNA pellet air dried and finally dissolved in 50 μl distilled water.

PCR oligonucleotide primers for *fla*A were *F1* (5′ATGGCTTTTCAGGTCAATAC at 1-20 nt in the coding sequence [7]) and *F2* (5′CCTTAAGATATTTTGTT-GAACG at 1500–1521 nt [7]), and for *fla*B *F3* (5′CAAAACAACAGAGACCTT at 157–174 nt). Primer *F4* was universal for both genes (5′TTCAGCAAAAT-

https://doi.org/10.1017/S0950268800057927 Published online by Cambridge University Press

Table 1. *Restriction enzyme patterns of* fla*A and* fla*B PCR amplification products*

	Patterns in *fla*A				Patterns in *fla*B			
Strains*	*Msp*I	*Hin*dIII	*Mbo*I	*Alu*I	*Msp*I	*Hin*dIII	*Mbo*I	*Alu*I
D32	1	1	8	2	4	1	2	2
D34	1	1	8	2	—	—	—	—
D101	1	2	2	11	—	—	—	—
D11	1	2	2	11	—	—	—	—
D33	2	3	2	4	—	—	—	—
G13	1	1	1	1	1	1	2	1
G34	1	1	1	1	—	—	—	—
G35	1	1	1	1	—	—	—	—
G1–4	1	1	1	2	3	1	5	1
G36	1	1	1	2	—	—	—	—
G11	1	1	1	3	—	—	—	—
G642	1	1	2	2	—	—	—	—
G44	1	1	2	6	—	—	—	—
G46	1	1	2	9	1	1	2	3
G14	1	1	3	1	—	—	—	—
G8	1	1	3	2	2	1	3	1
G979	1	1	4	1	1	1	4	5
G2–6	1	1	4	2	1	1	2	6
G32	1	1	6	1	—	—	—	—
G16	1	1	7	2	—	—	—	—
G7	1	1	8	2	1	1	1	1
G20	1	1	9	5	—	—	—	—
G4	1	2	1	1	—	—	—	—
G987	1	4	8	8	—	—	—	—
G42	2	1	1	1	1	1	2	3
G942	2	1	2	3	—	—	—	—
G41	2	1	3	4	—	—	—	—
G533	2	1	3	4	—	—	—	—
G9	2	1	4	1	1	1	1	1
G6	2	1	5	1	2	1	3	1
G313	2	2	4	7	1	2	4	4
L439	1	1	1	14	—	—	—	—
L354	1	1	10	12	—	—	—	—
L387	1	1	11	12	—	—	—	—
L428	1	1	11	12	—	—	—	—
L376	1	1	3	14	—	—	—	—
L394	1	1	3	14	—	—	—	—
L343	1	1	6	13	—	—	—	—
L441	2	1	2	1	—	—	—	—
L355	2	1	4	4	—	—	—	—
L366	2	1	5	12	—	—	—	—
L416	2	1	6	12	—	—	—	—
L271	2	1	9	12	—	—	—	—
L275	2	2	8	7	—	—	—	—
NCTC12385	1	1	1	2	—	—	—	—
NCTC12384	2	1	2	4	—	—	—	—
NCTC11637	2	1	4	1	—	—	—	—
NCTC11916	2	1	4	1	—	—	—	—
NCTC11638	2	3	3	10	—	—	—	—

RFLP digestion patterns of *fla*A amplification product (primers *F1.F2*) and *fla*B amplification product (primers *F3.F4*).

* Strains were isolated in Dublin (D), Grampian (G), London (L), or were National Collection of Type Cultures type strains.

† Only those strains shown have been tested.

https://doi.org/10.1017/S0950268800057927 Published online by Cambridge University Press

CCACATC at 1411–1428 nt in *fla*A and 1422–1439 nt in *fla*B, this work). Amplification using a Perkin-Elmer DNA thermal cycler was for 30 cycles of denaturation at 94 °C for 1 min, annealing at 58 °C for 1 min, and extension at 72 °C for 2 min, all of which was preceded by an additional denaturation step at 94 °C for 4 min and terminated by a protracted extension step at 72 °C for 8 min.

Restriction enzymes to digest the *fla* PCR products were chosen which would cut them into a few (*Hin*dIII: six base recognition site) or many (*Alu*I, *Mbo*I, *Msp*I: four base recognition sites) fragments. Digests in a 20 μl final volume comprised 4 μl PCR product, 10 U of enzyme and 2 μl of the appropriate 10 × buffer were incubated at 37 °C overnight. Electrophoresis in 2·5 % agarose at 3 V cm^{-1} for 2 h was followed by ethidium bromide staining, UV transillumination and the fingerprint patterns recorded by video camera and output as a photographic image and stored digitally on computer for future reference. For each enzyme the different patterns were given serial numbers.

Statistical comparison of the associations between the different enzyme patterns was by visual examination of the results in Table 1 and for the more frequent patterns comparison of the observed and expected frequencies in a contingency table (Table 2).

RESULTS

Specificity of primers

The *fla* primers are located in the conserved regions at the ends of the gene [11]. The specificity of *F1* primer to *fla*A and of *F3* to *fla*B was confirmed in PCR reactions using the downstream primer *F4* whose sequence is exactly complementary over its 18 nt to both genes. *Alu*I digestion of these amplification products confirmed that a different DNA sequence was amplified with each primer pair, for each of the six different strains tested. The product digestion patterns using *F1.F4* primers were identical to those obtained using *F1.F2* primers, allowing for the slightly longer product of the latter. Thus *F1.F2* primers and *F1.F4* primers selectively amplified *fla*A whereas *F3.F4* primers selectively amplified *fla*B.

Allelic variation in fla*A and* fla*B*

All 49 isolates studied here yielded a PCR product of 1521 nt in length using *fla*A primers *F1.F2*, with no detectable variation in product lengths indicative of either insertions or deletions in the coding sequence. The number of restriction enzyme sites in *fla*A for each of the four enzymes used varied from strain to strain but the 2 *Msp*I sites, 3 *Hin*dIII sites, 5 *Mbo*I sites and 6 *Alu*I sites of the *fla*A DNA sequence of Leying and colleagues [10] were typical. In the strains studied here these enzymes generated 2, 4, 11 and 14 different patterns respectively: thus enzymes with more restriction sites in *fla*A generated more variant patterns (Fig. 1). In conjunction, these four enzymes gave an indication of the sequence diversity of *fla*A in each of the isolates, and allowed the recognition of many different allelic variants. Within these 49 isolates there were 37 differentiable alleles with 8 of these alleles being found in more than 1 isolate (Table 1). However, instances where two or more strains do have identical pattern combinations always simply comprised the most frequent patterns of each individual enzyme. Thus amongst

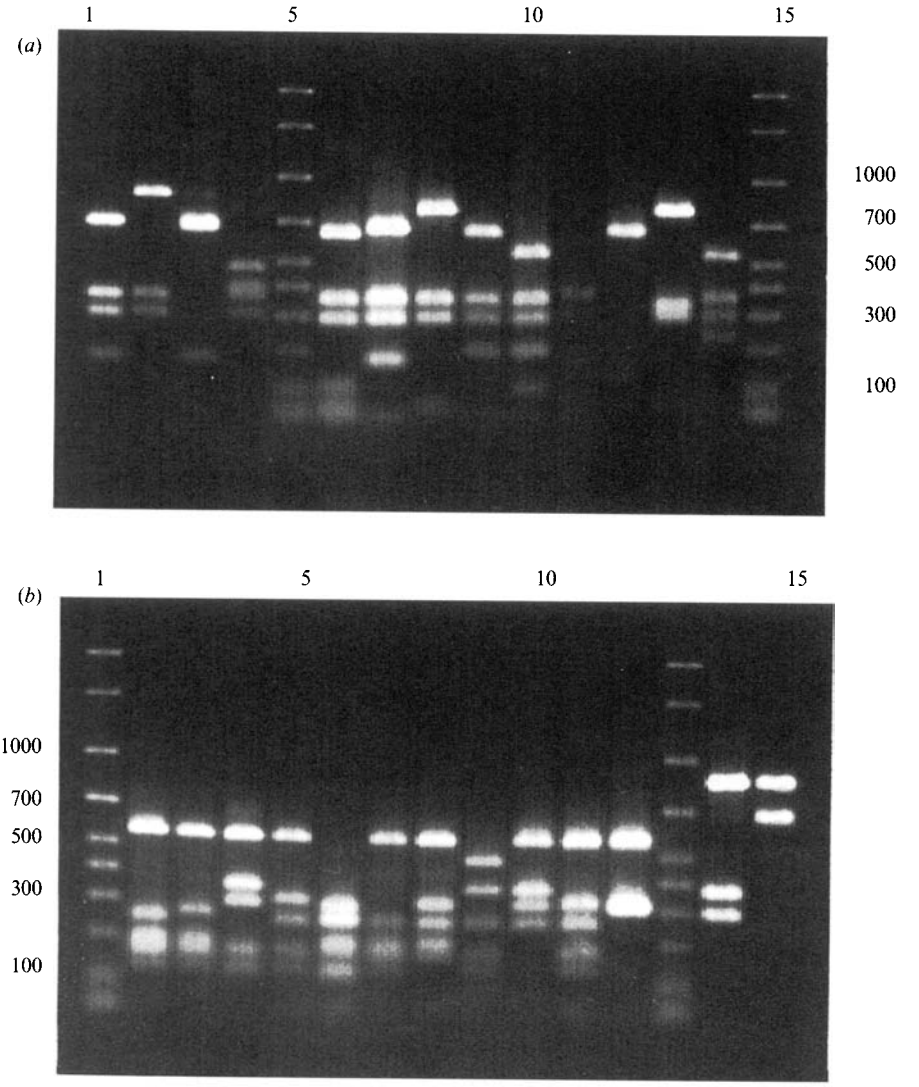

Fig. 1. Representative restriction fragment length polymorphism patterns in *fla*A PCR products. RFLP digestion patterns of *fla*A amplification products using primers *F1.F2* (a) *Hin*dIII patterns 1–4 (lanes 1–4), *Mbo*I patterns 1–9 (lanes 6–14). (*b*) *Alu*I patterns 1–11 (lanes 2–12), *Msp*I patterns 1 and 2 (lanes 14, 15). Size standards are 2000, 1500, 1000, 700, 500, 400, 300, 200, 100, 50 bp in length.

Grampian strains the individual enzyme patterns *Msp*I #1, *Hin*dIII #1, *Mbo*I #1 and *Alu*I #1 and #2 predominate in the sample collection and the two most popular combinations of patterns are #1,#1,#1,#1 with three representative strains and #1,#1,#1,#2 with two representative strains. Assuming a random association between these different patterns then expected frequencies for these two combinations of patterns would be 2·0 and 1·4 respectively. The infrequency of most of the different patterns makes statistical testing difficult, however, comparisons of the frequencies of combinations of the more frequent patterns are possible using the more extensive data for the Grampian strains (Table 2). There

https://doi.org/10.1017/S0950268800057927 Published online by Cambridge University Press

Table 2. *Observed and predicted frequencies of restriction patterns of* fla*A in Grampian strains*

(Frequencies of observed patterns, abstracted from Table 1, compared to predicted frequencies (in parentheses) calculated from the product of the probabilities of each of the individual patterns)

	*Msp*I patterns		
*Alu*I patterns	1	2	Total
1	7	3	10
	(8·2)	(1·8)	—
2	7	0	7
	(5·8)	(1·2)	—
Total	14	3	17

	*Msp*I patterns		
*Mbo*I patterns	1	2	Total
1	7	1	8
	(5·6)	(2·4)	—
2	3	1	4
	(2·8)	(1·2)	—
3	2	2	4
	(2·8)	(1·2)	—
4	2	2	4
	(2·8)	(1·2)	—
Total	14	6	20

	*Alu*I patterns		
*Mbo*I patterns	1	2	Total
1	5	2	7
	(4·7)	(2·3)	—
2	0	1	1
	(0·7)	(0·3)	—
3	1	1	2
	(1·3)	(0·7)	—
4	2	0	2
	(1·3)	(0·7)	—
Total	8	4	12

is good agreement between observed and expected frequencies. The geographical distributions of the different *fla*A digestion patterns are shown in Fig. 2.

The *fla*B gene from 12 randomly selected isolates was also amplified, using primers *F3.F4*, and the products digested with each of the four restriction enzymes (Table 1). Extensive diversity in this gene was also evident with 9 differentiable alleles amongst the 12 patterns scored. Three pairs of indistinguishable alleles were noted (G6, G8; G7, G9; G42, G46); however, the *fla*A alleles for each of these pairs of strains were all different. Conversely, strains G7 and D32 which were indistinguishable by *fla*A had completely different *fla*B alleles. Thus not only is there a lack of association between markers at a single locus, *fla*A, in the strain collection, this lack of association is also seen between two different loci, *fla*A and *fla*B.

https://doi.org/10.1017/S0950268800057927 Published online by Cambridge University Press

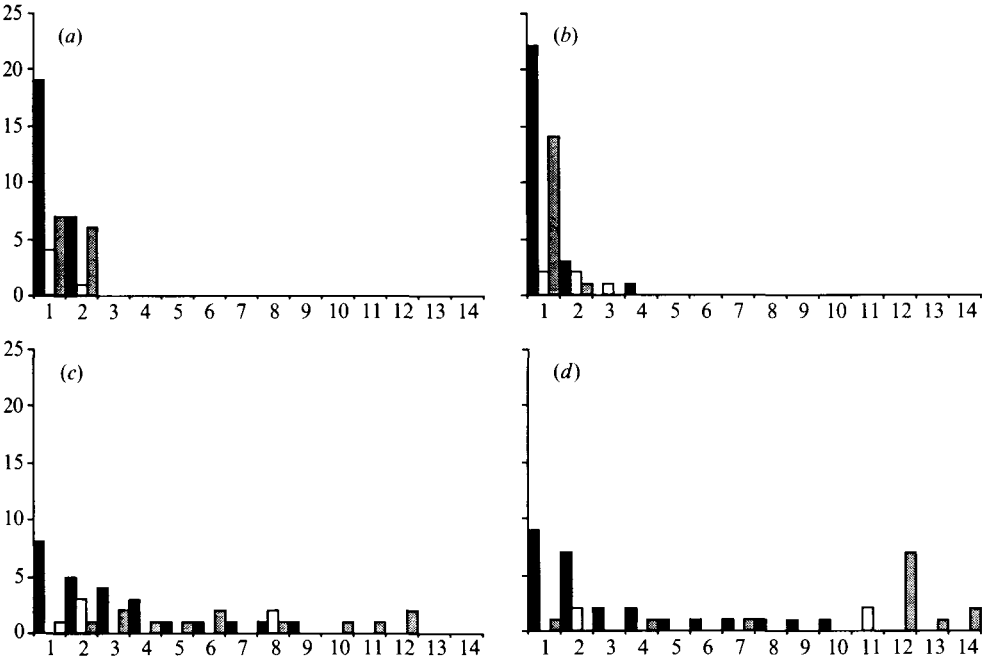

Fig. 2. Frequency distribution of *fla*A digestion patterns by geographical region. (*a*) Distribution of *Msp*I patterns. (*b*) Distribution of *Hin*dIII patterns. (*c*) Distribution of *Mbo*I patterns of strains. (*d*) Distribution of *Alu*I patterns of strains. Frequency (abscissa) of *Msp*I, *Alu*I, *Hin*dIII, and *Alu*I restriction patterns (ordinate) in *fla*A. ■, Grampian; □, Dublin; ▨, London.

DISCUSSION

The *fla*A and *fla*B digestion patterns were reproducible, easily interpreted and conveniently subdivided the enormous allelic variety of the genes into discrete and more manageable information sets. Although restriction maps could have been constructed for the *fla* alleles in each of the stains examined here to determine which particular restriction sites were present in each allele the simpler approach of using the digestion patterns from each of the enzymes themselves was used to indicate the presence of different mutations in the alleles. Each restriction enzyme therefore effectively sampled a different selection of mutations in each allele. Regional differences were noted in the relative frequencies of the different patterns in *fla*A with the most common patterns in one geographical region often being much rarer in another region (Fig. 2).

The extensive diversity observed here readily suggests this method as a tool for the differentiation of *H. pylori* strains and it may be particularly useful in the assessment of the efficacy of antimicrobial chemotherapy in individual patients who later suffer from further colonization by *H. pylori* or where an epidemiological link is suspected [12]. Caution in the interpretation of epidemiological results is however, indicated since many of the *fla*A alleles observed here, which were from a wide variety of locations, showed extensive similarity if only one or two enzyme patterns were used to compare strains. The number of differentiable alleles

https://doi.org/10.1017/S0950268800057927 Published online by Cambridge University Press

amongst the 49 isolates was therefore largely a function of the number of different enzymes used and it seems highly probable that if more enzymes were employed then the eight alleles which were found to be represented in more than one isolate could also be further subdivided. Conversely, the groupings of indistinguishable isolates with one set of enzymes will be different from the groups formed if a different set of enzymes is used (Table 1). The implications of this for a *H. pylori* typing scheme designed to identify 'related' strains would seriously compromise the technique.

Numerous typing methods for *H. pylori* have been reported which range from those using chromosomal restriction enzyme fragments which span the whole genome (PFGE) [4] or a smaller portion of it using smaller genomic fragments [13, 14], to typing methods which select individual genes for PCR amplification and restriction enzyme fingerprinting [7, 15] as here. Many of these studies have noted extensive genetic diversity in this species. Taylor and colleagues [4] have suggested five mechanisms for this extensive genetic diversity: (i) strain variable protective methylation of DNA restriction sites; (ii) amplification and/or deletion of chromosomal sequences, (iii) movement of short repetitive sequences (translocation), (iv) higher frequencies of mutation at silent sites (synonymous substitutions) which would be selectively neutral, and (v) horizontal genetic exchange through natural transformation. The diversity that has been observed here in the different *fla* alleles cannot be subject to methylation-based variability as the DNA is *in vitro* amplified. The PCR products are located in coding sequence and no variation in the length of the PCR product was noted in the strains examined here or by others. Neither the acquisition of new sequences from elsewhere in the genome (e.g. cassettes) nor phase variation have been reported in the *fla* genes. These would be expected to be restricted to particular regions of the gene but since digestion fragment length polymorphisms of all sizes are observed here the restriction site polymorphisms must also be located throughout the gene. Thus the variation observed here is most probably the result of mutational alteration (option iv). Whether these need be exclusively neutral as Taylor and colleagues [4] have suggested will depend on the contribution of each mutation to each strain's overall genetic fitness. This variability could be dramatically extended by the exchange of genetic material between strains and in this case markers will be in linkage equilibrium in the population (option v).

The restriction enzyme patterns of *fla*A with the four enzymes used here give a picture of the genetic sequence of *fla*A in each strain. In general the observed combinations of the enzyme patterns seem to be random combinations of the patterns that are available; there was no association between patterns. This suggested that these mutations can combine together to form new allelic combinations. Clearly there is much sequence diversity in *fla*A and *fla*B and this may reflect host immune selection against particular flagellar phenotypes. The genetic source of the diversity in this gene seems to be predominantly point mutations as no polymorphisms were noted here in the length of the PCR products. These allelic variants could have arisen by two mechanisms either by a progressive accumulation of point mutations in the different strains over time or by horizontal gene transfer of sequences between strains allowing numerous reassortments of a few variant sequences. Mutational events alone would generate

different *fla*A sequences which were unique to each individual strain or clone of strains but the constituent mutations would not be generally dispersed throughout the bacterial population as is observed here. If mutational accumulation were the only source of diversity then the combinations of mutations in the different alleles would be expected to be in linkage disequilibrium. On the other hand, if there were transfer of genomic DNA between strains then a comparatively small number of mutations could be reassorted into many different, novel combinations. In this latter case combinations of different mutations would be in equilibrium in the population as a whole with the more frequent individual mutations most commonly being found together simply as a function of their higher individual frequency in the population. Similarly there will be a lack of association between alleles at different loci. Linkage disequilibrium was not observed here and suggests that transfer of genetic material between strains may be disrupting underlying clonal frameworks.

Most bacteria have a population structure where a few clones of related strains predominate in the population [16, 17]. Significant horizontal gene transfer would disrupt the hierarchical relationships between strains, reduce the importance of dominant clones, and increase the variety of different strain types. Although *H. pylori* is known to be naturally transformable under laboratory conditions [18, 19] whether this occurs *in vivo* is unknown; however, many other transformable bacterial species are also extremely diverse [20, 21]. A non-clonal population structure has also been proposed for *Neisseria gonorrhoeae* where large clusters of apparently identical strains (clones) have been readily subdivided when further typing data was applied, leading to the conclusion that the original groups were not comprised of phylogenetically related organisms but simply of strains which had markers which were common in a randomly assorted pool [22].

ACKNOWLEDGEMENTS

This research was supported by the British Council through a grant to Z. F. This research benefited from the SEQNET facility run by the SERC.

REFERENCES

1. Marshall BJ, Warren JR. Unidentified curved bacilli in the stomach of patients with gastritis and peptic ulceration. Lancet 1984; **i**: 1311–15.
2. Forman D, Newell DG, Fullerton F, *et al*. Association between infection with *Helicobacter pylori* and risk of gastric cancer: evidence from a prospective investigation. BMJ 1991; **302**: 1302–5.
3. Banatvala N, Feldman R. The epidemiology of *Helicobacter pylori*: missing pieces in the jigsaw. Commun Dis Rep; **3**: R56–9.
4. Taylor DE, Eaton M, Chang M, Salam SM. Construction of a *Helicobacter pylori* genome map and demonstration of diversity at the genome level. J Bacteriol 1992; **174**: 6800–6.
5. Majewski SIH, Goodwin CS. Restriction endonuclease analysis of the genome of *Helicobacter pylori* with a rapid extraction method: evidence for considerable genomic variation. J Infect Dis 1988; **157**: 465–71.
6. Akopyanz N, Bukanov NO, Westblom TU, Kresovich S, Berg DE. DNA diversity among clinical isolates of *Helicobacter pylori* detected by PCR-based RAPD fingerprinting. Nucl Acids Res 1992; **20**: 5137–42.

7. Akopyanz N, Bukanov NO, Westblom TU, Berg DE. PCR-based RFLP analysis of DNA sequence diversity in the gastric pathogen *Helicobacter pylori*. Nucl Acids Res 1992; **20**: 6221–5.
8. Desai M, Linton D, Owen RJ, Stanley J. Molecular typing of *Helicobacter pylori* isolates from asymptomatic, ulcer and gastritis patients by urease polymorphism. *Epidemiol Infect* 1994; **112**: 151–60.
9. Eaton KA, Morgan DR, Krakowla SA. *Helicobacter pylori* virulence factors in gnotobiotic piglets. Infect Immun 1989; **57**: 1119–25.
10. Leying H, Suerbaum S, Geis G, Haas R. Cloning and characterisation of a *Helicobacter pylori* flagellin gene. Mol Microbiol 1992; **6**: 2863–74.
11. Suerbaum S, Josenhans C, Labigne, A. Cloning and genetic characterization of the *Helicobacter pylori* and *Helicobacter mustelae fla*B flagellin genes and construction of *H. pylori fla*A- and *fla*B-negative mutants by electroporation-mediated allelic exchange. J Bacteriol; 1993; **175**: 3278–88.
12. Nwokolo CU, Bickley J, Attard AR, Owen RJ, Costas M, Fraser IA. Evidence of clonal variants of *Helicobacter pylori* in three generations of a duodenal ulcer disease family. Gut 1992; **33**: 1323–7.
13. Owen RJ, Fraser J, Costas M, Morgan D, Morgan DR. Signature patterns of DNA restriction fragments of *Helicobacter pylori* before and after treatment. *J Clin Pathol* 1990; **43**: 646–9.
14. Lagenberg W, Rauws EAJ, Widjojokusumo A, Tytgat GNJ, Zanen HC. Identification of *Campylobacter pyloridis* isolates by restriction endonuclease DNA analysis. J Clin Microbiol 1986; **24**: 414–7.
15. Moore, RA, Kureishi A, Wong S, Bryan LE. Categorization of clinical isolates of *Helicobacter pylori* on the basis of restriction digest analyses of PCR-amplified *ure*C genes. J Clin Microbiol 1993; **31**: 1334–5.
16. Selander RK, Caugant DA, Whittam TS. Genetic structure and variation in natural populations of *Escherichia coli*. In: Neiderhardt FC, Ingraham JL, Low KB, Magasanik B, Schaechter M, Umbarger HE, eds. *Escherichia coli* and *Salmonella typhimurium* cellular and molecular biology. Washington, DC: Am Soc Microbiol, **XX**: 1625–48.
17. Musser JM, Kroll JS, Granoff DM, *et al*. Global genetic structure and molecular epidemiology of encapsulated *Haemophilus influenzae*. Rev Infect Dis 1990; **12**: 75–111.
18. Nedenskov-Sorensen P, Bukholm G, Bovre K. Natural competence for genetic transformation in *Campylobacter pyloridis*. J Infect Dis 1990; **161**: 365–6.
19. Wang,Y, Roos KP, Taylor DE. Transformation of *Helicobacter pylori* by chromosomal metronidazole resistance and by a plasmid with a selectable chloramphenical resistance marker. J Gen Microbiol 1993; **139**: 2485–93.
20. O'Rourke M, Stevens E. Genetic structure of *Neisseria gonorrhoeae* populations: a non-clonal pathogen. J General Microbiol 1993; **139**: 2603–11.
21. Forbes KJ, Bruce KD, Ball A, Pennington TH. Variation in length and sequence of porin (*omp*P2) alleles of non-capsulate *Haemophilus influenzae*. Mol Microbiol 1992; **6**: 2107–12.
22. O'Rourke M, Spratt BG. Further evidence for the non-clonal population structure of *Neisseria gonorrhoeae*: extensive genetic diversity within isolates of the same electrophoretic type. Microbiol 1994; **140**: 1285–90.