

A Novel Glucosyltransferase Involved in O-Antigen Modification of *Shigella flexneri* Serotype 1c^{∇†}

Robert M. Stagg,¹ Swee-Seong Tang,^{1,5} Nils I. A. Carlin,² Kaisar A. Talukder,³
Phung D. Cam,⁴ and Naresh K. Verma^{1*}

Biochemistry and Molecular Biology, Research School of Biology, The Australian National University, Canberra, ACT 0200, Australia¹; Department of Molecular Biology, SBL Vaccin AB, SE-105 21 Stockholm, Sweden²; International Centre for Diarrhoeal Disease Research, Dhaka, Bangladesh³; National Institute of Hygiene and Epidemiology, Department of Microbiology, Hanoi, Vietnam⁴; and University Malaya, Kuala Lumpur, Malaysia⁵

Received 13 May 2009/Accepted 23 August 2009

The O antigen of serotype 1c differs from the unmodified O antigen of serotype Y by the addition of a disaccharide (two glucosyl groups) to the tetrasaccharide repeating unit. It was shown here that addition of the first glucosyl group is mediated by the previously characterized *gtrI* cluster, which is found within a cryptic prophage at the *proA* locus in the bacterial chromosome. Transposon mutagenesis was performed to disrupt the gene responsible for addition of the second glucosyl group, causing reversion to serotype 1a. Colony immunoblotting was used to identify the desired revertants, and subsequent sequencing, cloning, and functional expression successfully identified the gene encoding serotype 1c-specific O-antigen modification. This gene (designated *gtrIC*) was present as part of a three-gene cluster, similar to other *S. flexneri* glucosyltransferase genes. Relative to the other *S. flexneri* *gtr* clusters, the *gtrIC* cluster is more distantly related and appears to have arrived in *S. flexneri* from outside the species. Analysis of surrounding sequence suggests that the *gtrIC* cluster arrived via a novel bacteriophage that was subsequently rendered nonfunctional by a series of insertion events.

Shigella flexneri is a pathovar of *Escherichia coli* that is the main causative agent of endemic bacillary dysentery (shigellosis). It is estimated that *S. flexneri* is responsible for approximately 100 million shigellosis cases annually, resulting in hundreds of thousands of deaths, predominantly in young children (11). Currently no vaccine is available, although there is evidence to suggest that serotype-specific immunity occurs following infection and that induction of immunity can be replicated with vaccines (9). *Shigella* serotype diversity arises due to differences in the chemical structure of the O-antigen repeating unit in the lipopolysaccharide, which is the main target of the adaptive host immune response following infection.

Because immunity to *S. flexneri* can be conferred by the induction of antibodies directed against the O antigen, an understanding of the prevalence of different serotypes and the underlying basis of serotype diversity can inform appropriate vaccine design. All *S. flexneri* serotypes (with the exception of serotype 6) share a common O-antigen backbone, consisting of a repeating tetrasaccharide unit that is comprised of one N-acetylglucosamine residue (GlcNAc) and three rhamnose residues (RhaI, RhaII, and RhaIII) (14). The 12 traditionally recognized *S. flexneri* serotypes differ by the presence or absence of just six different chemical modifications (glucosylations or O acetylations) of the O antigen. The genes responsible for these O-antigen modifications are introduced into the

bacterial genome via bacteriophages (3). Glucosylation of the *S. flexneri* O antigen is mediated by three genes [*gtrA*, *gtrB*, and *gtr(type)*] that are arranged in a single operon known as a *gtr* cluster. *gtrA* and *gtrB* are highly conserved between different *gtr* clusters and encode proteins involved in transferring the glucosyl group from the cytoplasm into the periplasm, where O-antigen modification is thought to take place. *gtr(type)* is unique to each *gtr* cluster and encodes a glucosyltransferase that is responsible for attaching the glucosyl group to a specific sugar unit of the O antigen via a specific linkage (3).

Investigations of *S. flexneri* have typically focused on serotypes for which commercially available typing sera are available. More recently, it has become clear that other serotypes are also epidemiologically important. In Bangladesh in the late 1980s, two novel *S. flexneri* strains that did not agglutinate with antibodies specific for the traditionally recognized serotypes were isolated (4). Chemical analysis of the O antigen revealed that these strains belonged to a new serotype, which was named serotype 1c due to the similarity its O antigen shares with the O antigens of serotype 1a and 1b strains (19). Serotype 1c has since been isolated in Egypt, Indonesia, Pakistan, and Vietnam (6, 15, 18). Serotype 1c was shown to be the most prevalent *S. flexneri* serotype in a northern province of Vietnam, accounting for more than a third of all *S. flexneri* strains isolated from 1998 to 1999 (15). Identification of serotype 1c currently relies on agglutination testing using monoclonal antibody MASF 1c (19).

The O antigen of serotype 1c is distinguished by the presence of a disaccharide (two glucosyl groups) linked to the GlcNAc in the tetrasaccharide repeating unit of the O antigen. The first glucosyl group is joined to GlcNAc via an $\alpha 1 \rightarrow 4$ linkage, as occurs in the O antigen of serotype 1a and serotype

* Corresponding author. Present address: Biochemistry and Molecular Biology, Research School of Biology, The Australian National University, Canberra, ACT 0200, Australia. Phone: 61 2 6125 2666. Fax: 61 2 6125 0313. E-mail: naresh.verma@anu.edu.au.

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

∇ Published ahead of print on 28 August 2009.

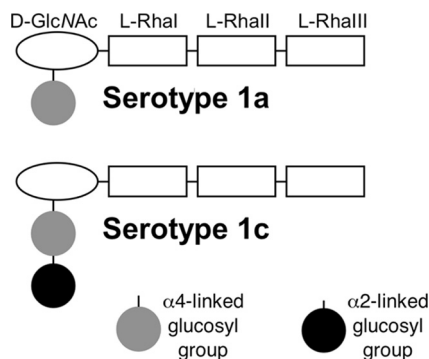


FIG. 1. Chemical structure of the tetrasaccharide repeat units in the O antigens of *S. flexneri* serotypes 1a and 1c. Note that the O antigen of serotype 1b (not shown) differs from that of serotype 1a by the O acetylation of L-RhaIII.

1b strains (type I modification). The O antigen of serotype 1c is distinguished by the presence of a second glucosyl group that is linked to the first via an $\alpha 1 \rightarrow 2$ linkage (Fig. 1). Type Ia modification is prerequisite to type 1c modification.

In this study, the genetic basis of O-antigen modification in serotype 1c was elucidated. Serotype 1c strains isolated from different locations and times were compared to gain insight into the evolution of this serotype. This is the first report of the identification of a glucosyltransferase gene that is responsible for addition of the second glucosyl group, causing serotype conversion from serotype 1a to serotype 1c.

MATERIALS AND METHODS

Bacterial strains. Bacteria were grown aerobically (~150 to 200 rpm) at 37°C in Luria Bertani broth (LB) or on LB agar plates (LBA). Antibiotics (Sigma-Aldrich) were added at the following final concentrations: ampicillin, 100 $\mu\text{g/ml}$; chloramphenicol, 25 $\mu\text{g/ml}$; tetracycline, 10 $\mu\text{g/ml}$; kanamycin, 50 $\mu\text{g/ml}$. Bacterial strains used in the study are listed in Table 1.

Agglutination. Serotyping was performed by slide agglutination. A sterile loop was used to mix bacteria from LB agar plates with a drop of antibody on a glass slide. The slide was gently agitated while being monitored for agglutination. Negative controls were performed using 0.9% NaCl instead of antibody. Isolates were tested using the monoclonal antibody MASF 1c (19) and also commercial antisera (Denka Seiken, Tokyo, Japan) directed against *S. flexneri* type and group factor antigens.

DNA techniques. Plasmid DNA and genomic DNA were prepared according to the methods of Sambrook et al. (13). Plasmids used in this study are listed in Table 1. Oligonucleotide primers used for PCR were purchased from Sigma Proligo (Lismore, Australia) and are listed in Table S1 in the supplemental material. PCR was performed using *Pfu* polymerase (Promega, Madison, WI) according to the manufacturer's instructions. DNA sequencing was performed using the Big Dye version 3.1 sequencing protocol and analyzed with the ABI 3730 capillary sequence analyzer at the Biomolecular Resources Facility, John Curtin School of Medical Research, Australian National University. Digestion of DNA was performed using enzymes supplied by Fermentas and ligations performed using T4 DNA ligase (Promega).

Colony immunoblotting. Bacteria were grown on sterile Nitrobind nitrocellulose transfer membranes (Micron Separations) overlaid on LBA plates containing the appropriate antibiotics. When colonies had grown to approximately 1 mm in diameter, bacterial cells were fixed to the membrane by baking at 80°C for 1 h. Excess, nonbound bacteria were washed off the membrane with Tris-saline solution (20 mM Tris-HCl, 150 mM NaCl; pH 7.4).

The membrane was blocked by incubation with 5% (wt/vol) skim milk powder in Tris-saline for 1 h. Primary and secondary antibodies were diluted in 1% skim milk in Tris-saline and incubated overnight and for 1 h, respectively. Each antibody incubation was followed by three washes with Tris-saline. Signal from the alkaline phosphatase-conjugated secondary antibody was detected using a SuperSignal West Pico detection kit (Pierce). Membrane signals were detected using Hyperfilm ECL film (Amersham Biosciences), which was developed with a Kodak X-Omat 1000 processor.

Southern blotting. Following agarose gel electrophoresis of digested genomic DNA samples, DNA was transferred to Hybond N⁺ nucleic acid transfer membranes (Amersham Biosciences) via capillary action. A Gigaprime DNA labeling kit (Bresatec) was used for generating [α -³²P]dCTP-labeled DNA probe.

Bacterial conjugation. Bacterial conjugation was used for shifting the transposon vector pUT-TnphoA from the host *E. coli* strain (E1212) into the serotype 1c strain Y394. E1212 and Y394 cultures were grown aerobically at 37°C in LB (with ampicillin for B1950) until the optical density at 600 nm reached 0.8. A 50- μl portion of each culture was then added to 5 ml of 10 mM MgSO₄. The mixture was vortexed and then passed through a sterile Millipore type GS

TABLE 1. Bacterial strains and plasmids

Strain or plasmid	Reference or source	Description
Strains		
ALX0559	6	Serotype 1c strain isolated in Egypt
K-265	16	Serotype 1c strain isolated in Bangladesh
K-212	16	Serotype 1c strain isolated in Bangladesh
Y394	19	Serotype 1c strain isolated in Bangladesh
Y53	2	Serotype 1a strain
E1212	Renato Marona	<i>E. coli</i> host strain of pUT-TnphoA
SFL124	10	Serotype Y strain containing a deletion of <i>aroD</i>
SFL1569	15	Serotype 1c strain isolated in Vietnam
SFL1622	This study	SFL124 expressing <i>gtrI</i> cluster (pNV1242) from Y394; agglutinates with type I sera
SFL1875	This study	Y394 where type 1c modification has been disrupted by transposon mutagenesis
SFL1905	This study	Recombinant Y53 strain expressing <i>gtrIC</i> (pNV1650); agglutinates with MASF 1c and type I antisera
SFL1925	This study	Knockout strain SFL1875 expressing <i>gtrIC</i> (pNV1650); agglutinates with MASF 1c
Plasmids		
pUC18	Fermentas	Commercial cloning vector
pBCSK ⁺	Stratagene	Commercial cloning vector
pNV1242	This study	<i>gtrI</i> cluster amplified from Y394 using <i>gtrAF</i> -(BamHI)pTRG and <i>gtrR</i> -(HindIII) and cloned into BamHI and HindIII sites of pBCSK ⁺
pUT-TnphoA	5, 8	AmpR plasmid pUT containing the KmR-encoding TnphoA
pNV1650	This study	<i>gtrIC</i> amplified from Y394 with <i>gtrIc-F</i> -(BamHI) and <i>gtrIc-R2</i> -(BamHI) and cloned into the BamHI site of pBCSK ⁺ in the same orientation as the <i>lac</i> promoter

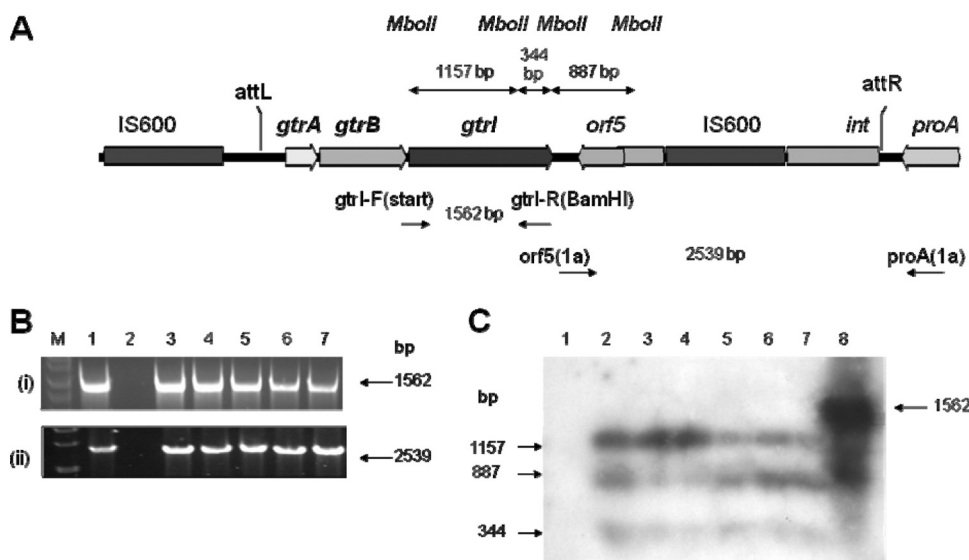


FIG. 2. Genetic arrangement of the *gtrI* cluster in serotype 1c strains. (A) *gtrI* cluster and surrounding sequence in serotype 1a strain Y53 as determined by Adhikari et al. (2). Primer binding sites and MboII cutting sites are shown. (B) PCR products amplified from different serotype 1c strains using two different primer pairs. M, marker. Lanes: 1, Y53 (serotype 1a control); 2, SFL1253 (serotype 4a negative control); 3, Y394 (serotype 1c); 4, K-265 (serotype 1c); 5, K-212 (serotype 1c); 6, ALX0559 (serotype 1c); 7, SFL1569 (serotype 1c). (C) Southern blot of MboII-digested genomic DNA probed with *gtrI*. Lanes: 1, SFL1253 (serotype 4a negative control); 2, Y394 (serotype 1c); 3, K-265 (serotype 1c); 4, K-212 (serotype 1c); 5, ALX0559 (serotype 1c); 6, SFL1569 (serotype 1c); 7, SFL1564 (serotype 1c); 8, *gtrI* (probe template).

0.22- μ m membrane. The membrane was transferred to an LBA plate (with the side containing the cells facing up) and incubated at 37°C overnight to allow conjugation. The following day, the bacterial lawn that had grown on the membrane was resuspended in 5 ml of 10 mM MgSO₄. Cells were plated out on LBA plates containing tetracycline to repress growth of B1950 and kanamycin to repress growth of Y394 cells that did not contain *TnphoA*.

Nucleotide sequence accession number. The DNA sequence determined in this study has been deposited in GenBank under accession number FJ905303.

RESULTS

Addition of the α 1 \rightarrow 4-linked glucosyl group to GlcNAc (type I modification). It was determined previously that the type I modification in serotype 1a strains is encoded by the *gtrI* cluster, which is present within a cryptic prophage that lies adjacent to the *proA* locus in the bacterial chromosome (2). It was hypothesized that the same genetic arrangement is responsible for type I modification in serotype 1c strains. The *gtrI* cluster was successfully amplified from the serotype 1c strain Y394 and cloned into the vector pBCSK⁺ to create the plasmid pNV1242. Sequencing demonstrated that the *gtrI* cluster in Y394 was identical to that of the previously published serotype 1a strain (2). pNV1242 was transformed into the serotype Y strain SFL124 (to create the strain SFL1622), and positive agglutination with type I antisera demonstrated functional expression of the *gtrI* cluster.

Genetic analysis was performed on a range of serotype 1c strains to determine if they each contained the *gtrI* cluster within the cryptic prophage adjacent to the *proA* locus (Fig. 2A). Successful PCR amplification indicated that *gtrI* was present in all serotype 1c strains (Fig. 2B, panel i), and in all cases the *gtrI* cluster was present in the cryptic prophage at the *proA* locus (Fig. 2B, panel ii). Southern blotting was performed to confirm results obtained by PCR analysis. Genomic DNA from serotype 1c strains was digested with MboII and probed

with [α -³²P]dCTP-labeled *gtrI*. Three bands of expected sizes were observed, consistent with each strain containing the *gtrI* cluster within a cryptic prophage (Fig. 2C).

Addition of the α 1 \rightarrow 2-linked glucosyl group (type Ic modification). A random transposon mutagenesis approach was used to identify the genetic locus responsible for type Ic modification. A library of knockout mutants was constructed, and clones where type Ic modification had been disrupted—resulting in reversion to serotype 1a—were identified by colony immunoblotting.

The vector pUT-*TnphoA*, which carries the transposon *TnphoA*, was introduced into the serotype 1c strain Y394 by conjugation. Following selection of cells that had taken up *TnphoA*, approximately 10,000 colonies were screened by colony immunoblotting using type I antiserum as the primary antibody. Eleven colonies that produced strong immunoblot signals were subjected to agglutination testing with type I antiserum, which confirmed that reversion to serotype 1a had occurred. One isolate (SFL1875) was selected for further study.

Genomic DNA from SFL1875 was digested separately with BamHI and EcoRI, and libraries were constructed by cloning the resulting fragments into pBCSK⁺ and pUC18, respectively. Ligation mixes were transformed into *E. coli* (OmniMAX 2-t1; Invitrogen) and clones containing SFL1875 sequence with the *TnphoA* insertion were selected by growth in the presence of kanamycin. The insert was sequenced by first using plasmid- and transposon-specific primers and then primer walking. Sequence from the cloned BamHI and EcoRI fragments of SFL1875 was collated.

The sequence revealed that in SFL1875, *TnphoA* had inserted into a *gtrB* homologue within a novel *gtr* cluster (Fig. 3). Upstream of this *gtrB* homologue is a *gtrA* homologue, and

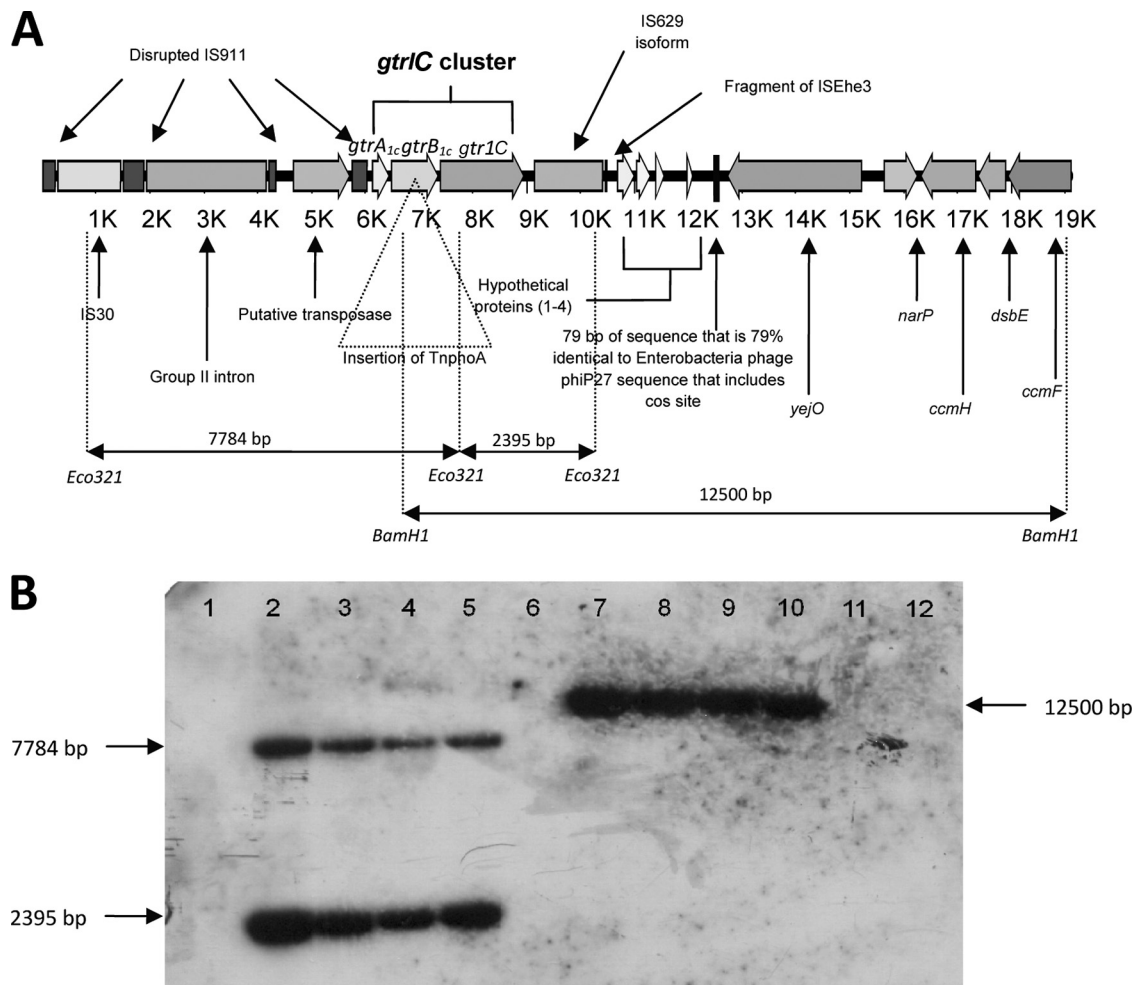


FIG. 3. Genetic arrangement of the *gtrIC* cluster in serotype 1c strains. (A) *gtrI* cluster and surrounding sequence in serotype 1c strain Y394. Eco321 and BamHI cutting sites are shown. (B) Southern blot of separately Eco321- and BamHI-digested genomic DNA probed with *gtrIC*. Lanes: 1, marker SPP1 EcoRI fragments; 2 and 7, Y394 (serotype 1c); 3 and 8, K-265 (serotype 1c); 4 and 9, SFL1569 (serotype 1c); 5 and 10, ALX0559 (serotype 1c); 6, Y53/SFL1416 (serotype 1a); 12, marker SPP1 XbaI fragments.

downstream is an open reading frame (ORF) that is 1,581 bp in length. It was postulated that the 1,581-bp ORF encoded the glucosyltransferase responsible for type 1c modification, and the ORF was tentatively named *gtrIC*. In other *S. flexneri gtr* clusters, expression of the specific glucosyltransferase is dependent on a promoter upstream of *gtrA*. It was hypothesized that insertion of *TnphoA* into *gtrB* had disrupted expression of *gtrIC*, leading to reversion to serotype 1a.

Functional testing of *gtrIC*. *gtrIC* was amplified from Y394 genomic DNA with the primers GtrIc-F(BamHI) and GtrIcR2-BamHI and cloned into the BamHI site of pBCSK⁺ in the same orientation as the *lac* promoter. The resultant plasmid (pNV1650) was transformed into the knockout strain SFL1875 (which created the strain SFL1925), and agglutination testing with MASF 1c confirmed that *gtrIC* complemented SFL1875 and conferred type 1c modification. pNV1650 was also transformed into the wild-type serotype 1a strain Y53 to create the strain SFL1905, which also showed positive agglutination when tested with MASF 1c. Type 1c modification in SFL1905 was

confirmed by colony immunoblotting with MASF 1c as the primary antibody (Fig. 4).

Colony immunoblotting was also performed using type I antiserum as the primary antibody. Similar to the finding of Wehler and Carlin (19), we could not detect the type I modification in the wild-type serotype 1c strain Y394. Presumably type 1c modification in the O antigen of Y394 is complete, and this prevents type I antibodies from binding to the type I modification (the first glucosyl group). In contrast, type I modification was detected in SFL1905, suggesting that type 1c modification in this strain was only partial (Fig. 4).

Analysis of the *gtrIC* cluster. *gtrIC* encodes a protein (GtrIc) of 526 amino acids that has a predicted molecular mass of 59.9 kDa. Searching the NCBI nonredundant protein database with Blastp revealed GtrIc homologues from 18 different bacterial species, but none of these were *E. coli* or *Shigella*. The highest scoring match was a hypothetical protein from *Citrobacter koseri* (a gram-negative enterobacterium) that is encoded within a *gtr* cluster that includes *gtrA* and *gtrB* homologues.

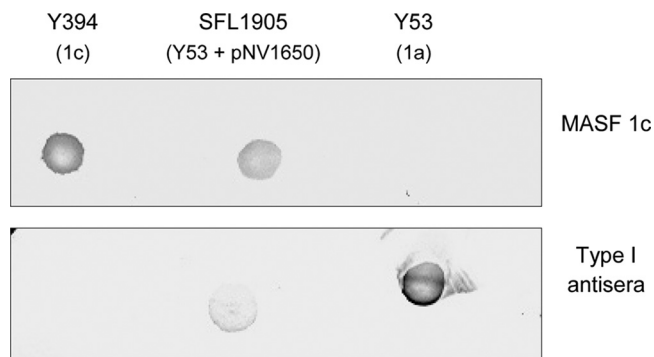


FIG. 4. Colony immunoblots showing the extent of O-antigen modification in SFL1905, compared to control strains. The primary antibodies used for each blot are indicated. All antibodies were diluted in 1% skim milk. MASF 1c was diluted 1/500 and detected using an anti-mouse immunoglobulin M peroxidase conjugate (Sigma Aldrich) secondary antibody, diluted 1/1,000. Type I was diluted 1/2,000 and detected using an anti-rabbit immunoglobulin G peroxidase conjugate (Sigma Aldrich), diluted 1/2,000.

Multiple alignments of GtrIc and a selection of closely related proteins (identified via Blastp), as well as the other *S. flexneri* Gtr's, were performed. The phylogenetic tree suggests that GtrIc is more similar to proteins from species such as *C. koseri*, *Streptococcus mutans*, and *Lactococcus lactis* than it is to Gtr's from *S. flexneri* (Fig. 5).

*gtrA*_{1c} and *gtrB*_{1c} share <75% identity at the nucleotide level with the corresponding genes in the five previously identified *S. flexneri* *gtr* clusters (*gtrI*, *gtrII*, *gtrIV*, *gtrV*, and *gtrX* clusters). *gtrA* and *gtrB* sequences from the other five *gtr* clusters show higher levels of conservation, sharing between 84% and 99% identity with each other.

gtrIC was amplified from a further three serotype 1c strains that had been isolated from patients in Vietnam (SFL1569), Egypt (ALX0559), and Bangladesh (K-265). All *gtrIC* sequences were identical to that found for Y394, except for *gtrIC* amplified from strain K-265, which contained a 6-bp deletion that resulted in the omission of a tryptophan and a lysine residue at the 3' terminus of GtrIc.

Analysis of the region surrounding *gtrIC* cluster. Approximately 7 kb of sequence upstream of the *gtrIC* cluster and 12 kb of sequence downstream of the cluster were obtained (Fig. 3A). The sequence upstream of the *gtrIC* cluster in strain Y394 is comprised of insertion elements. The insertion sequence IS911 has been disrupted three times, by IS30, a group II intron, and another putative insertion sequence.

Immediately downstream of *gtrIC* in Y394 is an isoform of the insertion element IS629 (82% identity) that has inserted into the end of another insertion element, ISEhe3. This is followed by approximately 2 kb of sequence that contains four putative ORFs and a 79-bp stretch of sequence that shares homology (79% identity) with a region of the enterobacterial phage ϕ P27 that contains a Cos site. The Cos site is not intact in Y394, and the functions of the four hypothetical proteins are unknown.

A further 6,537 bp of sequence downstream of *gtrIC* was found to correspond to *S. flexneri* housekeeping genes. This region includes *yejO* (putative ATP-binding component of a transport system), *narP* (nitrate/nitrite response regulator),

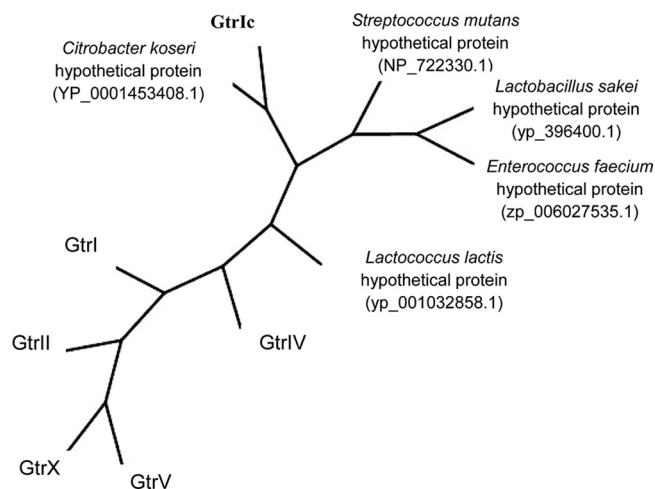


FIG. 5. Unrooted phylogenetic tree showing the relatedness of GtrIc to other *S. flexneri* glucosyltransferases and the five most closely related proteins identified by Blastp search (accession numbers are shown). Branch lengths are proportional to estimated divergence along each branch. Multiple protein alignment was performed using the default settings of the ClustalW program (17). Phylogenetic tree was constructed using the Protml program (1). Both programs were accessed via the Biomanager website (<http://www.angis.org.au>).

ccmH (possible subunit of heme lyase), *dsbE* (disulfide oxidoreductase), and *ccmF* (cytochrome *c*-type biogenesis protein). This stretch of Y394 sequence is 99% identical to that found in the serotype 2a strain 2457T. Of note, a tRNA^{Pro} lies adjacent to the *yejO* gene in serotype 2a (2457T) and serotype 5a (8401) strains (12, 20), but this has been disrupted in Y394.

Southern blotting was performed to analyze serotype 1c strains to determine whether each of them contained the *gtrIC* cluster adjacent to the *yejO* locus. Genomic DNA from serotype 1c strains (isolates from Bangladesh, Egypt, and Vietnam) were digested with Eco321 and BamHI separately and probed with an [α -³²P]dCTP-labeled *gtrIC* sequence that had been amplified using the primers GtrIc-F(BamHI) and GtrIcR2-BamHI. Two bands (7,784 and 2,395 bp) were observed as expected for Eco321 digestion. One band (12,500 bp) was observed from BamHI digestion as predicted (Fig. 3). These results clearly show that the same gene arrangement exists in all 1c strains and in each strain containing the *gtrIC* cluster adjacent to *yejO* locus.

DISCUSSION

In serotype 1a strains, the type I modification (glucosylation of the GlcNAc in the O antigen, via an α 1 \rightarrow 4 linkage) is encoded by the *gtrI* cluster (2). It was shown here that the *gtrI* cluster is also responsible for type I modification in serotype 1c strains. The arrangement of the *gtrI* cluster was investigated for serotype 1c strains isolated from Bangladesh, Vietnam, and Egypt. PCR and Southern blot analysis showed that in all serotype 1c strains evaluated, the *gtrI* cluster is present in a cryptic prophage at the *proA* locus in the bacterial chromosome (identical to serotype 1a strains). All serotype 1a and 1c strains examined to date are descended from a single parental

strain, in which a *gtrI*-encoding bacteriophage was disrupted and largely deleted by insertion elements.

The gene responsible for type 1c modification (*gtrIC*) was identified here for the first time. *gtrIC* was present within a *gtr* cluster that was arranged similarly to the *gtr* clusters present in other *S. flexneri* serotypes. However, phylogenetic analysis indicated that the *gtrIC* cluster was more closely related to *gtr* clusters from non-*S. flexneri* species. This suggests that the *gtrIC* cluster did not arise following a duplication of the *gtrI* cluster, or any of the other known *S. flexneri gtr* clusters, but arrived separately in *S. flexneri* from outside the species.

The *gtrI* and *gtrIC* clusters (located at the *proA* and *yejO* loci, respectively) are probably located on opposite sides of the chromosome of serotype 1c strain Y394. In other *S. flexneri* and *E. coli* strains, these loci are separated by 2 Mb (the entire *S. flexneri* chromosome is approximately 4.5 Mb). It is probable that the *gtrI* and *gtrIC* clusters were introduced via two separate bacteriophages. It was previously concluded that the *gtrI* cluster was introduced into *S. flexneri* by a bacteriophage that integrated adjacent to the *proA* locus (2, 7). Analysis of sequence adjacent to the *gtrIC* cluster suggests that it may also have been the site of a bacteriophage insertion.

Due to the large number of insertion events surrounding both the *gtrI* and *gtrIC* clusters, it will not be possible to definitely determine the origins of these sequences unless serotype 1a and 1c strains without the insertion elements are identified. All serotype 1a and serotype 1c strains identified to date appear to be derived from a single ancestral serotype 1a strain in which the *gtrI* cluster was introduced via a bacteriophage that was subsequently disrupted by insertion elements. Based on currently available sequence, it is hypothesized that serotype 1c strains arose following the introduction of the *gtrIC* cluster via a bacteriophage that inserted into a separate location on the chromosome of the ancestral serotype 1a, with subsequent disruption of this bacteriophage by insertion elements.

Colony immunoblotting showed that when the serotype 1a strain Y53 expressed *gtrIC* (SFL1905), a high proportion of type I modifications underwent type 1c modification. However, in contrast to the wild-type serotype 1c strain Y394, some of the type I modifications did not undergo further modification and could be detected by type I antiserum in an immunoblot. The reason for incomplete type 1c modification in SFL1905 has not been determined. It is possible that the introduction of *gtrIC* via a plasmid leads to suboptimal production of GtrIC and this results in incomplete type 1c modification. Alternatively, it is possible that *gtrIC* alone is insufficient to mediate complete type 1c modification. It was assumed here that *gtrA_{1c}* and *gtrB_{1c}* in Y394 are redundant, due to the presence of *gtrA₁* and *gtrB₁* in this strain. Indeed, a high level of type 1c modification was observed in SFL1905 in the absence of these genes. However, it is possible that GtrIc operates most efficiently when one or both of GtrA_{1c} and GtrB_{1c} are present and that at least one of these is required for complete conversion to serotype 1c.

In summary, the gene that encodes serotype 1c-specific O-antigen modification (*gtrIC*) was identified for the first time in this study. Relative to the other *S. flexneri gtr* clusters, the *gtrIC* cluster is more distantly related and appears to have arrived in *S. flexneri* from outside the species. Analysis of the surrounding sequence suggests that the *gtrIC* cluster arrived via a novel bacteriophage. The findings from this study give insight into the evolution of this

serotype and may potentially have important implications for the development of a serotype-specific vaccine.

ACKNOWLEDGMENTS

We thank Renato Marona for providing *E. coli* host strain E1212 of plasmid pUT-Tn*phoA*, and A. El-Gendy for *S. flexneri* serotype 1c strain.

This work was supported by a grant from the National Health and Medical Research Council of Australia.

REFERENCES

- Adachi, J., and M. Hasegawa. 1996. Computer science monographs, no. 28, MOLPHY version 2.3: programs for molecular phylogenetics based on maximum likelihood. Institute of Statistical Mathematics, Tokyo, Japan.
- Adhikari, P., G. Allison, B. Whittle, and N. K. Verma. 1999. Serotype 1a O-antigen modification: molecular characterization of the genes involved and their novel organization in the *Shigella flexneri* chromosome. *J. Bacteriol.* **181**:4711–4718.
- Allison, G. E., and N. K. Verma. 2000. Serotype-converting bacteriophages and O-antigen modification in *Shigella flexneri*. *Trends Microbiol.* **8**:17–23.
- Carlin, N. L., M. Rahman, D. A. Sack, A. Zaman, B. Kay, and A. A. Lindberg. 1989. Use of monoclonal antibodies to type *Shigella flexneri* in Bangladesh. *J. Clin. Microbiol.* **27**:1163–1166.
- de Lorenzo, V., M. Herrero, U. Jakubzik, and K. N. Timmis. 1990. Mini-Tn5 transposon derivatives for insertion mutagenesis, promoter probing, and chromosomal insertion of cloned DNA in gram-negative eubacteria. *J. Bacteriol.* **172**:6568–6572.
- El-Gendy, A., N. El-Ghorab, E. M. Lane, R. A. Elyazed, N. I. Carlin, M. M. Mitry, B. A. Kay, S. J. Savarino, and L. F. Peruski, Jr. 1999. Identification of *Shigella flexneri* subserotype 1c in rural Egypt. *J. Clin. Microbiol.* **37**:873–874.
- Guan, S., and N. K. Verma. 1998. Serotype conversion of a *Shigella flexneri* candidate vaccine strain via a novel site-specific chromosome-integration system. *FEMS Microbiol. Lett.* **166**:79–87.
- Herrero, M., V. de Lorenzo, and K. N. Timmis. 1990. Transposon vectors containing non-antibiotic resistance selection markers for cloning and stable chromosomal insertion of foreign genes in gram-negative bacteria. *J. Bacteriol.* **172**:6557–6567.
- Jennison, A. V., and N. K. Verma. 2004. *Shigella flexneri* infection: pathogenesis and vaccine development. *FEMS Microbiol. Rev.* **28**:43–58.
- Karnell, A., B. A. Stocker, S. Katakura, F. P. Reinhold, and A. A. Lindberg. 1992. Live oral auxotrophic *Shigella flexneri* SFL124 vaccine with a deleted *aroD* gene: characterization and monkey protection studies. *Vaccine* **10**:389–394.
- Kotloff, K. L., J. P. Winickoff, B. Ivanoff, J. D. Clemens, D. L. Swerdlow, P. J. Sansonetti, G. K. Adak, and M. M. Levine. 1999. Global burden of *Shigella* infections: implications for vaccine development and implementation of control strategies. *Bull. W. H. O.* **77**:651–666.
- Nie, H., F. Yang, X. Zhang, J. Yang, L. Chen, J. Wang, Z. Xiong, J. Peng, L. Sun, J. Dong, Y. Xue, X. Xu, S. Chen, Z. Yao, Y. Shen, and Q. Jin. 2006. Complete genome sequence of *Shigella flexneri* 5b and comparison with *Shigella flexneri* 2a. *BMC Genomics* **7**:173.
- Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Simmons, D. A., and E. Romanowska. 1987. Structure and biology of *Shigella flexneri* O antigens. *J. Med. Microbiol.* **23**:289–302.
- Stagg, R. M., P. D. Cam, and N. K. Verma. 2008. Identification of newly recognized serotype 1c as the most prevalent *Shigella flexneri* serotype in northern rural Vietnam. *Epidemiol. Infect.* **136**:1134–1140.
- Talukder, K. A., Z. Islam, M. A. Islam, D. K. Dutta, A. Safa, M. Ansaruz-zaman, A. S. Faruque, S. N. Shahed, G. B. Nair, and D. A. Sack. 2003. Phenotypic and genotypic characterization of provisional serotype *Shigella flexneri* 1c and clonal relationships with 1a and 1b strains isolated in Bangladesh. *J. Clin. Microbiol.* **41**:110–117.
- Thompson, J. D., D. G. Higgins, and T. J. Gibson. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**:4673–4680.
- von Seidlein, L., D. R. Kim, M. Ali, H. Lee, X. Wang, V. D. Thiem, D. G. Canh, W. Chaicumpa, M. D. Agtini, A. Hossain, Z. A. Bhutta, C. Mason, O. Sethabutr, K. Talukder, G. B. Nair, J. L. Deen, K. Kotloff, and J. Clemens. 2006. A multicentre study of *Shigella* diarrhoea in six Asian countries: disease burden, clinical manifestations, and microbiology. *PLoS Med.* **3**:e353.
- Wehler, T., and N. I. Carlin. 1988. Structural and immunochemical studies of the lipopolysaccharide from a new provisional serotype of *Shigella flexneri*. *Eur. J. Biochem.* **176**:471–476.
- Wei, J., M. B. Goldberg, V. Burland, M. M. Venkatesan, W. Deng, G. Fournier, G. F. Mayhew, G. Plunkett III, D. J. Rose, A. Darling, B. Mau, N. T. Perna, S. M. Payne, L. J. Runyen-Janecky, S. Zhou, D. C. Schwartz, and F. R. Blattner. 2003. Complete genome sequence and comparative genomics of *Shigella flexneri* serotype 2a strain 2457T. *Infect. Immun.* **71**:2775–2786.