JOURNAL OF BACTERIOLOGY, Dec. 2007, p. 9145–9149 Vol. 189, No. 24
0021-9193/07/$08.00+0 doi:10.1128/JB.00722-07

Genetic Diversity of the Gene Cluster Encoding Longus, a Type IV Pilus of Enterotoxigenic *Escherichia coli*▽†

Oscar G. Gomez-Duarte,[1]* Sujay Chattopadhyay,[2] Scott J. Weissman,[3] Jorge A. Giron,[4] James B. Kaper,[5] and Evgeni V. Sokurenko[2]*

Division of Infectious Diseases, Department of Pediatrics, University of Iowa Children's Hospital, Iowa City, Iowa[1]; Department of Microbiology, University of Washington, Seattle, Washington[2]; Division of Infectious Disease, Immunology, and Rheumatology, Department of Pediatrics, University of Washington, Seattle, Washington[3]; Department of Immunobiology, University of Arizona, Tucson, Arizona[4]; and Department of Microbiology, University of Maryland, Baltimore, Maryland[5]

Received 7 May 2007/Accepted 20 September 2007

Enterotoxigenic *Escherichia coli* (ETEC) strains produce a type IV pilus named Longus. We identified a 16-gene cluster involved in the biosynthesis of Longus that has 57 to 95% identity at the protein level to CFA/III, another type IV pilus of ETEC. Alleles of the Longus structural subunit gene *lngA* demonstrate a diversity of 12 to 19% at the protein level with strong positive selection for point replacements and horizontal transfer.

Enterotoxigenic *Escherichia coli* (ETEC) is an important cause of infant diarrhea in developing countries (5, 18), a leading cause of traveler's diarrhea (1, 14), and a re-emergent diarrheal pathogen in the United States (2, 35). One of the ETEC putative colonization factors is Longus, a type IV pilus (T4P) composed of a 22-kDa major structural subunit designated LngA (11, 13), which is estimated to be encoded or expressed by 10 to 35% of ETEC strains (12, 15, 20, 21). Antibodies reacting with LngA were found in stool from patients with ETEC infections (22). LngA shares homology with the major subunit of various T4Ps (9, 13), including the toxin-coregulated pilin of *Vibrio cholerae* (29) and the bundle-forming pilin of enteropathogenic *E. coli* (10). The highest homology (79%) is shared by LngA with CofA, the major subunit of another T4P of ETEC—CFA/III pili (13, 16, 27). Here we identify the genes involved in the assembly and regulation of Longus and describe their genetic and evolutionary variability.

Three plasmid libraries were constructed by partial restriction endonuclease digestion of the Longus-encoding virulence plasmid from ETEC strain E9034A. Plasmids were screened by PCR with *lngA*-derived primers and then other *lng*-specific primers (the plasmids used in this study are listed in Table 1). The Longus gene cluster obtained from ETEC strain E9034A is 14 kb in length and contains 16 open reading frames. Fourteen genes share considerable homology, as well as cluster topology, with CFA/III genes and are thus designated, in homology with the *cof* cluster, *lngR*, *lngS*, *lngT*, *lngA*, *lngB*, *lngC*, *lngD*, *lngE*, *lngF*, *lngG*, *lngH*, *lngI*, *lngJ*, and *lngP* (Fig. 1). Gene length, homology to known proteins, and putative functions are presented for each gene in Table 2. Two open reading frames, *lngX1* and *lngX2*, were identified that have no homology to CFA/III pilin genes.

We screened a collection of ETEC strains by PCR with a previously described *lngA*-specific primer pair (13) and a *cofA*-specific primer pair based on the published *cofA* sequence (28). The bacterial strains used in this study belong to different serotypes and were collected at different time points and in different regions of the world. Among the 56 ETEC strains studied, we identified two (4%) *cofA*-positive strains and 21 (38%) *lngA*-positive strains (see Table S1 in the supplemental material). We also screened the collection with specific primers based on other gene sequences of the Longus or CFA/III cluster (R, S, T, B, C, E, F, G, H, I, J, and P). The *lngA*-positive strains tested positive only with Longus-derived primers, while the *cofA*-positive strains tested positive only with *cof*-derived primers.

The *lngA* genes from the different ETEC strains were sequenced, multiple sequence alignments were performed with ClustalX 1.83 (30), and then PAUP* 4.0b (26) was used to construct maximum-likelihood-based phylogenetic trees. The *lngA* alleles segregated into three distinctive phylogenetic groups defined as groups 1, 2, and 3 (Fig. 2a). Primary (the most basal) nodes of groups 1 and 2 differed in 103 nucleotides (19%), leading to 22 amino acid changes (12%). Groups 1 and 2 differed from group 3 in 142 (26%) and 125 (23%) nucleotides, respectively, each leading to 34 (19%) amino acid changes. Alleles of the *cofA* gene in the two *cofA*-positive strains were also highly divergent from each other, with 137 nucleotide differences (24.8%) leading to 46 residue changes (25%). Although both allelic variants of *cofA* were more closely related to each other than to any of the *lngA* variants, the phylogenetic separation was not distinct (Fig. 2a). For example, the *lngA* group 1 and group 3 variants are phylogenetically more distant from each other than the *lngA* group 3

* Corresponding author. Mailing address for Oscar G. Gomez-Duarte: Division of Pediatric Infectious Diseases, University of Iowa Children's Hospital, 200 Hawkins Drive, Iowa City, IA 52242. Phone: (319) 356-2270. Fax: (319) 356-4855. E-mail: oscar-gomez@uiowa.edu. Mailing address for Evgeni V. Sokurenko: Department of Microbiology, University of Washington, Seattle, WA 98195-7242. Phone: (206) 685-2162. Fax: (206) 543-8297. E-mail: evs@u.washington.edu.

† Supplemental material for this article may be found at http://jb.asm.org/.

▽ Published ahead of print on 19 October 2007.

TABLE 1. Host strains and plasmids used or generated in this study

Strain or plasmid	Description	Reference or source
pE9034A	ETEC strain E9034A virulence plasmid; carries Longus, CS3, ST, and LT genes	13
E. coli DH5α	*supE44* Δ*lacU169* (ϕ80 *lacZ*Δ*M15*) *hsdR17 recA1 endA1 gyrA96 thi-1 relA1*	Laboratory collection
E. coli HB101	*supE44* Δ(*mcrC-mrr*) *recA13 ara-14 proA2 lacY1 galK2 rpsL20 xyl-5 mtl-1 leuB6 thi-1*	Laboratory collection
pBluescript SKII	Cloning vector (Ap^r)	Stratagene, La Jolla, CA
pCR2.1	Cloning vector (Ap^r Km^r)	Invitrogen, Carlsbad, CA
pACYC184	Cloning vector (Cm^r Tet^r)	New England BioLabs, Ipswich, MA
pJAG1	5.0-kb BamHI fragment containing Longus genes cloned into pACYC177	11
pOG140	pBluescript SK with 1.0-kb EcoRV insert containing entire *lngA* gene	13
pOG341	pACYC184 with 8-kb SalI insert containing *lngF* to *lngP*	This study
pOG347	pACYC184 with 5-kb EcoRI insert containing *lngR* to *lngA*	This study
pOG349	pACYC184 with 3-kb EcoRI insert containing *lngG* to *lngH*	This study
pOG378	pACYC184 with 3-kb EcoRI insert containing *lngT* to *lngB*	This study
pOG380	pACYC184 with 5-kb NcoI insert containing *lngJ* to *lngP*	This study
pOG381	Same as pOG380 but with inverted 5-kb NcoI insert containing *lngJ* to *lngP*	This study
pOG371	pCR2.1 with 1.5-kb *lngR-lngS* insert	This study

and *cofA* group 2 variants, obscuring evolutionary distinction of the Longus and CFA/III fimbrial structures.

Within-group analysis of *lngA* sequences identified three distinct allelic variants within group 1 and three variants within group 2 that differed from each other by three to seven nucleotides (Fig. 2a). However, all 15 within-group mutations resulted in amino acid substitutions, indicating that these point replacements accumulated under strong positive selection. While immune escape may drive the accumulation of these point mutations, the changes may also be functionally adaptive for ETEC isolates by modifying some functional (e.g., putative adhesive) properties of Longus fimbriae, as shown for mannose- and di-galactose-specific fimbrial adhesins of uropathogenic *E. coli* (24, 25, 32), *bfpA* of EPEC (3), decay-accelerating factor-specific Dr adhesins of uropathogenic and diarrheal *E. coli* (6, 17, 31), and type 1 fimbrial adhesin of *Salmonella enterica* serovar Typhimurium (4).

Alignment of predicted LngA protein sequences (182 amino acids long) revealed that the between-group diversity of the protein is clustered in three regions, amino acids 56 to 77, 104 to 121, and 148 to 182 (see Fig. S1 in the supplemental material). Most of the within-group point replacements are located in these diversity regions as well. Unlike the amino acid diversity, the diversity of silent changes across *lngA* does not exhibit distinct clustering (see Fig. S1 in the supplemental material). The distribution of amino acid polymorphisms between the LngA and CofA protein sequences (as well as between the two CofA groups) shows less distinct clustering but reflects the general pattern of diversity seen among the LngA groups. Both LngA and CofA represent major surface subunits and thus antigens expected to be under strong selection for structural diversification. In particular, surface epitope regions would be prone to accumulate extensive structural changes while regions that are critical for proper tertiary structure or fimbrial morphology and/or function would remain conserved.

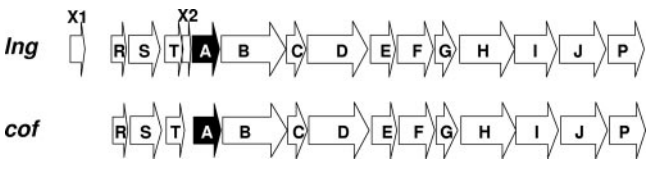

FIG. 1. Homology of Longus and CFA/III gene clusters. Arrows indicate the direction of transcription. Genes in black correspond to major structural subunit genes.

Unlike the structural subunit genes, nonstructural Longus genes are highly conserved (determined for the entire gene cluster in a subset of six strains), with more than 98% identity among homologs at the DNA (Fig. 2b) and protein (data not shown) levels. Homologs of nonstructural genes of the two CFA/III clusters (determined for *cofC*, *-E*, *-H*, and *-P*) were also conserved (99% identical to each other, on average), and after concatenation, they formed a distinct outgroup relative to the Longus genes (Fig. 2b), which is in contrast to the *lngA*/*cofA* tree (Fig. 2a). This indicates that while Longus and CFA/III pili are evolutionarily related, the *cof* and *lng* genes are paralogous in nature; i.e., they belong to evolutionarily distinct types of fimbriae.

To assess phylogenetic congruence among different *lng* loci, we performed an incongruence length difference test (8) with PAUP* 4.0b (26) that compared the sum of the branch lengths of a given pair of trees with the sum of the branch-lengths obtained through 1,000 random partitionings of the original data sets. Individual phylograms for the nonstructural genes were congruent with each other, providing no evidence of recombinational shuffling of alleles among different clusters (not shown). In contrast, a gene tree of concatenated nonstructural genes (Fig. 2b) was not congruent with the tree of corresponding *lngA* alleles ($P = 0.001$), suggesting that the major structural subunit gene is moving horizontally among the fimbrial gene clusters.

To determine the populational and intergenomic dynamics of the Longus gene clusters, we also sequenced two housekeeping loci—*adk* and *fumC*—in different ETEC strains. Housekeeping genes constitute the "backbone" of the bacterial chromosome, subject to recombinational events only infrequently in the species *E. coli*, providing the basis for phylogenetic and clonal grouping of *E. coli* strains (34). Both loci demonstrated high levels (98.1% for *fumC* and 99.4% for *adk*) of identity at the DNA level (Fig. 2c) that are within the range shown for housekeeping genes of *E. coli*

TABLE 2. Biosynthetic Longus genes, putative functions, gene homologues, and allelic variations

Gene	Length (bp)	% Identity to CFA/III gene/protein	Putative function	Type IV pilus gene homologs	D_n/D_s[d] of allelic variants
lngX1	378	NA[a]	Unknown	None	NA
lngR	303	72/72	Gene regulation	*cofR*, *clpB*, *pefB*; PapB regulator family	0.011/0.010
lngS	894	68/70	Gene regulation	*cofS*, *perA*, *toxT*; AraC regulator family	0.004/0.015
lngT	441	79/82	Transglycosylase	*cofT*, *bfpH*	0.005/0.012
lngX2	219	NA	Unknown	None	NA
lngA	708	74/79	Structural subunit	*cofA*, *cfcA*, *bfpA*, *tcpA*	0.053/0.571
lngB	1,572	74/78	Pilin-like subunit	*cofB*, *cfcB*, *tcpB*	0.005/0.068
lngC	411	7370	Unknown	*cofC*, *cfcC*, *tcpQ*, *bfpG*	0.001/0.016
lngD	1,476	81/91	Outer membrane lipoprotein	*cofD*, *cfcD*, *bfpB*, *tcpC*, *pulD*	0.001/0.010
lngE	558	75/75	Inner membrane	*cofE*, *cfcE*, *tcpR*	0.003/0.010
lngF	771	68/66	Inner membrane	*cofF*, *cfcF*, *tcpD*	0.004/0.014
lngG	483	73/75	Periplasmic protein	*cofG*, *cfcG*, *bfpU*, *tcpS*	0.003/0.016
lngH	1,332	73/75	Nucleotide binding, twitching motility	*cofH*, *cfcH*, *bfpD*, *tcpT*, *pulE*	0.001/0.016
lngI	969	83/96	Inner membrane	*cofI*, *cfcI*, *bfpE*, *tcpE*, *pulF*	0.002/0.026
lngJ	1,119	63/57	ATPase, twitching motility	*cofJ*, *cfcJ*, *bfpF*, *tcpF*	0.002/0.033
lngP	852	62/59	Prepilin peptidase	*cofP*, *cfcP*, *bfpP*, *tcpJ*, *pulO*	0.010/0.038
adk[b]	534[c]				0.0005/0.025
fumC[b]	465[c]				0.004/0.072

[a] NA, not applicable.
[b] Chromosomal housekeeping gene used for evolutionary analysis.
[c] Partial locus.
[d] D_n/D_s, ratio of nonsynonymous changes (as determined according to reference 40 by DnaSP software) to synonymous, silent changes.

at the species level (23). No congruence was detected between the concatenated *adk*/*fumC* tree and the *lngA* groups ($P = 0.001$), indicating frequent horizontal transfer of the *lngA* genes among ETEC strains. The *lngA* alleles have shown incongruence even with the major phylogenetic clusters of *E. coli*—ECOR groups A and B1—to which the vast majority of the ETEC strains studied belong (determined by the Clermont method [7]; see Table S1 in the supplemental material), indicating gene movement among distant clones as well. At the same time, the trees of *adk*/*fumC* and nonstructural *lng* genes were congruent with each other (the incongruence length difference test *P* value equals 1) for the same sample set, indicating that the nonstructural genes are significantly less prone to transfer than *lngA*. Thus, it appears that the main mechanism for horizontal transfer of *lngA* among ETEC strains involves single-gene recombination rather than transfer of the entire Longus-carrying plasmid.

Sequence analysis of the fimbrial genes also provides insight into the evolutionary history of Longus. It appears that the highly conserved nonstructural *lng* genes evolved in a manner similar to that of housekeeping genes. It involved accumulation of point mutations (rather than recombination) and purifying selection against amino acid replacement changes, the latter based on the low ratio of nonsynonymous changes (D_n as determined according to reference 35 by DnaSP software) to synonymous, silent changes (D_s) seen in most of the genes (Table 2). Silent changes are considered to be functionally neutral and accumulate randomly over time, reflecting the molecular clock. We used two different molecular clock rates, 3E-8 changes/year (16) and 6E-9 changes/year (33). Thus, on the basis of nonstructural gene diversity, Longus and CFA/III fimbriae were acquired by *E. coli* around 0.73 to 3.64 and 0.97 to 4.83 million years ago, respectively (16, 33), i.e., probably at some time after *E. coli* speciation, which is estimated to have occurred around 5 to 6 million years ago (19).

The synonymous and nonsynonymous diversity of *lngA* and *cofA* alleles is approximately 10-fold higher than that of corresponding nonstructural genes (Table 2). The high rates of both nonsynonymous and synonymous changes in the major subunit genes could be due to a high rate of either nonhomologous intragenic recombination (gene shuffling) or point mutations in the structural genes comparative to the nonstructural ones. Alternatively, this could be explained by parallel evolution of the structural allelic variants for extended periods of evolutionary time, with the *lngA* (and *cofA*) gene groups starting to diverge at the approximate time of the divergence of the Longus and CFA/III gene clusters from each other 20 to 100 million years ago, i.e., well before *E. coli* speciation. We propose that *lng* and *cof* clusters were acquired by *E. coli* in separate unique events after the time of speciation, with this evolutionary bottleneck explaining both the distinct phylogeny and low diversity of the *lng* and *cof* nonstructural genes. Afterwards, there were continuous horizontal-transfer events acting only on the major subunit genes *lngA* and *cofA*, both between and within species, driven by strong pressure for antigenic diversity.

Despite the antigenic diversity of *lngA* and *cofA*, the present study indicates that there is significant structural conservation between the group variants, allowing for likely cross-reactivity among them and thus the development of a Longus-CFA/III immunoprotective multicomponent vaccine against ETEC diarrheal disease.

Nucleotide sequence accession number. Newly described DNA sequences reported in this study were deposited in the GenBank database. The entire Longus cluster DNA sequence was assigned accession number EF595770. The *cofA* and *lngA* gene DNA sequences were assigned accession numbers EU107087 to EU107107.

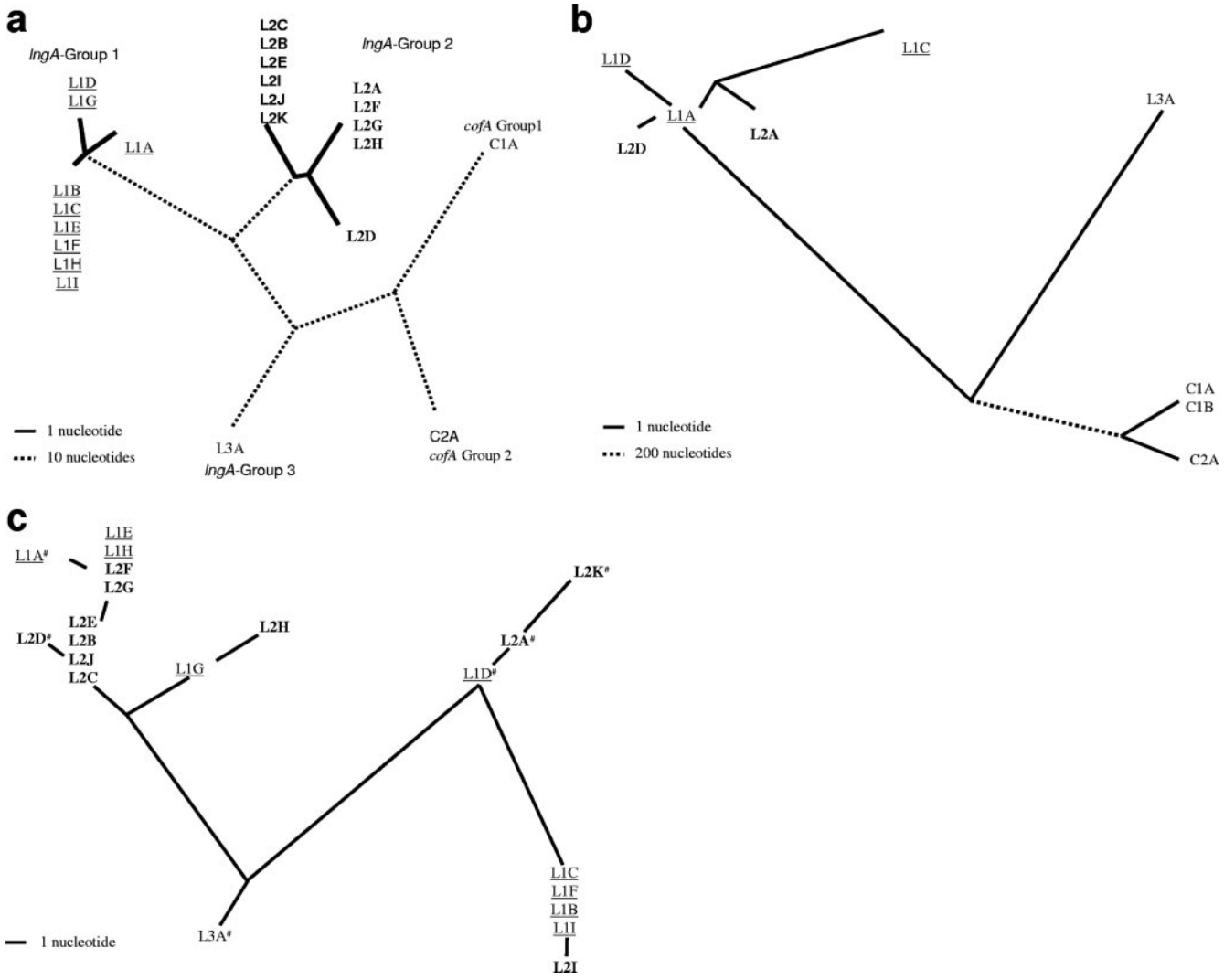

FIG. 2. Maximum-likelihood phylograms of Longus sequences. (a) Unrooted phylogram of *lngA* sequences. *lngA* sequences from 21 ETEC strains along with *cofA* sequences from CFA/III ETEC strains were included. (b) Phylogram of concatenated nonstructural genes in the Longus (based on a subset of six strains) and CFA/III gene clusters. The intra-*cof* (260-1/E2528C/M403-C1) comparison of nonstructural genes is based on 4 out of 13 loci (*cofC*, *-E*, *-H*, and *-P*). (c) Phylogram of concatenated *adk* and *fumC* genes from 21 ETEC strains. The branch lengths refer to numbers of nucleotide differences. For panels A and B, two different length scales are used. The symbol # indicates strains in which the non-*lngA* genes were sequenced (see panel B). Homologous *lngA* allelic groups are denoted by underlined (group 1), bold (group 2), and plain (group 3) text. Strains from which sequences were derived are represented by three-letter designations. *lngA* group 1 strains: E9034A, L1A; 01117-5, L1B; B2C, L1C; M104, L1D; M145-C2, L1E; M424-C1, L1F; M526-C6B, L1G; P307, L1H; M408-C1, L1I. *lngA* group 2 strains: 10159-a, L2A; 11381a, L2B, 2108-2, L2C; B7A, L2D; G1026, L2E; M445-C1, L2F; M452-C1, L2G; M626-C, L2H; M633-C1, L2I; MP215-1, L2J; BR5, L2K. *lngA* group 3 strain ECOR27, L3A. *cofA* group 1 strains: 2528C, C1A; 260-1, C1B. *cofA* group 2 strain M403-C1, C2A.

This work was supported in part by NIH grant GM60731 to E.V.S., grants AI21657 and DK58957 to J.B.K., a K08 award to S.J.W., and a K12 award to O.G.G.-D. It was also supported by grant 1892-2007 from the Children's Miracle Network—University of Iowa Children's Hospital to O.G.G.-D.

We are grateful to Steve L. Moseley for providing *E. coli* strains for this study. We thank Jing Bai for technical assistance with DNA experiments.

REFERENCES

1. **Adachi, J. A., C. D. Ericsson, Z. D. Jiang, M. W. DuPont, S. R. Pallegar, and H. L. DuPont.** 2002. Natural history of enteroaggregative and enterotoxigenic *Escherichia coli* infection among US travelers to Guadalajara, Mexico. J. Infect. Dis. **185:**1681–1683.
2. **Beatty, M. E., P. M. Adcock, S. W. Smith, K. Quinlan, L. A. Kamimoto, S. Y. Rowe, K. Scott, C. Conover, T. Varchmin, C. A. Bopp, K. D. Greene, B. Bibb, L. Slutsker, and E. D. Mintz.** 2006. Epidemic diarrhea due to enterotoxigenic *Escherichia coli*. Clin. Infect. Dis. **42:**329–334.
3. **Blank, T. E., H. Zhong, A. L. Bell, T. S. Whittam, and M. S. Donnenberg.** 2000. Molecular variation among type IV pilin (*bfpA*) genes from diverse enteropathogenic *Escherichia coli* strains. Infect. Immun. **68:**7028–7038.
4. **Boddicker, J. D., N. A. Ledeboer, J. Jagnow, B. D. Jones, and S. Clegg.** 2002. Differential binding to and biofilm formation on, Hep-2 cells by *Salmonella enterica* serovar Typhimurium is dependent upon allelic variation in the *fimH* gene of the *fim* gene cluster. Mol. Microbiol. **45:**1255–1265.
5. **Cama, R. I., U. D. Parashar, D. N. Taylor, T. Hickey, D. Figueroa, Y. R. Ortega, S. Romero, J. Perez, C. R. Sterling, J. R. Gentsch, R. H. Gilman, and R. I. Glass.** 1999. Enteropathogens and other factors associated with severe disease in children with acute watery diarrhea in Lima, Peru. J. Infect. Dis. **179:**1139–1144.
6. **Carnoy, C., and S. L. Moseley.** 1997. Mutational analysis of receptor binding mediated by the Dr family of *Escherichia coli* adhesins. Mol. Microbiol. **23:**365–379.

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. **Farris, J. S., M. Kallersjo, A. G. Kluge, and C. Bult.** 1995. Testing significance of incongruence. Cladistics **10:**315–319.
9. **Girón, J. A., O. G. Gómez-Duarte, K. G. Jarvis, and J. B. Kaper.** 1997. Longus pilus of enterotoxigenic *Escherichia coli* and its relatedness to other type 4 pili—a mini review. Gene **192:**39–43.
10. **Girón, J. A., A. S. Y. Ho, and G. K. Schoolnik.** 1991. An inducible bundle-forming pilus of enteropathogenic *Escherichia coli*. Science **254:**710–713.
11. **Girón, J. A., M. M. Levine, and J. B. Kaper.** 1994. Longus: a long pilus ultrastructure produced by human enterotoxigenic *Escherichia coli*. Mol. Microbiol. **12:**71–82.
12. **Girón, J. A., G. I. Viboud, V. Sperandio, O. G. Gómez-Duarte, D. Maneval, M. J. Albert, M. M. Levine, and J. B. Kaper.** 1995. Prevalence and association of Longus pilus structural gene (*lngA*) with colonization factor antigens, ST/LT enterotoxin types, and serotypes of enterotoxigenic *Escherichia coli*. Infect. Immun. **63:**4195–4198.
13. **Gómez-Duarte, O. G., A. Ruiz-Tagle, D. C. Gomez, G. I. Vivoud, K. G. Jarvis, J. B. Kaper, and J. A. Giron.** 1999. Longus and colonization factor antigen III (CFA/III): two highly related but distinct type 4 pili produced by enterotoxigenic *Escherichia coli*. Microbiology **145:**1809–1816.
14. **Gorbach, S. L., B. H. Kean, D. G. Evans, D. J. Evans, Jr., and D. Bessudo.** 1975. Travelers' diarrhea and toxigenic *Escherichia coli*. N. Engl. J. Med. **292:**933–936.
15. **Gutiérrez-Cázares, Z., F. Qadri, M. J. Albert, and J. A. Giron.** 2000. Identification of enterotoxigenic *Escherichia coli* harboring Longus type IV pilus gene by DNA amplification. J. Clin. Microbiol. **38:**1767–1771.
16. **Guttman, D. S., and D. E. Dykhuizen.** 1994. Clonal divergence in *Escherichia coli* as result of recombination, not mutation. Science **266:**1380–1383.
17. **Korotkova, N., S. Chattopadhyay, T. A. Tabata, V. Beskhlebnaya, V. Vigdorovich, B. K. Kaiser, R. K. Strong, D. E. Dykhuizen, E. V. Sokurenko, and S. L. Moseley.** 2007. Selection for functional diversity drives accumulation of point mutations in Dr adhesins of *Escherichia coli*. Mol. Microbiol. **64:**180–194.
18. **Kosek, M., C. Bern, and L. Guerrant.** 2003. The global burden of diarrhoeal disease, as estimated from studies published between 1992 and 2000. Bull. W. H. O. **81:**197–204.
19. **Le Gall, T., P. Darlu, P. Escobar-Paramo, B. Picard, and E. Denamur.** 2005. Selection-driven transcriptome polymorphism in *Escherichia coli*/Shigella species. Genome Res. **15:**260–268.
20. **Nishimura, L. S., J. A. Giron, S. L. Nunes, and B. E. Guth.** 2002. Prevalence of enterotoxigenic *Escherichia coli* strains harboring the Longus pilus gene in Brazil. J. Clin. Microbiol. **40:**2606–2608.
21. **Pichel, M. G., N. Binsztein, F. Qadri, and J. A. Giron.** 2002. Type IV Longus pilus of enterotoxigenic *Escherichia coli*: occurrence and association with toxin types and colonization factors among strains isolated in Argentina. J. Clin. Microbiol. **40:**694–697.
22. **Qadri, F., J. A. Giron, A. Helander, Y. A. Begum, M. Asaduzzaman, J. Xicohtencatl-Cortes, A. Negrete, and M. J. Albert.** 2000. Human antibody response to Longus type IV pilus and study of its prevalence among enterotoxigenic *Escherichia coli* in Bangladesh by using monoclonal antibodies. J. Infect. Dis. **181:**2071–2074.
23. **Rocha, E. P., J. M. Smith, L. D. Hurst, M. T. Holden, J. E. Cooper, N. H. Smith, and E. J. Feil.** 2006. Comparisons of dN/dS are time dependent for closely related bacterial genomes. J. Theor. Biol. **239:**226–235.
24. **Sokurenko, E. V., H. S. Courtney, D. E. Ohman, P. Klemm, and D. L. Hasty.** 1994. FimH family of type 1 fimbrial adhesins: functional heterogeneity due to minor sequence variations among *fimH* genes. J. Bacteriol. **176:**748–755.
25. **Sokurenko, E. V., M. A. Schembri, E. Trintchina, K. Kjaergaard, D. L. Hasty, and P. Klemm.** 2001. Valency conversion in the type 1 fimbrial adhesin of *Escherichia coli*. Mol. Microbiol. **41:**675–686.
26. **Swofford, D. L.** 2000. PAUP*: Phylogenetic analysis using parsimony and other methods. Sinauer Associates, Sunderland, MA.
27. **Taniguchi, T., N. F. Uchima-Senaga, Y. Takarada, S. Shibata, T. Tsukamoto, K. Yamamoto, and T. Honda.** 1995. Use of a new oligonucleotide probe for detection of colonization factor antigen III gene in enterotoxigenic *Escherichia coli*. Eur. J. Clin. Microbiol. Infect. Dis. **14:**713–716.
28. **Taniguchi, T., Y. Akeda, A. Haba, Y. Yasuda, K. Yamamoto, T. Honda, and K. Tochikubo.** 2001. Gene cluster for assembly of pilus colonization factor antigen III of enterotoxigenic *Escherichia coli*. Infect. Immun. **69:**5864–5873.
29. **Taylor, R. K., V. L. Miller, D. B. Furlong, and J. J. Mekalanos.** 1987. Use of *phoA* gene fusions to identify a pilus colonization factor coordinately regulated with cholera toxin. Proc. Natl. Acad. Sci. USA **84:**2833–2837.
30. **Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins.** 1997. The CLUSTAL_X Windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. **25:**4876–4882.
31. **Van Loy, C. P., E. V. Sokurenko, R. Samudrala, and S. L. Moseley.** 2002. Identification of amino acids in the Dr adhesin required for binding to decay-accelerating factor. Mol. Microbiol. **45:**439–452.
32. **Weissman, S. J., S. Chattopadhyay, P. A. Aprikian, M. Obata-Yasuoka, Y. Yarova-Yorovaya, A. Stapleton, W. Ba-Thein, D. Dykhuizen, J. R. Johnson, and E. V. Sokurenko.** 2006. Clonal analysis reveals high rate of structural mutations in fimbrial adhesins of extraintestinal pathogenic *Escherichia coli*. Mol. Microbiol. **59:**975–988.
33. **Whittam, T. S.** 1996. Genetic variation and evolutionary processes in natural populations of *Escherichia coli*, p. 2708–2720. *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, 2nd ed., vol. 2. ASM Press, Washington, DC.
34. **Wirth, T., D. Falush, R. Lan, F. Colles, P. Mensa, L. H. Wieler, H. Karch, P. R. Reeves, M. C. Maiden, H. Ochman, and M. Achtman.** 2006. Sex and virulence in *Escherichia coli*: an evolutionary perspective. Mol. Microbiol. **60:**1136–1151.
35. **Yoder, J. S., S. Cesario, V. Plotkin, X. Ma, K. Kelly-Whannon, and M. S. Dworkin.** 2006. Outbreak of enterotoxigenic *Escherichia coli* infection with an unusually long duration of illness. Clin. Infect. Dis. **42:**1513–1517.