Distribution, Functional Expression, and Genetic Organization of Cif, a Phage-Encoded Type III-Secreted Effector from Enteropathogenic and Enterohemorrhagic *Escherichia coli*[▽]

Estelle Loukiadis,[1] Rika Nobe,[1] Sylvia Herold,[2] Clara Tramuta,[3] Yoshitoshi Ogura,[4,5]
Tadasuke Ooka,[4,5] Stefano Morabito,[6] Monique Kérourédan,[1] Hubert Brugère,[1]
Herbert Schmidt,[7] Tetsuya Hayashi,[5,6] and Eric Oswald[1]*

*INRA, UMR 1225, Ecole Nationale Vétérinaire de Toulouse, Toulouse 31076, France[1]; Institute of Medical Microbiology and
Hygiene, Technical University of Dresden, 01307 Dresden, Germany[2]; Dipartimento di Produzioni Animali, Epidemiologia ed
Ecologia, Facoltà di Medicina Veterinaria, Università degli Studi di Torino, 10095 Grugliasco (TO), Italy[3]; Division of
Bioenvironmental Science, Frontier Science Research Center, University of Miyazaki, Miyazaki 889-1692, Japan[4];
Division of Microbiology, Department of Infectious Diseases, Faculty of Medicine, University of Miyazaki,
Miyazaki 889-1692, Japan[5]; Istituto Superiore di Sanità, Dipartimento di Sanità Alimentare e Animale,
00161 Rome, Italy[6]; and Institute of Food Science and Biotechnology, Department of
Food Microbiology, University of Hohenheim, 70599 Stuttgart, Germany[7]*

Enteropathogenic *Escherichia coli* (EPEC) and enterohemorrhagic *E. coli* (EHEC) inject effector proteins into host cells via a type III secretion system encoded by the locus of enterocyte effacement (LEE). One of these effectors is Cif, encoded outside the LEE by a lambdoid prophage. In this study, we demonstrated that the Cif-encoding prophage of EPEC strain E22 is inducible and produces infectious phage particles. We investigated the distribution and functional expression of Cif in 5,049 *E. coli* strains of human, animal, and environmental origins. A total of 115 *E. coli* isolates from diverse origins and geographic locations carried *cif*. The presence of *cif* was tightly associated with the LEE, since all the *cif*-positive isolates were positive for the LEE. These results suggested that the Cif-encoding prophages have been widely disseminated within the natural population of *E. coli* but positively selected within the population of LEE-positive strains. Nonetheless, 66% of *cif*-positive *E. coli* strains did not induce a typical Cif-related phenotype in eukaryotic cells due to frameshift mutations or insertion of an IS element in the *cif* gene. The passenger region of the prophages carrying *cif* was highly variable and showed various combinations of IS elements and genes coding for other effectors such as *nleB*, *nleC*, *nleH*, *nleG*, *espJ*, and *nleA/espI* (some of which were also truncated). This diversity and the presence of nonfunctional effectors should be taken into account to assess EPEC and EHEC pathogenicity and tropism.

Enteropathogenic *Escherichia coli* (EPEC) and enterohemorrhagic *E. coli* (EHEC) are important causative agents of diarrhea worldwide (16). EPEC is responsible for pediatric protracted diarrhea and is a significant health threat in the developing world, whereas EHEC causes sporadic but deadly outbreaks of hemorrhagic colitis and hemolytic and uremic syndrome in Europe, North America, and Japan. EPEC and EHEC are closely related pathogens, but only EHEC can produce Shiga toxins (Stx) that are encoded by lysogenic bacteriophages (14). Both EPEC and EHEC colonize the intestinal mucosa and produce a characteristic histopathological feature at the microvillus brush border of enterocytes known as an "attaching and effacing" (A/E) lesion. This lesion is characterized by intimate bacterial adhesion to the host-cell membrane resulting from the hijacking of the host cytoskeleton by EPEC and EHEC (11, 16). This alteration results from the action of a repertoire of effector proteins that are delivered into host

cells via a type III secretion system (TTSS) encoded by a pathogenicity island named the locus of enterocyte effacement (LEE).

The identification of a novel EPEC and EHEC type III-secreted effector, the cycle-inhibiting factor (Cif), was recently reported (19). Upon injection into epithelial cells, Cif triggers an irreversible cytopathic effect (CPE) characterized by the inhibition of the cell cycle G_2/M-phase transition and, at least in HeLa cells, the progressive recruitment of focal adhesion plaques, leading to the assembly of stress fibers (19, 25, 33). Cif is not encoded by the LEE but by a lambda-like prophage present in most of the EPEC and EHEC serovars. However, the *cif* gene is absent or truncated in EHEC reference strain RIMD 0509952 (O157 Sakai) and in EPEC strain E2348/69. In the EPEC strain E22 genome, the Cif-encoding prophage was found to be integrated near the *bio* operon, which is part of the core genome present in every *E. coli* organism, and was also found to be similar to a cryptic prophage found in EHEC O157:H7 reference strains EDL 933 and Sakai (19). The location of the Cif-encoding prophage in other EPEC and EHEC serovars remains to be determined.

The availability of at least seven complete *E. coli* genome sequences to date has highlighted the importance of bacterio-

---

TABLE 1. Pathotype, serotype, and virulence factors of reference strains used for genotypic and phenotypic study of *E. coli* strains

Strain	Pathotype (species[a])	Serotype[b]	Virulence factor(s) screened	Reference or source[c]
MG 1655	*E. coli* K-12[d]	Orough:H48		U00096
C600	*E. coli* K-12[d]	Orough:H48		2
E22	EPEC (rabbit)	O103:H2	*eae*-β_1, *cif*	19
E22 *cif*::FRT	EPEC (rabbit)	O103:H2	*eae*-β_1, *cif*-	19
E22 *escN*::FRT	EPEC (rabbit)	O103:H2	*eae*-β_1, *cif*, *escN*-	19
B171-8	EPEC (human)	O111:H-	*cif*, *eae*-β_1	12, 33; this study
RIMD0509952	EHEC (human)	O157:H7	*eae*-γ_1, *stx*$_1$, *stx*$_2$	13
E2348/69	EPEC (human)	O127:H6	*eae*-α_1	29
CF11201	EPEC (human)	O125:H-	*eae*-η	42
CL37	EHEC (human)	O111:H8	*eae*-θ	42
EDL 933	EHEC (human)	O157:H7	*eae*-γ_1, *stx*$_1$, *stx*$_2$	29
EF73	EPEC (human)	O125:H6	*eae*-α_2	29
H-19	EHEC (human)	O26:H11	*eae*-β_1	29
ICC95	EPEC (human)	O86:H34	*eae*-β_2/δ	29
PMK5	EHEC (human)	O103:H2	*eae*-ϵ, *stx*$_1$	29
95NR1	EHEC (human)	O111:H-	*eae*-γ_2, *stx*$_1$, *stx*$_2$	29
4795/95	STEC (human)	O84:H4	*eae*-ζ	42
6044/95	EHEC (human)	O118:H5	*eae*-κ	42
7476/97	EHEC (human)	O145:H4	*eae*-ι	42

[a] Species from which the strain was isolated.
[b] O:H serotypes of *E. coli* strains. H- indicates the presence of nonmotile strains.
[c] First description of the strain virulence factors.
[d] Laboratory strain.

phages in the genetic diversity of bacteria as an important mechanism for evolution and adaptability to new hosts or ecological niches by direct horizontal transfer of new genes (13, 14, 27, 39). In EPEC and EHEC, acquisition of new effector proteins such as Cif that can use the preexisting TTSS for injection into target cells could confer specific advantages in their pathogenic schemes.

In this study, we observed that the *cif*-encoding prophage of reference EPEC strain E22 could be induced and spread to other strains. As populations of *E. coli* from environmental and animal origins may be important natural reservoirs of phages (at least of Stx-encoding phages) that infect *E. coli* (14, 18, 23, 24), we next investigated the distribution of *cif*-encoding prophages in a large collection of *E. coli* isolates from human, animal, and environmental origins. The Cif-encoding prophages were found ubiquitously in *E. coli* isolates from different origins, countries, and time periods and were always found associated with the LEE. Of note, in the passenger region of the prophage, *cif* is associated with other genes coding for other type III-secreted effectors. However, Cif was often not functional due to frameshift mutations leading to truncations in the N- and C-terminal regions.

MATERIALS AND METHODS

Bacterial strains. A total of 5,049 *E. coli* isolates from diverse origins and geographic locations were used in this study. Many of these strains were described in previous studies. Pathogenic strains included 14 EHEC and 30 EPEC strains isolated worldwide from human and animal stools (13, 19–22, 26, 29, 37, 41, 42). Wild-type strains from animal and environmental origins that have not been associated with disease were represented by a collection of 5,001 isolates from slurry, wastewater, and river water collected in and near 12 French slaughterhouses; the collection includes 91 LEE-positive *E. coli* isolates and 31 Stx-producing isolates (18). Four reference *E. coli* strains (EHEC strains RIMD 0509952 [O157 Sakai] and EDL 933, EPEC strain B171-8, and rabbit EPEC strain E22) were also included in this study. Control strains used for genotypic and phenotypic studies of Cif-encoding prophages are listed in Table 1.

Isolation of Cif-encoding ΦE22 bacteriophages and transduction experiments. In order to determine the ability of the Cif-encoding prophages from EPEC strain E22 to produce infectious phage particles, we performed transduction experiments as follows. For preparing high-titer phage stock lysates of 10^5 to 10^6 PFU/ml, EPEC strain E22 was cultured in Luria-Bertani broth (LB; Invitrogen, Paisley, Scotland) supplemented with 5 mM CaCl$_2$ to reach log phase (optical density at 600 nm of 0.6). Mitomycin C (Sigma-Aldrich, Deisenhofen, Germany) was added to achieve a final concentration of 0.5 μg/ml, and the culture was incubated for 16 h at 37°C under conditions of constant agitation (180 rpm). The phage particles were separated from the cell debris by centrifugation (7,500 × *g* for 30 min at 4°C) followed by filtration through 0.22-μm-pore-size filters (Schleicher and Schuell GmbH, Dassel, Germany). To remove bacterial nucleic acids, the phage lysate was treated with 0.5 μg/ml DNase and RNase (Sigma) for 45 min at 37°C. NaCl (5.8% [wt/vol]) was added, and after 1 h incubation at 4°C the solution was centrifuged at 7,500 × *g* for 30 min at 4°C. Phage particles were precipitated by adjusting the supernatant to 10% (wt/vol) polyethylene glycol and harvested by centrifugation (10,000 × *g* for 30 min at 4°C). The resulting pellet was suspended in SM buffer (5.8 g of NaCl, 2 g of MgSO$_4$·7H$_2$O, 50 ml of 1 M Tris-HCl [pH 7.5], and 0.1 g of gelatin in a final volume of 1 liter of double-distilled water). Plaque assays were performed as described previously (31) using *E. coli* K-12 derivative C600 as the recipient strain.

For transduction experiments, 100 μl of prepared phage stock solution was mixed with 100 μl of log-phase culture of *E. coli* C600 and 2 μl of 1 M CaCl$_2$ and incubated for 4 h at 37°C. After addition of 4 ml LB broth and 20 μl of 1 M CaCl$_2$, the mix was further incubated for 48 h at 37°C. The bacterial cells were collected by centrifugation at 7,500 × *g* for 30 min at 4°C and grown on LB agar overnight at 37°C. Colonies were investigated by PCR for the presence of ΦE22 with *cif*-specific primers: cif-int-s/cif-int-as and cif-rbs-s/cif-as (Fig. 1) (19). A PCR-positive colony was selected and designated *E. coli* TE22/C600. To compare restriction patterns of E22 and TE22/C600 phage genomes, the phage DNA was isolated as described previously (31) and subjected to restriction analysis by single and double digestions with EcoRI and/or HindIII according to the manufacturer's recommendations (Fermentas GmbH, St.Leon-Rot, Germany).

DNA sequence analysis. Entire or partial sequences of the Cif-encoding prophages of strain E22 (GenBank accession number AAJV01000058), strain B171-8 (accession number AAJX01000049), and strain E2348/69 (contig Scaffold9_cons; http://www.sanger.ac.uk/Projects/Escherichia_Shigella) are available in online databases. The sequence of Sp3, a prophage of strain O157 Sakai which corresponds to the Cif-encoding prophages of strains E22, B171-8, and E2348/69, is also available in the DDBJ/EMBL/GenBank database (accession number BA000007). The sequences of phages of EHEC O26:H11 strain 11368,
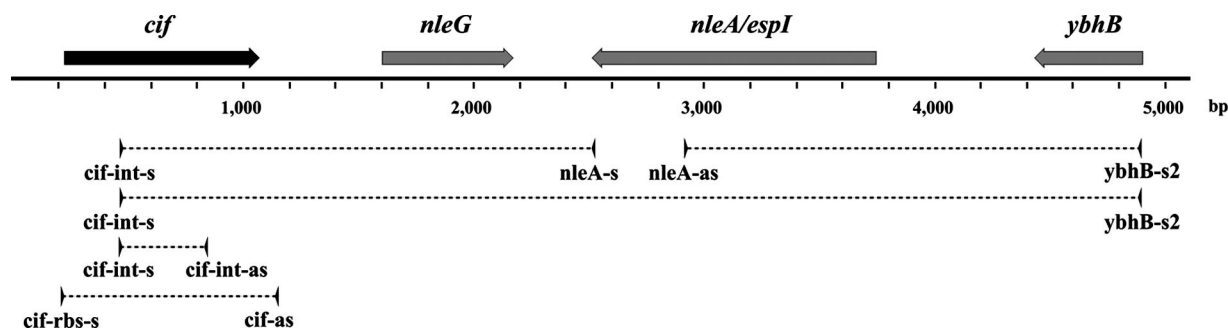
FIG. 1. Physical map and genetic organization of the *cif* gene locus of EPEC O103:H2 strain E22. The line shows a chromosomal fragment of 3,527 bp isolated from EPEC strain E22 (GenBank accession number AAJV01000058). The genes *nleG* and *nleA/espI* are part of the Cif-encoding prophage ΦE22. Gene *ybhB* belongs to the *bio* operon, which is part of the core genome present in every *E. coli*. The positions of the PCR primers used in this study are shown (solid triangles).

EHEC O111:H- strain 11128 (where H- indicates the presence of nonmotile strains), and EHEC O103:H2 strain 12009 were determined as a part of the genome sequencing project involving these three EHEC strains that is in progress in T. Hayashi's laboratory. The sequence assembly for each prophage region containing the *cif* gene was confirmed by PCR using two primers, one targeting the *lom* family gene of the lambdoid phages (lom-s; 5′-GTCTGAAC ATCAGTCCACGC-3′) and the other the *ybhB* gene (ybhB-s2; 5′-CTCATCA GTAACGATCTGCG-3′).

Searches for open reading frames and prediction of translation start positions were performed using Open Reading Frame Finder software (http://www.ncbi .nlm.nih.gov/gorf/gorf.html). Searches for homologous DNA and protein sequences were conducted using BLAST software (1) and the nonredundant GenBank database (http://www.ncbi.nlm.nih.gov/BLAST/BLAST/). The Mac Vector software package (Mac Vector, Inc.) was used to align sequence data and to create phylogenetic trees from a distance matrix created by the unweighted pair group method with arithmetic averages.

Assignment to phylogenetic groups and detection of virulence genes by PCR. Assignment of *E. coli* isolates to one of the four main phylogenetic groups A, B1, B2, and D was performed by multiplex PCR as described previously (7).

DNA from the isolates of *E. coli* was subjected to multiplex PCR for the detection of virulence genes encoding intimin (*eae*) and Shiga toxins (*stx*$_1$ and *stx*$_2$) as described previously (5).

Eight intimin types (α, β, γ, σ, ζ, η, ι, and κ) were investigated by PCR as described previously (42). To establish further differentiation within the different types of intimin genes, PstI restriction analysis of the PCR amplification products was performed to differentiate α$_1$, α$_2$, β$_1$, β$_2$/δ, γ$_1$, and γ$_2$/θ variants as described previously (29).

The presence of the internal region of *cif* gene and the *cif* full gene in the EPEC and EHEC strains was investigated by PCR using primer pair cif-int-s and cif-int-as and primer pair cif-rbs-s and cif-as, respectively, as described previously (19) (Fig. 1).

To examine the chromosomal insertion site of the Cif-encoding prophage in *E. coli* strains, the ybhB-s2 primer was designed in the *ybhB* gene and used in combination with the cif-int-s forward primer. The expected sizes of the PCR products were 4,437 bp (for strains B171 and E22; Fig. 1), 3,337 bp (E2348/69), 4,436 bp (12009), 4,626 bp (11128), and 7,334 bp (11368).

To examine the genetic organization of the Cif-encoding prophage in *E. coli* strains, the nleA-s primer (5′-ATCCAAGAAACAAGAGTCTAA-3′) was designed in the conserved region of *nleA/espI* gene. A 2,068-bp fragment was amplified by PCR using the cif-int-s forward and nleA-s reverse primers (Fig. 1). The nleA-as (5′-TACGAGGATAGTCAGTGAATC-3′) and the espJ-as (5′-G GTTCTCCCAGTCTCAGAG-3′) primers were also designed in the conserved region of the *nleA/espI* and *espJ* genes, respectively. A fragment with an expected size of 2,167 bp was amplified by use of the ybhB-s2 and nleA-as primers (Fig. 1) and a fragment of 1,564 bp by use of the ybhB-s2 and espJ-as primers (Fig. 1).

Western blot analysis. To detect the production and secretion of Cif proteins by EPEC and EHEC strains, protein samples from bacterial culture pellets and supernatants were prepared as described previously (19), separated by electrophoresis (12% Nupage; Invitrogen), and transferred to nitrocellulose membranes (Schleider and Schuell). Membranes were probed with rabbit anti-Cif antibodies (33). After being washed, bound antibodies were visualized with horseradish peroxidase-conjugated secondary antibodies (Zymed) followed by

chemiluminescence detection (LumigGLO; Cell Signaling Technology) and autoradiography (Biomax; Kodak).

Demonstration of the presence of the A/E lesion and assay for the Cif-related CPE. The expression of LEE-encoded genes leading to production of A/E lesions was tested by visualization of filamentous actin accumulation at sites of bacterial adhesion on cultured human epithelial cells. The fluorescent-actin staining (FAS) test was performed with HeLa cells (ATCC CCL-2) according to a protocol described previously (29).

The ability of *cif*-positive *E. coli* strains to induce Cif-related CPE was tested by visualization of enlarged nuclei and the multiplication of focal adhesions and stress fibers on cultured human epithelial cells. CPE assays were performed with HeLa cells as described previously (4).

Nucleotide sequence accession numbers. The sequences encompassing the region from the *lom* family gene to the *ybhB* gene of the indicated three EHEC strains have been deposited in the DDBJ/EMBL/GenBank database under accession numbers AB303062 (strain 12009), AB303061 (strain 11128), and AB303060 (strain 11368).

RESULTS

The Cif-encoding prophage of *E. coli* E22 can produce infectious phages in vitro. The *cif* gene in the reference EPEC strain E22 was previously shown to be located on a lambdoid prophage (19, 23). In order to determine the ability of the Cif-encoding prophage ΦE22 to produce infectious phage particles, we performed phage induction with mitomycin C and plaque assays. High-titer phage stock lysates were prepared, and the presumed ΦE22 phage solution was used for transduction experiments with the *E. coli* K-12 derivative C600 as the recipient strain with a multiplicity of infection of 1. Colonies of transduced *E. coli* C600 bacteria were tested by PCR for the presence of Cif-encoding ΦE22 by use of the *cif*-specific primers cif-int-s/cif-int-as and cif-rbs-s/cif-as (Fig. 2A). One PCR-positive colony was selected and designated *E. coli* TE22/C600. In the TE22/C600 chromosome, ΦE22 was found to be inserted close to the *ybhB* gene in the same site as in the parental strain E22 (data not shown). The phage DNA from *E. coli* TE22/C600 exhibited the same EcoRI and HindIII digestion restriction patterns as that from strain E22 (data not shown). TE22/C600 was induced with mitomycin C. Electron microscopy observation of precipitated lysates of TE22/C600 revealed phage particles with hexagonal heads and long tails, resembling phage lambda in their morphology (14) (Fig. 2B). The head had a diameter of 60 nm, and the mean dimensions of the tail were about 140 nm (length) and 10 and 12 nm (width). We
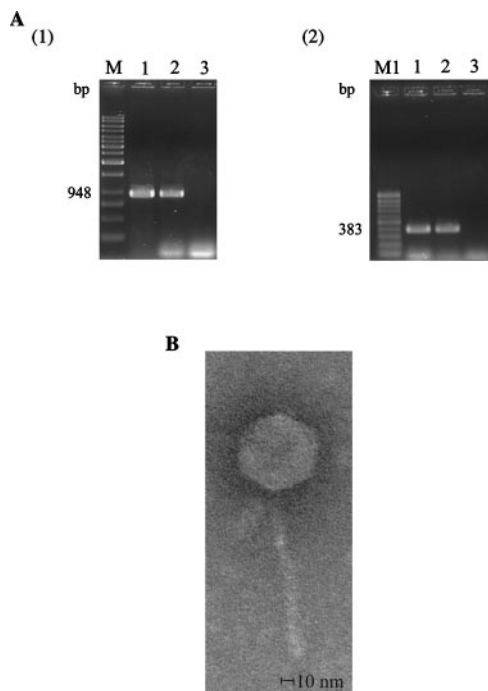
FIG. 2. A Cif-encoding prophage of *E. coli* E22 can produce infectious phages ΦE22 in vitro. (A) Agarose gel electrophoresis of PCR products obtained using primer pair cif-rbs-s and cif-as (1) and primer pair cif-int-s and cif-int-as (2), showing the presence of the *cif* gene in *E. coli* K-12 derivate strain C600 following transduction with Cif-encoding phage ΦE22. Lanes 1, strain *E. coli* C600 after transduction with Cif-encoding phage ΦE22 (designated *E. coli* TE22/C600); lanes 2, parental strain E22; lanes 3, strain C600 (negative control); lane M, GeneRuler 1-kb DNA ladder molecular mass marker (Fermentas GmbH, Germany) (range, 250 bp to 10,000 bp); lane M1, GeneRuler 100-bp DNA ladder molecular mass marker (Fermentas GmbH, Germany) (range, 100 bp to 1,000 bp). (B) Electron micrograph of Cif-encoding bacteriophage ΦE22. Phages were obtained from an overnight culture of EPEC strain E22 induced with mitomycin C. Electron microscopy of phage ΦE22 was performed by Manfred Rohde, GBF, Braunschweig, Germany.

conclude that the Cif-encoding prophage of *E. coli* E22 can produce infectious phages in vitro.

Cif-encoding prophages are strongly associated with the LEE pathogenicity island. Since the Cif effector uses the TTSS encoded on the LEE for injection into host cells (4, 19), we examined the link between the presence of the Cif-encoding prophages and the presence of the LEE island. A collection of 5,049 *E. coli* isolates from human, animal, and environmental origins was screened by PCR for the presence of the *cif* and *eae* genes. Among these 5,049 *E. coli* isolates, 115 isolates were positive for the internal region of the *cif* gene.

All the 115 *cif*-positive *E. coli* isolates harbored the *eae* gene, suggesting that *cif* is strongly associated with the LEE in *E. coli* strains. The *cif*-positive *E. coli* isolates represented 71% (34/48) of the *eae*-positive isolates associated with disease and 89% (81/91) of the *eae*-positive isolates that have not been associated with disease. Altogether, 83% (115/139) of the *eae*-positive isolates harbored the *cif* gene. We also performed intimin typing (29, 42) and found that the *cif*-positive strains were not restricted to a specific intimin variant, since α1, β1, β2/δ, γ2/θ, ε, κ, and η intimins were detected. Likewise, the *cif*-positive strains were not restricted to one of the four main phylogenetic groups of LEE-positive strains A, B1, B2, and E (9), as the *cif*-positive and *eae*-positive strains belonged mainly to phylogenetic group B1 but also to groups A and B2 (Table 2 and data not shown). However, none of the *cif*-positive and *eae*-positive strains tested belonged to group E, in which most of EHEC strains of serotype O157:H7 are concentrated (9).

Cif-encoding phages were integrated between the *ybhC* and *ybhB* genes and have coevolved with housekeeping genes in *E. coli* strains. In strain E22, the Cif-encoding lambdoid prophage was found integrated between the *gal* and *bio* operons, precisely between the *ybhC* and *ybhB* genes. This integration site is the same as that for phage lambda, and the *att* core sequences were also identical to those of phage lambda (15, 34, 40). A lambdoid prophage, Sp3, is also integrated at the same chromosome locus in EHEC O157:H7 strain Sakai, although it does not encode the *cif* gene (13). To test whether the chromosomal insertion site of the Cif-encoding prophages was the same in all of the 115 *cif*-positive *E. coli* isolates, we screened by PCR the physical association between the *cif* and *ybhB* genes by use of the cif-int-s and ybhB-s2 primers (Fig. 1). We obtained PCR products with 72% (83/115) of the *cif*-positive *E. coli* isolates, thus suggesting that most of the Cif-encoding lambdoid prophages were integrated in the same chromosomal region between the *ybhC* and *ybhB* genes (Table 2 and data not shown).

We next examined the genomic sequences of a gene on the *cif*-encoding phage (putative tail fiber protein) and a host housekeeping gene (*aroE*) from seven EHEC and EPEC strains to construct phylogenetic trees. We obtained similar phylogenetic trees for the two genes (Fig. 3). This observation may indicate that Cif-encoding phage and housekeeping genes have undergone a coevolution and suggests an uptake of Cif-encoding phages early in the evolution of EPEC and EHEC.

Cif-encoding prophages do not always encode a functional Cif protein. To determine whether the presence of the internal region of the *cif* gene in *E. coli* strains as studied above correlated with the induction of the Cif-related CPE, infection assays using HeLa cells were performed. Among the 115 *cif*-positive *E. coli* isolates (including the two reference EPEC strains B171-8 and E22), 39 (34%) induced the typical Cif-related CPE. Infected HeLa cells exhibited irreversible growth arrest, assembly of actin stress fibers, and enlargement of the nuclei and cell body, as described previously (19, 33). Notably, the *cif*-positive and CPE-positive strains were almost equally distributed among pathogenic isolates (34%; 11/32) (Table 2) and among nonpathogenic isolates (32%; 26/81) (data not shown).

To further examine the *cif* gene in CPE-positive and CPE-negative strains, the integrity of *cif* gene was investigated in all strains by PCR using a primer pair able to amplify the full length of the gene (19). All the 39 CPE-positive strains and 69 of the 76 CPE-negative strains were positive for the full *cif* gene (Table 2 and data not shown).

To further test the functionality of the *cif* gene, production of Cif protein was examined by Western blotting and TTSS function was tested by FAS assay (4, 19, 33). In the randomly selected 21 *cif*-positive but CPE-negative EPEC and EHEC strains, Cif could not be detected in bacterial lysates, indicating that they were unable to produce Cif (Fig. 4 and Table 2). All

TABLE 2. Phenotypic and genotypic characteristics of 48 EPEC and EHEC strains from human and animal origins (data for the 5,001 *E. coli* isolates from slurry, wastewater, and river water are not shown)

| Strain | Pathotypes (species[a]) | Phenotype | | | Genotype[n] | | | | | | | Phylogenetic group |
		Serotype[b]	Cif-related CPE[c]	Cif production and secretion[d]	*cif* (internal)[e]	*cif* (full)[f]	*eae*	*cif-ybhB* association[g]	*cif-nleA* association[h]	*nleA-ybhB* association[i]	*espJ-ybhB* association[j]	
EF53	EPEC (human)	O111:H-	+	+	+	+	γ2/θ	–	–	–	–	A
EF58	EPEC (human)	O127:H40	+	+	+	+	γ2/θ	–	–	–	–	A
6044/95	EPEC (human)	O118:H5	+	+	+	+	κ	+	–	–	+	A
DEC9a	EHEC (human)	O26:H11	+	+	+	+	β1	+	–	–	+	B1
DEC11a	EPEC (human)	O128:H2	+	+	+	+	β1	+	+	+	–	B1
EF138	EPEC (human)	O128:H(ND)[k]	+	+	+	+	β1	+	+	+	–	B1
E22	EPEC (rabbit)	O103:H2	+	+	+	+	β1	+	+	+	–	B1
RDEC-1	EPEC (rabbit)	O15:H-	+	+	+	+	β1	+	–	–	+	B1
B171-8	EPEC (human)	O111:H-	+	+	+	+	β1	+	+	+	–	B1
C/15333	EPEC (bovine)	O26	+	+	+	+	β1	+	–	–	+	B1
E5	EPEC (bovine)	O26	+	+	+	+	β1	+	–	–	+	B1
E6468/62	EPEC (human)	O86:H34	+	+	+	+	β2/δ	–	–	–	–	B2
CF11201	EPEC (human)	O125:H-	+	+	+	+	η	+	–	–	–	B2
EF70	EPEC (human)	O127:H40	–	+	+	+	γ2/θ	–	–	–	–	A
EF44	EPEC (human)	O127:H40	–	–	+	+	γ2/θ	–	–	–	–	A
DEC8b	EHEC (human)	O111:H8	–	–	+	+	γ2/θ	+	–	–	+	B1
CL37	EHEC (human)	O111:H8	–	–	+	+	γ2/θ	+	–	–	+	B1
95NR1	EHEC (human)	O111:H-	–	–	+	+	γ2/θ	+	–	–	+	B1
11128	EHEC (human)	O111:H-	–	–	+	+	γ2/θ	+	–	–	+	B1
EF39	EPEC (human)	O128:H-	–	–	+	+	γ2/θ	+	–	–	+	B1
DEC12a	EPEC (human)	O111:H2	–	–	+	+	β1	+	–	+	–	B1
H-19	EHEC (human)	O26:H11	–	–	+	+	β1	+	–	–	+	B1
11368	EHEC (human)	O26:H11	–	–	+	+	β1	+	–	–	+	B1
EF21	EPEC (human)	O128:H2	–	–	+	+	β1	–	–	–	+	B1
1390	EPEC (porcine)	O45	–	–	+	+	β1	+	–	–	+	B1
PMK5	EHEC (human)	O103:H2	–	–	+	+	ε	+	+	+	–	B1
12009	EHEC (human)	O103:H2	–	–	+	+	ε	+	+	+	–	B1
EF109	EPEC (human)	O142:H (ND)	–	–	+	+	α1	+	–	–	+	B2
EF105	EPEC (human)	O142:H (ND)	–	–	+	–	α1	–	–	–	–	B2
EF104	EPEC (human)	O142:H (ND)	–	–	+	–	α1	–	–	–	–	B2
E2348/69	EPEC (human)	O127 H6	–	–	+	–	α1	+	–	–	+	B2
DEC1a	EPEC (human)	O55:H6	–	–	+	–	α1	+	–	–	+	B2
DEC2a	EPEC (human)	O55:H6	–	–	+	–	α1	+	–	–	+	B2
DEC10a	EHEC (human)	O26:H11	–	–	+	–	β1	–	–	–	+	B1
EDL 933	EHEC (human)	O157:H7	NA[l]	NA	–	–	γ1	NA	NA	–	–	E[m]
RIMD 0509952	EHEC (human)	O157:H7	NA	NA	–	–	γ1	NA	NA	–	–	E[m]
ECOR37	EPEC (marmoset)	O (ND):H (ND)	NA	NA	–	–	γ1	NA	NA	–	–	E[m]
EDL 931	EHEC (human)	O157:H7	NA	NA	–	–	γ1	NA	NA	–	–	E[m]
DEC5d	EPEC (human)	O55:H7	NA	NA	–	–	γ1	NA	NA	–	–	E[m]
DEC3a	EHEC (human)	O157:H7	NA	NA	–	–	γ1	NA	NA	–	–	E[m]
DEC4a	EHEC? (bovine)	O157:H7	NA	NA	–	–	γ1	NA	NA	–	–	E[m]
EF60	EPEC (human)	O55:H7	NA	NA	–	–	γ1	NA	NA	–	–	E[m]
EF59	EPEC (human)	O55:H7	NA	NA	–	–	γ1	NA	NA	–	–	E[m]
EF28	EPEC (human)	O55:H7	NA	NA	–	–	γ1	NA	NA	–	–	E[m]
EF50	EPEC (human)	O128:H8	NA	NA	–	–	γ2/θ	NA	NA	–	+	B1
EF22	EPEC (human)	O128:H2	NA	NA	–	–	γ2/θ	NA	NA	–	–	A
4795/97	EHEC (human)	O84:H4	NA	NA	–	–	ζ	NA	NA	–	–	B1
7476/96	EPEC (human)	O145:H4	NA	NA	–	–	ι	NA	NA	–	–	B2

[a] Species from which the strain was isolated.

[b] O:H serotypes of *E. coli* isolates. H- indicates the presence of nonmotile strains.

[c] Functionality of Cif was determined by CPE assays on cultured epithelial cells as described previously (4).

[d] The production and secretion of Cif protein was assayed by Western blotting as described previously (33).

[e] The presence of the internal region of *cif* gene was investigated by PCR using cif-int-s and cif-int-as primers as described previously (19).

[f] The presence of the *cif* full gene was investigated by PCR using cif-rbs-s and cif-as primers as described previously (19).

[g] The physical association between conserved regions of the *cif* gene and the *ybhB* gene was screened by PCR using cif-int-s and ybhB-s2 primers. PCR products that ranged from about 3 kb to about 7.5 kb (which correspond to size intervals of PCR products obtained with EPEC O103:H2 strain E22, EPEC O111:H- strain B171-8, EPEC O127:H6 strain E2348/69, EHEC O103:H2 strain 12009, EHEC O111:H- strain 11128, and EHEC O26:H11 strain 11368) were considered to represent positive results.

[h] The physical association between conserved regions of genes *cif* and *nleA/espI* was screened by PCR using cif-int-s and nleA-s primers.

[i] The physical association between conserved regions of genes *nleA/espI* and *ybhB* was screened by PCR using nleA-as and ybhB-s2 primers.

[j] The physical association between conserved regions of the *espJ* gene and the *ybhB* gene was screened by PCR using espJ-as and ybhB-s2 primers.

[k] (ND), the specific serotype was not determined.

[l] NA, not applicable.

[m] Assignment of *E. coli* isolates to phylogenetic group E was performed as described previously (9).

[n] +, presence; –, absence.

but 4 of these 21 strains were FAS positive, thus demonstrating that their TTSS was functional in the conditions used for the assay (data not shown).

To link variations of amino-acid sequences with Cif functionality, predicted sequences of Cif proteins of six EHEC and EPEC strains were aligned using the ClustalW algorithm (Fig. 5). Amino-acid sequences of Cif proteins were found to be highly conserved, but in CPE-negative strains, frameshift mutations or insertion of a partial IS2 led to truncation in the N-terminal and C-terminal regions.

Gene encoding putative tail fiber protein   *aroE* gene
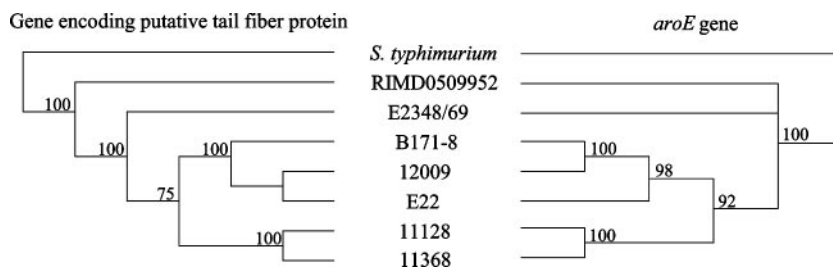

FIG. 3. Phylogenetic analysis and comparison of sequences of the phage-encoded gene that encodes a putative tail fiber protein and the housekeeping gene *aroE* from seven EPEC and EHEC strains (EPEC O103:H2 strain E22, EPEC O111:H- strain B171-8, EPEC O127:H6 strain E2348/69, EHEC O103:H2 strain 12009, EHEC O111:H- strain 11128, EHEC O26:H11 strain 11368, and EHEC O157:H7 strain RIMD 0509952). The phylogenetic tree of genes that encode putative tail fiber proteins was rooted with the gene (locus tag STM2588) that encodes a tail fiber-like protein in *Salmonella enterica* serovar Typhimurium (*S. typhimurium*) strain LT2 as the outgroup (GenBank accession number AE006468). In serovar Typhimurium strain LT2, this gene is encoded by Gifsy-1 lambdoid prophage (10). The phylogenetic tree of the housekeeping gene *aroE*, encoding the shikimate 5-dehydrogenase, was rooted using the *aroE* gene (locus tag STM1359) of serovar Typhimurium strain LT2 as the outgroup. Numbers in the dendrogram indicate bootstrap values.

Mosaic organization of the passenger region of the Cif-encoding prophages. The genetic organizations of seven prophages integrated in the same chromosomal region in EPEC and EHEC strains (EPEC O103:H2 strain E22, EPEC O111:H- strain B171-8, EPEC O127:H6 strain E2348/69, EHEC O103:H2 strain 12009, EHEC O111:H- strain 11128, EHEC O26:H11 strain 11368, and EHEC O157:H7 strain RIMD 0509952) were compared. These seven prophages are all lambda-like phages resembling each other. The regions downstream of a putative tail fiber gene, the passenger regions, also contained similar sets of TTSS effector genes but exhibited significant variations such as insertion or deletion of a specific effector gene(s) and insertion of various insertion sequence (IS) elements and a part of integrase
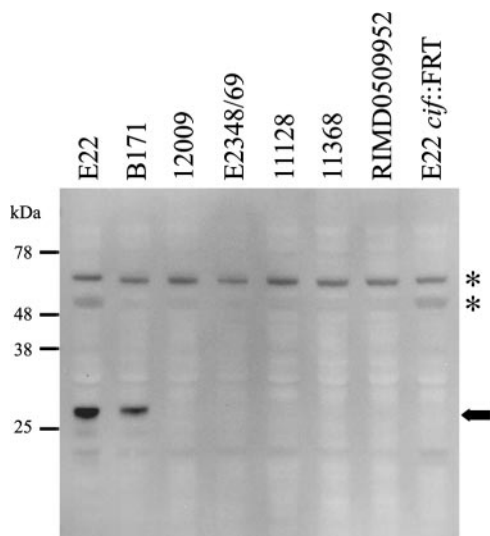

FIG. 4. Western blot analysis of the production of Cif proteins by seven EPEC and EHEC reference strains (EPEC O103:H2 strain E22, EPEC O111:H- strain B171-8, EHEC O103:H2 strain 12009, EPEC O127:H6 strain E2348/69, EHEC O111:H- strain 11128, EHEC O26:H11 strain 11368, and EHEC O157:H7 strain RIMD 0509952). E22 mutated with respect to the *cif* gene (E22 *cif*::FRT) (19) was used as a negative control. Bacterial cells were lysed and analyzed by Western blotting for the production of Cif proteins (indicated by the arrow). The asterisks indicate nonspecific bands. Molecular masses are indicated on the left.

gene. These prophages were classified into three types according to the genetic organizations and compositions of TTSS effector genes of this variable region (Fig. 6). Type 1 phages contained *nleB*, truncated *nleC*, *nleH*, *cif*, *nleG*, and *nleA/espI* genes in succession. This organization was found in the two reference strains that encode functional Cif proteins (strains E22 and B171-8) and in strain 12009, which does not encode a functional Cif protein. Type 2 phages contained *nleH*, *cif*, *nleG* (deleted in strain E2348/69 and truncated in strain 11128), and *espJ* genes in succession. This organization was found to be associated with the three strains that encode nonfunctional Cif proteins (strains 11128, 11368, and E2348/69). Interestingly, *cif* was always found associated with the *nleH* gene, and neighboring sequences, including the partial IS*604* sequence, were conserved. Furthermore, a fragment of integrase gene was always found immediately downstream *cif* (except in strain E2348/69). Altogether, these results suggest that both insertion sequences and the integrase have played a pivotal role in the acquisition of *cif*. Sp3, a lambdoid prophage which is integrated in the same chromosomal region of EHEC O157:H7 strain Sakai but does not harbor the *cif* gene, was classified as type 3, where *nleB*, *nleC*, *nleH*, and *nleD* genes are successively encoded.

To test whether functional Cif proteins were preferably encoded by prophages showing type 1 organization, we screened by PCR the physical association between the *cif* and *nleA/EspI* genes by use of the cif-int-s and nleA-s primers (Fig. 1). However, this association was found in a mere 18% (7/39) EHEC and EPEC strains that had a functional *cif* gene. Furthermore, this association was also detected in the strains that did not have a functional *cif* gene (Table 2 and data not shown). These results indicated that Cif functionality is not associated with the organization of the Cif-encoding prophage but is the result of more recent event(s). To determine the distribution of prophage types in the 115 *cif*-positive EPEC and EHEC strains, we screened by PCR the physical associations between the *nleA/espI* and *ybhB* genes and between the *espJ* and *ybhB* genes. Physical association between the *nleA/espI* and *ybhB* genes was expected to be found only in type 1 prophages, whereas association between the *espJ* and *ybhB* genes was found in type 2 prophages. The most frequent prophage type was type 2 (detected in 64% [74/115] of the *cif*-positive EPEC and EHEC strains). Type 1 prophages were detected only in

A

```
E22        1 MKDITLPPPTSASCLTGAISVNTEAVLSPMQHTSALHVRDFASLCSQNLK  50
B171-8     1 MKDITLPLPTSASCLTGAISVNTEAVLSPMQHTSALHVRDFASLCSQNLK  50
12009      1                       MQHTSALHVRDFASLCSQNLK  21
11128      1                       MQHTSALHVRDFASLCSQNLK  21
11368      1 MKDITLPPPTSASCLTGAISVNTEAVLSPMQHTSALHVRDFASLCSQNLK  50
E2348/69   1 MKDITLPPPTSASCLTGAISVNTEAVLSPMQHTSALHVRDFASLCSQNLK  50
                                  *********************

E22       51 ANVLLNSDDHEVPIHQKNPAAIMQNIDSNIKQMATDWGMSIEEVEVIIGR 100
B171-8    51 ANVLLNSDDHEVPIHQKNPAAIMQNIDSNIKQMATDWGMSIEEVEVIIGR 100
12009     22 ANVLLNSDDHEVPIHQKNPAAIMQNIDSNIKQMATDWGMSIEEVEVIIGR  71
11128     22 ANVLLNSDDHEVPIHQKNPAAIMQNIDSNIKQMATDWGMSIEEVEVIIGR  71
11368     51 ANVLLNSDDHEVPIHQKNPAAIMQNIDSNIKQMATDWGMSIEEVEVIIGR 100
E2348/69  51 ANVLLNSDDHEVPIHQKNPAAIMQNIDSNIKQMATDWGMSIEEVEVIIGR 100
             *************************************************

E22      101 EKGIVEPSCGVTANAIMKLFLDKDGFSYCFENEQTLSLEQLQERLSCMPE 150
B171-8   101 EKGIVEPSCGVTANAIMKLFLDKDGFSYCFENEQTLSLEQLQERLSCMPE 150
12009     72 EKGIVEPSCGVTANAIMKLFLDKDGFSYCFENEQTLSLEQLQERLSCMPE 121
11128     72 EKGIVEPSCGVTANAIMKLFLDKDGFSYCFENEQTLSLEQLQERLSCMPE 121
11368    101 EKGIVEPSCGVTANAIMKLFLDKDGFSYCFENEQTLSLEQLQERLSCMPE 150
E2348/69 101 EKGIVEPSCGVTANAIMKLFLDKDGFSYCFENEQTLSLEQLQERLSCMPE 150
             *************************************************

E22      151 CKSFVLRVNDGALGHAYIVDIPKGENSCRPAFLYQSDLGEGVTRKLRFED 200
B171-8   151 CKSFVLRVNDGALGHAYIVDIPKGENSCRPAFLYQSDLGEGVTRKLRFED 200
12009    122 CKSFVLRVNDGALGHAYIVDIPKGENSCRPAFLYQSDLGEGVTRKLRFED 171
11128    122 CKSFVLRVNDGALGHAYIVDIPKGENSCRPAFLYQSDLGEGVTRKLRFED 171
11368    151 CKSFVLRVNDGALGHAYIVDIPKGENSCRPAFLYQSDLGEGVTRKLRFED 200
E2348/69 151 CKSFVLRVNDGALGHAYIVDIPKGENSCRPAFLYQSDLGEGVTRKLRFED 200
             *************************************************

E22      201 WMTHKALTPILLDDICNYFSCMSQNKTDLEQIATLFDIDGNVKMLRKENI 250
B171-8   201 WMTHKALTPILLDDICNYFSCMSQNKTDLEQIATLFDIDGNVKMLRKENI 250
12009    172 WMTHKALTPILLDDICNYFSCMSQNKTDLEQIATLFDIDGNVKMLRKENI 221
11128    172 WMTHKALTPILLDDICNYFSCMSQNKTDLEQIATLFDIDGNVKMLRKENI 221
11368    201 WMTHKALTPILLDDICNYFSCMSQNKTDLEQIATLFDIDGNVKMLRKENI 250
E2348/69 201 WMTHKALTPILLDDICNY*                                 218
             ******************

E22      251 QYQKHDNFSFQLFEYDTDNIEKNIEIIKSLCS* 282
B171-8   251 QYQKHDNFSFQLFEYDTDNIEKNIEIIKSLCS* 282
12009    222 QYQKHDNFSFQLFEYDTDNIEKKH*         245
11128    222 QYQKHDNFSFQLFEYDTDNIEKTLR*        246
11368    251 QYQKHDNFSFQLFEYDTDNIEKTLR*        275
```

B

(1)

```
E22        1 ATGAAAGACATTACCCTTCCCCCCCC----GACGTCCGCGTCCTGTCTGA  46
B171-8     1 ATGAAAGACATTACCCTTCCCCTCCC----GACGTCCGCGTCCTGTCTGA  46
12009      1 ATGAAAGACATTACCCTTCCCCCCCCC---GACGTCCGCGTCCTGTCTGA  47
11128      1 ATGAAAGACATTACCCTTCCCCCCCCCCCCGACGTCCGCGTCCTGTCTGA  50
11368      1 ATGAAAGACATTACCCTTCCCCCCCC----GACGTCCGCGTCCTGTCTGA  46
E2348/69   1 ATGAAAGACATTACCCTTCCCCCCCC----GACGTCCGCGTCCTGTCTGA  46
             ********************** ***    *******************

E22       47 CAGGGGCCATATCTGTAAATACTGAAGCTGTATTATCTCCCATGCAACAC  96
B171-8    47 CAGGGGCCATATCTGTAAATACTGAAGCTGTATTATCTCCCATGCAACAC  96
12009     48 CAGGGGCCATATCTGTAAATACTGAAGCTGTATTATCTCCCATGCAACAC  97
11128     51 CAGGGGCCATATCTGTAAATACTGAAGCTGTATTATCTCCCATGCAACAC 100
11368     47 CAGGGGCCATATCTGTAAATACTGAAGCTGTATTATCTCCCATGCAACAC  96
E2348/69  47 CAGGGGCCATATCTGTAAATACTGAAGCTGTATTATCTCCCATGCAACAC  96
             *************************************************
```

(2)

```
E22      646 TGTAATTACTTCTCCTG 662
B171-8   646 TGTAATTACTTCTCCTG 662
12009    647 TGTAATTACTTCTCCTG 663
11128    650 TGTAATTACTTCTCCTG 666
11368    646 TGTAATTACTTCTCCTG 662
E2348/69 646 TGTAATTACTAGACTGG 662
             **********  *  *
```

(3)

```
E22      810 TGAAAAAAA-CATTGAGATAATAAAATCACTATGTAGTTAG 849
B171-8   810 TGAAAAAAA-CATTGAGATAATAAAATCACTATGTAGTTAG 849
12009    811 TGAAAAAAAACATTGAGATAATAAAATCACTATGTAGTTAG 851
11128    814 TGAAAAAA--CATTGAGATAATAAAATCACTATGTAGTTAG 852
11368    810 TGAAAAAA--CATTGAGATAATAAAATCACTATGTAGTTAG 848
             ********  **************************
```
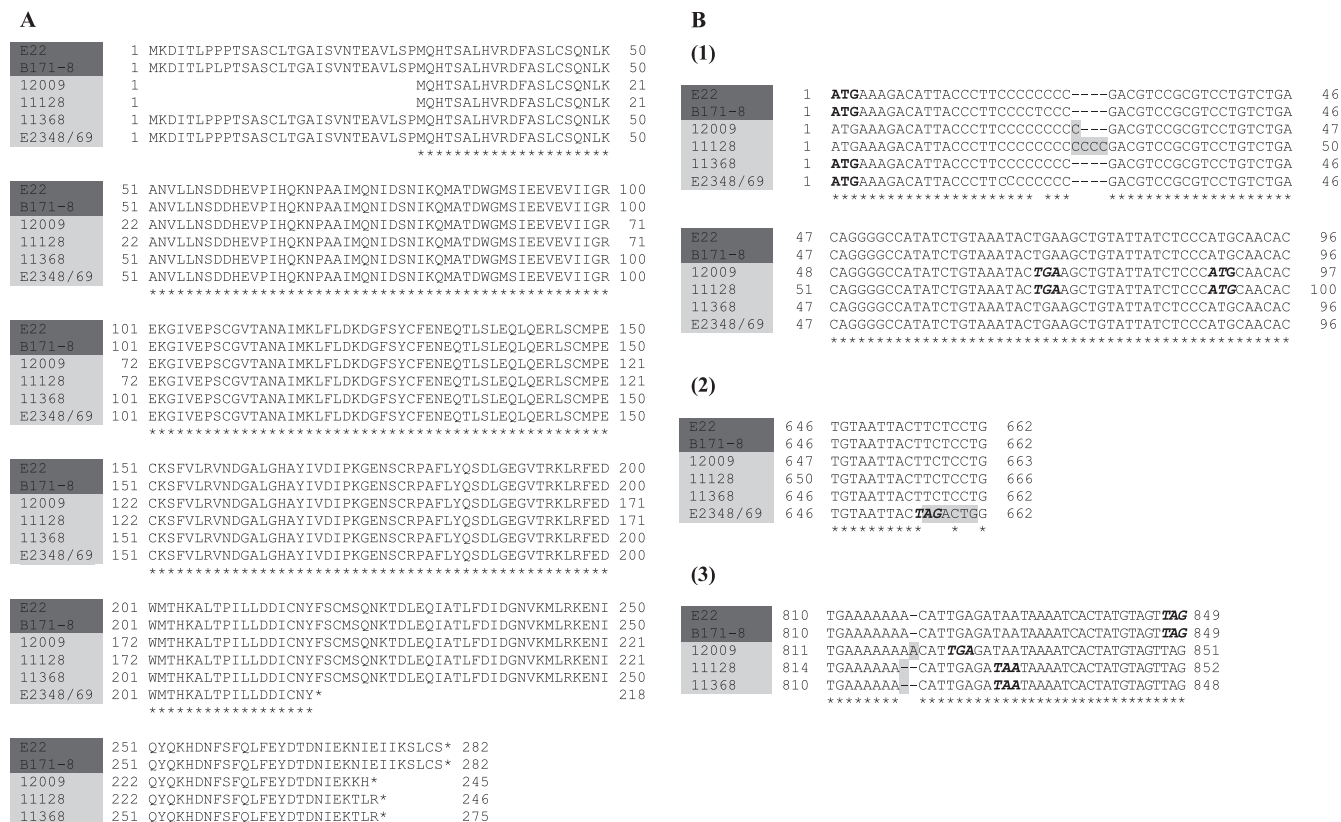
FIG. 5. The functionality of Cif proteins correlates with the presence of a *cif* full gene. (A) Alignments of the predicted amino-acid sequences of Cif proteins of six EPEC and EHEC reference strains (EPEC O103:H2 strain E22, EPEC O111:H- strain B171-8, EHEC O103:H2 strain 12009, EHEC O111:H- strain 11128, EHEC O26:H11 strain 11368, and EPEC O127:H6 strain E2348/69). Stars indicate conserved amino acids. The presence of functional Cif protein was tested by examining the CPE responses of the challenged strains upon interaction with HeLa cells. CPE-positive strains are indicated by black characters on a dark-gray background, and CPE-negative strains are indicated by black characters on a light-gray background. In each CPE-negative strain, Cif is truncated in the N- and/or C-terminal region. (B) Mutations in the *cif* gene nucleotide sequences. Identical residues are indicated by stars. Mutations that impact the amino-acid sequence of Cif proteins are indicated by black characters in light-gray boxes in the sequences. Stop codons are represented by bold italic characters. Start codons are represented by bold roman characters. Numbers indicate the positions of nucleotides from the start codon of the *cif* gene of EPEC strain E22. The presence of functional Cif protein was tested by examination of the CPE responses of the challenged strains upon interaction with HeLa cells. CPE-positive strains are indicated by black characters on a dark-gray background, and CPE-negative strains are indicated by black characters on a light-gray background. In each CPE-negative strain, the *cif* gene is mutated in the N- and/or C-terminal sequence. (Panel 1) In strains 12009 and 11128, addition of one or four nucleotides C has occurred after nucleotide 26, and the genes are interrupted at nucleotides 72 and 69, respectively, explaining the deletion of the first 29 amino acids of Cif protein in these two strains. (Panel 2) The *cif* gene in strain E2348/69 is interrupted at nucleotide 654 by a sequence identical to that of the terminal inverted repeat of IS2 (GTCAGA). (Panel 3) In strain 12009, addition of one nucleotide (an A nucleotide) has occurred after nucleotide 819, and the gene is interrupted at nucleotide 823. In strains 11128 and 11368, deletion of one nucleotide (an A nucleotide) has occurred after nucleotides 821 and 817, respectively, and the genes are interrupted at nucleotides 829 and 825, respectively. These mutations explain the deletion of the C-terminal part of the Cif protein in these three strains.

10% (11/115) of the *cif*-positive strains investigated (Table 2 and data not shown). Thirty *cif*-positive EPEC and EHEC strains (26% of the investigated strains) did not show any physical associations between the *nleA*/*espI* and *ybhB* genes or between the *espJ* and *ybhB* genes, suggesting that other types of genetic organization could be found in the passenger regions of these *cif*-positive EPEC and EHEC strains.

DISCUSSION

We observed that the Cif-encoding prophage isolated from EPEC strain E22 was able to infect and lysogenize *E. coli* K-12 derivate C600 strain in vitro, suggesting that the *cif* gene may have been widely spread by phage conversion within the natural population of *E. coli*. We investigated the distribution of

cif in a large collection of 5,049 natural *E. coli* strains of human, animal, and environmental origins. Within this collection, 115 *E. coli* isolates from different origins, countries, and isolation dates were positive by PCR for the internal region of the *cif* gene, confirming that it has been widely spread within the natural population of *E. coli*, presumably by lysogenic conversion. It can be speculated that populations of *E. coli* from environmental and animal origins may be natural reservoirs of Cif-encoding phages and might play an important role in the emergence of new pathogenic *E. coli* strains.

Interestingly, all the *cif*-positive *E. coli* isolates also harbored the *eae* gene, indicating that Cif-encoding prophages are tightly associated with the LEE pathogenicity island. This result suggests that Cif-encoding prophages and the LEE islands have
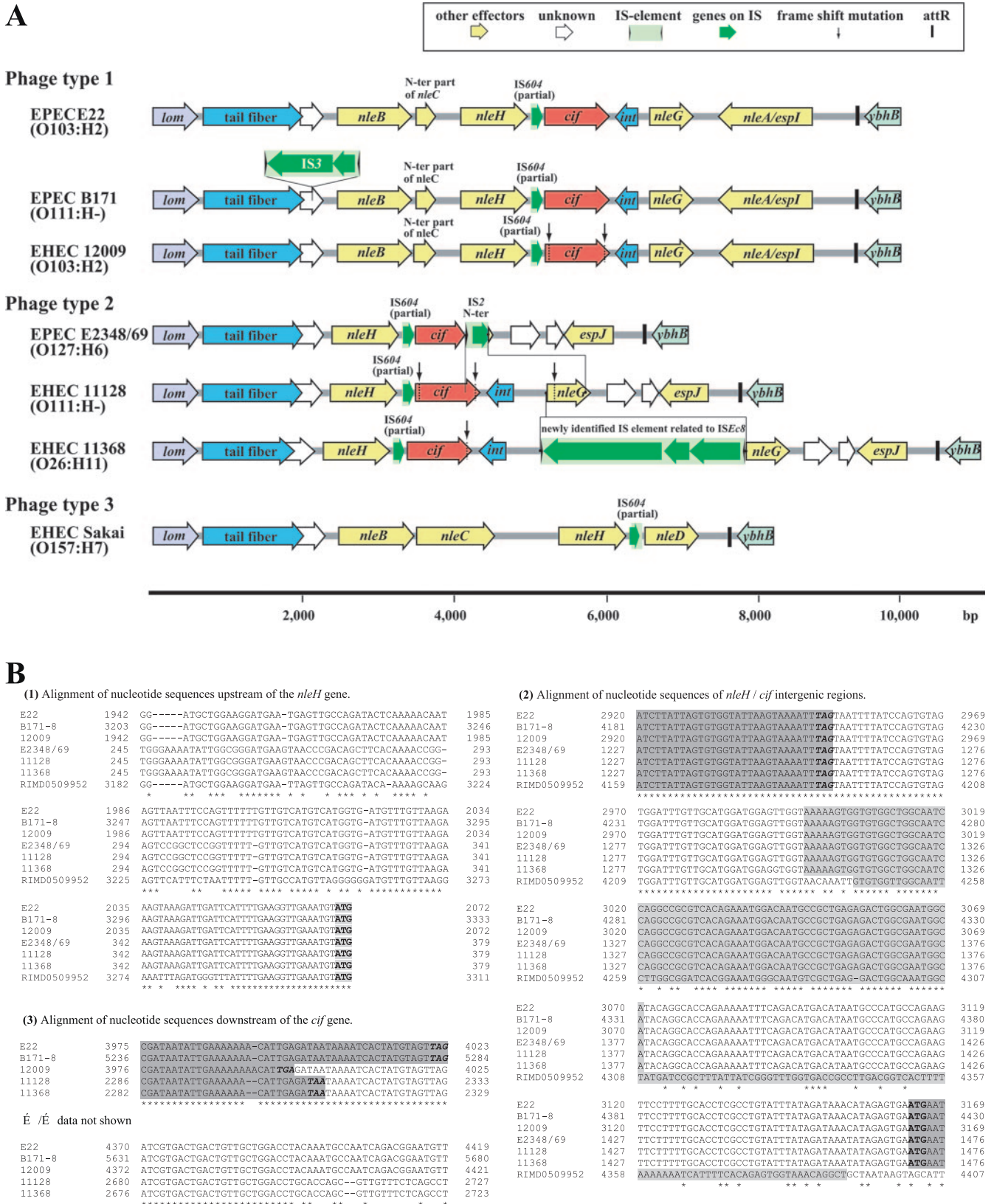
FIG. 6. Genomic comparisons of Cif-encoding prophages or similar prophages integrated in the same chromosomal region of seven EPEC and EHEC strains reveal a mosaic organization of their passenger regions encoding non-LEE-encoded effectors. (A) Schematic representation of the organization of Cif-encoding or similar prophages after the *lom* gene in the seven EPEC and EHEC reference strains (EPEC O103:H2 strain E22, EPEC O111:H- strain B171-8, EHEC O103:H2 strain 12009, EPEC O127:H6 strain E2348/69, EHEC O111:H- strain 11128, EHEC O26:H11

been positively coselected in *E. coli* strains. Our results suggest that TTSS effectors such as Cif can be added by horizontal transfer in a plug-and-play manner once the TTSS and its appropriate regulation are acquired. Tobe et al. (35) proposed that the major function of lambdoid prophages in EHEC is to carry TTSS effectors and that the type III secretion in *E. coli* may be connected to a vast phage "metagenome." This ability of phage-encoded TTSS effectors to "plug in" the preexisting TTSS is not limited to *E. coli* species and has also been described as present in *Salmonella enterica* serovar Typhimurium: lambda prophages Gifsy-1, Gifsy-2, and Gifsy-3 encode TTSS effectors GogB, SseI, and SspH1, respectively, which can use the preexisting TTSS to gain access directly to the target cell (8). Such phage-encoded effectors could confer selective advantages over competing bacteria, especially in environmental niches that are quickly changing. Phage conversion can provide the necessary genetic material on a time scale that allows bacteria to quickly exploit ecological opportunities, as any combination that improved fitness in the host or that allowed colonization of new niches would be positively selected. Thus, Cif-encoding phages could play a key role in the short-term adaptation processes of LEE-positive *E. coli* and the *cif* gene could be defined as a "moron" that is not required for the phage life cycle but rather serves as a fitness-improving factor by contributing to the survival of the bacterial host (3).

We showed that the phylogenetic tree generated by the alignment of phage-encoded sequences extensively overlaps that built up by the *aroE* gene sequences, suggesting that the Cif-encoding prophages could have been acquired early in the evolution of EPEC and EHEC. However, the observation that Cif-encoding phages have been found in all EPEC and EHEC phylogenetic groups but not in group E containing most of the O157:H7 EHEC strains (9) suggests that interaction between the *cif* gene and the genomic background of O157:H7 EHEC strains might prevent integration and/or retention of Cif-encoding prophages. One could also speculate that the Cif-encoding prophage of O157:H7 EHEC has been forced by selective pressure to lose *cif* but conserves other phage-related determinants. Such remodeling in EHEC strains of group E has already been reported for other virulence-associated genes such as *efa1/lifA* located in a pathogenicity island termed O122 in strain EDL 933 and SpLE3 in strain Sakai. *efa1/lifA* has been

detected in EHEC strains belonging to serogroups other than O157 and in EPEC strains, while EHEC O157 strains possess only a fragment of this gene corresponding to the first 1,300 bp of the 5′ region (22). Nevertheless, in EHEC O157 strains, the function of the *efa1/lifA* gene product might be complemented by ToxB, a homologous protein that is encoded by the pO157 virulence plasmid (38). Similarly, the function of Cif protein might be complemented by another unknown protein in O157:H7 EHEC strains. Strikingly, none of the *cif*-positive EHEC strains induced the Cif-related cytopathic effect under the conditions used for the assay. Alignment of predicted amino-acid sequences of Cif proteins of the six EHEC and EPEC strains for which sequences are available indicates that, in the strains that did not induce the CPE, frameshift mutations have led to truncation in the N-terminal and/or C-terminal regions of the protein. Since the N terminus of Cif is required for secretion and translocation (4) and since its C terminus may have a role in protein stability and/or folding (4, 19), these truncations explain the inability of *cif*-mutated strains to produce a Cif-related CPE on epithelial cells. Our results showed that the presence of a TTSS effector as tested by PCR does not fully correlate with functionality, highlighting the need of functional assays. In prokaryotes, the prevalence of pseudogenes (disabled copies of genes characterizable by disruptions of their reading frames due to frameshifts and premature stop codons) has been estimated to be low, representing between 1 to 5% of the genome (with a few well-known exceptions such as *Mycobacterium leprae*) (17). Generation of pseudogenes might have played pivotal roles in the adaptive microevolution of pathogens such as *Yersinia pestis* (36) or enhanced the virulence and antimicrobial resistance of *Salmonella enterica* serovar Choleraesuis (6). Therefore, the efforts to elucidate the biological roles or significances of these pseudogenes, particularly in pathogens, might be vital not only to assess their pathogenicities and tropisms but also to better understand their evolution pathways. To our knowledge, this is the first report of work in which the presence and functionality of a TTSS effector-encoding gene were systematically investigated using a large collection of natural *E. coli* isolates.

Remarkably, most Cif-encoding prophages examined in this study were found to be integrated in the same chromosomal region between the *ybhC* and *ybhB* genes, which contrasts with

---

strain 11368, and EHEC O157:H7 strain RIMD 0509952). A careful comparison of the sequences of strains 11128 and 11368 suggests that, in strain 11128, an IS element was once inserted in the upstream region of *nleG* but has since been deleted by homologous recombination between two short repeated sequences created by the IS insertion (data not shown). (B) Alignment of the nucleotide sequences upstream of the *nleH* gene between the *nleH* and *cif* genes and downstream of the *cif* gene. Numbers indicate the positions of nucleotides after the stop codon of the gene encoding a putative tail fiber protein. The nucleotide sequences of *nleH* and *cif* genes are indicated by black characters on a light-gray background and by black characters on a dark-gray background, respectively. Stop codons are represented by bold italic characters, and start codons are represented by bold roman characters. Stars indicate the presence of identical nucleotides. (Panel 1) Alignment of nucleotide sequences upstream of the *nleH* gene. Nucleotide sequences of the *nleH* genes are indicated by black characters on a light-gray background. Nucleotide sequences from nucleotide 2005 to nucleotide 2069 (numbering from those of the E22 strain) are conserved except in RIMD 0509952, in which nucleotide substitutions or additions have occurred. The nucleotides sequences upstream of this region in strains E22, B171-8, and 12009 are identical, and those in strains E2348/69, 11128, and 11368 are identical. (Panel 2) Alignment of nucleotide sequences of the *nleH*/*cif* intergenic regions. Nucleotide sequences of the *nleH* and *cif* genes (when present) are indicated by black characters on a light-gray background and by black characters on a dark-gray background, respectively. The IS*604*-derived sequences are also indicated by black characters on a light-gray background. In *cif*-positive strains, nucleotide sequences are conserved between the *nleH* and *cif* genes. In strain 11368, a substitution of nucleotide 1299 has occurred. (Panel 3) Alignment of nucleotide sequences downstream of the *cif* gene. Nucleotide sequences of the *cif* genes are indicated by black characters on a dark-gray background. Nucleotide sequences are conserved from the stop codons of the *cif* gene to nucleotide 4394 (numbering based on the E22 sequence) except in E2348/69 (data not shown).

previous results showing that Stx-encoding lambdoid prophages can be found integrated in various chromosome regions (28, 30). Cif-encoding prophages harbored a variable passenger region that encodes several non-LEE-encoded effectors: the *nleB*, *nleC*, *nleH*, *cif*, *nleG*, *espJ*, and *nleA/espI* genes. Hence, the persistence of a nonfunctional Cif protein could be explained by coselection with these other prophage-encoded morons that may act as fitness-improving factors. However, this passenger compartment of morons shows a mosaic structure and contains many insertion sequence elements, parts of integrase genes, and truncated forms of not only *cif* but also *nleC* and *nleG*. This could be significant from the perspective of evolution of phage-encoded type III-secreted effectors in *E. coli*. The presence of IS elements and integrase gene fragments suggests that ISs and integrases have played a pivotal role in the creation of the passenger region of the Cif-encoding prophages. Accumulation of TTSS effector genes downstream of phage tail fiber genes could result either from recombination events between several lambda-like prophages carrying different sets of TTSS effector genes that have been integrated in the same chromosomal region or from integration in the lambda-like prophage of TTSS effector genes encoded elsewhere in the chromosome by the ISs or integrases. As homologs of *nleG* and *nleA/espI* genes that are encoded downstream of *cif* in the Cif-encoding prophage are also found in Sp9, a lambdoid prophage of EHEC O157:H7 strain Sakai (13), one could speculate that at least two prophages might have been integrated in the same chromosomal region in tandem and that deletion of one of the two prophage genomes might have led to the creation of the passenger region of the Cif-encoding prophage.

Our results may also provide support for the idea of the evolutionary mechanism proposed by Stavrinides et al. (32). The authors have identified numerous mosaic or truncated type III-secreted effectors in eight different genera, including plant pathogens, animal pathogens, and mutualists, and proposed a evolutionary mechanism, called "terminal reassortment," in which the termini of type III-secreted effectors reassort with other genetic information to create new chimeric proteins. In this evolutionary process, phages may serve as mixing vessels for type III-secreted effectors or may provide the genetic variation and homologous sequence for recombination. Unfortunately, no large-scale study of functionality of phage-encoded effectors is available to confirm whether this shuffling terminal reassortment mechanism drives the evolution of phage-encoded type III-secreted effectors in *E. coli* and plays a role in the ongoing arms race between pathogen and host.

ACKNOWLEDGMENTS

REFERENCES

1. **Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman.** 1990. Basic local alignment search tool. J. Mol. Biol. **215:**403–410.
2. **Bachmann, B.** 1996. Derivations and genotypes of some mutant derivatives of *Escherichia coli* K12, p. 2460–2488. *In* F. C. Neidhardt, R. Curtiss III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella*: cellular and molecular biology. ASM Press, Washington, DC.
3. **Brüssow, H., C. Canchaya, and W. D. Hardt.** 2004. Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. Microbiol. Mol. Biol. Rev. **68:**560–602.
4. **Charpentier, X., and E. Oswald.** 2004. Identification of the secretion and translocation domain of the enteropathogenic and enterohemorrhagic *Escherichia coli* effector Cif, using TEM-1 β-lactamase as a new fluorescence-based reporter. J. Bacteriol. **186:**5486–5495.
5. **China, B., V. Pirson, and J. Mainil.** 1996. Typing of bovine attaching and effacing *Escherichia coli* by multiplex in vitro amplification of virulence-associated genes. Appl. Environ. Microbiol. **62:**3462–3465.
6. **Chiu, C. H., P. Tang, C. Chu, S. Hu, Q. Bao, J. Yu, Y. Y. Chou, H. S. Wang, and Y. S. Lee.** 2005. The genome sequence of *Salmonella enterica* serovar Choleraesuis, a highly invasive and resistant zoonotic pathogen. Nucleic Acids Res. **33:**1690–1698.
7. **Clermont, O., S. Bonacorsi, and E. Bingen.** 2000. Rapid and simple determination of the *Escherichia coli* phylogenetic group. Appl. Environ. Microbiol. **66:**4555–4558.
8. **Ehrbar, K., and W. D. Hardt.** 2005. Bacteriophage-encoded type III effectors in *Salmonella enterica* subspecies 1 serovar Typhimurium. Infect. Genet. Evol. **5:**1–9.
9. **Escobar-Páramo, P., O. Clermont, A. B. Blanc-Potard, H. Bui, C. Le Bouguenec, 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.
10. **Figueroa-Bossi, N., E. Coissac, P. Netter, and L. Bossi.** 1997. Unsuspected prophage-like elements in *Salmonella typhimurium*. Mol. Microbiol. **25:**161–173.
11. **Garmendia, J., G. Frankel, and V. F. Crepin.** 2005. Enteropathogenic and enterohemorrhagic *Escherichia coli* infections: translocation, translocation, translocation. Infect. Immun. **73:**2573–2585.
12. **Girón, J. A., A. S. Ho, and G. K. Schoolnik.** 1991. An inducible bundle-forming pilus of enteropathogenic *Escherichia coli*. Science **254:**710–713.
13. **Hayashi, T., K. Makino, M. Ohnishi, K. Kurokawa, K. Ishii, K. Yokoyama, C. G. Han, E. Ohtsubo, K. Nakayama, T. Murata, M. Tanaka, T. Tobe, T. Iida, H. Takami, T. Honda, C. Sasakawa, N. Ogasawara, T. Yasunaga, S. Kuhara, T. Shiba, M. Hattori, and H. Shinagawa.** 2001. Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12. DNA Res. **8:**11–22.
14. **Herold, S., H. Karch, and H. Schmidt.** 2004. Shiga toxin-encoding bacteriophages—genomes in motion. Int. J. Med. Microbiol. **294:**115–121.
15. **Hoess, R. H., and A. Landy.** 1978. Structure of the lambda *att* sites generated by int-dependent deletions. Proc. Natl. Acad. Sci. USA **75:**5437–5441.
16. **Kaper, J. B., J. P. Nataro, and H. L. Mobley.** 2004. Pathogenic *Escherichia coli*. Nat. Rev. Microbiol. **2:**123–140.
17. **Liu, Y., P. M. Harrison, V. Kunin, and M. Gerstein.** 2004. Comprehensive analysis of pseudogenes in prokaryotes: widespread gene decay and failure of putative horizontally transferred genes. Genome Biol. **5:**R64.
18. **Loukiadis, E., M. Kerouredan, L. Beutin, E. Oswald, and H. Brugere.** 2006. Characterization of Shiga toxin gene (*stx*)-positive and intimin gene (*eae*)-positive *Escherichia coli* isolates from wastewater of slaughterhouses in France. Appl. Environ. Microbiol. **72:**3245–3251.
19. **Marchès, O., T. N. Ledger, M. Boury, M. Ohara, X. Tu, F. Goffaux, J. Mainil, I. Rosenshine, M. Sugai, J. De Rycke, and E. Oswald.** 2003. Enteropathogenic and enterohaemorrhagic *Escherichia coli* deliver a novel effector called Cif, which blocks cell cycle G$_2$/M transition. Mol. Microbiol. **50:**1553–1567.
20. **Morabito, S., H. Karch, H. Schmidt, F. Minelli, P. Mariani-Kurkdjian, F. Allerberger, K. A. Bettelheim, and A. Caprioli.** 1999. Molecular characterisation of verocytotoxin-producing *Escherichia coli* of serogroup O111 from different countries. J. Med. Microbiol. **48:**891–896.
21. **Morabito, S., R. Tozzoli, A. Caprioli, H. Karch, and A. Carattoli.** 2002. Detection and characterization of class 1 integrons in enterohemorrhagic *Escherichia coli*. Microb. Drug Resist. **8:**85–91.
22. **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.
23. **Muniesa, M., J. E. Blanco, M. De Simon, R. Serra-Moreno, A. R. Blanch, and J. Jofre.** 2004. Diversity of stx2 converting bacteriophages induced from Shiga-toxin-producing *Escherichia coli* strains isolated from cattle. Microbiology **150:**2959–2971.
24. **Muniesa, M., and J. Jofre.** 2004. Abundance in sewage of bacteriophages infecting *Escherichia coli* O157:H7. Methods Mol. Biol. **268:**79–88.
25. **Nougayrède, J. P., F. Taieb, J. De Rycke, and E. Oswald.** 2005. Cyclomodu-

lins: bacterial effectors that modulate the eukaryotic cell cycle. Trends Microbiol. **13:**103–110.

26. **Ochman, H., and R. K. Selander.** 1984. Standard reference strains of *Escherichia coli* from natural populations. J. Bacteriol. **157:**690–693.
27. **Ohnishi, M., K. Kurokawa, and T. Hayashi.** 2001. Diversification of *Escherichia coli* genomes: are bacteriophages the major contributors? Trends Microbiol. **9:**481–485.
28. **Ohnishi, M., J. Terajima, K. Kurokawa, K. Nakayama, T. Murata, K. Tamura, Y. Ogura, H. Watanabe, and T. Hayashi.** 2002. Genomic diversity of enterohemorrhagic *Escherichia coli* O157 revealed by whole genome PCR scanning. Proc. Natl. Acad. Sci. USA **99:**17043–17048.
29. **Oswald, E., H. Schmidt, S. Morabito, H. Karch, O. Marchès, and A. Caprioli.** 2000. Typing of intimin genes in human and animal enterohemorrhagic and enteropathogenic *Escherichia coli*: characterization of a new intimin variant. Infect. Immun. **68:**64–71.
30. **Recktenwald, J., and H. Schmidt.** 2002. The nucleotide sequence of Shiga toxin (Stx) 2e-encoding phage φP27 is not related to other Stx phage genomes, but the modular genetic structure is conserved. Infect. Immun. **70:**1896–1908.
31. **Schmidt, H., M. Bielaszewska, and H. Karch.** 1999. Transduction of enteric *Escherichia coli* isolates with a derivative of Shiga toxin 2-encoding bacteriophage φ3538 isolated from *Escherichia coli* O157:H7. Appl. Environ Microbiol. **65:**3855–3861.
32. **Stavrinides, J., W. Ma, and D. S. Guttman.** 2006. Terminal reassortment drives the quantum evolution of type III effectors in bacterial pathogens. PLoS Pathog. **2:**e104.
33. **Taieb, F., J. P. Nougayrede, C. Watrin, A. Samba-Louaka, and E. Oswald.** 2006. *Escherichia coli* cyclomodulin Cif induces G arrest of the host cell cycle without activation of the DNA-damage checkpoint-signalling pathway. Cell Microbiol. **8:**1910–1921.
34. **Thomas, M. S., and W. T. Drabble.** 1986. Secondary attachment site for

bacteriophage lambda in the *guaB* gene of *Escherichia coli*. J. Bacteriol. **168:**1048–1050.
35. **Tobe, T., S. A. Beatson, H. Taniguchi, H. Abe, C. M. Bailey, A. Fivian, R. Younis, S. Matthews, O. Marchès, G. Frankel, T. Hayashi, and M. J. Pallen.** 2006. An extensive repertoire of type III secretion effectors in *Escherichia coli* O157 and the role of lambdoid phages in their dissemination. Proc. Natl. Acad. Sci. USA. **103:**14941–14946.
36. **Tong, Z., D. Zhou, Y. Song, L. Zhang, D. Pei, Y. Han, X. Pang, M. Li, B. Cui, J. Wang, Z. Guo, Z. Qi, L. Jin, J. Zhai, Z. Du, X. Wang, J. Yu, P. Huang, H. Yang, and R. Yang.** 2005. Pseudogene accumulation might promote the adaptive microevolution of *Yersinia pestis*. J. Med. Microbiol. **54:**259–268.
37. **Tóth, I., F. Herault, L. Beutin, and E. Oswald.** 2003. Production of cytolethal distending toxins by pathogenic *Escherichia coli* strains isolated from human and animal sources: establishment of the existence of a new *cdt* variant (Type IV). J. Clin. Microbiol. **41:**4285–4291.
38. **Tozzoli, R., A. Caprioli, and S. Morabito.** 2005. Detection of *toxB*, a plasmid virulence gene of *Escherichia coli* O157, in enterohemorrhagic and enteropathogenic *E. coli*. J. Clin. Microbiol. **43:**4052–4056.
39. **Wagner, P. L., and M. K. Waldor.** 2002. Bacteriophage control of bacterial virulence. Infect. Immun. **70:**3985–3993.
40. **Wang, H., C. H. Yang, G. Lee, F. Chang, H. Wilson, A. del Campillo-Campbell, and A. Campbell.** 1997. Integration specificities of two lambdoid phages (21 and e14) that insert at the same *attB* site. J. Bacteriol. **179:**5705–5711.
41. **Whittam, T. S., M. L. Wolfe, I. K. Wachsmuth, F. Orskov, I. Orskov, and R. A. Wilson.** 1993. Clonal relationships among *Escherichia coli* strains that cause hemorrhagic colitis and infantile diarrhea. Infect. Immun. **61:**1619–1629.
42. **Zhang, W. L., B. Kohler, E. Oswald, L. Beutin, H. Karch, S. Morabito, A. Caprioli, S. Suerbaum, and H. Schmidt.** 2002. Genetic diversity of intimin genes of attaching and effacing *Escherichia coli* strains. J. Clin. Microbiol. **40:**4486–4492.