JOURNAL OF CLINICAL MICROBIOLOGY, Sept. 1997, p. 2386–2392 Vol. 35, No. 9
0095-1137/97/$04.00+0

Evidence for Recombination in the Flagellin Locus of *Campylobacter jejuni*: Implications for the Flagellin Gene Typing Scheme

CLARE S. HARRINGTON,[1]* FIONA M. THOMSON-CARTER,[2] AND PHILIP E. CARTER[1]

Department of Medical Microbiology, University of Aberdeen,[1] and Scottish Campylobacter Reference Laboratory,[2] Aberdeen, Scotland, United Kingdom

Received 24 March 1997/Returned for modification 9 May 1997/Accepted 4 June 1997

The flagellin subunit of the flagellar filament in *Campylobacter jejuni* is encoded by two highly homologous tandem genes, *flaA* and *flaB*. The *flaA* gene was sequenced in 18 strains of *C. jejuni*, including isolates from three outbreak groups. Sequences obtained were compared with *flaA* sequences available in the GenBank database, and all were analyzed for mosaic gene structure by using recently described statistical tests for detecting gene conversion among aligned sets of sequences. Strong evidence was found supporting recombination between *flaA* genes of different strains (i.e., intergenomic recombination). Intragenomic recombination between the *flaA* and *flaB* genes of *C. jejuni* TGH9011 was also demonstrated. Both mechanisms of recombination may act as a potential means by which pathogenic strains can generate increased antigenic diversity, so allowing them to escape the immunological responses of the host. Furthermore, demonstration of recombination within and between flagellin loci of natural strains suggests that flagellin gene typing (restriction fragment length polymorphism analysis of PCR-amplified flagellin genes) cannot be considered a stable method for long-term monitoring of pathogenic *Campylobacter* populations.

Campylobacter jejuni, one of the world's most common enteropathogens (19), is a gram-negative organism with a single polar flagellum at one or both ends of its spiral cells. Flagella are unsheathed and consist solely of flagellin, whose gene organization has been determined for *C. jejuni* 81116 (13), *C. jejuni* 1N1 (4), and *C. jejuni* TGH9011 (8) and for one strain (VC167) of the closely related *Campylobacter coli* (7). In each case, two highly homologous flagellin genes (*flaA* and *flaB*) of approximately 1,730 bp have been found to be located adjacent to one another in a tandem (head-to-tail) orientation, joined by an intervening segment of approximately 200 bp.

Expression of *flaA* and *flaB* is differentially regulated, with each gene containing its own distinct promoter (7, 13, 23). Inactivation of the *flaA* gene results in mutants with severely impaired motility both for *C. jejuni* 81116 and for *C. coli* VC167 (6, 24). However, *flaB* inactivation has no effect on motility for *C. jejuni* 81116 (24) and leads to mutants with only slightly reduced motility for *C. coli* VC167 (6). With the *flaB* product playing an apparently minimal role in flagellin structure and cell motility, the reason for flagellin gene duplication in *Campylobacter* spp. is yet to be established. However, it has been suggested that *flaB* may serve as a gene donor (1), of which parts could be introduced (through homologous recombination) into the *flaA* gene, either to compensate for deleterious mutations or possibly to increase the immunogenic repertoire of a given *C. jejuni* strain.

There are reports in the literature of spontaneous intragenomic flagellin recombination both for *C. coli* (1) and for *C. jejuni* (25), both of which are known to be naturally transformable (21). Likewise, intergenomic recombination (horizontal gene transfer) has recently been demonstrated between the flagellin loci of two mutant *C. jejuni* strains, each containing different antibiotic resistance markers in their flagellin genes (25). However, there are currently no reports of recombination within natural populations of *C. jejuni* or *C. coli*. Demonstration of genetic recombination between flagellin genes of natural populations would be of significance not only because it hints at a role for flagellin gene duplication (i.e., increased immunogenic repertoire) but also because it would raise questions about the long-term stability of flagellin gene typing, a system increasingly employed for subtyping both *C. jejuni* and *C. coli* (2, 11, 12, 14).

This study set out to examine *flaA* sequences for mosaic structure by using recently described statistical tests for detecting gene conversion among aligned sets of sequences (10). We report an analysis of the *flaA* gene from 18 *C. jejuni* strains, including isolates from three outbreak groups. Sequences obtained were compared with *flaA* sequences from the GenBank database, and strong evidence is presented for recombination within the flagellin locus.

MATERIALS AND METHODS

Bacterial strains. The properties and sources of the 18 strains of *C. jejuni* used in this work are listed in Table 1. Sixteen strains were isolated from infected human patients, including strains from three outbreak groups. All bacteria were cultivated at 42°C for 48 h on 5% (vol/vol) defibrinated sheep blood agar under microaerobic conditions. Strains were preserved at −70°C in storage medium (pH 7.4) containing 2.5% (wt/vol) nutrient broth no. 2 (Oxoid, Basingstoke, United Kingdom); 0.15% agar no. 1 (Oxoid); 0.05% (each) $FeSO_4$, $K_2O_5S_2$, and CH_3COCO_2Na; and 15% (vol/vol) glycerol.

Serotyping. Serotyping was performed according to the somatic O (formerly heat-stable) antigenic scheme of Penner and Hennessey (15) with a panel of 42 O antisera.

DNA purification. Chromosomal DNA was extracted by resuspending bacteria from a 2-day-old plate into 1 ml of SET buffer (150 mM NaCl, 15 mM EDTA, 10 mM Tris-HCl [pH 8.0]), centrifuging to pellet the cells (3,000 × *g*, 5 min), resuspending pellets in 600 μl of SET buffer containing 0.5% (wt/vol) sodium dodecyl sulfate and 100-μg/ml proteinase K, and incubating at 50°C for 2 h. Proteins were removed from cell lysates by standard phenol-chloroform extrac-

* Corresponding author. Mailing address: Department of Medical Microbiology, University of Aberdeen, Medical School Buildings, Foresterhill, Aberdeen, AB25 2ZD, Scotland, United Kingdom. Phone: (44) 1244-681818, ext. 52864. Fax: (44) 1224-685604. E-mail: c.s.harrington@abdn.ac.uk.

TABLE 1. *Campylobacter* strains used for *flaA* sequence analysis

Code[a]	Strain[b,c]	Outbreak group	Penner serotype(s)	*flaA* sequence length (bp)	*flaA* sequence identity[d]
01	PO1*	1	4c	1,722	1
02	PO2*		4c	1,722	2
	PO3		4c	1,722	(2)
04	PO4*		2	1,713	3
05	PO5*		2	1,713	4
	PO6	1	4c	1,722	1
	PO7	1	4c	1,722	1
	PO8	1	4c	1,722	1
09	PO9*	2	8, 17	1,722	5
	P10	2	8, 17	1,722	5
	P25	1	4c	1,722	1
26	P26*	3	8, 17	1,713	6
	P27	3	8, 17	1,713	6
	P37	2	8, 17	1,722	5
	P39	3	8, 17	1,713	6
	P40	3	8, 17	1,713	6
A	NCTC 11351*		23	1,713	7
B	NCTC 12107*		19	1,722	8
C	*C. jejuni* 81116 (j05635)			1,725	9
D	*C. jejuni* TGH9011 (m74578)			1,722	10
E	*C. jejuni* TGH9011 (z29327)			1,722	11
F	*C. jejuni* 1N1 (x57173)			1,725	12
G	*C. coli* VC167 (m64670)			1,713	13

[a] These 13 strains represent unique *flaA* sequences, some shared by one or more outbreak strains (see final column).

[b] NCTC, National Collection of Type Cultures; P, human isolates obtained from D. R. A. Wareing, Public Health Laboratory, Royal Preston Hospital; remaining strains, sequences obtained from GenBank (accession numbers given in parentheses).

[c] Strains used for design of internal *flaA* primers (Table 2) are marked with an asterisk.

[d] Each number represents a unique sequence. Strain 3 has the same sequence as strain 2 except for one nucleotide position (see text).

tion techniques (16), after which DNA was precipitated from the aqueous phase by addition of a 1/10 volume of 3 M sodium acetate (pH 5.3) and 2 volumes of ethanol at −20°C. DNA was collected by centrifugation (13,000 × *g*, 10 min), washed in 70% ethanol, dried, and finally suspended in TE buffer (10 mM Tris-HCl, 1 mM EDTA [pH 8.0]) at a concentration of ~50 to 100 ng/μl.

Primer design and synthesis. Oligonucleotides used for amplification and sequencing (Table 2) were synthesized on an Applied Biosystems (Warrington, United Kingdom) model 891 DNA synthesizer. Primers were designed against upstream and downstream *flaA* sequences (A_{up} and A_{down}, respectively) and against an upstream *flaB* sequence (B_{up}) by aligning appropriate regions of known *C. jejuni* and *C. coli* flagellin sequences. Primers internal to *flaA* were designed by alignment of *flaA* sequences from eight isolates (indicated in Table 1) as information became available.

PCR and sequencing. Primer pairs used for amplification of *flaA* and/or *flaB* template are indicated in Fig. 1A. PCRs (50 μl) consisted of 50 to 100 ng of genomic DNA; 0.5 U of *Taq* polymerase (Boehringer Mannheim, Poole, United Kingdom); and final concentrations of 1× PCR buffer (Boehringer Mannheim), 200 μM deoxynucleoside triphosphates, and 0.25 μM (each) forward and reverse primers. Thermal cycling conditions were 30 cycles at 94°C for 1 min, 50°C for 1 min, and 72°C for 1.5 min, all preceded by a denaturation step at 94°C for 5 min and followed by an extended elongation step at 72°C for 7 min. PCR amplification products were purified for sequencing by using Centricon C-100 columns (Amicon, Gloucester, United Kingdom). Purified templates were sequenced directly with an Applied Biosystems model 373A automated DNA sequencer and a *Taq* dye-deoxy terminator sequencing kit (Applied Biosystems), following the instructions provided by the manufacturer.

RESULTS

PCR and sequencing. The *flaA* gene (with flanking regions) was amplified from all strains by using two sets of PCR primers (A_{up}/A6 and A1/A_{down}). Certain strains were also sequenced over the 5′ region of their *flaB* gene, for which a third PCR product (B_{up}/A6) was generated. All three primer pairs (A_{up}/A6, A1/A_{down}, and B_{up}/A6) generated products of the expected size (~700, 1,950, and 700 bp, respectively), as shown in Fig. 1B for strain P01. Both strands of the *flaA* gene were sequenced over the entire coding region, while *flaB* sequence was determined largely for one strand only. Sequencing primers are listed in Table 2.

Extent of sequence diversity among *flaA* sequences. Table 1 summarizes the *flaA* sequencing data obtained from the 18 strains sequenced during this study, as well as for the five sequences obtained from GenBank under the accession numbers given. Strains within any one outbreak group were all found to share the same sequence. In total, 13 unique sequences were identified, ranging in length from 1,713 to 1,725 bp (571 to 575 amino acids). Strains P02 and P03 actually differed by one nucleotide position, a synonymous substitution located at base 1248 of the submitted sequences, but are considered as identical in the following analyses. Repeated PCR and sequencing confirmed the presence of this point mutation.

Pairwise comparison of aligned *flaA* sequences (Table 3) showed that sequence divergence (percentage of different nucleotide positions) ranged from 0.3% (between 09 and E) to a

TABLE 2. Primers used for flagellin PCR amplification and sequencing

Primer[a]	Priming site[b]	Orientation	Primer sequence (5′ to 3′)[c]
A_{up}	*flaA*: −66 to −42	Forward	WWKWAACGATATAGYWTWTAACAAG
A_{down}	*flaB*: 56 to 36	Reverse	AACAACTGAATTTGCATGTGC
A1	*flaA*: 1 to 26	Forward	GGATTTCGTATTAACACAAATGGTGC
A2	*flaA*: 1725 to 1702	Reverse	CTGTAGTAATCTTAAAACATTTTG
A3	*flaA*/*B*: 317 to 338	Forward	AAACAAGAACCATGCTTCAAGC
A4	*flaA*/*B*: 338 to 317	Reverse	GCTTGAAGCATGGTTCTTGTTT
A5	*flaA*/*B*: 625 to 644	Forward	GTTGGAACAGGWCTTGGAGC
A6	*flaA*/*B*: 644 to 625	Reverse	GCTCCAAGWCCTGTTCCAAC
A7	*flaA*/*B*: 1456 to 1475	Forward	ATGGCWGTRATGGAYATAGC
A8	*flaA*/*B*: 1475 to 1456	Reverse	GCTATRTCCATYACWGCCAT
A9	*flaA*/*B*: 988 to 1007	Forward	TCTTTAGTTAAAAATGATGG
A10	*flaA*/*B*: 1007 to 988	Reverse	CCATCATTTTTAACTAAAGA
B_{up}	(*flaA*: 1844 to 1863)	Forward	AAACTTGGAACACTTCTTGC

[a] A1 and A2 were described previously (11); A_{up}, A_{down}, and B_{up} were designed from flagellin sequences available in the GenBank database; all remaining primers were designed from a consensus of the eight sequences listed in Table 1.

[b] *C. jejuni* 81116 numbering (GenBank accession no. j05635) referring to *flaA* or *flaB* alone or to both genes (*flaA*/*B*) as indicated; number 1 corresponds to the first coding nucleotide; negative numbers are related to nucleotides upstream of the coding region, with −1 corresponding to the first nucleotide (G of the initiation codon ATG) upstream of the first expressed codon; the position of B_{up} is given in parentheses since it falls in the intergenic spacer downstream of *flaA*.

[c] Base codes are standard International Union of Biochemistry codes for bases and ambiguity.

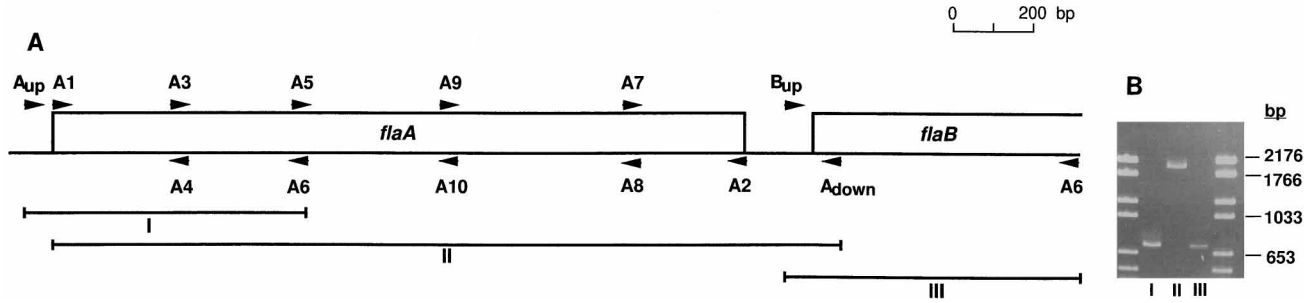

FIG. 1. (A) Schematic representation of the flagellin genes of *C. jejuni* 81116 showing locations of primers used for PCR amplification and sequencing. The positions of the three PCR products (I, II, and III) used for sequencing are also indicated. (B) PCR fragments of strain 01 obtained with primer sets A_{up}/A6 (I), A1/A_{down} (II), and B_{up}/A6 (III). Flanking tracks show molecular size markers with lengths as indicated.

maximum of 23.9% (between G and 26), with most pairs of sequences showing divergences of between 5 and 12%. Figure 2 shows the evolutionary tree constructed from the distance data and confirms that only three clusters of sequences (01/02, 09/D/E, and 04/05/A) were diverged at less than 5%. In one of these groups, sequences D and E diverge at 2.5%, despite representing the same *C. jejuni* strain (TGH9011). Three sequences (01, 02, and C) formed a distinct subgroup, clustering with the only *C. coli* sequence (G) represented in this study.

Identification of common and variable *flaA* domains. Alignment of deduced amino acid sequences from the 13 unique *flaA* sequences identified regions of high conservation corresponding to the N and C termini of the flagellin protein, while the internal region showed far greater variability, including all sites of amino acid insertion-deletion. Only 12 of the first (N-terminal) 176 amino acid residue positions (6.8%) were found to be informative polymorphic sites (those sites where at least two different residues are expressed by at least two strains each), and similarly, of the last (C-terminal) 100 residues, only 4 sites were informatively polymorphic (4%). Conversely, the internal region (302 sites, including alignment gaps) was informatively polymorphic at ~40% of its sites.

Detection of mosaic structure in the *flaA* gene. Alignment of the 13 distinct *flaA* nucleotide sequences identified 649 polymorphic sites from a total of 1,737 sites (including 42 alignment gaps). Examination of aligned sequences (not shown) identified two clear regions of mosaic gene structure, illustrated in Fig. 3. In comparing strains 01 and 02 (Fig. 3A), the first 347 sites were identical, as were all sites from 490 to the end (i.e., 0% divergence), but between these identical regions lay a segment differing by 21 of the 142 sites (14.8% divergence). Similarly, in comparing strains 09, D, and E (Fig. 3B), sequences from strains D and E both differed from that of strain 09 by just 5 of the first 1,548 positions (0.3% divergence), after which strain D differed from the other two sequences by 42 of the remaining 189 sites (22.2%).

The maximum chi-square method (10) was used to assess the significance of the mosaic structures identified in Fig. 3. This method involves comparison of two sequences, and identification of crossover points (cuts) which maximize the differences between proportions of sites occupied by the same and by different bases, both before and after the putative crossover site. Significance is determined by comparison of real data against T trial pairs, with observed mosaic structure being significant at the level $P < 1/T$. The cut sites identified by maximum chi-square were found to be the same as those identified by visual examination of the aligned sequences, and all were significant at $P < 10^{-4}$.

The maximum chi-square test is designed to assess the significance of mosaics between specific pairs of strains, leaving the data set as a whole unsampled. In contrast, Sawyer's method (18) allows an assessment of the overall recombinogenic nature of the sequences, based on pairwise examination of silent polymorphic sites between all strains. An arbitrary P value is assigned by comparison of the observed test output against 10,000 random permutations of the data, with P defined as the proportion of permuted data sets having an output greater than or equal to the original score. By applying the

TABLE 3. Distance matrix for *Campylobacter flaA* genes

Strain[b]	% Distance[a] compared with strain:											
	02	04	05	09	26	A	B	C	D	E	F	G
01	1.24	22.10	22.34	22.28	23.05	22.64	23.35	14.54	23.46	22.52	23.46	14.24
02		22.22	22.46	22.46	23.23	22.93	23.64	14.30	23.64	22.70	23.76	14.13
04			1.89	6.80	5.91	3.84	8.45	21.16	9.28	6.97	9.57	22.52
05				7.27	5.32	4.43	8.04	21.28	9.63	7.45	9.10	23.17
09					7.74	7.86	10.22	20.98	2.78	0.30	10.64	22.28
26						6.74	7.33	22.28	10.11	8.04	8.69	23.88
A							6.21	21.10	10.34	8.04	8.33	22.75
B								21.99	12.71	10.52	5.38	22.87
C									22.64	21.22	22.52	17.44
D										2.48	13.06	23.82
E											10.93	22.52
F												23.52

[a] Percentages of simple nucleotide distances (uncorrected for multiple base changes), with bases opposite alignment gaps omitted from analysis.
[b] Strains are identified by codes given in Table 1.

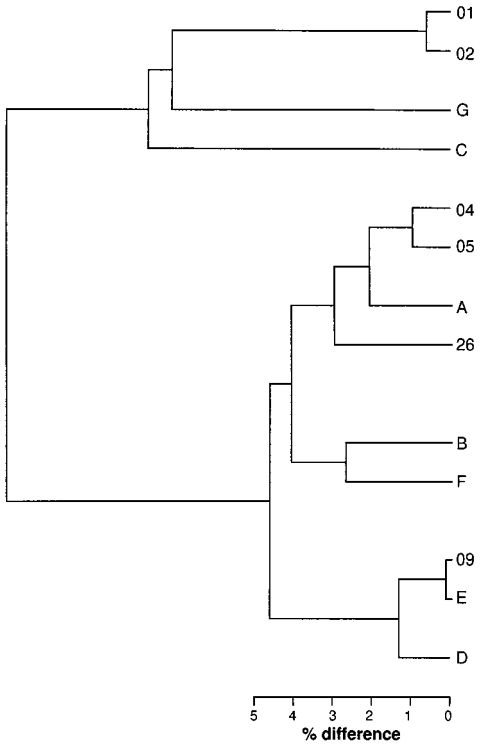

FIG. 2. Unrooted evolutionary tree for strains representing the 13 unique *flaA* sequences identified in Table 1. The tree was constructed from a matrix of pairwise genetic distances by the unweighted pair group average-linkage method of clustering. The scale bar represents a 5% difference in nucleotide sequences, as determined by measuring total lengths of horizontal lines connecting any two strains.

Sawyer test to the full data set, P was found to be $<10^{-4}$, confirming at very high significance that recombination had occurred between the sequences. Sawyer's test was rerun with a reduced data set (01, 02, 09, D, and E were all excluded) in order to determine whether the remaining sequences had any residual mosaic structure. The P value was still highly significant at 10^{-3}, suggesting that recombination had also occurred within this reduced data set.

In order to try to identify additional mosaic gene structures, the maximum chi-square test was used for pairwise comparison of all 13 *flaA* sequences. Many pairs of sequences were found to give block structures, consisting of two or three blocks. Each block represented a region of differing sequence divergence, and all cut points were significant at $P < 10^{-4}$. In no case (other than those already described) did any pair of sequences have total or near-total sequence identity over any block, although divergences of less than ~5% were sometimes observed.

Comparison of each sequence with the sequence for B gives a good indication of the range of block structures identified (Fig. 4). For comparison, conserved and variable *flaA* regions are also shown, superimposed on the gene for strain B. Over the entire set of sequences, the most common block structure identified consisted of three blocks, an upstream and downstream relatively conserved region, separated by a more diverged central region. The most common cut points for this block structure were found to be between nucleotides 473 and 528 for the upstream cut and between nucleotides 1420 and 1441 for the downstream cut. This quite clearly correlates with the conserved-variable-conserved domains of the *flaA* gene and is exemplified by comparison of sequence C or G with B (Fig. 4). A variation on this type of block structure occurred when either cut point (or both) was different from those defined by the conserved-variable-conserved domains (as seen for sequence 04, 05, A, or F compared with B). In some cases, the central diverged region could itself be subdivided into two regions of differing divergence (compare sequence 01, 02, 09, or E with B). One final type of block structure was identified, consisting of just two blocks across the whole sequence, as for sequences D and 26 compared with B.

Origin of altered blocks in the *flaA* gene. Strain D has clearly acquired an altered block of the *flaA* gene at its 3′ terminus compared to strains 09 and E (Fig. 3B). A database search with the downstream segment of the strain D *flaA* gene (189 bases) identified it as 100% identical to the 3′-most segment of the *flaB* gene from strain E. Figure 5 shows a clear transition (at base 1534) of strain D *flaA* sequence from strain E *flaA* to *flaB* sequence, when the three sequences are aligned (polymorphic sites only are given), confirming the identity of this 189-base segment. Since strains D and E actually represent the same (TGH9011) *C. jejuni* strain, this suggests that intragenomic recombination between the highly homologous *flaA* and *flaB* genes of strain E (only 6.1% diverged over the entire sequence) led to the mosaic structure seen for the *flaA* sequence of strain D. The exact position of the upstream crossover point between *flaA* and *flaB* cannot be determined since these sequences are identical between nucleotides 793 and 1533 (Fig. 5), and the recombination site could be anywhere within this region. The downstream crossover site must occur downstream of the coding regions for *flaA* and *flaB* sequences, since the introduced gene segment represents the 3′-most region of the *flaB* gene. The actual crossover site cannot be determined without additional sequence information, but a possible location is within the first 50 bases downstream of the coding region, since these are identical for both *flaA* and *flaB* of strain E (data not shown).

The altered region (Fig. 3A) of *flaA* sequence for strain P02 (shared also by strain P03) compared to strain P01 (shared also by strains P06, P07, P08, and P25) might also have been introduced from the *flaB* gene as a consequence of intragenomic recombination within any of these strains. To test for this

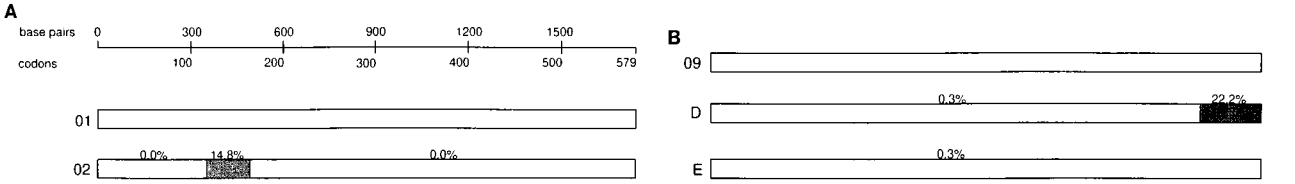

FIG. 3. Mosaic *flaA* genes for strain 01 compared to strain 02 (A) and for strains D and E compared to strain 09 (B). Each line represents the *flaA* gene (nucleotides 1 to 1737) of the strains indicated on the left. Unshaded regions differ from the corresponding regions in the reference sequence (top line) by ≤0.3%. Shaded regions represent diverged blocks of DNA compared to the reference strains, with percent nucleotide sequence divergence as shown.

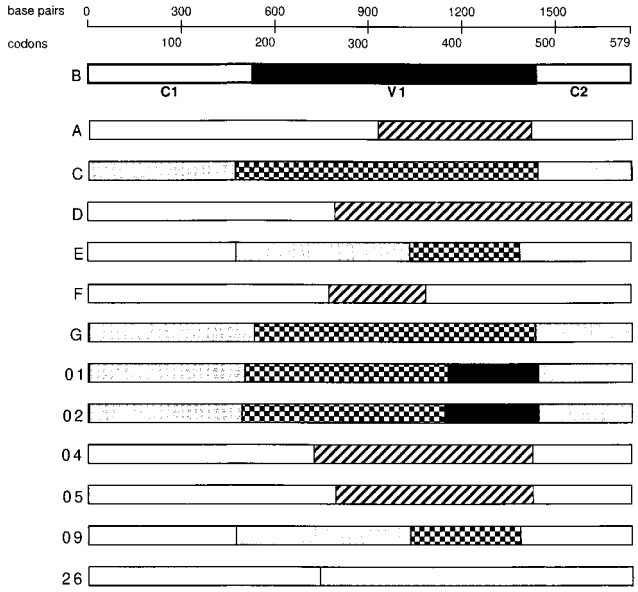

FIG. 4. Mosaic structure of the *flaA* genes of the *Campylobacter* strains identified in Table 1. The different shadings represent regions of the *flaA* gene of each strain that differ from the corresponding regions in strain B by differing amounts: unshaded, ≤5.2% diverged; light grey, 9 to 13% diverged; diagonal stripes, 15 to 20% diverged; grid shading, 28 to 35% diverged; and black, 46 to 47% diverged. Conserved (unshaded) and variable (shaded) domains of the flagellin A protein (see text) are shown superimposed on the gene for strain B.

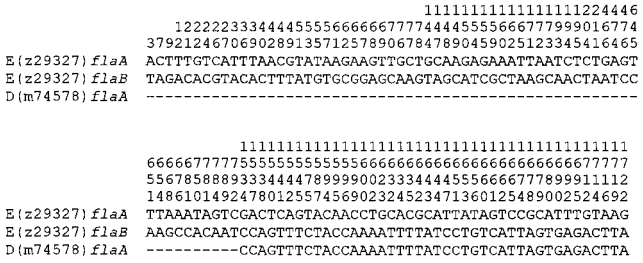

FIG. 5. Alignment of the *flaA* and *flaB* sequences of strain E with the *flaA* sequence of strain D, showing only polymorphic sites. The numbers (in vertical format) above the sequences correspond to the aligned nucleotide position of the polymorphic site.

possibility, the *flaB* genes of all these strains were sequenced over approximately the first 600 bases, which includes the altered block (nucleotides 348 to 489) of DNA seen in P02 compared to P01. Alignment of *flaB* sequences showed them all to be identical. Alignment of this conserved *flaB* sequence with the *flaA* sequences of strains P01 and P02 (Fig. 6) (polymorphic sites only) showed it to be identical to P01 *flaA* sequence over the altered block of DNA, such that the altered block in the *flaA* gene of strain P02 cannot have derived from the *flaB* gene of strain P01, and intragenomic recombination must therefore be discounted. The actual origin of the altered block in strain P02 cannot be identified, since none of the other sequences used in this study share the altered sequence, and neither is it found in any *flaB* sequence present in the GenBank database.

DISCUSSION

***flaA* sequence variation and identification of domain structure.** Alignment of *flaA* genes sequenced during this study identified common (C) and variable (V) domains very similar to those described previously for *Campylobacter* flagellin sequences (4), except that the C1 domain was found to extend over the first 176, rather than the first 170, amino acids. In other bacterial flagellins, conserved regions are known to contain sequences essential for transport and assembly of flagellin monomers into flagellin filaments (26), while the internal region lacks this functional constraint and is therefore more highly variable. In this respect, it is interesting to note that most of the changes seen between the *flaA* and *flaB* genes of strain E (Fig. 5) occurred within conserved rather than within variable regions of the gene, as has been noted previously both for *C. jejuni* (13) and for *C. coli* (6).

Recombination in the flagellin locus of *C. jejuni*. Recombination in the flagellin locus has been described previously in the literature for mutant *Campylobacter* strains. Alm et al. (1) described isolation of revertant motile *C. coli* strains in which intragenomic recombination resulted either in deletion or in repositioning (into *flaB*) of Km^r cassettes initially located in the *flaA* gene. Wassenaar et al. (25) subsequently isolated similar motility revertants from immotile mutant *C. jejuni* strains. The same authors also provided direct evidence for intergenomic recombination, when they isolated double-resistant transformants from coculture of two different mutants, each containing a different antibiotic resistance marker in their *flaA* or *flaB* genes. To our knowledge, the present study cites the first direct evidence for intragenomic flagellin recombination in a nonmutant (*C. jejuni* TGH9011) strain and for intergenomic recombination within natural populations of *C. jejuni*. However, it should be noted that similar studies have provided clear evidence for horizontal transfer of flagellin phase 1 (*fliC*) gene sequences in natural populations of *Salmonella* serovars (9, 20), while evidence has also been cited supporting recombinational reassortment of flagellin gene sequences in *Helicobacter pylori* (5).

Strains D and E represent the same *C. jejuni* type strain (TGH9011), and identification of the 3′-most diverged region (Fig. 3B) between their *flaA* genes was therefore unexpected. The diverged region of strain D was subsequently found to have 100% sequence identity with the equivalent region of the *flaB* gene of strain E (Fig. 5), suggesting that intragenomic recombination within the flagellin locus of strain E had generated the mosaic *flaA* gene structure seen for strain D. Similarly, comparison of strain P01 with P02 found them to have identical *flaA* sequences aside from a diverged region (14.8%) between bases 348 and 489. Strain P01 represents five strains constituting outbreak group 1, whereas strain P02 (and P03) was isolated in the same regional public health laboratory, but 1 year later than the outbreak strains (22). There is strong evidence that strains P02 and P03 are in fact clonally related to the outbreak group since they have identical Penner serotypes, biotypes, phagetypes, and *Hae*III and *Pst*I ribotypes and also have the same *Sma*I pulsed-field gel electrophoresis and mul-

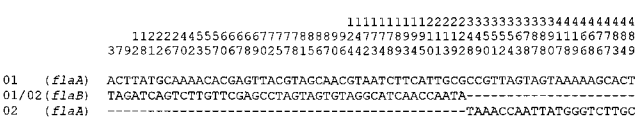

FIG. 6. Alignment of the first 600 bases of the *flaA* and *flaB* sequences of strains 01 and 02, showing only polymorphic sites. The numbers (in vertical format) above the sequences correspond to the aligned nucleotide position of the polymorphic site. Sequences are identical over the region 490 to 600. Strains 01 and 02 have identical *flaB* sequences, and these are therefore represented as one sequence.

tilocus enzyme electrophoresis profiles (22). Strains P01 and P02 therefore have common ancestry, making recombination, rather than genetic drift, by far the most likely explanation for the diverged region of the *flaA* gene, particularly since divergence was limited to one small region of the overall gene sequence. We have shown that intragenomic recombination was not the source of this diverged region, since it was not derived from the *flaB* gene of any of the seven strains involved, and that it must therefore have been introduced from an external source, involving horizontal (intergenomic) transfer following natural transformation. Presumably this occurred during the year following isolation of the original outbreak group.

Mosaic *flaA* gene structures were also identified between pairs of clonally unrelated strains (Fig. 4). Sources of individual blocks could not be identified, since none of them had close sequence identity with other known *flaA* genes over the same region. The most common block structure consisted of three blocks, in which the central block showed greater divergence than the two outside blocks and in which boundaries between blocks were closely matched to conserved and variable amino acid domains (4). These structures can be explained in terms of the differing degrees of sequence constraint imposed within the different domains, although the possibility of recombination as a means for generating the more diverse central regions cannot be excluded. We also identified several block structures which were not superimposable on the domain structure. In these cases, the block structures cannot be explained in terms of differing levels of constraint along the flagellin gene and are in fact difficult to explain by any mechanism other than that they represent regions derived from different evolutionary lineages and were brought together by recombination.

Stability of flagellin gene typing for epidemiological investigations. Flagellin gene typing (PCR amplification of *flaA* genes followed by restriction enzyme digestion, most commonly with *Dde*I or *Hin*fI) has been described recently as a means to differentiate between thermophilic *Campylobacter* strains and has proved to be both quick and discriminatory for epidemiological typing (2, 11, 12, 14). Consequences of recombination between flagellin genes for interpretation of population structures based on flagellin gene typing have been discussed previously with regard to *H. pylori* strains (5). Recombination within or between flagellin loci has the effect of altering the flagellin sequence of a strain, while leaving it otherwise genotypically unaltered. Thus, in the long term, relationships determined by flagellin gene typing will not accurately reflect true clonal relationships, as determined by overall genotypic similarity methods.

Previous studies have noted the occurrence of conserved flagellin profiles across different pulsed-field gel electrophoresis types and ribotypes (i.e., across different genotypes), further suggesting that flagellin gene typing should not be used as the sole basis for epidemiological grouping of strains (17). Common possession of identical (or similar) flagellin gene sequences by genotypically dissimilar strains could be a result of horizontal gene transfer between such strains, as has been shown previously with regard to phase 1 flagellin genes of different *Salmonella* serovars (20). Conversely, in the case of strains D and E in this study, two effectively identical strains (both *C. jejuni* TGH9011) would be grouped separately on the basis of flagellin gene typing, since introduction of the 3′-most 189 bases of *flaB* into the *flaA* gene of strain D results in loss of two *Dde*I sites and acquisition of one new *Hin*fI site.

This study has identified specific examples of both intra- and intergenomic recombination between flagellin genes of *C. jejuni* and also provides evidence for mosaic *flaA* gene structure between other pairs of strains, indicative of past recombinational events between these strains. There are several well-documented examples of intergenomic gene transfer in the literature, most noticeably between the penicillin-binding protein genes of different *Streptococcus* and *Neisseria* species (3) and between the phase 1 flagellin genes of *Salmonella* serovars (9, 20). This sets a strong precedent for the possibility of flagellin recombination occurring by natural transformation of *Campylobacter* strains in the environment. Reassortment of flagellin gene sequences by horizontal gene transfer and recombination argues against the long-term stability of flagellin gene typing, thereby making accurate long-term epidemiological surveillance by this system an unlikely prospect.

ACKNOWLEDGMENTS

This work was supported by Scottish Office grant K/MRS/SO/C2192.

We thank Edward Feil and Brian Spratt for their helpful discussions concerning interpretation of statistical gene conversion tests. This research benefited from the use of the SEQNET facility run by the SERC.

REFERENCES

1. **Alm, R. A., P. Guerry, and T. J. Trust.** 1993. Significance of duplicated flagellin genes in *Campylobacter*. J. Mol. Biol. **230:**359–363.
2. **Birkenhead, D., P. M. Hawkey, J. Heritage, D. M. Gascoyne-Binzi, and P. Kite.** 1993. PCR for the detection and typing of campylobacters. Lett. Appl. Microbiol. **17:**235–237.
3. **Dowson, C. G., T. J. Coffey, and B. G. Spratt.** 1994. Origin and molecular epidemiology of penicillin-binding-protein-mediated resistance to β-lactam antibiotics. Trends Microbiol. **2:**361–366.
4. **Fischer, S. H., and I. Nachamkin.** 1991. Common and variable domains of the flagellin gene, *flaA*, in *Campylobacter jejuni*. Mol. Microbiol. **5:**1151–1158.
5. **Forbes, K. J., Z. Fang, and T. H. Pennington.** 1995. Allelic variation in the *Helicobacter pylori* flagellin genes *flaA* and *flaB*: its consequences for strain typing schemes and population structure. Epidemiol. Infect. **114:**257–266.
6. **Guerry, P., R. A. Alm, M. E. Power, S. M. Logan, and T. J. Trust.** 1991. Role of two flagellin genes in *Campylobacter* motility. J. Bacteriol. **173:**4757–4764.
7. **Guerry, P., S. M. Logan, S. Thornton, and T. J. Trust.** 1990. Genomic organization and expression of *Campylobacter* flagellin genes. J. Bacteriol. **172:**1853–1860.
8. **Khawaja, R., K. Neote, H. L. Bingham, J. L. Penner, and V. L. Chan.** 1992. Cloning and sequence analysis of the flagellin gene of *Campylobacter jejuni* TGH9011. Curr. Microbiol. **24:**213–221.
9. **Li, J., K. Nelson, A. C. McWhorter, T. S. Whittam, and R. K. Selander.** 1994. Recombinational basis of serovar diversity in *Salmonella enterica*. Proc. Natl. Acad. Sci. USA **91:**2552–2556.
10. **Maynard Smith, J.** 1992. Analyzing the mosaic structure of genes. J. Mol. Evol. **34:**126–129.
11. **Nachamkin, I., K. Bohachick, and C. M. Patton.** 1993. Flagellin gene typing of *Campylobacter jejuni* by restriction fragment length polymorphism analysis. J. Clin. Microbiol. **31:**1531–1536.
12. **Nachamkin, I., H. Ung, and C. M. Patton.** 1996. Analysis of HL and O serotypes of *Campylobacter* strains by the flagellin gene typing system. J. Clin. Microbiol. **34:**277–281.
13. **Nuijten, P. J. M., F. J. A. M. van Asten, W. Gaastra, and B. A. M. van der Zeijst.** 1990. Structural and functional analysis of two *Campylobacter jejuni* flagellin genes. J. Biol. Chem. **265:**17798–17804.
14. **Owen, R. J., C. Fitzgerald, K. Sutherland, and P. Borman.** 1994. Flagellin gene polymorphism analysis of *Campylobacter jejuni* infecting man and other hosts and comparison with biotyping and somatic antigenic serotyping. Epidemiol. Infect. **113:**221–234.
15. **Penner, J. L., and J. N. Hennessey.** 1980. Passive hemagglutination technique for serotyping *Campylobacter fetus* subsp. *jejuni* on the basis of soluble heat stable antigens. J. Clin. Microbiol. **12:**732–737.
16. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
17. **Santesteban, E., J. Gibson, and R. J. Owen.** 1996. Flagellin gene profiling of *Campylobacter jejuni* heat-stable serotype 1 and 4 complex. Res. Microbiol. **147:**641–649.
18. **Sawyer, S.** 1989. Statistical tests for detecting gene conversion. Mol. Biol. Evol. **6:**526–538.
19. **Skirrow, M. B., and M. J. Blaser.** 1992. Clinical and epidemiologic considerations, p. 3–9. *In* I. Nachamkin, M. J. Blaser, and L. S. Tompkins (ed.), *Campylobacter jejuni*: current status and future trends. American Society for Microbiology, Washington, D.C.

20. **Smith, N. H., P. Beltran, and R. K. Selander.** 1990. Recombination of *Salmonella* phase 1 flagellin genes generates new serovars. J. Bacteriol. **172:**2209–2216.
21. **Wang, Y., and D. E. Taylor.** 1990. Natural transformation in *Campylobacter species*. J. Bacteriol. **172:**949–955.
22. **Wareing, D. R. A.** Personal communication.
23. **Wassenaar, T. M., N. M. C. Bleumink-Pluym, D. G. Newell, P. J. M. Nuitjen, and B. A. M. van der Zeijst.** 1994. Differential flagellin expression in a *flaA* $flaB^+$ mutant of *Campylobacter jejuni*. Infect. Immun. **62:**3901–3906.
24. **Wassenaar, T. M., N. M. C. Bleumink-Pluym, and B. A. M. van der Zeijst.** 1991. Inactivation of *Campylobacter jejuni* flagellin genes by homologous recombination demonstrates that *flaA* but not *flaB* is required for invasion. EMBO J. **10:**2055–2061.
25. **Wassenaar, T. M., B. N. Fry, and B. A. M. van der Zeijst.** 1995. Variation of the flagellin gene locus of *Campylobacter jejuni* by recombination and horizontal gene transfer. Microbiology **141:**95–101.
26. **Wilson, D. R., and T. J. Beveridge.** 1993. Bacterial flagellar filaments and their component flagellins. Can. J. Microbiol. **39:**451–472.