

CONCISE COMMUNICATION

Genetic and Evolutionary Analysis of Mutations in the *gusA* Gene That Cause the Absence of β -Glucuronidase Activity in *Escherichia coli* O157:H7Steven R. Monday,¹ Thomas S. Whittam,²
and Peter C. H. Feng¹¹Division of Microbiological Studies, Food and Drug Administration,
Washington, DC; ²National Food Safety and Toxicology Center,
Michigan State University, East Lansing

Escherichia coli serotype O157:H7 do not exhibit β -glucuronidase (GUD) activity but carry the *gusA* gene (*uidA*) that encodes for GUD. In trans-complementation, the *gusA* gene cloned from the GUD-positive variant strain 493-89 effectively restored GUD activity in O157:H7 strain 35150. Comparison of *gusA* sequences from the GUD-negative 35150 strain to that of 493-89 revealed several base mutations, including a guanosine (G) dinucleotide insertion that caused a frameshift in the 35150 *gusA* gene and introduced a predicted premature termination codon. This explains the absence of GUD activity in O157:H7. A 35150 *gusA* construct from which the G-G insertion was deleted restored activity in GUD-negative O157:H7 transformants. The G-G insertion was present in all GUD-negative O157:H7 strains but was absent in their GUD-positive variants. The G-G insertion that produced the characteristic GUD-negative phenotype to O157:H7 strains appeared later than the other *gusA* mutations in the evolutionary emergence of O157:H7.

The enzyme β -glucuronidase (GUD) is prevalent in *Escherichia coli* [1], including most pathogenic strains [2], except for enterohemorrhagic *E. coli* O157:H7 serotype [3]. Consequently, the GUD-negative phenotype often is used to differentiate O157:H7 from other *E. coli*. The *E. coli* GUD is encoded by the chromosomal *gusA* (*uidA*) gene, which is regulated by 2 repressors [4]. GUD expression also is sensitive to cAMP-dependent catabolite repression [5] and to other physiologic factors [6]. Although O157:H7 strains do not exhibit GUD, they carry the entire *gusA* gene, including regulatory regions [7].

The absence of GUD activity in O157:H7 is not due to the lack of enzyme induction, catabolite repression by lactose, or other physiologic factors, which suggests that the defect may be genetic. In an earlier study [7], the *gusA* gene of O157:H7 was compared with that of *E. coli* K12 [8] (GenBank accession no. M14641) and was found to contain several mutations, including an A→T substitution in the putative -10 promoter site [7]. Analysis of the regulatory region, however, showed that the *gusA* promoter of O157:H7 was functional [7], hence ruling out transcriptional repression of GUD expression. Most of the mutations in the O157:H7 *gusA* structural gene are degenerate,

except for a few that alter amino acid sequence and charges, including a putative frameshift at the 3' terminus. Although these mutations could affect GUD activity, that aspect was not fully elucidated. The *E. coli* K-12 *gusA* sequence has since been updated, prompting us to reexamine the O157:H7 *gusA* sequence. In this study, the *gusA* gene from O157:H7 and its GUD-positive variants were compared to determine whether these genetic mutations might explain the absence of GUD in O157:H7 and to determine whether the occurrence of these mutations are consistent with the proposed evolutionary model for the emergence of O157:H7 serotype [9].

Materials and Methods

Bacteria. *E. coli* 35150 (American Type Culture Collection) and 13B88 (a clinical isolate) are GUD-negative O157:H7 strains. Strain G5101, an atypical O157:H7 isolate [10], and strain 493-89, an O157:H⁻ strain [11] (both clinical isolates), are GUD-positive phenotypic variants. All strains produce both Shiga toxin 1 (Stx1) and Stx2, except 493-89, which produces only Stx2. All cloning and manipulation was facilitated in *E. coli* DH5 α [12].

Cloning of *gusA*. The *gusA* gene was amplified by polymerase chain reaction (PCR) by use of the high-fidelity exonuclease-positive Vent DNA polymerase (New England Biolabs). A 50- μ L reaction contained 1 \times Thermopol buffer, 6 mM MgSO₄, 200 nM *gusA*-specific primers PF20 and PF28 [7], 200 μ M dNTP, and $\sim 10^5$ bacterial cells as template. A hot start amplification was done for 32 cycles (95°C for 1 min, 53°C for 45 s, and 75°C for 2.5 min), with a final extension of 10 min at 75°C. The desired PCR product was excised from a 1% agarose gel, was ligated into *EcoRV*-digested pBluescript SK+, and was electroporated into competent DH5 α cells [12]. By using vector-specific T3/T7 promoter primers, trans-

Received 9 April 2001; revised 7 June 2001; electronically published 15 August 2001.

Presented in part: 101st general meeting, American Society for Microbiology, Orlando, Florida, May 2001 (poster P-13).

Reprints or correspondence: Dr. Peter Feng, HFS-516, Food and Drug Administration, 200 C St. SW, Washington, DC 20204 (pfeng@cfsan.fda.gov).

The Journal of Infectious Diseases 2001; 184:918–21

© 2001 by the Infectious Diseases Society of America. All rights reserved.
0022-1899/2001/18407-0016\$02.00

formants were screened for the proper plasmid construct, which then was digested with *Hin*DIII and *Bam*HI, to determine insert orientation.

Complementation. The 35150 strain was transformed with plasmid constructs containing either the nonfunctional 35150 *gusA* gene or the functional GUD-positive 493-89 *gusA* gene. Transformants were selected on tryptic soy agar (TSA) plates containing 100 µg/mL ampicillin (Amp) and 75 µg/mL 4-methylumbelliferyl-β-D-glucuronide (MUG). Amp resistant colonies that exhibited GUD (bluish fluorescence around colonies under longwave UV) [1] were tested serologically for the O157 and H7 antigens and by multiplex PCR for characteristic O157:H7 markers [13].

Sequencing. *gusA* plasmid constructs derived from the GUD-negative strain 35150 and GUD-positive strain 493-89 were CsCl purified. Using an automated system, sequencing was accomplished with the universal T3/T7 promoter-specific primers, *gusA*-specific primers described elsewhere [7], and 3 additional primers—SM1 (5'-GGTGATGTCAGCGTTGAAC-3'), SM2 (5'-CTAGTGCCTTGTCCAGTTGC-3'), and SM3 (5'-CAGGAGAACTGCATCAGC-3'). Both strands of the *gusA* sequence were analyzed by GCG software (Genetics Computer Group).

Inverse PCR. The guanosine-guanosine (G-G) dinucleotide insertion in the 35150 *gusA* gene was deleted by inverse PCR [12]. Using pBluescript SK+ :35150 *gusA* as template and 300 nM each of these site-specific primers (up, 5'-CTTGTCAGTTGCAACCA-CC-3'; down, 5'-GCACCAGCGGGACTTTGC-3'), the PCR was done as described above, except that primers were annealed at 52°C and were extended for 5.5 min. The PCR product was purified and electroporated back into 35150. Transformants were selected for on TSA-amp-MUG plates. To verify the presence or absence of the G-G insertion, the plasmid-borne *gusA* insert from 2 GUD-positive and -negative amp-resistant transformants were sequenced, using the PF25 and PF26 primers [7].

Results

Cloning of *gusA*. The *gusA* amplicon from both GUD-negative O157:H7 strains (35150 and 13B88) and GUD-positive variants (493-89 and G5101) was ~2100 bp in size. Restriction of constructs with *Hin*DIII revealed a single ~5000-bp fragment, which confirmed the presence of properly sized insert within the vector (data not shown). Both *gusA* and the vector contain single *Bam*HI sites so that constructs that produced 4400- and 700-bp *Bam*HI fragments had the 5' of the *gusA* gene adjacent to the vector T3/LacZ promoter, whereas those that produced the 3700- and 1400-bp fragments had the insert in the opposite orientation (data not shown).

Complementation. The GUD-negative 35150 strain transformed with *gusA* cloned from itself remained GUD negative. In contrast, 35150 transformed with *gusA* from 493-89 exhibited GUD activity, regardless of the *gusA* insert orientation within the plasmid vector, indicating the presence of a functional endogenous promoter. These GUD-positive transformants expressed O157 and H7 antigens and had all genetic markers for O157:H7, including *stx1* and *stx2* genes, and the +93 *gusA* mutation that is unique to O157:H7 (data not shown).

Sequencing. The *gusA* sequences obtained were submitted to the GenBank database. The *gusA* sequence of strain 35150 (GenBank accession no. AF305917) and strain 493-89 (GenBank accession no. AF305918) were compared to each other and to those of *E. coli* K12, independently submitted by the Japan *E. coli* Genome Sequencing Project (GenBank accession no. D90805) and by Jefferson et al. [8] (GenBank accession no. M14641).

The *gusA* of 493-89 was 2103-bp long, with the open-reading frame (ORF) initiating at nucleotide (nt) 179 and terminating at the TGA codon at +1988–1990. Comparison with the same sequence from K12 (D90805) showed that the 493-89 *gusA* differed by 19 nt, including a base change at –23 that was outside the ORF (table 1). In comparison, the 35150 *gusA* was 2105 bp and differed from that of 493-89 at only 4 sites (+686, +776, +1370, and +1671) within the ORF. Of these, the +1671 mutation was cryptic, but those at +776 and +1370 altered amino acid sequences in 35150 (table 1). Of more importance, the +686 mutation in the *gusA* of 35150 was a G-G dinucleotide insertion that caused a frameshift in the ORF, which severely altered amino acid sequence downstream of the lesion and introduced a TGA stop codon at +711 that is predicted to cause premature termination in the GUD translation. Sequencing of the frameshift mutation region from 2 other O157:H7 strains (G5101 and

Table 1. Comparison of codon differences in the *gusA* sequence of the β-glucuronidase (GUD)-positive *Escherichia coli* K12 and 493-89 (O157:H-) and the GUD-negative ATCC 35150 (O157:H7).

Base-change position ^a	K12 ^b	493-89 ^c	ATCC 35150 ^c
-23	A	T	T
+93	GAT	GAG (D31E)	GAG (D31E)
+191	GCG	GTG (A64V)	GTG (A64V)
+393	ACC	ACA	ACA
+501	TAT	TAC	TAC
+686 ^d	—	—	GG (frameshift)
+690	ACT	ACC	ACC
+758	TGC	TAC (C253Y)	TAC (C253Y)
+776 ^d	CAG	CAG	CGG (Q259R)
+801	CTT	CTG	CTG
+849	ATT	ATC	ATC
+882	GGT	GGC	GGC
+900	GAC	GAT	GAT
+906	CGT	CGC	CGC
+942	GAC	GAT	GAT
+1059	GCT	GCA	GCA
+1158	CAA	CAG	CAG
+1308	GAC	GAT	GAT
+1370 ^d	AGC	AGC	ATC (S457I)
+1410	GGA	GGT	GGT
+1671	TCG	TCG	TCA
+1704	AAA	AAG	AAG
+1716	ACT	ACC	ACC

^a With respect to *gusA* open-reading frame ATG start codon. Sites in bold represent regions where substitution resulted in alteration of the promoter sites or incorporation of alternative amino acids into the primary structure of the protein. Resulting amino acid substitutions are shown in parentheses.

^b GenBank accession no. D90805 submitted by the Japan *E. coli* Genome DNA Sequencing Project group.

^c Determined in this study.

^d Nucleotide substitution that is unique to the ATCC 35150 strain.

13B88) confirmed that both isolates contained the +776 mutation described for O157:H7 strains; however, the G-G dinucleotide insertion at +686 was absent in GUD-positive strain G5101, although it was present in GUD-negative strain 13B88.

Inverse PCR. We deleted the G-G insertion in pBluescript SK:35150 *gusA* by site-directed mutagenesis with inverse PCR. The resulting PCR product consisted of a mixture of the original pBluescript SK:35150 *gusA* template that still had the G-G insertion and the linear PCR product that contained the altered *gusA* allele (without the G-G insertion). When this mixture was electroporated back into 35150, which also has a non-functional *gusA* allele, a few GUD-positive, but mostly GUD-negative, transformants were obtained, as expected. Sequencing of the *gusA* insert from 2 GUD-positive and -negative transformants showed that all 4 isolates had the +776 mutation unique to O157:H7; however, the +686 frameshift mutation was present in only the *gusA* insert obtained from GUD-negative transformants. PCR analysis of additional strains by use of primers specific for the G-G insertion confirmed that the frameshift was found in all GUD-negative O157:H7 strains but was absent in their GUD-positive variants (data not shown).

Discussion

The GUD-negative O157:H7 strain 35150 transformed with the *gusA* gene from its GUD-positive variant 493-89 exhibited GUD phenotype. This confirms that GUD inactivity in O157:

H7 is due to genetic defects and that the *gusA* ORF of 493-89 has the necessary transcriptional elements for GUD expression.

The *gusA* sequences of 493-89 and *E. coli* K12 (GenBank accession no. D90805) were both 2103 nt long and carried the 1812-bp *gusA* ORF, but the 493-89 sequence had 19 point mutations (table 1). The mutation at -23, which is within the putative -10 promoter, was described elsewhere but was ruled out as a cause for GUD inactivity in O157:H7, because the promoter was functional [7]. Of the remaining mutations, only 3 (+93, +191, and +758) resulted in amino acid changes. However, since both K12 and 493-89 are GUD positive, none of these mutations inhibited GUD activity.

The regulatory sequences upstream of the *gusA* start codon were identical in both 35150 and 493-89. The *gusA* ORF of both strains also had the +93 T to G mutation that was absent in *E. coli* K12. Because the +93 mutation is unique and conserved in O157:H7 and its Stx-producing phenotypic variants [7, 14], it is often used to detect O157:H7 [9, 13, 14]. Because of its close association with O157:H7, a common misconception is that this mutation is the cause of GUD inactivity in O157:H7 [15]. However, since the +93 mutation is also present in GUD-positive O157:H7 variants (493-89 and G5101), it has no aberrant effect on GUD.

Previously, a putative frameshift caused by an insertion at +1258 was reported in the O157:H7 *gusA* ORF [7] and was deduced on the basis of comparison with an earlier *gusA* sequence for K12 [8]. However, the *gusA* sequence of this region determined in this study for 493-89 and 35150 was consistent

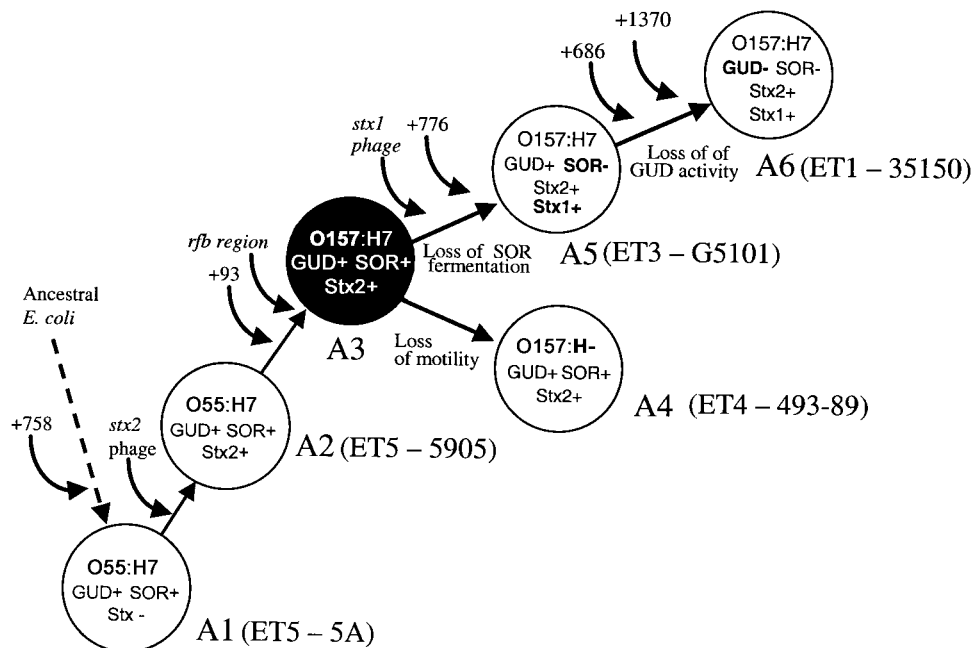


Figure 1. Postulated occurrence of *gusA* mutations in the evolutionary emergence of *Escherichia coli* serotype O157:H7. The amino acid substitutions that result from these nucleotide changes are listed in table 1. The electrophoretic types (ET) and the representative strains, shown in parentheses, are derived from the published emergence model for O157:H7 [9]. Figure is modified from [9]. GUD, β -glucuronidase; SOR, sorbitol; Stx, Shiga toxin.

with the updated K12 sequence (GenBank accession no. D90805). Thus, no frameshift exists in this region, as postulated elsewhere [7]. However, the O157:H7 *gusA* ORF had 3 mutations (+686, +776, and +1370) that changed primary amino acid sequence; only +776 was reported elsewhere [7]. Of these, +686 was a G-G dinucleotide insertion, which produced a frameshift that not only altered amino acid sequence but also introduced a premature stop codon identical to the one that terminates the translation of the functional *gusA*. Deletion of the G-G dinucleotide insertion without altering the +776 and +1370 mutations restored GUD activity in O157:H7 strains, which confirms that GUD inactivity in O157:H7 strains was due solely to the G-G insertion mutation at +686.

Because these amino acid altering *gusA* mutations seem to be highly conserved in O157:H7, we examined other *E. coli* to see how these mutations fit into the stepwise evolutionary model of O157:H7 [9]. The +758 probably appeared the earliest. It is not present in K12 but is found in the O55:H7 type strain at A1 and in all subsequent strains (figure 1). The +93 mutation is postulated to have occurred during transition from O55:H7 (A2) to a hypothetical O157:H7 intermediate (A3; figure 1), because it is found in all strains that follow the A3 intermediate. The +776 mutation probably appeared after A3, because strains before A3 and those in A4 divergence (represented by 493-89) did not carry +776, unlike all strains in the A5 branch (figure 1). *gusA* mutations in A3 have not been determined, because this intermediate has not yet been encountered. Finally, the GUD-positive O157:H7 variant (G5101) at A5 acquired the +1370 mutation that is not found in preceding strains and the G-G insertion at +686, which predictively caused the loss of GUD function, to give rise to the typical GUD-negative O157:H7 at A6 (35150). Thus, the occurrence of these *gusA* mutations is consistent with and supports the proposed evolutionary model for O157:H7 emergence [9].

References

- Feng P, Hartman PA. Fluorogenic assays for the immediate confirmation of *Escherichia coli*. *Appl Environ Microbiol* **1982**;43:1320-9.
- McDaniels AE, Rice EW, Reyes AL, Johnson CH, Haugland RA, Stelma GN Jr. Confirmational identification of *Escherichia coli*: a comparison of genotypic and phenotypic assays for glutamate decarboxylase and β -D-glucuronidase. *Appl Environ Microbiol* **1996**;62:3350-4.
- Doyle MP, Shoeni JL. Survival and growth characteristics of *Escherichia coli* associated with hemorrhagic colitis. *Appl Environ Microbiol* **1984**;48:855-6.
- Novel M, Novel G. Regulation of β -glucuronidase synthesis in *Escherichia coli* K-12: pleiotropic constitutive mutations affecting *uxu* and *uidA* expression. *J Bacteriol* **1976**;127:418-32.
- Blanco C, Ritzenthaler P, Mata-Gilsinger M. Nucleotide sequence of a regulatory region of the *uidA* gene in *Escherichia coli* K-12. *Mol Gen Genet* **1985**;199:101-5.
- Fujisawa T, Mori M. Influence of bile salts on β -glucuronidase activity of intestinal bacteria. *Lett Appl Microbiol* **1996**;22:271-4.
- Feng P, Lampel KA. Genetic analysis of *uidA* gene expression in enterohemorrhagic *Escherichia coli* serotype O157:H7. *Microbiology* **1994**;140:2101-7.
- Jefferson RA, Burgess SM, Hirsh D. β -glucuronidase from *Escherichia coli* as a gene-fusion marker. *Proc Natl Acad Sci USA* **1986**;83:8447-51.
- Feng P, Lampel KA, Karch H, Whittam TS. Sequential genotypic and phenotypic changes in the emergence of *Escherichia coli* O157:H7. *J Infect Dis* **1998**;177:1750-3.
- Hayes PS, Blom K, Feng P, Lewis J, Strockbine NA, Swaminathan B. Isolation and characterization of a β -D-glucuronidase-producing strain of *Escherichia coli* O157:H7 in the United States. *J Clin Microbiol* **1995**;33:3347-8.
- Gunzer F, Bohm H, Russman H, Bitzan M, Aleksic S, Karch H. Molecular detection of sorbitol-fermenting *Escherichia coli* O157 in patients with hemolytic-uremic syndrome. *J Clin Microbiol* **1992**;30:1807-10.
- Sambrook J, Fritsch EF, Maniatis T. *Molecular cloning: a laboratory manual*. 2d ed. Cold Springs Harbor, NY: Cold Springs Harbor Laboratory Press, **1989**.
- Feng P, Monday SR. Multiplex PCR for detection of trait and virulence factors in enterohemorrhagic *Escherichia coli* serotypes. *Mol Cell Probes* **2000**;14:333-7.
- Feng P, Fields PI, Swaminathan B, Whittam TS. Characterization of non-motile *Escherichia coli* O157 and other serotypes by using an anti-flagellin monoclonal antibody. *J Clin Microbiol* **1996**;34:2856-9.
- Paton AW, Paton JC. Detection and characterization of Shiga toxin-producing *Escherichia coli* by using a multiplex PCR assay for *stx*₁, *stx*₂, *eaeA*, enterohemorrhagic *E. coli hlyA*, *rfb*_{O111}, and *rfb*_{O157}. *J Clin Microbiol* **1998**;36:598-602.