Molecular characterisation of urease genes from urease-positive thermophilic campylobacters (UPTC)

Y. KAKINUMA, H. IIDA, T. SEKIZUKA, I. TANEIKE, S. TAKAMIYA*, J. E. MOORE†, B. C. MILLAR† and M. MATSUDA

Laboratory of Molecular Biology, School of Environmental Health Sciences, Azabu University, Fuchinobe 1-17-71, Sagamihara 229-8501; †Department of Molecular and Cellular Parasitology, Juntendo University School of Medicine, Bunkyoku Hongo 2-1-1, Tokyo 113-8421, Japan; and ‡Department of Bacteriology, Northern Ireland Public Health Laboratory, Belfast City Hospital, Belfast BT9 7AD, N. Ireland, UK

Accepted: 29 February 2008

ABSTRACT

This study aims to clarify the molecular characteristics of the urease gene operon from urease-positive thermophilic campylobacters (UPTC) obtained from different sources and in various countries. Sequence heterogeneity was observed for the promoter structures at the -35-like region among the 12 isolates examined. The most probable TTG start codon was suggested for the *ureB* and *ureH* genes, and for the *ureA, E, F* and *G* genes, ATG was suggested among all the isolates examined. Overlap was detected between *ureA* and *ureB* and between *ureB* and *ureE* among all the isolates examined. UPTC is the first example of an overlap between the two structural genes *ureA* and *ureB*. When the completely sequenced open reading frames (ORFs) for *ureE*, *ureF, ureG* and *ureH* were identified, non-coding regions between *ureE* and *ureF, ureF* and *ureG*, and *ureG* and *ureH* were also demonstrated. All six start codons of the six urease genes were demonstrated to be preceded by Shine-Dalgarno sequences among all the isolates examined. The Cys-His sequence corresponding to urease active sites were aligned perfectly and fully conserved among the three UPTC isolates examined. A putative and intrinsic ρ-independent transcriptional terminator was identified to be identical among all the isolates examined. A partial and putative ORF of about 200 bp in length showing high sequence similarity to GTP cyclohydrolase I was observed downstream of *ureH*.

KEY WORDS: Campylobacter.
Genes.
Molecular biology.
Urease.

Introduction

Campylobacter lari is a relatively recently discovered thermophilic *Campylobacter* species.[1,2] An atypical group of urease-positive thermophilic campylobacters (UPTC) was first identified in England in 1985.[3,4] After the original description of UPTC, isolates have been reported in France, Northern Ireland, The Netherlands and Japan.[5–13]

The UPTC taxon has been suggested to belong within *C. lari*, possibly as a biovar, based on numerical analysis of high-resolution polyacrylamide gel electrophoresis of proteins.[4] The characterisation of UPTC as a variant of *C. lari* has also been described by hybridisation dot blot assay, as well as by biochemical characterisation.[3] Thus, the UPTC organism is a biochemically atypical taxon, which produces urease among the genus *Campylobacter*, as well as by *C. sputorum* biovar *paraureolyticus*.[14]

In relation to the urease genes of UPTC, the authors recently described the genetic heterogeneity of urease gene loci, consisting of parts of *ureA* and *ureB* (approximately 1.96 kb), following polymerase chain reaction (PCR) amplification using new degenerate PCR primers constructed *in silico*, and analysis of the nucleotide sequences and sequence similarities, as well as the genetic organisation of the urease gene operon from a Japanese UPTC isolate (CF89-12).[16]

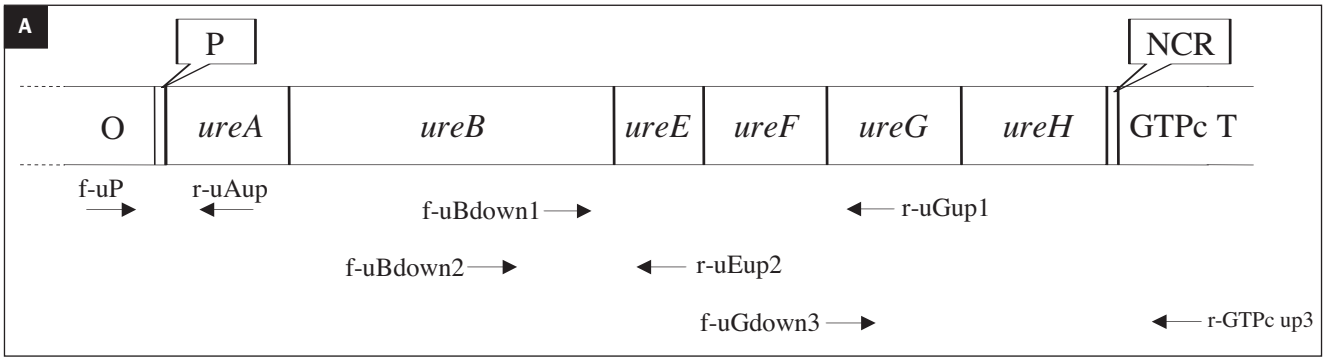
The aim of the present study is to clarify the molecular characteristics of the urease gene operon, including accessory genes of UPTC isolates obtained from different sources and in various countries.

Materials and methods

Fifteen isolates of UPTC were examined and their details are shown in Table 1. These organisms were isolated from the natural environment, including from seagulls and humans, in Japan, England, Northern Ireland and France. Culture and genomic DNA preparation were carried out as described previously.[15] Amplification, TA cloning and sequencing of the amplicons were carried out using procedures described previously.[15] Primer pairs were constructed for the amplification of the promoter region, the structural genes, the accessory genes of UPTC urease and a partial and putative open reading frame (ORF), including a putative transcriptional terminator (Fig. 1). This was based on the nucleotide sequence information following cloning, sequencing and analysis of a urease gene cluster (operon) from the recombinant plasmid DNA of a genomic DNA library of a Japanese UPTC isolate (CF89-12).[16]

Nucleotide sequencing of the amplicons was performed as described previously.[15] Sequence analysis was carried out using the Genetyx-Mac (version 9; Genetyx Co., Tokyo, Japan) computer software. For accuracy, multiple TA-cloned PCR products were sequenced repeatedly.

Correspondence to: Dr. Motoo Matsuda
Email: matsuda@azabu-u.ac.jp



B

Primer	Nucleotide sequence (5'→3')	Reference
f-uP	TCTTATTCCAAAAAAACTAC	15
r-uAup	YAATATGCTCATCAATTCAC	15, 16
f-uBdown1	GGTSARMKWAKTARCTCTAA	16, 20
r-uGup1	ACCASWACCWACAGGACCAC	16, 21–24
f-uBdown2	GATTTAGTGCTTTGGCAACC	15, 16
r-uEup2	GTATCAAACCAACTAAGCTC	16
f-uGdown3	GGGTTATAATGGTTAAAATAGGC	16
r-GTPc up3	GATTGCTTTAAGTTCTACTAG	16

Fig. 1. Schematic representation of a urease structural and accessory gene cluster for UPTC, including (A) the locations of primers for PCR amplification and (B) primer details (O: other region; P: promoter region; NCR: non-coding region; GTPc: GTP cyclohydrolase I; T: terminator region).

Results

Accession numbers of the nucleotide sequence data obtained in the present and previous studies[15,16] are accessible in the DDBJ/EMBL/GenBank (Table 1). In the current study, first the molecular characteristics were analysed from the promoter structure region upstream of the *ureA* to the 3′ end region of the *ureB* within the urease gene operon.

In Figure 2, the nucleotide sequences of about 120 bp (375–495 bp, CF89-12 in AB201709) from the 12 isolates of UPTC examined are shown. The nucleotide positions used are for those of the UPTCCF89-12 isolate (AB201709). Two putative promoter structures consisting of consensus sequences at the –35-like region (TTGCC[T]T; nucleotide position 428–433 bp) and the –10 region (TATAAT; 453–458), as well as the start codon ATG (483–485 bp) for the *ureA* gene were observed among all the 12 isolates examined (Fig. 2).

Based on the sequence similarity analysis, ATG (483–485 bp) was suggested to be the start codon for the *ureA* gene. Probable ribosome binding sites (5′-AGGAGG-3′; Shine-Dalgarno [SD] sequence)[17,18] were also demonstrated at 473–477 bp for the *ureA* gene. The sequences at the –35-like region were 5′ - TTGCTT - 3′ for five isolates (A1, A2, A3, 89049 and 92251) and 5′- TTGCCT - 3′ for the other seven isolates. Thus, sequence heterogeneity was observed for the promoter structures at the –35-like region.

The nucleotide sequences of the 65-bp regions (1137–1201 bp, AB201709) between the 3′ end of *ureA* and the 5′ end of *ureB* of urease genes from the 12 UPTC isolates are shown in Figure 3. The corresponding regions of urease genes from the seven *Helicobacter* species were also aligned for comparison purposes (Fig. 3). The nucleotide sequences of the region from all 12 UPTC isolates contained only three base differences, at nucleotide positions 1139, 1170 and 1180 bp, but not so few for the *Helicobacter* isolates, as shown in Figure 3.

Interestingly, the TTG codon (1151–1153 bp) is the most probable start codon for *ureB* of all the 12 UPTC isolates, whereas ATG is the probable start codon for *Helicobacter* (Fig. 3). Therefore, TTG (1151–1153 bp) would be the most

```
UPTC CF89-12       375 TATAGAAATTCAAAAAGACTATGAAACT-GAATTAAAAAATGCAATAGAATTTTTTGCCTAAAAGCAAGTAAATAGTTATATAATATCTAAAAAA-TTAT 472
UPTC CF89-14           ............................-.....................................................................-....
UPTC NCTC12892         ..................T.........-.....................................................................-....
UPTC NCTC12893         ............................-.....................................................................-....
UPTC NCTC12894         ............................-.....................T...............................................-....
UPTC NCTC12895         C...........................-.....................................A.............................-....
UPTC NCTC12896         ............................-.....................................................................-....
UPTC A1                .......................-...A.A.............T................T..........C.........................T....
UPTC A2                .......................-...A.A.............T................T..........C.........................T....
UPTC A3                .......................-...A.A.............T................T..........C.........................T....
UPTC 89049             .......................-...A.A.............T................T..........C.........................T....
UPTC 92251             .......................-...A.A.............T................T..........C.........................T....
                        ***************** *** *** * ************* ****** ******** ****** *** ************************ ****

UPTC CF89-12       473 AAGGATTAAAATGCATTTTACTC 495
UPTC CF89-14           ........................
UPTC NCTC12892         ........................
UPTC NCTC12893         ........................
UPTC NCTC12894         ........................
UPTC NCTC12895         ........................
UPTC NCTC12896         ........................
UPTC A1                ........................
UPTC A2                ........................
UPTC A3                ........................
UPTC 89049             ........................
UPTC 92251             ........................
                       ************************
```

Fig. 2. Nucleotide sequence alignments of the 120-bp region including the putative promoter and 5′ end region of *ureA* urease genes from the 12 isolates of UPTC. Dots indicate identical bases, changes are indicated, positions identical in all isolates are marked by asterisks, numbers at the left and right refer to the nucleotide position in AB 201709.

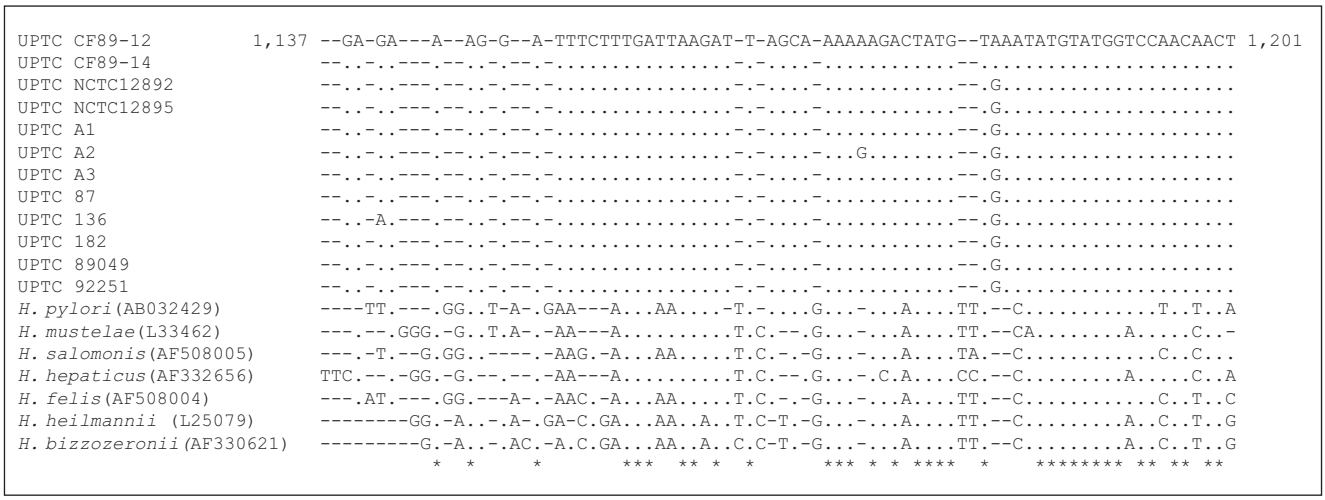


Fig. 3. Nucleotide sequence alignments of the 60-bp regions between the 3' end of *ureA* and the 5' end of *ureB* urease genes from the 12 isolates of UPTC. The corresponding regions of urease genes from the seven *Helicobacter* species are aligned for comparison. The accession numbers for the sequences from the genus *Helicobacter* used in the present study are shown in parentheses.

probable start codon for the *ureB* of the UPTC urease gene operon. In relation to the most probable TTG start codon for *ureB*, an overlap was detected between *ureA* and *ureB*, namely among the four letters including the two between the first U and the second G of the stop codon (UGA, 1152–1154 bp) of the structural *ureA* gene and the second U and the third G of the most probable start codon, TTG (1151–1153 bp) of *ureB* among all the 12 isolates examined.

The nucleotide sequences of about 3000 bp (between the 3' end region of *ureB* and about 200-bp in length of a partial and putative ORF [GTP cyclohydrolase I][16]) from six UPTC isolates (CF89-12, CF89-14, NCTC12892, NCTC12893, A3 and 89049; Table 1) were aligned and analysed. In these regions, completely sequenced ORFs of the accessory genes *ureE, ureF, ureG* and *ureH* were identified for the six isolates. A possible overlap was detected between the third letter A of the stop codon (TG[A]A; 2846–2848 bp) of *ureB* and the first letter of the start codon (ATG; 2848–2850 bp) of the accessory gene *ureE* among all six isolates of UPTC. In addition, three non-coding regions of two bps (3316 and 3317 bp) between *ureE* and *ureF*, 10 bps (3987–3996 bp) between *ureF* and *ureG* and 1 bp (4597 bp) between *ureG* and *ureH* were identified among the six isolates.

In the present study, all six start codons for the six urease genes were preceded by Shine-Dalgarno (SD) sequences: AAGGA (473–477 bp) for *ureA*, as described above, AGGA (1142–1145 bp) for *ureB*, AGAA (2828–2831 bp) for *ureE*, AAGAA (3308–3312 bp) for *ureF*, AAGG (3987–3990 bp) for *ureG* and AGGA (4589–4592 bp) for *ureH* in all six isolates. A non-coding region of approximately 45 bp occurred between *ureH* and the partial and putative ORF (GTP cyclohydrolase I) among all six UPTC isolates examined.

Discussion

Nucleotide sequence similarities (120-bp region in Fig. 2) were 93.4–100% among the 12 UPTC isolates investigated in this study. A region of approximately 250 bp was also cloned and sequenced, which lay upstream of these typical promoter structures, and this appeared not to contain any partial and possible ORFs for all 12 isolates examined, based on the forward and reverse nucleotide sequence similarity analyses.

With respect to *ureB* in all 12 UPTC isolates examined, there is a possible ATG start codon at 1184–1186 bp, and

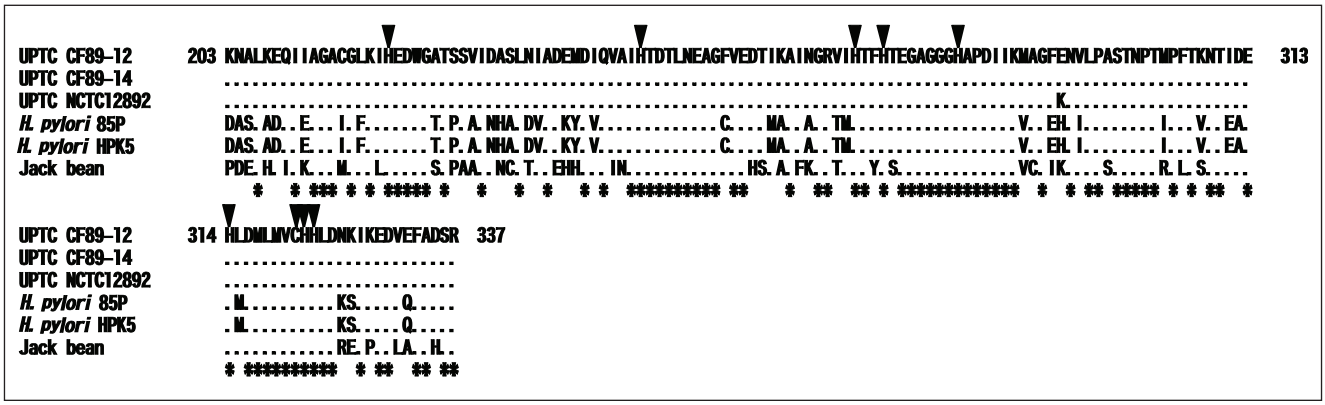


Fig. 4. Partial amino acid sequence alignments of the possible ORFs of *ureB* from the three UPTC isolates (CF89-12, CF89-14 and NCTC12892). The corresponding regions of the possible ORFs from *H. pylori* (X57162; AB032429) and the jack bean (*Canavalia ensiformis*, P07374) are aligned for comparison. Numbers at the left and right refer to the amino acid positions in the putative ORF for *ureB* of CF89-12 (AB201709). Arrows point to eight histidine residues and one cysteine residue.

about 30 nucleotide sequences immediately upstream from this possible ATG codon showed high sequence similarity to the 5′ end of *ureB* from seven species of *Helicobacter* (Fig. 3). The length of the deduced *ureB* polypeptides from all 12 UPTC isolates is similar to those of the *Helicobacter* organisms. Therefore, TTG (1151–1153 bp) would be the most probable start codon for the *ureB* of the UPTC urease gene operon. UPTC is the first example of an occurrence of the TTG start codon for bacterial urease structural genes. In addition, UPTC is also the first example of an overlap for two bacterial structural genes, *ureA* and *ureB*, whose urease comprises two subunits, A and B, similar to those of the genus *Helicobacter*.[19]

The complete and combined nucleotide sequences of the possible ORFs for the structural genes *ureA* and *ureB* (nine isolates of CF89-12, CF89-14, NCTC12892, NCTC12895, A1, A2, A3, 89049 and 92251 for *ureA*; three isolates of CF89-12, CF89-14 and NCTC12892 for *ureB*; Table 1) have been deposited. The nucleotide sequence similarities were 95.8–100% for *ureA* among nine isolates and 98.1–100% for *ureB* among three isolates. When positions of the polymorphic sites and nucleotide sequences were examined, 44 and 32 heterogeneous sites of all substitutions were located in the possible ORFs of *ureA* (669 bp) from the nine isolates and *ureB* (1692 bp) from the three isolates, respectively.

Previously, when nucleotide sequence alignment analysis of the approximately 1.96 kb regions consisting of parts of *ureA* (about 570 bp) and *ureB* (about 1390 bp) from 12 UPTC isolates was carried out, 144 heterogeneous sites of all substitutions were located throughout this region, the substitution ratio being higher in the *ureA* region (about 10 bases) than in the *ureB* region (about 15 bases). Thus, the substitution ratio was higher in the *ureA* region than in the *ureB* region when both whole *ureA* and *ureB* gene regions were compared.

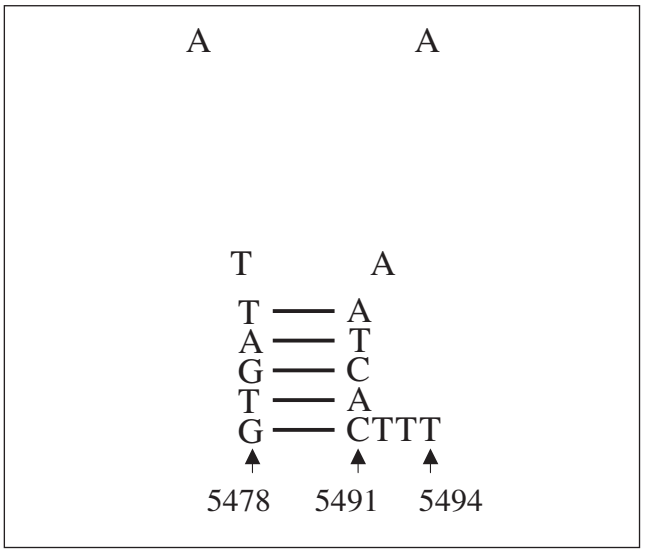


Fig. 5. A hypothetical intrinsic transcription terminator structure containing a stem (5478–5491 bp) and a single-strand run of T residues (5492–5494 bp).

Figure 4 shows the partial amino acid sequence alignments of eight histidine residues and one cysteine residue in the possible ORFs of *ureB* from three UPTC isolates (CF89-12, CF89-14 and NCTC12892), corresponding to *H. pylori*[20,25] and the jack bean[26] urease active sites. These nine residues are perfectly aligned and fully conserved among the six sequences.

Regarding the completely sequenced ORFs, the sequence similarities of *ureE, ureF, ureG* and *ureH* were 97.4–100%, 97.2–100%, 96.5–99.9% and 95.7–100% among the six UPTC isolates (CF89-12, CF89-14, NCTC12892, NCTC12894, A3 and 89049), respectively. In addition, 18, 27, 28 and 32 heterogeneous sites of all substitutions were located in the

Table 1. Isolates of UPTC analysed in the present study and accession numbers of the nucleotide sequences of the urease genes accessible in the DDBJ/EMBL/GenBank.

Organism	Isolate No.	Country	Source	Accession number		
				Promoter-*ureA*	*ureA-ureB*	*ureB, E, F, G, H*-GTPc
UPTC	CF89-12	Japan	River water	AB201709	AB201709	AB201709
					AB182111	
UPTC	CF89-14	Japan	River water	AB204540	AB182112	AB204551
UPTC	NCTC12892	England	River water	AB204541	AB182115	AB204552
UPTC	NCTC12893	England	River water	AB204542	NA	AB204553
UPTC	NCTC12894	England	Sea water	AB204543	NA	NA
UPTC	NCTC12895	England	Mussel	AB204544	AB182122	NA
UPTC	NCTC12896	England	Mussel	AB204545	NA	NA
UPTC	A1	N. Ireland	Seagull	AB204546	AB182117	NA
UPTC	A2	N. Ireland	Seagull	AB204547	AB182118	NA
UPTC	A3	N. Ireland	Seagull	AB204548	AB182119	AB204554
UPTC	87	N. Ireland	Sea water	NA	AB182116	NA
UPTC	136	N. Ireland	Scallop	NA	AB182121	NA
UPTC	182	N. Ireland	Sea water	NA	AB182120	NA
UPTC	89049	France	Human	AB204549	AB182113	AB204555
UPTC	92251	France	Human	AB204550	AB182114	NA

possible ORFs of *ureE* (465 bp), *ureF* (666 bp), *ureG* (597 bp) and *ureH* (750 bp) from the six UPTC isolates, respectively.

As previously demonstrated for the UPTC CF89-12 isolate, a partial and putative ORF (approximately 200 bp), which was shown to have a high sequence similarity to the GTP cyclohydrolase I gene (AAFK 01000005), was also identified downstream of the *ureH* gene among the other five isolates (CF89-14, NCTC12892, NCTC12893, A3 and 89049), based on the sequence similarity analysis in the present study. The nucleotide sequence similarities were 98–100% for the partial and putative ORF (GTP cyclohydrolase I) among the six UPTC isolates, including UPTC CF89-12.

A hypothetical intrinsic ρ-independent transcriptional terminator sequence (5'-GTGATTAAAATCACTTT-3'; 5478–5494 bp) was found to be identical for the urease gene operon among all six UPTC isolates examined. It contains a stem and a single-strand run of T residues (Fig. 5; CF89-12, CF89-14, NCTC12892, NCTC12894, A3 and 89049). The terminator sequences were located in the partial and putative ORF (GTP cyclohydrolase I) and were identical among all the isolates examined. □

JEM was supported by the Research & Development Office, Department of Health, Northern Ireland (Infectious Disease Recognised Research Group [RRG] 9.9).

References

1 Skirrow MB, Benjamin J. '1001' campylobacters: cultural characteristics of intestinal campylobacters from man and animals. *J Hyg (Camb)* 1980; **85**: 417–42.

2 Benjamin J, Leaper S, Owen R J, Skirrow MB. Description of *Campylobacter laridis*, a new species comprising the nalidixic acid-resistant thermophilic *Campylobacter* (NARTC) group. *Curr Microbiol* 1983; **8**: 231–8.

3 Bolton FJ, Holt AV, Hutchinson DN. Urease-positive thermophilic campylobacters. *Lancet* 1985; **i**: 1217–8.

4 Owen RJ, Costas M, Sloss L, Bolton FJ. Numerical analysis of electrophoretic protein patterns of *Campylobacter laridis* and allied thermophilic campylobacters from the natural environment. *J Appl Bacteriol* 1988; **65**: 69–78.

5 Mégraud F, Chevrier D, Desplaces N, Sedallian A, Guesdon JL. Urease-positive thermophilic *Campylobacter* (*Campylobacter laridis* variant) isolated from an appendix and from human faeces. *J Clin Microbiol* 1988; **26**: 1050–1.

6 Bezian MC, Ribou G, Barberis-Giletti C, Megraud F. Isolation of a urease-positive thermophilic variant of *Campylobacter lari* from a patient with urinary tract infection. *Eur J Clin Microbiol Infect Dis* 1990; **9**: 895–7.

7 Wilson IG, Moore JE. Presence of *Salmonella* spp. and *Campylobacter* spp. in shellfish. *Epidemiol Infect* 1996; **116**: 147–53.

8 Mastuda M, Kaneko A, Fukuyama M *et al*. First finding of urease-positive thermophilic strains of *Campylobacter* in river water in the Far East, namely in Japan, and their phenotypic and genotypic characterization. *J Appl Bacteriol* 1996; **81**: 608–12.

9 Endtz HP, Vliegenthart JS, Vandamme P *et al*. Genotypic diversity of *Campylobacter lari* isolated from mussels and oysters in the Netherlands. *Int J Food Microbiol* 1997; **34**: 79–88.

10 Kaneko A, Matsuda M, Miyajima M, Moore JE, Murphy PG. Urease-positive thermophilic strains of *Campylobacter* isolated from seagulls (*Larus* spp.). *Lett Appl Microbiol* 1999; **29**: 7–9.

11 Mastuda M, Shibuya Y, Itoh Y *et al*. First isolation of urease-positive thermophilic *Campylobacter* (UPTC) from crows (*Corvus levaillantii*) in Japan. *Int J Hyg Environ Health* 2002; **205**: 321–4.

12 Matsuda M, Kaneko A, Stanley T *et al*. Characterization of urease-positive thermophilic *Campylobacter* subspecies by multilocus enzyme electrophoresis typing. *Appl Environ Microbiol* 2003; **69**: 3308–10.

13 Matsuda M, Moore JE. Urease-positive thermophilic *Campylobacter* species. *Appl Environ Microbiol* 2004; **70**: 4515–8.

14 On SLW, Atabay HI, Corry JEL, Harrington CS, Vandamme P. Amended description of *Campylobacter sputorum* and revision of its infrasubspecific (biovar) division, including *C. sputorum* biovar *paraureolyticus*, a urease-producing variant from cattle and humans. *Int J Syst Bacteriol* 1998; **48**: 195–206.

15 Usui K, Iida H, Ueno H *et al*. Genetic heterogeneity of urease gene loci in urease-positive thermophilic *Campylobacter* (UPTC). *Int J Hyg Environ Health* 2006; **209**: 541–5.

16 KakinumaY, Iida H, Sekizuka T *et al*. Cloning, sequencing and characterization of a urease gene operon from urease-positive thermophilic *Campylobacter* (UPTC). *J Appl Microbiol* 2007; **103**: 252–60.

17 Ofengand J, Denman R, Nurse K *et al*. Affinity labeling of tRNA-binding sites on ribosomes. *Methods Enzymol* 1988; **164**: 372–97.

18 Stiege W, Stade K, Schuler D, Brimacombe R. Covalent cross-linking of poly (A) to *Escherichia coli* ribosomes, and localization of the cross-link site within the 16S RNA. *Nucleic Acids Res* 1988; **16**: 2369–88.

19 Mobley HLT, Hausinger RP. Microbial urease: significance, regulation and molecular characterization. *Microbiol Rev* 1989; **53**: 85–108.

20 Labigne A, Cussac V, Courcoux P. Shuttle cloning and nucleotide sequences of *Helicobacter pylori* genes responsible for urease activity. *J Bacteriol* 1991; **173**: 1920–31.

21 Mulrooney SB, Hausinger RP. Sequence of the *Klebsiella aerogenes* urease genes and evidence for accessory proteins facilitating nickel incorporation. *J Bacteriol* 1990; **172**: 5837–43.

22 Cussac V, Ferrero RL, Labigne A. Expression of *Helicobacter pylori* urease genes in *Escherichia coli* grown under nitrogen-limiting conditions. *J Bacteriol* 1992; **174**: 2466–73.

23 Maeda M, Hidaka M, Nakamura A, Masaki H, Uozumi T. Cloning, sequencing and expression of thermophilic *Bacillus* sp. strain TB-90 urease gene complex in *Escherichia coli. J Bacteriol* 1994; **176**: 432–42.

24 Zhu J, Teng C-H, Chang C-F *et al*. Cloning and characterization of a *Helicobacter bizzozeronii* urease gene cluster. *DNA Seq* 2002; **13**: 321–31.

25 Akada JK, Shirai M, Takeuchi H, Tsuda M, Nakazawa T. Identification of the urease operon in *Helicobacter pylori* and its control by mRNA decay in response to pH. *Mol Microbiol* 2000; **36**: 1071–84.

26 Takashima K, Suga T, Mamiya G. The structure of jack bean urease. The complete amino acid sequence, limited proteolysis and reactive cysteine residues. *Eur J Biochem* 1988; **175**: 151–65.