Use of the *tna* Operon as a New Molecular Target for *Escherichia coli* Detection[▽]

Camilla Bernasconi,[1]† Giorgio Volponi,[2]† Claudia Picozzi,[2]* and Roberto Foschino[2]

*European Commission, Joint Research Centre, Institute for Health and Consumer Protection, Ispra (VA),
Italy,[1] and DiSTAM, Università degli Studi di Milano, Milano, Italy[2]*

A quantitative real-time PCR targeting the *tnaA* gene was studied to detect *Escherichia coli* and distinguish *E. coli* from *Shigella* spp. These microorganisms revealed high similarity in the molecular organization of the *tna* operon.

Council Directives 2006/7/EC (1) and 98/83/EC (2) on water quality and Commission Regulation 2073/2005 (7) on microbiological criteria for foodstuffs require the determination of *Escherichia coli* to assure hygiene control and consumer safety. Nowadays, gene probe technology provides rapid and highly sensitive techniques for the specific detection of pathogenic microorganisms (6, 8, 11, 15). *E. coli* is able to use tryptophan as a carbon source, and this phenotypic trait is normally employed as a test for identification in conventional culture-based techniques. The tryptophanase operon (*tna*) is a 3,049-bp region consisting of two major structural genes: *tnaA* (1,431 bp), coding for the tryptophanase that catalyzes the degradation of L-tryptophan to indole, pyruvate, and ammonia; and *tnaB* (1,248 bp), coding for a tryptophan permease (21). *tnaA* is preceded by a transcribed regulatory leader region containing a short open reading frame, *tnaL*, specifying a 25-residue leader peptide. In between there is a 205-bp spacer region that contains several transcription pause sites. Moreover, in silico analysis of complete genomes of eight *E. coli* strains available in GenBank confirmed that the *tna* operon is present in single copy, ranking it as a promising molecular biomarker for quantification of the target organism. The aim of this study was to develop a quantitative real-time PCR (Q-PCR) assay to detect pathogenic and nonpathogenic strains of *E. coli*, differentiating them from *Shigella* spp., which are closely related bacteria. Actually the two taxa are often difficult to distinguish by phenotypic traits, while, at genome level, some authors consider them as belonging to the same species (4, 5, 14). In the current study the molecular basis of the indole phenotype in *Shigella* and *Escherichia* species was examined by an investigation of the molecular organization of the *tna* operon. One-hundred eighteen *E. coli* strains and 100 non-*E. coli* strains representative of other enteric and environmental species, collected from different sources and international collections, were studied (Table 1). All the new isolates were identified by the Vitek system (bioMérieux, Marcy l'Etoile, France) and/or through

the sequencing of the 16S rRNA gene operon. After a preliminary screening, the gene encoding tryptophanase (*tnaA*) was chosen as target for the design of *E. coli*-specific PCR primers and of a TaqMan MGB probe by comparing relative gene sequences of *E. coli* K-12 with those of *Shigella flexneri* 2a strain 2457[T], *E. coli* O157:H7 EDL 933, and *E. coli* CFT073 (Table 2). The rationale behind this selection was to identify genomic regions able to discriminate the two species, since published protocols are not suitable for this purpose (3, 9).

Real-time PCR amplification was performed in a 25-μl volume containing 1× TaqMan Universal PCR Master Mix (Applied Biosystems, Foster City, CA), 1 μM of primers, 200 nM of probe, and 100 ng of DNA template. Tests were performed in triplicate in an ABI Prism 7900 thermal cycler (Applied Biosystems), using MicroAmp optical tubes and caps (Applied Biosystems). Amplification conditions are reported in Table 2. One hundred five out of 118 *E. coli* strains (89%) gave positive signal in the Q-PCR test, while no amplification signal was observed in 98 out of 100 non-*E. coli* strains (98%). In particular, all *Shigella* sp. isolates were negative. According to ISO/FDS 16140:2003 (12), the new assay showed an accuracy of 93.1% and a specificity of 98.0%. Thirteen *E. coli* strains, 12 with an indole-positive phenotype, did not produce the expected fragment: CD3, DSM 682, E01, E05, E16, E22, E81 (indole negative), EC52, ED172 (O103:H2), ED226 (O113:H21), EF1 (O111:H⁻), EscCol1, and EscCol2. On the other hand, *Escherichia albertii* DSM 17582[T] (indole negative) and *Escherichia fergusonii* DSM 13698[T] (indole positive) strains gave a false-positive amplification signal. Then, the *tna* operon was entirely sequenced through a gene walking analysis using primers tnaOP_F and tnaOP_R, designed on the 5′ region of *tnaL* and on the 3′ region of *tnaB* of *E. coli* K-12, respectively (Table 2). The same survey was extended to *Shigella boydii* DSM 7532[T], *Shigella flexneri* DSM 4782[T], and *Shigella sonnei* ATCC 29930[T], SHIson1, and PO2. PCR was performed as described above, and the amplification cycles are reported in Table 2. Both strands of the amplification products of the *tna* operon were sequenced, analyzed by the BLAST 2 sequence program, and deposited in the GenBank database. A multiple-alignment distance, determined by the unweighted pair group method using arithmetic averages was used to draw a tree based on *tna* operon sequences by the Kodon software (Applied Maths, Kortrijk, Belgium). *S. sonnei* strains did not am-

TABLE 1. Strains used in this work

Species	Strain(s) (collection[a])	No. of strains isolated in this work	Source(s)
Aeromonas hydrophila	DSM 30187[T]		
Aeromonas veronii		1	Surface water
Aeromonas agilis	ATCC 966[T]		
Bacillus cereus	ATCC 14579[T]		
Bacillus mycoides	ATCC 6462[T]		
Citrobacter amalonaticus	DSM 4593[T]		
Citrobacter braaki		1	Surface water
Citrobacter farmeri	DSM 17655[T]		
Citrobacter freundii	DSM 30039[T], CitFre1 (DiSTAM)	2	Surface water
Citrobacter koseri		1	Surface water
Citrobacter sp.		1	Surface water
Enterobacter aerogenes		1	Surface water
Enterobacter agglomerans	ENTagg4 (DiSTAM)	1	Surface water
Enterobacter amnigenus		1	Surface water
		1	Groundwater
Enterobacter cloacae	DSM 30054[T]	1	Surface water
		1	Groundwater
Enterobacter gergoviae	DSM 9245[T]		
Enterobacter intermedius		1	Surface water
		1	Groundwater
Enterococcus faecalis	ATCC 19433[T], ATCC 27332, NCDO 611		
Enterococcus faecium	ATCC 19434[T]		
Enterococcus hirae	ATCC 8043[T]		
Escherichia albertii	DSM 17582[T]		
Escherichia blattae	DSM 4481[T]		
Escherichia coli	ATCC 11229, 11775[T], 25922, 35150	19	Human stool
	DSM 682, 1576, 6255, 10650, 11250	6	Groundwater
	ED173, ED172, EF1, EF28, ED226 (ISS)	58	Surface water
	NCTC 12079	2	Wastewater
	EscCol1, EscCol2, EC43a2, EC43b, EC44, EC45, EC46, EC50, EC51, EC52, EC55, CD2, CD3, CL10, CM11, 393 (DiSTAM)	2	Treated wastewater
Escherichia fergusonii	DSM 13698[T]		
Escherichia hermannii	DSM 4560[T]	2	Surface water
Escherichia vulneris	DSM 4564[T]	3	Surface water
Flavimonas sp.		1	Surface water
Hafnia alvei	DSM 30163[T]	1	Surface water
Klebsiella ornithinolytica		1	Surface water
Klebsiella oxytoca	DSM 5175[T]	2	Surface water
	KleOxy1 (DiSTAM)	2	Groundwater
Klebsiella pneumoniae	KlePne1 (DiSTAM)		
Lactobacillus acidophilus	ATCC 4356[T]		
Lactobacillus buchneri	CNRZ 214		
Lactococcus lactis	ATCC 19435[T]		
Leclercia adecarboxylata	DSM 5077[T]	1	Surface water
Listeria innocua	ATCC 33090[T]		
Morganella morganii	DSM 30164[T]; CE6, MorMorg1 (DiSTAM)	2	Surface water
Pantoea agglomerans		1	Surface water
Pasteurella haemolitica		1	Surface water
Pasteurella multocida		1	Surface water
Plesiomonas shigelloides	DSM 8224[T]	2	Surface water
Proteus hauseri	ATCC 13315		
Proteus mirabilis	DSM 4479[T]	1	Surface water
Providencia stuartii		1	Surface water
Pseudomonas aeruginosa	ATCC 10145[T], 27853		
Pseudomonas fluorescens	DSM 50106, 50148		
Pseudomonas mendocina		1	Surface water
Pseudomonas putida	ATCC 12633[T]		
Rahnella aquatilis	RahAcq1 (DiSTAM)		
Salmonella enterica serovar Enteritidis	ATCC 13076[T]		
Salmonella enterica serovar Typhimurium	ATCC 13311[T]		
Serratia fonticola		1	Groundwater
Serratia liquefaciens		1	Groundwater
Serratia marcescens	DSM 30121[T]		
	SERmar2, SERmar3, SERmar4 (DiSTAM)		
Shigella boydii	DSM 7532[T]		
Shigella flexneri	DSM: 4782[T], ShiFlex (DiSTAM)		
Shigella sonnei	ATCC 29930[T]; PO2, ShiSon1 (DiSTAM)		
Sphingomonas paucimobilis		1	Surface water
Staphylococcus aureus	ST11 (DiSTAM)		
Vibrio agarivorans	DSM 13756[T]		
Vibrio parahaemoliticus	DSM 10027[T]		
Yersinia enterocolitica	DSM 4780[T], 13030[T]; YerEnt (DiSTAM)		
Yersinia intermedia		1	Surface water

[a] ISS, Istituto Superiore di Sanitá, Rome, Italy; DiSTAM, Dipartimento di Scienze e Tecnologie Alimentari e Microbiologiche, Milan, Italy.

plify the *tna* operon. *E. coli* E81, *E. albertii* DSM 17582[T], and *S. boydii* DSM 7532[T] isolates amplified regions of 4,387 bp, 3,827 bp, and 2,814 bp, respectively. The other strains, including *E. fergusonii* DSM 13698[T] and *S. flexneri* DSM 4782[T],

amplified a region of 3,065 bp as expected. The dendrogram obtained by the alignment of *tna* operon sequences allows discrimination of two major clusters of nucleotide distance at the 72.1% level (Fig. 1). Cluster A includes all the *Escherichia*

TABLE 2. Primers and probe designed in this work and relative amplification conditions

Target gene	Primer or probe	Sequence[a] (5′→3′)	Position on gene[b]	Thermal profile of amplification (no. of cycles)	Product size (bp)
tnaA	tnaA_F	GGGGCGGTGACGCAG	464–478	95°C, 10 min (1); 95°C, 15 s (40); 60°C, 1 min (40)	136
	tnaA_R	CCTGGTGAGTCGGAATGGTG	599–580		
	Probe	FAM-CGATGATGCGCGGCG-MGB	492–506		
tnaL	tnaOP_F	CGAGGATAAGTGCATTATGAATATCT	−16–10	94°C, 10 min (1); 94°C, 1 min (35);	3,065
tnaB	tnaOP_R	TTAGCCAAATTTAGGTAACACGTT	3049–3026	58°C, 1 min (35); 72°C, 1 min (35)	

[a] FAM, 6-carboxyfluorescein.
[b] Positions on genes are given according to the *tnaL* sequence of *E. coli* K-12 (U00096).

spp. and four strains of *Shigella* spp. without insertion sequences (*S. boydii* B12 and *S. flexneri* 2a strain 2457T, 5 strain 8401, and 2a strain 301); cluster B groups all the other *Shigella* sp. strains presenting insertion sequence (IS) elements and showing an indole-negative phenotype. *E. coli* K-12 W3110 has two complete IS5 elements (1,195 bp): one in the intragenic *tnaL-tnaA* region and the other in the *tnaB* gene (Fig. 2).

Isolate E81, from a healthy individual, had a *tnaA* gene that was interrupted by a 1,329-bp IS, which showed a 100% homology with IS10 present in *Salmonella enterica* serovar Typhimurium (17). *E. albertii* DSM 17582[T] has an IS of 768 bp (IS1A) in the spacing region between *tnaL* and *tnaA*; in addition, two nucleotide deletions (T_{1785} and C_{2587}) and one insertion (A_{1718}) with respect to *E. coli* K-12 were detected. As a
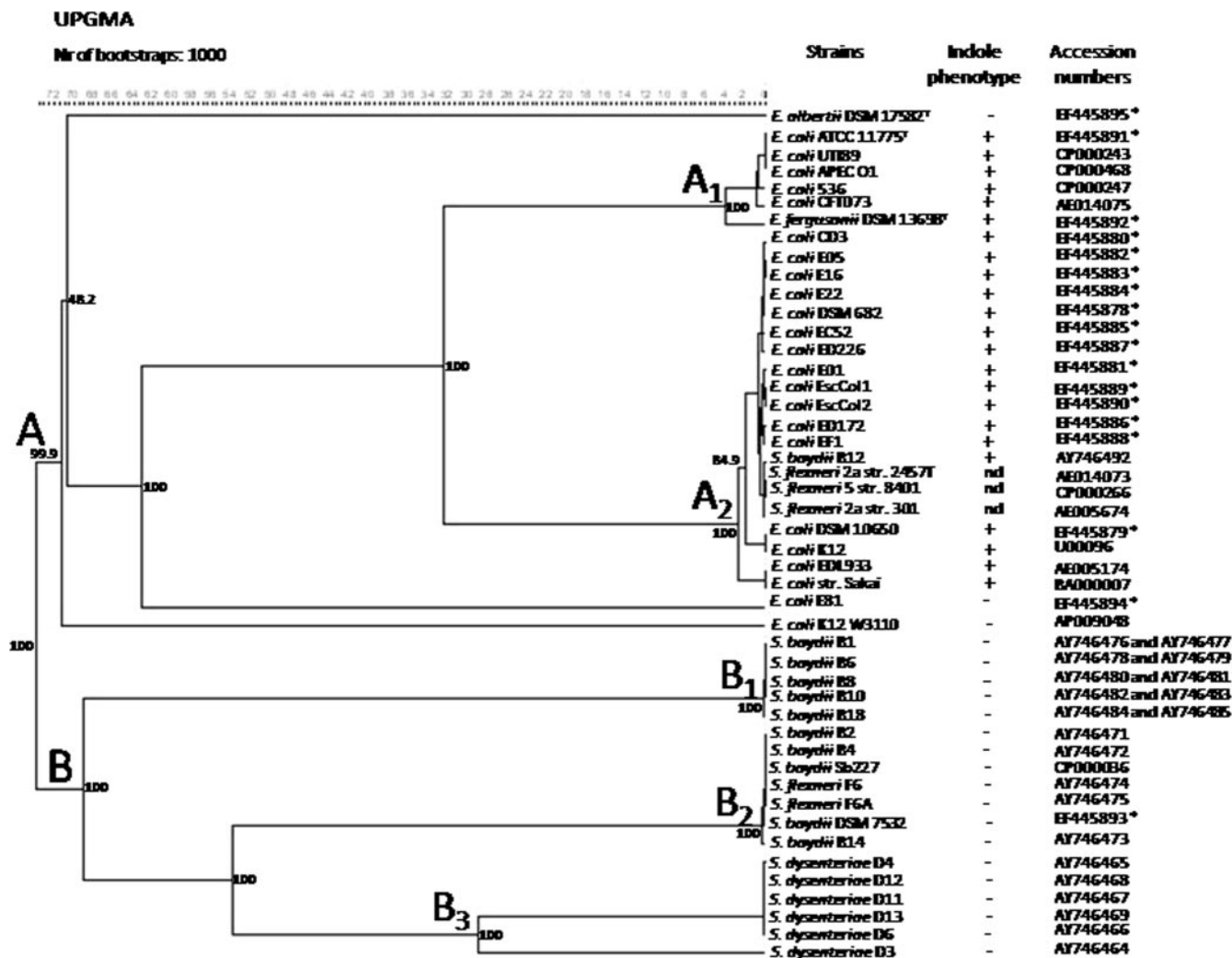

FIG. 1. Tree obtained by the unweighted-pair group method using arithmetic averages (UPGMA) for *tna* operon sequences of *Escherichia* and *Shigella* species. Bootstrap values are indicated on each node of the tree (1,000 pseudoreplicates). ∗, sequences obtained in this work; nd, not determined.
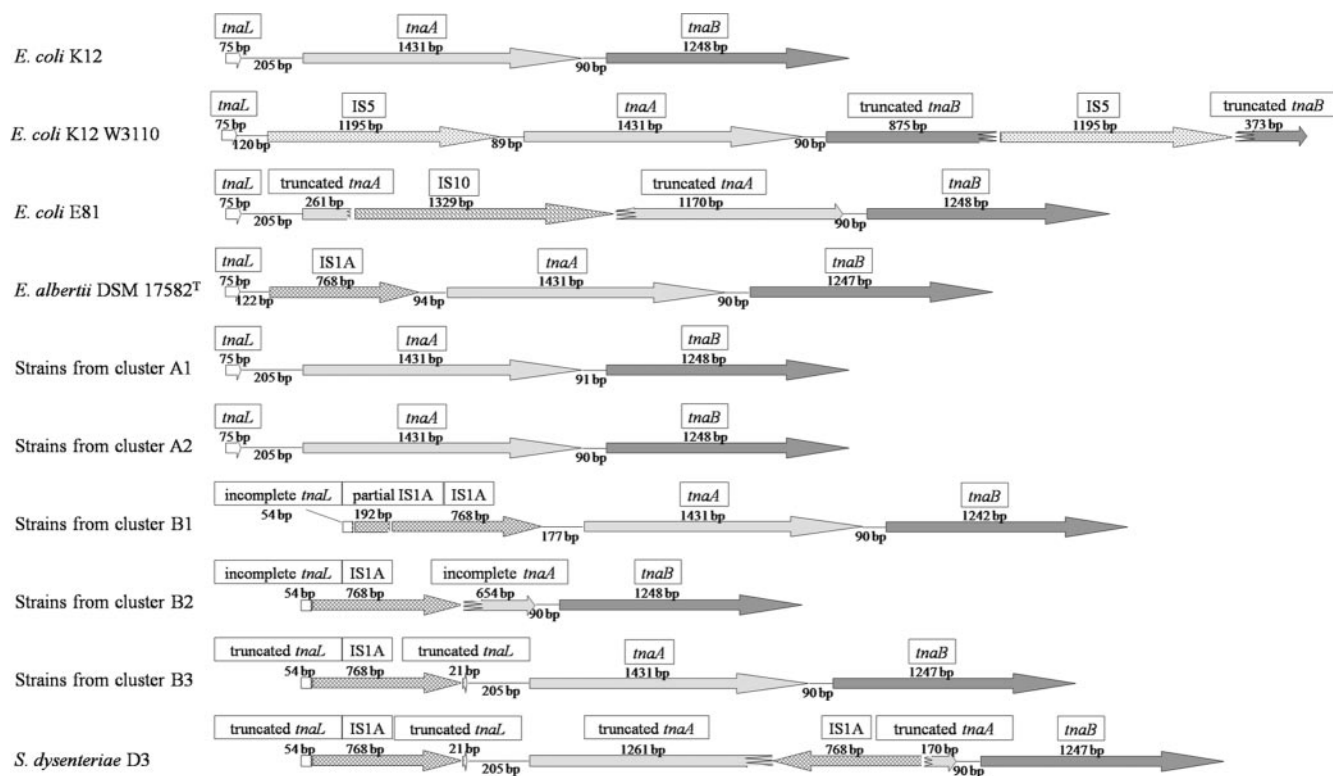
FIG. 2. Comparison of gene organizations of *tna* operons from *Escherichia* and *Shigella* species.

consequence, these strains are unable to produce a functional enzymatic system for tryptophan degradation. In *E. coli* ATCC 11775[T], *E. coli* UTI89, avian pathogenic *E. coli* O1, *E. coli* 536, *E. coli* CFT073, and *E. fergusonii* DSM 13698[T] (cluster A_1) the *tna* operon has the same organization as in *E. coli* K-12 but with a single nucleotide insertion (A_{1718}) in the region between *tnaA* and *tnaB*. Since this insertion falls in a noncoding region, open reading frames are not affected and strains are phenotypically indole positive. With the exception of DSM 10650 and two O157:H7 isolates (EDL933 and strain Sakai), all the *E. coli* strains of cluster A_2 present, compared to *E. coli* K-12, three point mutations in the region from which primer tnaA_F was designed: G_{472} to A, G_{475} to C, and G_{788} to A. Moreover, apart from strains ED226 (O113:H21), ED172 (O103:H2), and EF1 (O111:H[−]), they also present two point mutations in correspondence to primer tnaA_R: C_{580} to T and C_{583} to T. All these mutations take place on the third base of the codon, and, since the amino acid does not change, they do not affect the translation into a functional tryptophanase. However, they decrease the annealing efficiency during PCR amplification, giving rise to a false-negative signal. It is noteworthy that the same point mutations are present in *Shigella* strains grouped in this cluster. *Shigella boydii* B12 has an indole-negative phenotype since, as reported by Rezwan et al. (20), it presents an IS before the region we have investigated that disrupts the expression of the *tna* operon. Cluster B groups *Shigella* spp. whose *tna* operons are affected by insertion elements causing the indole-negative phenotype. In particular, *S. boydii* strains in cluster B_1 show one partial 192-bp IS*1* sequence followed by a full IS*1* sequence; these sequences determine a deletion of 49

bp, including 21 bp of *tnaL*, and they present the same point mutations in *tnaA* primers annealing regions already described for *E. coli* strains of cluster A_2. *S. boydii* and *S. flexneri* strains of cluster B_2 displayed instead a full IS*1* of 768 bp starting from base 55 of *tnaL*; the insertion was followed by the deletion of the complete 205-bp interspace region between *tnaL* and *tnaA* and of the first 777 bp in the 5′ start sequence of the *tnaA* gene; therefore, 235 bp is missing. *Shigella dysenteriae* strains (cluster B_3) presented the same 768-bp IS, IS*1*, at base 55 of *tnaL* of cluster B_2, but without any deletion. Moreover, strain D3 showed the presence of another IS*1* located in an opposite orientation on the *tnaA* gene, giving rise to a lower nucleotide homology (71.0%) compared to other strains (>99.9%) of the same cluster. Furthermore, these isolates presented, in the regions from which primers were drawn, the same point mutations detected on cluster B_1 and A_2 strains. The high frequency of IS elements is known to mediate various genetic rearrangements, including inversions and deletions, that could play an important role in the evolution of the taxa. Besides, the occurrence of gene transfer by conjugation, transduction, and formation of recombinants between *S. flexneri* and *E. coli*, particularly for pathogenic serotypes, has already been demonstrated (16, 18). IS elements have often been associated with negative phenotypes since they can disrupt the functionality of the genes (20); in our study this trait was highlighted through the sequence analysis of the E81 strain, the sole *E. coli* isolate with a negative indole phenotype. The other *E. coli* strains that do not gave amplification signals in this Q-PCR assay reveal the same point mutations and the same organization of the operon as some *S. boydii* and *S. flexnerii* strains, corroborating

the hypothesis that the *Shigella* pathotype arose from *E. coli* ancestors (7, 19). Anyway, it is not possible to find other regions in the *tna* operon that could allow the discrimination between the species that we considered in this work. Although *E. fergusonii* and *E. albertii* gave a false-positive signal, this fact does not invalidate the test since they have the same habitat as *E. coli* (10). Finally, this PCR assay can be employed as a rapid preliminary tool, even if it should be integrated with phenotypic results.

Nucleotide sequence accession numbers. Nucleotide sequences obtained in this work have been deposited in the GenBank database under accession numbers EF445878 to EF445895.

REFERENCES

1. **Anonymous.** 2006. Council Directive 2006/7/EC of 15 February 2006 concerning the management of bathing water quality and repealing Directive 76/160/EEC. Off. J. Eur. Union **L64:**37–51.
2. **Anonymous.** 1998. Council Directive 98/83/EC of 3 November 1998 on the quality of water intended for human consumption. Off. J. Eur. Communities **L330:**32–54.
3. **Bej, A. K., J. L. DiCesare, L. Haff, and R. M. Atlas.** 1991. Detection of *Escherichia coli* and *Shigella* spp. in water by using the polymerase chain reaction and gene probes for *uid*. Appl. Environ. Microbiol. **57:**1013–1017.
4. **Brenner, D. J.** 1984. Family I. *Enterobacteriaceae*, p. 408–420. *In* N. R. Krieg (ed.), Bergey's manual of systematic bacteriology, vol. 1. Williams & Wilkins, Baltimore, MD.
5. **Cilia, V., B. Lafay, and R. Christen.** 1996. Sequence heterogeneities among 16S ribosomal RNA sequences, and their effect on phylogenetic analysis at the species level. Mol. Biol. Evol. **13:**451–461.
6. **Deisingh, A. K., and M. Thompson.** 2004. Strategies for the detection of *Escherichia coli* O157:H7 in foods. J. Appl. Microbiol. **96:**419–429.
7. **European Commission.** 2005. Commission Regulation (EC) no. 2073/2005 of 15 November 2005 on microbiological criteria for foodstuffs. Off. J. Eur. Union **L338:**1–26.
8. **Fode-Vaughan, K. A., J. S. Maki, J. A. Benson, and M. L. Collins.** 2003. Direct PCR detection of *Escherichia coli* O157:H7. Lett. Appl. Microbiol. **37:**239–243.
9. **Hsu, S. C., and H. Y. Tsen.** 2001. PCR primers designed from malic acid dehydrogenases gene and their use for detection of *Escherichia coli* in water and milk samples. Int. J. Food Microbiol. **64:**1–11.
10. **Huys, G., M. Cnockaert, J. M. Janda, and J. Swings.** 2003. *Escherichia albertii* sp nov., a diarrhoeagenic species isolated from stool specimens of Bangladeshi children. Int. J. Syst. Evol. Microbiol. **53:**807–810.
11. **Ibekwe, A. M., P. M. Watt, P. J. Shouse, and C. M. Grieve.** 2004. Fate of *Escherichia coli* O157:H7 in irrigation water on soils and plants as validated by culture method and real-time PCR. Can. J. Microbiol. **50:**1007–1014.
12. **International Organization for Standardization.** 2003. Microbiology of food and animal feeding stuffs—protocol for the validation of alternative methods, ISO 16140, 1st ed. International Organization for Standardization, Geneva, Switzerland.
13. Reference deleted.
14. **Kingombe, C. I., M. L. Cerqueira-Campos, and J. M. Farber.** 2005. Molecular strategies for the detection, identification, and differentiation between enteroinvasive *Escherichia coli* and *Shigella* spp. J. Food Prot. **68:**239–245.
15. **Kuhnert, P., P. Boerlin, and J. Frey.** 2000. Target genes for virulence assessment of *Escherichia coli* isolates from water, food and the environment. FEMS Microbiol. Rev. **24:**107–117.
16. **Lan, R., B. Lumb, D. Ryan, and P. R. Reeves.** 2001. Molecular evolution of large plasmid in *Shigella* clones and enteroinvasive *Escherichia coli*. Infect. Immun. **69:**6303–6309.
17. **Mahillon, J., and M. Chandler.** 1998. Insertion sequences. Microbiol. Mol. Biol. Rev. **62:**725–774.
18. **Parsot, C.** 2005. *Shigella* spp. and enteroinvasive *Escherichia coli* pathogenicity factors. FEMS Microbiol. Lett. **252:**11–18.
19. **Regnault, B., S. Martin-Delautre, M. Lejay-Collin, M. Lefevre, and P. A. Grimont.** 2000. Oligonucleotide probe for the visualization of *Escherichia coli/Escherichia fergusonii* cells by in situ hybridization: specificity and potential applications. Res. Microbiol. **151:**521–533.
20. **Rezwan, F., R. Lan, and P. R. Reeves.** 2004. Molecular basis of the indole-negative reaction in *Shigella* strains: extensive damages to the *tna* operon by insertion sequences. J. Bacteriol. **186:**7460–7465.
21. **Yanofsky, C., V. Horn, and P. Gollnick.** 1991. Physiological studies of tryptophan transport and tryptophanase operon induction in *Escherichia coli*. J. Bacteriol. **173:**6009–6017.