JOURNAL OF CLINICAL MICROBIOLOGY, Nov. 2010, p. 4107–4114 Vol. 48, No. 11
0095-1137/10/$12.00 doi:10.1128/JCM.00677-10

Distribution of Non-Locus of Enterocyte Effacement Pathogenic Island-Related Genes in *Escherichia coli* carrying *eae* from Patients with Diarrhea and Healthy Individuals in Japan▽

Hiroshi Narimatsu,[1,2]* Kikuyo Ogata,[1] Yoshihiro Makino,[2] and Kenitiro Ito[3]

Laboratory of Microbiology, Oita Prefectural Institute of Health and Environment,[1] and Department of Public Health, Faculty of Medicine, Oita University,[2] Oita, and Infectious Diseases Surveillance Center, National Institute of Infectious Diseases, Tokyo,[3] Japan

Received 31 March 2010/Returned for modification 19 May 2010/Accepted 7 September 2010

The relationship to diarrhea of genes located on the pathogenicity islands (PAI) other than the locus of enterocyte effacement (LEE) was investigated. Enteropathogenic *Escherichia coli* (EPEC), the retention of *espC* on the EspC PAI, the OI-122 genes (*efa1*/*lifA*, *nleB*), the phylogenetic marker gene *yjaA*, and the bundle-forming pilus gene *bfpA* on the EPEC adherence factor (EAF) plasmid were studied. *E. coli* strains carrying the intimin gene (*eae*) without the Shiga toxin gene, isolated from patients with diarrhea ($n = 83$) and healthy individuals ($n = 38$) in Japan, were evaluated using PCR. The genotypes of *eae* and *espC* were identified by heteroduplex mobility assay (HMA). The proportions of strains isolated from individuals with and without diarrhea that carried these genes were as follows: *bfpA*, 13.3 and 7.9%, respectively; *espC*, 25.3 and 36.8%; *efa1*/*lifA*, 32.5 and 13.2%; *nleB*, 63.9 and 60.5%; *yjaA*, 42.2 and 55.3%. Statistical significance ($P < 0.05$) was achieved only for *efa1*/*lifA*. The proportion of strains lacking *espC* and carrying *efa1*/*lifA* was higher for patient-derived strains (30.1%) than for strains from healthy individuals (13.2%), but the difference was not significant. Strains carrying both *espC* and *efa1*/*lifA* were rare (2 strains from patients). Statistical analyses revealed significant relationships between *espC* and *yjaA* and between *efa1*/*lifA* and *nleB*, as well as significant inverse relationships between *espC* and *efa1*/*lifA* and between *efa1*/*lifA* and *yjaA*. *espC* was found in *eae* HMA types a1, a2, and c2, whereas *efa1*/*lifA* was found in types b1, b2, and c1. In addition, 6 polymorphisms of *espC* were found. The *espC*, *yjaA*, *efa1*/*lifA*, and *nleB* genes were mutually dependent, and their distributions were related to *eae* type, findings that should be considered in future epidemiological studies.

Enteropathogenic *Escherichia coli* (EPEC) (16), the first pathotype found in *E. coli*, induces diarrhea and can be life-threatening in infants. Histopathology of EPEC infections, known as "attaching and effacing" (A/E) lesions, shows that the bacteria attach intimately to intestinal epithelial cells, causing striking cytoskeletal changes, effacing the microvilli of the intestines. It is thought that effector and regulator genes for various functions that induce this A/E histopathology are located in a 35-kb pathogenicity island (PAI) (9), called the locus of enterocyte effacement (LEE), on the EPEC chromosome and on a 70- to 100-kb plasmid, EPEC adherence factor (EAF). The 94-kDa outer membrane protein intimin, which contributes to the adhesion of EPEC to the intestinal epithelial cells, is encoded by the *eae* gene in the LEE. On the other hand, a type IV pilus, the bundle-forming pilus (BFP), which contributes to interbacterial adherence and possibly to adherence to the intestinal epithelial cells, is encoded by the *bfpA* gene on the EAF plasmid.

In developing countries, in most cases, EPEC isolates recovered from humans with diarrhea are typical EPEC (t-EPEC) strains that have an EAF plasmid; however, in developed industrial countries, such as Japan, most strains lack the EAF plasmid and are identified as atypical EPEC (a-EPEC) (34).

Although a-EPEC has been isolated in large outbreaks involving both adults and children, it is often isolated from healthy individuals, and these strains are thought to be nonpathogenic. In the past, in case-control studies from both Japan and elsewhere comparing the retention ratios of adherence-related genes, such as *eae*, in *E. coli* isolates obtained from patients with diarrhea and from healthy individuals, no statically significant differences were found. It was therefore determined that the probability of diarrhea inducement could not be estimated solely on the basis of adherence-related genes (12, 27, 31). Hence, elucidation of the closeness of the relation of LEE and non-LEE genes as an indicator of the pathogenicity of a-EPEC was sought around the world.

In recent years, various PAIs other than the LEE PAI have been found on some EPEC chromosomes. There have been reports of genes putatively related to pathogenicity (13, 26). Some EPEC strains secrete a 110-kDa protein with serine protease activity called EspC, which has been found to result in increases in short-circuit current and a potential difference in rat jejunal tissue mounted in Ussing chambers, causing cytoskeletal damage as well as hemoglobin proteolysis (5, 25, 29). EspC is encoded by the *espC* gene in the EspC PAI on the chromosome, outside LEE (25), and belongs to the subfamily of serine protease autotransporters of *Enterobacteriaceae*

* Corresponding author. Mailing address: Laboratory of Microbiology, Oita Prefectural Institute of Health and Environment, 2-8 Takaenishi, Oita-shi, Oita, 870-1117, Japan. Phone: 81-97554-8980. Fax: 81-97554-8987. E-mail: narimatsu-hiroshi@pref.oita.lg.jp.

▽ Published ahead of print on 15 September 2010.

TABLE 1. Oligonucleotide primers used in this study

Gene	Primer name	Sequence (5′ → 3′)[a]	Annealing temp (°C)	No. of cycles	Product length (bp)	GenBank accession no.	Position (bp)	Reference or source of primers
eae	eaek1	(F) GCTTAGTGCTGGTTTAGGAT				M58154	1709–1728	Nakazawa and Itoh, 1995 (28)
	eaek4	(R) TCGCCGTTCAGAGATCGC	55	30	489 (eaek1/eaek4)		2197–2180	Yatsuyanagi et al., 1996 (35)
	EA-2	(R) CTCTGCAGATTAACCTCTGC			591 (eaek1/EA-2)		2299–2280	
	mSK1	(F) CCGGCACAAGCATAAGC	55	30	310	M58154	1673–1689	This study
	eaekas_a	(R) TGGCAAAATGATCTGCTG					1982–1965	
espC	espC 2L	(F) AAACATTGGCTCTGGTGGTC			216	AF297061	6376–6395	This study
	espC 2R	(R) CCATCCCCCGTTTTAAGATT	55	30			6591–6572	
	espCseq_A2	(R) AGATGTAGCATGTCCCTGCAT			634 (espC 2L/ espCseq_A2)		7009–6989	This study
yjaA	yjaA L	(F) GATCTTGTTCTGCAACTCCAC	55	30	222	CP000948	4311097–4311117	This study
	yjaA R	(R) TGCCATGACCGCACTATC					4311318–4311301	
nleB	nleB L	(F) GCTTTCACCGATAAGGACAAC	55	30	273	AB303062	2428–2448	This study
	nleB R	(R) TCGCCATCAACAAAAATACC					2700–2681	
efa1/lifA	efa1lifA L	(F) AGAATGGAAGATCACACCAG	55	30	310	AJ459584	9993–10012	This study
	efa1lifA R	(R) ATAATGCCTTTCATCCACAC					10302–10283	

[a] (F), forward; (R), reverse.

(SPATEs) (6, 10, 11). However, the target of EspC in living organisms is unclear, and its distribution in a-EPEC and contribution to the symptoms of diarrhea have not been investigated. The finding of secretory immunoglobulin A antibodies responding to EspC in breast milk from Mexican women living under poor sanitation conditions suggests that EspC probably participates as an enterotoxin in EPEC infections (24).

efa1 (30) encodes the large, 385-kDa adhesin protein Efa1 and is located on a PAI that is similar to genomic O island 122 (OI-122) (18, 21) of the enterohemorrhagic *E. coli* (EHEC) O157:H7 strain EDL933. This gene is the same as *lifA*, which encodes lymphostatin (LifA) (19, 20), in the EPEC strain E2348/69. LifA inhibits the proliferation of peripheral blood lymphocytes and gastrointestinal lymphocytes as well as the production of lymphokines. The gene was designated *efa1/lifA*.

In a case-control study with Norwegian children (2), where a statistical comparison comprehensively investigated the presence of pathogenicity-related gene sequences in a-EPEC strains isolated from children with and without diarrhea, Afset et al., found that *efa1/lifA* and *nleB* (the gene encoding non-LEE effector protein B) in OI-122 were the genes most strongly linked to diarrhea and that the phylogenetic marker gene *yjaA* was inversely related. The Norwegian study used a microarray composed of 242 different probes for the detection of 182 virulence genes or markers. Seven additional putative virulence genes that were not included in the microarray were detected by PCR.

To study the possibility of the relatedness of the genes encoded outside LEE to diarrhea, we investigated the retention of all the genes mentioned above in *eae*-containing *E. coli* strains isolated from healthy individuals and patients with diarrhea, in an attempt to establish whether they could provide an indicator for EPEC diarrhea symptoms, including those caused by a-EPEC, in Japan.

MATERIALS AND METHODS

Bacterial strains. *E. coli* strains containing the *eae* gene, isolated from healthy individuals and from patients with diarrhea and kept at the National Institute of Infectious Disease and the Oita Prefectural Institute of Health and Environment (except for EHEC), were used in the investigation. The samples included all serotypes; where the same serotype was selected, the samples were from multiple places in geographically disparate regions, belonged to different groups, or were isolated at different times. There were a total of 83 strains from diarrhea patients and 38 strains from healthy individuals exhibiting no symptoms of diarrhea. In addition to those strains, a reference stain, KI1317 (O127a:H6), isolated in Thailand, was used.

Serotyping. The commercial serotyping *E. coli* antisera Seiken Set1 and Set2 (Denka Seiken Co. Ltd., Tokyo, Japan) were used. Serotyping was performed according to the manufacturer's instructions. Serotypes that could not be distinguished by this method were designated OUT (O but untypeable) or HUT (H but untypeable).

PCR. The absence of the Shiga toxin gene (*stx*) was determined by the methods of Karch and Meyer (17) and Etoh et al. (8). The *bfpA* gene was confirmed using the primer set bfpAks/bfpAkcomas2 and the methods of Iida et al. (14).

Detection of each of the genes *eae*, *espC*, *efa1/lifA*, *nleB*, and *yjaA* was performed by the following PCR methods: bacterial strains, cultured overnight on normal agar plates at 36°C, were suspended in 100 μl of sterile distilled water, and the template DNA supernatant was prepared by boiling extraction for 10 min followed by centrifugation; 5 μl of the template was included in 50 μl of total reactant solution, and the final concentration of the primer was 0.2 μM. Each of the primers is shown in Table 1. Buffer solution and either TaKaRa Ex-*Taq* HS (Takara Bio Inc., Otsu, Japan) or GoTaq (Promega Corp., Madison, WI) DNA polymerase were added to the sample according to the instructions for the procedure.

The thermal cycler used was either the DNA Engine Tetrad PTC-225 (Bio-Rad Laboratories, Inc., Hercules, CA) or the GeneAmp PCR system 9600R (Applied Biosystems Ltd., Carlsbad, CA). The reaction conditions used, according to the calculation format, involved preheating at 94°C for 2 min, followed by 30 cycles of 5 s at 94°C, 5 s at 55°C, and 10 s at 72°C. The PCR reactants were added to ethidium bromide-stained gels, which were subjected to 2% agarose gel electrophoresis (E-Gel; Invitrogen, Carlsbad, CA), and the products were observed on a transilluminator. PCR of *eae* was performed using the eaek1/eaek4 (28) primer set and either eaek1/EA-2 (35) or mSK1/eaekas_a. The PCR primer set used for *espC* screening was espC 2L/espC 2R.

Genotyping. Intimin typing was performed by the heteroduplex mobility assay (HMA) method reported by Ito et al. (15). The type of cluster x, as reported by

Ito et al., was termed the e type. The PCR for the purpose of HMA typing employed either the eaek1/eaek4 or the mSK1/eaekas_a primer set.

The test for polymorphism of the *espC* gene was performed using the HMA method. The espC 2L/espCseq_A2 primer set was used for PCR of *espC* for the purpose of HMA typing.

HMA. Amplicons obtained from *eae* PCR or *espC* PCR were subjected to the HMA. Briefly, an appropriate amount of the amplicon from the isolate was mixed with 2 μl of the amplicon from an *eae* reference strain, KI1218 or KI1223, or an *espC* reference strain, 02G82D1, 2 μl of 50 mM EDTA (pH 8.0), and sterile distilled water to a final volume of 10 μl. The mixture was denatured at 94°C for 5 min, reannealed at 72°C for 3 min, and subjected to 50°C for 1 h. Heteroduplexes were separated by polyacrylamide gel electrophoresis (PAGE) on a 7.5% separation gel and a 5% stacking gel without sodium dodecyl sulfate (SDS).

Statistical analysis. The retention ratios of each gene in the strains from subjects with and without diarrhea were compared. Fisher's exact test was used to establish independence. Repeat multiple testing was performed using the Benjami and Hochberg method in order to consider the significance. The false discovery rate (FDR) was set at 0.05.

We performed Fisher's exact test to establish independence for the statistical analysis of mutual relatedness among the five types of genes investigated. We defined *P* values of <0.05 as statistically significant and calculated the odds ratio (OR) and the 95% confidence interval (95% CI).

RESULTS

Status of gene retention. Details regarding the retention statuses of the genes investigated in the strains used are shown in Table 2. When the existence of the EAF plasmid was investigated, as indicated by the presence of the *bfpA* gene, 11 strains from patients with diarrhea (13.3%) and 3 strains (7.9%) from normal subjects tested positive, indicating that most were a-EPEC.

Of the 83 strains from patients with diarrhea, 21 were *espC* positive (25.3%), 27 were *efa1*/*lifA* positive (32.5%), 53 were *nleB* positive (63.9%), and 35 were *yjaA* positive (42.2%). Of the 38 strains from healthy individuals, 14 were *espC* positive (36.8%), 5 were *efa1*/*lifA* positive (13.2%), 23 were *nleB* positive (60.5%), and 21 were *yjaA* positive (55.3%) (Table 2).

Statistical analysis showed that only the retention ratio of *efa1*/*lifA* was significant, and when the comparison was performed excluding the *bfpA*-positive strains, the result was the same; i.e., only the retention ratio of *efa1*/*lifA* was significant (Table 3).

The number of possible combinations of retention of the 4 genes discussed, excluding *bfpA*, is 16 in mathematical terms, but in this investigation only 12 were confirmed (Table 2). Of these, the *espC*-negative, *efa1*/*lifA*-positive, *yjaA*-negative, *nleB*-positive combination was presented in 21.7% (18/83) of the strains from patients with diarrhea and in 10.5% (4/38) in the strains from healthy individuals. Although the difference in distribution was the largest, no *bfpA*-positive strains were found in that group. Thirty-two strains were *efa1*/*lifA* positive, of which 27 were *nleB* positive, but only 4 retained *yjaA* (overlapping with *nleB* retention).

On the other hand, of the 35 *espC*-positive strains, 33 were *yjaA* positive, and 17 were *nleB* positive (16 of those were also *yjaA* positive). As demonstrated by these results, *efa1*/*lifA* tended to be retained in combination with *nleB*, and *espC* with *yjaA*, whereas, in contrast, *efa1*/*lifA* and *yjaA* or *bfpA* tended not to be retained together.

Accordingly, statistical analysis of the mutual relatedness between all 10 possible pairs of the five types of genes investigated in this study revealed that the relatedness was statistically significant (*P*, <0.05) in the following four combinations: *espC* and *efa1*/*lifA*, inversely related (*P* = 0.00059; OR = 0.113; 95% CI, 0.025 to 0.504); *espC* and *yjaA*, related (*P* = <0.00001; OR = 45.196;, 95% CI, 10.034 to 203.581); *efa1*/*lifA* and *nleB*, related (*P* = 0.003; OR = 4.408; 95% CI, 1.556 to 12.492); *efa1*/*lifA* and *yjaA*, inversely related (*P* = 0.00001; OR = 0.102; 95% CI, 0.033 to 0.314).

Patterns of serotypes and gene retention. When the retention status of each of the genes was observed (Table 2), there was no clear, recognizable pattern for the retention of *yjaA* and *nleB*; however, there was a trend for the retention of *efa1*/*lifA* and *espC* to be related to specific serotypes. Specifically, they were pattern (A), *espC* positive and *efa1*/*lifA* negative (e.g., O63:H6 and O157:H45); pattern (B), *espC* positive and *efa1*/*lifA* positive (1 strain each of O142:H6 and O161:HNM); pattern (C), *espC* negative and *efa1*/*lifA* positive (e.g., O55:H7); and pattern (D), *espC* negative and *efa1*/*lifA* negative (e.g., O128:H2).

Pattern (D) occurred in 37 strains from patients with diarrhea (44.6%) and in 19 strains from healthy individuals (50.0%), accounting for the largest proportion, irrespective of strain isolation source. Pattern (A), which occurred in 19 strains from patients with diarrhea (22.9%) and 14 strains from healthy individuals (36.8%), accounted for the second largest proportion for both groups. Pattern (C) occurred in 25 strains from patients with diarrhea (30.1%) and 5 strains from healthy individuals (13.2%), showing a tendency to be more frequent in strains from diarrhea patients. Pattern (B) was the rarest, occurring only in 2 strains from patients with diarrhea (2.4%).

Considering only the *bfpA*-positive strains, patterns (A) and (D) were the most common, as in the *bfpA*-negative strains, and pattern (B) was uncommon.

Pattern (A) occurred in 15 *bfpA*-negative strains (15/72 [20.8%]) and 4 *bfpA*-positive strains (4/11 [36.3%]) obtained from patients with diarrhea, as well as in 13 *bfpA*-negative strains (13/35 [37.1%]) and 1 *bfpA*-positive strain obtained from healthy individuals. Pattern (D) occurred in 32 *bfpA*-negative strains (32/72 [44.4%]) and 5 *bfpA*-positive strains (5/11 [45.5%]) obtained from patients with diarrhea, as well as in 17 *bfpA*-negative strains (17/35 [48.6%]) and 2 *bfpA*-positive strains (2/3 [66.7%]) obtained from healthy individuals. Pattern (B) occurred in 1 *bfpA*-negative strain (1/72 [1.4%]) and 1 *bfpA*-positive strain (1/11 [9.1%]), obtained only from patients with diarrhea. On the other hand, pattern (C) was uncommon among *bfpA*-positive strains. Pattern (C) occurred in 24 *bfpA*-negative strains (24/72 [33.3%]) and 1 *bfpA*-positive strain (1/11 [9.1%]) obtained from patients with diarrhea, as well as in 5 *bfpA*-negative strains (5/35 [14.3%]) and no *bfpA*-positive strains (0/3 [0%]) obtained from healthy individuals.

Pattern (C) showed the largest difference in distribution between patients with diarrhea and healthy individuals. However, statistical analysis of pattern (C) in patients with diarrhea (*n*, 25) and in healthy individuals (*n*, 5) resulted in a *P* value of 0.068, an OR of 2.845, and a 95% CI of 0.995 to 8.138, with no significant difference observed between the two groups. No statistically significant differences were observed in the other patterns between *E. coli* strains obtained from patients with diarrhea and those obtained from healthy individuals.

Intimin type and conservation of the gene. When we investigated the relation between the intimin type determined by HMA and the distribution of each gene (Table 2), *espC* was

TABLE 2. Profiles of genes harbored by *eae*-positive *E. coli* (without *stx*) strains isolated in Japan

Pattern	Presence or absence[a] of the following gene:					n^b	Intimin HMA	Patients		Healthy individuals	
	espC	*efa1/lifA*	*yjaA*	*nleB*	*bfpA*			*n*	Serotype(s)[c] (n^d)	*n*	Serotype(s) (no. of strains)
(A)	+	−	−	−	−	1	e group[e]	1	O153:HNM		
			+	−	−	15	a1	2	O127a:H4, OUT:HUT	4	O142:H34, O157:HUT (2), OUT:H34
							a2	2	O86a:H45, OUT:H45		
							c2	4	O63:H6 (2), OUT:HUT, OUT:H19	3	O63:H6 (2), O126:H6
			+	−	+	2	a1	1	O157:H45	1	O157:H45
			+	+	−	12	a1	3	O55:H6, OUT:H34 (2)	5	O63:HNM, OUT:H6, OUT:H10, OUT:H21, OUT:HUT
							c2	2	O63:H6, O125:H6		
							c3	1	O124:H40	1	O124:H40
			+	+	+	3	a1	3	O157:H45 (3)		
Subtotal						33		19		14	
(B)	+	+	−	+	−	1	e group	1	O161:HNM		
			+	+	+	1	a1	1	O142:H6		
Subtotal						2		2		0	
(C)	−	+	−	−	−	4	b1	1	O26:H11		
							d1	1	O115:HUT	1	O115:HUT
							d group[f]	1	O18:HUT		
			−	−	+	1	b2	1	O115:HUT		
			−	+	−	22	b1	9	O26:H8, O26:HNM, O103:HNM, O119:H2, O119:HNM, O153:H7, O167:H9 (2), OUT:HUT	3	O119:H2 (2), O153:H7
							b group[g]	1	OUT:HNM		
							c1	8	O55:H7 (7), O55:HNM	1	O55:H7
			+	+	−	3	b1	2	O119:H19, OUT:HNM		
							b2	1	OUT:HUT		
Subtotal						30		25		5	
(D)	−	−	−	−	−	14	b1	3	O128:HNM, OUT:HNM (2)		
							c2	1	O153:H19		
							d1	3	O119:HUT, OUT:H2, OUT:H28	3	OUT:H7, OUT:H21 (2)
							d2	1	OUT:H16		
							d group	1	OUT:H21		
							e1	1	OUT:HUT	1	OUT:H28
			−	−	+	3	d1	2	O153:H21 (2)	1	O153:H21
			−	+	−	17	a1	1	OUT:HUT		
							b1	5	O15:H2, O111:H2, O128:H2 (2), O26:HNM	3	O128:H2 (2), OUT:H2
							b group	1	O26:HNM		
							c1	1	O145:HNM		
							c2	2	O153:H19, O167:H19	1	OUT:H19
							d1	1	O115:HNM	2	O153:HUT, OUT:HUT
			−	+	+	2	a1	1	O86a:HUT	1	O86a:HUT
			+	−	−	5	a1	2	O114:HUT, OUT:H6		
							a group[h]			1	OUT:H6
							d1	1	OUT:H7		
							e1	1	OUT:HNM		
			+	+	−	13	a1	4	O20:H6, O20:HNM, OUT:H6 (2)	6	O20H6 (2), O20:HNM (2), OUT:H6 (2)
							b1	1	OUT:HNM		
							b2	1	O157:HN		
							b3	1	O157:H16		
			+	+	+	2	a1	1	OUT:H12		
							b2	1	O111:H27		
Subtotal						56		37		19	
Total						121		83		38	

[a] +, presence; −, absence.
[b] *n*, number of strains.
[c] OUT, O but untypeable; HUT, H but untypeable; HNM, H but nonmotile.
[d] Given only when greater than 1.
[e] Intimin was not e1, although the HMA pattern belonged to the e group.
[f] Intimin was neither d1 nor d2, although the HMA pattern belonged to the d group.
[g] Intimin was neither b1 nor b2, although the HMA pattern belonged to the b group.
[h] Intimin was neither a1 nor a2, although the HMA pattern belonged to the a group.

TABLE 3. Statistical analysis of genes associated with diarrhea

Strain group[a] and gene	No. of strains positive for the gene in:		P^b	Odds ratio	95% Confidence interval
	Patients	Healthy individuals			
All strains					
espC	21	14	0.2093	0.581	0.255–1.324
yjaA	35	21	0.2387	0.590	0.272–1.280
nleB	53	23	0.8397	1.152	0.523–2.538
efa1/lifA	27	5	0.0274	3.182	1.117–9.064
a-EPEC strains (*bfpA* negative)					
espC	17	13	0.1714	0.523	0.218–1.255
yjaA	28	20	0.0979	0.477	0.210–1.084
nleB	46	22	1.0000	1.000	0.435–2.300
efa1/lifA	26	5	0.0233	3.391	1.173–9.807

[a] The group of all strains comprised 83 strains from patients and 38 from healthy individuals. The a-EPEC strains comprised 72 strains from patients and 35 from healthy individuals.
[b] By Fisher's exact test.

distributed predominantly in intimin HMA type a1, a2, and c2 strains, and *efa1/lifA* was distributed in intimin HMA type b1, b2, and c1 strains. The *bfpA* gene was distributed in intimin HMA type a1, b2, and d1 strains, and, in particular, was concentrated in 9 of 14 type a1 strains. The *espC* and *efa1/lifA* genes were retained only in the 2 strains described above. The intimin HMA types were O142:H6, which was type a1, and O161: HNM, which was type e.

***espC* gene polymorphism.** As a result of investigation of the genetic polymorphism of the *espC* gene using HMA (Fig. 1; Table 4), 6 varieties—types a1, a2, a3, b1, b2, and c—were recognized, whereas type a4, observed in the Thai isolate KI1317 (O127a:H6), was not seen. Of the strains from patients with diarrhea, 85.7% (18/21) had type a1, while of the rest, 1 each had type a2, b2, or c. In isolates from healthy individuals, the occurrence of type a1 was most frequent, at 85.7% (12/14), followed by types a3 and b1, with 1 strain each. The *espC* HMA type a1 occurred in intimin HMA type a1, a2, c2, and c3 strains, and *espC* HMA types a3, b1, and b2 occurred in the intimin HMA type a1 strain, whereas *espC* HMA types a2 and c occurred in the intimin HMA type e strain. The O142:H6 strain, which simultaneously retained *espC* and *efa1/lifA*, was *espC* HMA type b2, and the O161:HNM strain was *espC* HMA type a2.

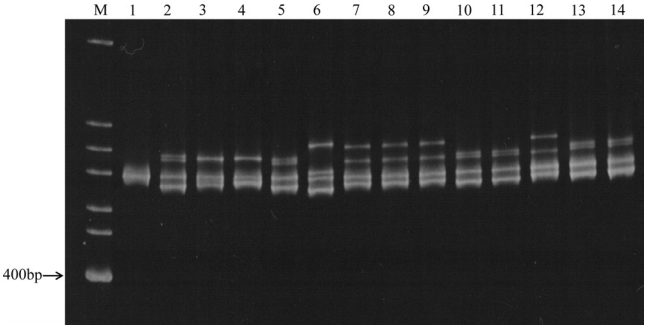

FIG. 1. *espC* HMA profiles of all HMA types with HMA reference strains. Each of the 7 HMA types, a1 to c, formed heteroduplexes with strain 02G82D1 (O153:HNM). Heteroduplexes were separated on homemade 7.5% polyacrylamide gels and were stained with ethidium bromide. Lanes: M, molecular size markers (100-bp ladder); 1, HMA type c, strain 02G82D1 (O153:HNM); 2, HMA type a1, KI1223 (O157:H45); 3, HMA type a2, 02G140S1 (O161:HNM); 4, HMA type a3, H37D1 (OUT:H10); 5, 10, and 11, HMA type a4, KI1317 (O127a:H6), isolated in Thailand; 6, HMA type b1, H72D5 (OUT:H6); 7, 8, 9, and 12, HMA type b2, KI1923 (O142:H6); 13, HMA type a1, 95G19D9 (O63:H6); 14, HMA type a1, KII1701 (O126:H6).

DISCUSSION

In our investigation of gene retention in strains from patients with diarrhea compared to strains from healthy individuals, the only gene for which a statistically meaningful difference was found was *efa1/lifA*. Afset et al. (2) found that *efa1/lifA* had the strongest statistical association with diarrhea. Other OI-122 genes, including *nleB*, were also observed to be associated with diarrhea, but *yjaA* was negatively associated with diarrhea. Our results for *efa1/lifA* were significant and consistent with those reported by Afset et al. However, our study results for *nleB* and *yjaA* differed from theirs. Although the investigation by Afset was a case-control study and thus our investigation methods differed, it is interesting that the trend we found with regard to *efa1/lifA* was similar to the trend in their report.

The results of our investigation show that *efa1/lifA* and *nleB* were retained as a set in many cases, whereas the combination of *efa1/lifA* and *yjaA* shows an inverse relationship. When *E. coli* strains were phylogenetically assigned to 6 groups (A, B1, B2, C, D, and E) by multilocus enzyme electrophoresis (MLEE) and multilocus sequence typing (MLST), *eae*-β type t-EPEC was distributed in group B1, *eae*-α type t-EPEC in group B2, EHEC in groups A and B1, and O157: H7 in group E (7). Because these methods are complex and time-consuming, Clermont et al. developed the simplified multiplex PCR method, which used a combination of 2 genes (*chuA* and *yjaA*) and an anonymous DNA fragment (TspE4.C2), phylogenetically classifying *E. coli* strains into 4 groups (A, B1, B2, and D) (3). As a result of the development of this method, because group D had not been distinguished from group E, EHEC O157: H7 was classified as group D and did not retain *yjaA*. On

TABLE 4. Relation between *espC* and intimin HMA types and other genes harbored

espC type by HMA	Intimin (*eae*) type by HMA	Presence or absence[a] of the following gene:			No. of strains				
					Total	From patients		From healthy individuals	
		efa1/lifA	*nleB*	*yjaA*		*bfpA* negative	*bfpA* positive	*bfpA* negative	*bfpA* positive
espC-a1	a1	−	−	+	8	2	1	4	1
		−	+	+	9	3	3	3	
	a2	−	−	+	2	2			
	c2	−	−	+	7	4		3	
		−	+	+	2	2			
	c3	−	+	+	2	1		1	
espC-a2	e	+	+	−	1	1			
espC-a3	a1	−	+	+	1			1	
espC-b1	a1	−	+	+	1			1	
espC-b2	a1	+	+	+	1		1		
espC-c	e	−	−	−	1	1			
All *espC* types					35	16	5	13	1
espC negative					86	56	6	22	2
Total					121	72	11	35	3

[a] +, presence; −, absence.

the other hand, group B2, to which *eae*-α type t-EPEC belonged, retained *yjaA*.

Moreover, it is not surprising that *efa1/lifA* is often retained in combination with *nleB*, because these genes are located in the same PAI, analogous to OI-122, which is found in EHEC O157:H7. The results of this study indicated that many strains that retain *efa1/lifA* constitute a phylogenetic group with a pattern of retaining *nleB* but not *yjaA*.

The fact that the *espC* gene was widely distributed in *E. coli* strains carrying the *eae* gene isolated from patients with diarrhea and healthy individuals in Japan became clear for the first time in this investigation. As noted in the introduction, EspC is a strong enterotoxin candidate; however, based on the statistical analysis, whether *espC* is retained or not is not thought to be an indicator of diarrhea induction. However, it was found (*de novo*) in this investigation that *espC* is present in 6 of the HMA types found in Japan. The existence of heteromorphisms of HMA types indicates that there are mutations in some percentage of *espC* genes, which would be reflected in the amino acid sequence of EspC, meaning that eventually there could be differences in toxicity. In the present results, most of the bacterial strains are type a1, but other HMA types of *espC* were found in strains from patients with diarrhea and 2 others were found in strains from healthy individuals.

Interestingly, the *espC* HMA types of strains that retained both *espC* and *efa1/lifA* were a2 and b2, not a1. Since there is a report that one of the toxins similar to EspC (SPATEs) is related to inflammatory bowel syndrome (22), EspC can be thought as a factor in worsening inflammation; therefore, the possibility of differences in toxicity between the polymorphs warrants further investigation.

Upon examination of the distribution of the genes of each intimin (*eae*) HMA type, it was found that *espC* was distributed in types a1, a2, c2, and c3, and *efa1/lifA* was in types b1, b2, and c1. These genes are asymmetrically distributed in the bacterial strains by intimin types. In this study, intimin typing was similar to the intimin (*eae*) HMA method of Ito et al. (15) and found that the 5′ terminal end of the *eae* gene was a relatively stable domain; similar domains can be easily grouped in a PAGE pattern. The *eae* types represented by Greek letters (1), which show a large domain with many mutations at the 3′ terminus of the *eae* gene, have also been investigated. The sequence of Greek letters shows the order in which the *eae* types were found and not their groups. However, according to a report by Ito et al. (15), there is a chimeric type *eae* with an α at the 5′ terminal end and a β at the 3′ terminus. For the most part, the types are identical, as shown by another method, and on referring to the report by Ito et al., it can be said that *espC* is found in *eae*-α type strains, whereas *efa1/lifA* is most often retained in *eae*-β, -γ, and -ε type strains.

Representative EPEC prototype strain E2348/69 (O127:H6) belongs to the *eae*-α type, whereas the *eae*-β type includes EHEC O26:H11 and the *eae*-γ type includes EHEC O157:H7 (1). Because *efa1* was discovered to be involved in the synthesis of adhesin protein in EHEC, and O157:H7 was assumed in phylogenetic research to have evolved from the O55:H7 serotype of a-EPEC (4), at least *efa1/lifA*-positive and *espC*-negative strain groups may have some relationship with EHEC. On the other hand, the *espC*-positive strain group is classified as the *eae*-α type, and because the work of Lacher et al. (23) showed that some a-EPEC strains evolved from t-EPEC by losing the EAF plasmid, it would appear to have emerged from t-EPEC, which is classified as *eae*-α.

The intimin (*eae*) type is probably related to the evolution of EPEC and EHEC. Based on these results, it appears that in the evolutionary processes, EPEC divided into the 4 groups observed as a result of acquisition or loss of the EspC PAI and the genes on OI-122 other than those on the LEE PAI. Strains presumed to be strongly pathogenic have an EspC PAI and an OI-122 type PAI, in addition to retaining the EAF plasmid, and are rare in Japan. Groups assumed to have comparatively reduced pathogenicity constitute the

majority of the strains from patients with diarrhea. The reason for the successive evolution toward a-EPEC with weakened pathogenicity in human populations may be that reduced impact is conducive to success in spreading. The reasons for the distribution are unclear and will be the subject of subsequent investigations.

Thus, among the genes investigated in this study, there were combinations of genes that tended to be related to each other and to be retained as a set instead of being independent (*efa1*/*lifA* with *nleB*; *espC* with *yjaA*), as well as combinations of genes that tended to be inversely related and showed a low frequency of simultaneous retention (*efa1*/*lifA* with *espC* or *yjaA*). We also determined gene distribution patterns for each serotype and accompanying intimin type. However, we could not find a statistically significant relationship between diarrhea and any of the four patterns of combinations of the presence and absence of *efa1*/*lifA* and *espC*, probably due to an insufficient number of isolates. In the future, we would like to reexamine this issue with a large sample size; nevertheless, *efa1*/*lifA* may not play a major role in diarrhea, because the 95% CI for the OR of *efa1*/*lifA* in the comparison of patients to healthy individuals is as low as 1.117 (Table 3). Afset et al. (2) attempted a statistical analysis of the contribution of individual genes to diarrhea. Our observations indicate that strains of the a-EPEC group, which simply retained *efa1*/*lifA* and *nleB*, but not *yjaA*, tend to be detected at a higher frequency in patients with diarrhea than in control subjects. However, the contribution of individual genes to the symptoms of diarrhea has not been proven.

Moreover, the fact that *bfpA*-positive strains tended not to retain *efa1*/*lifA* lends weight to the statistical analysis results of Afset et al. (2) in the investigation of a-EPEC for the gene retention pattern of *efa1*/*lifA*-positive strains. In a study of 67 a-EPEC strains obtained from patients with diarrhea in Australia and New Zealand, there were as few as 8 (12%) *efa1*/*lifA*-positive strains, whereas *nleB*-positive strains numbered 20 (30%) (33). In an epidemiological investigation of a-EPEC in Brazil, *efa1*/*lifA*-positive strains amounted to 30.4% and *nleB*-positive strains amounted to 36.9%, however, the retention ratio was not statistically significantly different from that for the control group (32).

As illustrated by these reports, there are several groups within a-EPEC, and the statistical results for the genetic retention patterns appear to reflect the pattern of the predominant a-EPEC group in the selected country and/or region for research. To date, the pathogenic mechanisms of Efa1/LifA responsible for causing diarrhea by directly affecting the intestines have not been identified. Therefore, we believe that *efa1*/*lifA* should be considered a specific marker in the a-EPEC group. The fact that the *efa1*/*lifA*-positive group was common in diarrhea-derived strains in Norway and Japan may reflect the existence of an epidemiologic factor that provides more infection opportunities to that group. This fact may also indicate that the *efa1*/*lifA*-positive group is linked either to a factor related to infectivity or retention of the group in humans or to an unknown entity that causes diarrhea.

The collection of data about the genesis of diarrhea in all regions and information about the relationship of the associated genes is important for future epidemiological studies. Standardization of the genes investigated would lead to more-effective evaluation and use of the information reported.

ACKNOWLEDGMENTS

We are grateful to Shigeru Matsushita (Tokyo Metropolitan Institute of Public Health, Tokyo, Japan), Mitugu Yamzaki (Aichi Prefectural Institute of Public Health, Nagoya, Japan), Kazuo Moriya (Saga Prefectural Institute of Public Health and Pharmaceutical Research, Saga, Japan), Takayuki Kurazono (Saitama Institute of Public Health, Saitama, Japan), Noriaki Hiruta (Yokosuka Institute of Public Health, Yokosuka, Japan), Jun Yatsuyanagi (Akita Prefectural Institute of Public Health, Akita, Japan) and Orn-Anong Ratchtrachenchai (Department of Medical Sciences, National Institute of Health, Nonthaburi, Thailand) for providing *E. coli* strains.

This work was partially accomplished as part of a training program for the enhancement of research capabilities in the field of epidemiology and public health science at the National Institute of Public Health, Wako, Japan. We thank the professors who advised us.

REFERENCES

1. **Adu-Bobie, J., G. Frankel, C. Bain, A. G. Goncalves, L. R. Trabulsi, G. Douce, S. Knutton, and G. Dougan.** 1998. Detection of intimins α, β, γ, and δ, four intimin derivatives expressed by attaching and effacing microbial pathogens. J. Clin. Microbiol. **36:**662–668.
2. **Afset, J. E., G. Bruant, R. Brousseau, J. Harel, E. Anderssen, L. Bevanger, and K. Bergh.** 2006. Identification of virulence genes linked with diarrhea due to atypical enteropathogenic *Escherichia coli* by DNA microarray analysis and PCR. J. Clin. Microbiol. **44:**3703–3711.
3. **Clermont, O., S. Bonacorsi, and E. Bingen.** 2000. Rapid and simple determination of the *Escherichia coli* phylogenetic group. Appl. Environ. Microbiol. **66:**4555–4558.
4. **Donnenberg, M. S., and T. S. Whittam.** 2001. Pathogenesis and evolution of virulence in enteropathogenic and enterohemorrhagic *Escherichia coli*. J. Clin. Invest. **107:**539–548.
5. **Drago-Serrano, M. E., S. G. Parra, and H. A. Manjarrez-Hernández.** 2006. EspC, an autotransporter protein secreted by enteropathogenic *Escherichia coli* (EPEC), displays protease activity on human hemoglobin. FEMS Microbiol. Lett. **265:**35–40.
6. **Dutta, P. R., R. Cappello, F. Navarro-Garcia, and J. P. Nataro.** 2002. Functional comparison of serine protease autotransporters of *Enterobacteriaceae*. Infect. Immun. **70:**7105–7113.
7. **Escobar-Páramo, P., O. Clermont, A. B. Blanc-Potard, H. Bui, C. Le Bouguénec, and E. Denamur.** 2004. A specific genetic background is required for acquisition and expression of virulence factors in *Escherichia coli*. Mol. Biol. Evol. **21:**1085–1094.
8. **Etoh, Y., K. Murakami, S. Ichihara, N. Sera, M. Hamasaki, S. Takenaka, K. Horikawa, K. Kawano, T. Takeishi, Y. Kuwana, A. Inoue, Y. Nagatsu, Y. Hira, M. Takahashi, and K. Ito.** 2009. Isolation of Shiga toxin 2f-producing *Escherichia coli* (O115:HNM) from an adult symptomatic patient in Fukuoka Prefecture, Japan. Jpn. J. Infect. Dis. **62:**315–317.
9. **Gal-Mor, O., and B. B. Finlay.** 2006. Pathogenicity islands: a molecular toolbox for bacterial virulence. Cell. Microbiol. **8:**1707–1719.
10. **Henderson, I. R., and J. P. Nataro.** 2001. Virulence functions of autotransporter proteins. Infect. Immun. **69:**1231–1243.
11. **Henderson, I. R., F. Navarro-Garcia, M. Desvaux, R. C. Fernandez, and D. Ala'Aldeen.** 2004. Type V protein secretion pathway: the autotransporter story. Microbiol. Mol. Biol. Rev. **68:**692–744.
12. **Hernandes, R. T., W. P. Elias, M. A. Vieira, and T. A. Gomes.** 2009. An overview of atypical enteropathogenic *Escherichia coli*. FEMS Microbiol. Lett. **297:**137–149.
13. **Iguchi, A., N. R. Thomson, Y. Ogura, D. Saunders, T. Ooka, I. R. Henderson, D. Harris, M. Asadulghani, K. Kurokawa, P. Dean, B. Kenny, M. A. Quail, S. Thurston, G. Dougan, T. Hayashi, J. Parkhill, and G. Frankel.** 2009. Complete genome sequence and comparative genome analysis of enteropathogenic *Escherichia coli* O127:H6 strain E2348/69. J. Bacteriol. **191:**347–354.
14. **Iida, M., M. Yamazaki, J. Yatsuyanagi, O. A. Ratchtrachenchai, S. Subpasu, N. Okamura, and K. Ito.** 2006. Typing of *bfpA* genes of enteropathogenic *Escherichia coli* isolated in Thailand and Japan by heteroduplex mobility assay. Microbiol. Immunol. **50:**713–717.
15. **Ito, K., M. Iida, M. Yamazaki, K. Moriya, S. Moroishi, J. Yatsuyanagi, T. Kurazono, N. Hiruta, and O. A. Ratchtrachenchai.** 2007. Intimin types determined by heteroduplex mobility assay of intimin gene (*eae*)-positive *Escherichia coli* strains. J. Clin. Microbiol. **45:**1038–1041.
16. **Kaper, J. B., J. P. Nataro, and H. L. Mobley.** 2004. Pathogenic *Escherichia coli*. Nat. Rev. Microbiol. **2:**123–140.
17. **Karch, H., and T. Meyer.** 1989. Single primer pair for amplifying segments of distinct Shiga-like-toxin genes by polymerase chain reaction. J. Clin. Microbiol. **27:**2751–2757.

18. **Karmali, M. A., M. Mascarenhas, S. Shen, K. Ziebell, S. Johnson, R. Reid-Smith, J. Isaac-Renton, C. Clark, K. Rahn, and J. B. Kaper.** 2003. Association of genomic O island 122 of *Escherichia coli* EDL 933 with verocytotoxin-producing *Escherichia coli* seropathotypes that are linked to epidemic and/or serious disease. J. Clin. Microbiol. **41:**4930–4940.
19. **Klapproth, J. M., M. S. Donnenberg, J. M. Abraham, H. L. Mobley, and S. P. James.** 1995. Products of enteropathogenic *Escherichia coli* inhibit lymphocyte activation and lymphokine production. Infect. Immun. **63:**2248–2254.
20. **Klapproth, J. M., I. C. Scaletsky, B. P. McNamara, L. C. Lai, C. Malstrom, S. P. James, and M. S. Donnenberg.** 2000. A large toxin from pathogenic *Escherichia coli* strains that inhibits lymphocyte activation. Infect. Immun. **68:**2148–2155.
21. **Konczy, P., K. Ziebell, M. Mascarenhas, A. Choi, C. Michaud, A. M. Kropinski, T. S. Whittam, M. Wickham, B. Finlay, and M. A. Karmali.** 2008. Genomic O island 122, locus for enterocyte effacement, and the evolution of virulent verocytotoxin-producing *Escherichia coli*. J. Bacteriol. **190:**5832–5840.
22. **Kotlowski, R., C. N. Bernstein, S. Sepehri, and D. O. Krause.** 2007. High prevalence of *Escherichia coli* belonging to the B2+D phylogenetic group in inflammatory bowel disease. Gut **56:**669–675.
23. **Lacher, D. W., H. Steinsland, T. E. Blank, M. S. Donnenberg, and T. S. Whittam.** 2007. Molecular evolution of typical enteropathogenic *Escherichia coli*: clonal analysis by multilocus sequence typing and virulence gene allelic profiling. J. Bacteriol. **189:**342–350.
24. **Manjarrez-Hernandez, H. A., S. Gavilanes-Parra, E. Chavez-Berrocal, A. Navarro-Ocaña, and A. Cravioto.** 2000. Antigen detection in enteropathogenic *Escherichia coli* using secretory immunoglobulin A antibodies isolated from human breast milk. Infect. Immun. **68:**5030–5036.
25. **Mellies, J. L., F. Navarro-Garcia, I. Okeke, J. Frederickson, J. P. Nataro, and J. B. Kaper.** 2001. *espC* pathogenicity island of enteropathogenic *Escherichia coli* encodes an enterotoxin. Infect. Immun. **69:**315–324.
26. **Morabito, S., R. Tozzoli, E. Oswald, and A. Caprioli.** 2003. A mosaic pathogenicity island made up of the locus of enterocyte effacement and a pathogenicity island of *Escherichia coli* O157:H7 is frequently present in attaching and effacing *E. coli*. Infect. Immun. **71:**3343–3348.
27. **Moriya, K., N. Sumi, M. Nakao, M. Yamazaki, M. Saito, and K. Ito.** 2000. The prevalence of virulence-related genes, *eaeA*, *aggR* and *astA*, of localized and aggregative-adherent *Escherichia coli* (EPEC and EAggEC) in healthy children and age-matched patients with diarrhea. J. Jpn. Assoc. Infect. Dis. **74:**134–142.
28. **Nakazawa, M., and K. Itoh.** 1995. Experimental infection of infant rabbits with verocytotoxin-producing *Escherichia coli* of bovine origin. J. Jpn. Assoc. Infect. Dis. **69:**772–776.
29. **Navarro-García, F., A. Canizalez-Roman, B. Q. Sui, J. P. Nataro, and Y. Azamar.** 2004. The serine protease motif of EspC from enteropathogenic *Escherichia coli* produces epithelial damage by a mechanism different from that of Pet toxin from enteroaggregative *E. col*i. Infect. Immun. **72:**3609–3621.
30. **Nicholls, L., T. H. Grant, and R. M. Robins-Browne.** 2000. Identification of a novel genetic locus that is required for in vitro adhesion of a clinical isolate of enterohaemorrhagic *Escherichia coli* to epithelial cells. Mol. Microbiol. **35:**275–288.
31. **Ochoa, T. J., F. Barletta, C. Contreras, and E. Mercado.** 2008. New insights into the epidemiology of enteropathogenic *Escherichia coli* infection. Trans. R. Soc. Trop. Med. Hyg. **102:**852–856.
32. **Scaletsky, I. C., K. R. Aranda, T. B. Souza, N. P. Silva, and M. B. Morais.** 2009. Evidence of pathogenic subgroups among atypical enteropathogenic *Escherichia coli* strains. J. Clin. Microbiol. **47:**3756–3759.
33. **Tennant, S. M., M. Tauschek, K. Azzopardi, A. Bigham, V. Bennett-Wood, E. L. Hartland, W. Qi, T. S. Whittam, and R. M. Robins-Browne.** 2009. Characterisation of atypical enteropathogenic *E. coli* strains of clinical origin. BMC Microbiol. **9:**117.
34. **Trabulsi, L. R., R. Keller, and T. A. Tardelli Gomes.** 2002. Typical and atypical enteropathogenic *Escherichia coli*. Emerg. Infect. Dis. **8:**508–513.
35. **Yatsuyanagi, J., Y. Kinouchi, S. Saito, H. Sato, and M. Morita.** 1996. Enteropathogenic *Escherichia coli* strains harboring enteroaggregative *Escherichia coli* (EAggEC) heat-stable enterotoxin-1 gene isolated from a foodborne like outbreak. J. Jpn. Assoc. Infect. Dis. **70:**73–79. (In Japanese.)