Emerging Microbes and Infections (2012) 1, e21; doi:10.1038/emi.2012.22

www.nature.com/emi

ORIGINAL ARTICLE

Identification of a divergent *O*-acetyltransferase gene oac_{1b} from *Shigella flexneri* serotype 1b strains

Qiangzheng Sun[1,*], Ruiting Lan[2,*], Yan Wang[1,*], Jianping Wang[1,*], Shengli Xia[3,*], Yiting Wang[1], Jin Zhang[3], Deshan Yu[4], Zhenjun Li[1], Huaiqi Jing[1] and Jianguo Xu[1]

Shigella flexneri **is a leading cause of bacterial dysentery in developing countries. Among the 15 known serotypes, four (1b, 3a, 3b and 4b) contain a group 6 epitope due to an acetyl group connected to the O-2 position of rhamnose III on the tetrasaccharide structure of the lipopolysaccharide. *O*-acetyltransferase encoded by a bacteriophage, Sf6, mediates the acetylation reaction. We found that the *oac* gene in serotype 1b strains was very different from that in serotypes 3a, 3b and 4b strains and is herein after referred to as oac_{1b} which shares with *oac* 88%–89% identity at the DNA level and 85% identity at the protein level. Considering that *S. flexneri* strains of serotypes 1–5 share a recent common ancestry, the divergent oac_{1b} is more likely to have been obtained from outside *S. flexneri* than to have undergone rapid divergence from the *oac* gene in the other serotypes (3a, 3b and 4b) within *S. flexneri*. The cloned oac_{1b} gene was found to perform the same acetylation function as *oac*. Analysis of the genomic regions flanking oac_{1b} showed that it was present in a prophage on the chromosome and the organizational structure is different from that of phage Sf6. Additionally, phage conversion assay showed that serotype 1b cannot be generated by infecting serotype 1a strains with Sf6. We conclude that oac_{1b} was carried by a non-Sf6 phage and is uniquely present in serotype 1b.**
Emerging Microbes and Infections (2012) **1,** e21; doi:10.1038/emi.2012.22; published online 5 September 2012

Keywords: *O*-acetyltransferase; serotype; serotype 1b; *Shigella flexneri*

INTRODUCTION

Shigella flexneri is the major pathogen responsible for bacterial dysentery in developing countries.[1] It is estimated that there are 164.7 million cases of shigellosis worldwide annually, resulting in 1.1 million deaths, most of which are children under the age of 5 years.[1] A more recent study estimated approximately 125 million annual shigellosis cases and 14 000 related deaths in Asia.[2] In China, *S. flexneri* is the most common *Shigella spp.*, accounting for up to 80% of shigellosis cases.[3–5]

Based on the O-antigen structure of the lipopolysaccharide (LPS), *S. flexneri* is currently divided into 15 serotypes.[5,6] With the exception of serotype 6, all share a common polysaccharide backbone comprised of repeating tetrasaccharide units (*N*-acetylglucosamine–rhamnose–rhamnose–rhamnose). Serologically, serotypes 1b, 3a, 3b and 4b strains may cross-react with group 6 antisera. Structural analysis also revealed that the LPS tetrasaccharide backbone in serotypes 1b, 3a, 3b and 4b all contain an acetyl group connected to the O-2 position in the rhamnose III, and this acetylation results in the appearance of the group 6 epitope.

Currently, it is generally believed that O-antigen acetylation is mediated by an *O*-acetyltransferase (Oac) encoded by the *oac* gene carried by the temperature bacteriophage Sf6.[7,8] In 1975, Gemski *et al.*[9] first reported that phage Sf6 could be isolated from a *S. flexneri* serotype 3a strain. Clark *et al.*[7] and Verma *et al.*[8] independently identified the *oac* gene from phage Sf6, which is 1002 bp in size, encoding a protein of 333 amino acids. Sequence comparison showed that Oac shares homology with a variety of proteins involved in *O*-acetylation.[10,11] Oac is an integral membrane protein with 10 transmembrane segments, and Oac function is associated with residues within cytoplasmic and periplasmic loops.[12] Residues R73, and R75R76 within cytoplasmic loop 3 are critical to the Oac function.[12] The *oac* gene cloned from phage Sf6 was shown to be capable of converting serotype X, Y, la and 4a to 3a, 3b, lb and 4b, respectively.[7]

In this study, we amplified and sequenced the *oac* gene from 36 serotype 1b, 3a, 3b and 4b strains, and found that the serotype 1b strains possess a divergent *oac* gene, herein named oac_{1b}. We further characterized the gene and genetic organization of the regions flanking oac_{1b} to show that $oac_{1b}$was not introduced by the phage Sf6, but by a potential novel phage.

[1]State Key Laboratory for Infectious Disease Prevention and Control, National Institute for Communicable Disease Control and Prevention, China CDC, Beijing 102206, China; [2]School of Biotechnology and Biomolecular Sciences, University of New South Wales, Sydney NSW 2052, Australia; [3]Branch for Enteric Disease Control and Prevention, Institute for Infectious Disease Control and Prevention, Henan Center for Disease Control and Prevention, Zhengzhou 450016, China and [4]Gansu Center for Disease Control and Prevention, Lanzhou 730000, China
*These authors contributed equally to this work.
Correspondence: JG Xu
E-mail: xujianguo@icdc.cn
Received 16 March 2012; revised 5 June 2012; accepted 3 July 2012

Table 1 Properties of *S. flexneri* serotype 3a, 3b, 4b and 1b strains analyzed in this study

Serotype	Strain	Regions of isolation	Host	Year of isolation	DNA sequence identity to *oac* of Sf6	*oac* genotype
3a	2002061	China/Henan	Human	2002	99	*oac*
	2002074	China/Henan	Human	2002	99	
	2002127	China/Henan	Human	2002	99	
	2002133	China/Henan	Human	2002	99	
	03HL12	China/Heilongjiang	Human	2003	99	
	06GS02	China/Gansu	Human	2006	99	
	06GS03	China/Gansu	Human	2006	99	
	HB05	China/Hubei	Human	2008	100	
	51575	China/Gansu	Human	—	100	
3b	2002110	China/Henan	Human	2002	99	*oac*
	20030363b	China/Henan	Human	2003	99	
4b	NCTC8336	England/London	Monkey	1947	100	*oac*
	NCTC8522	England/Birmingham	Human	1951	100	
	NCTC8598	England/London	Monkey	1953	100	
	NCTC9726	USA/Atlanta	—	1955	99	
	51577	China/Sichuan	Human	—	100	
1b	NCTC5	England/London	Human	1920	89	oac_{1b}
	1997005	China/Henan	Human	1997	89	
	1997019	China/Henan	Human	1997	89	
	1997020	China/Henan	Human	1997	89	
	1997021	China/Henan	Human	1997	89	
	1997022	China/Henan	Human	1997	89	
	2003070	China/Henan	Human	2003	89	
	2003071	China/Henan	Human	2003	89	
	2003072	China/Henan	Human	2003	89	
	2003073	China/Henan	Human	2003	89	
	2005020	China/Henan	Human	2005	89	
	06HN87	China/Henan	Human	2006	89	
	07GS73	China/Gansu	Human	2007	88	
	07HN57	China/Henan	Human	2007	89	
	09GS62	China/Gansu	Human	2009	89	
	09GS70	China/Gansu	Human	2009	88	
	09GS119	China/Gansu	Human	2009	88	
	51572	China/Shandong	Human	—	89	
	M1250	Australia	Human	—	89	
	M1349	Unknown[a]	Human	—	89	

[a] The source of M1349[18] is not known but not isolated from China.

MATERIALS AND METHODS

Bacterial strains and culture conditions

The *S. flexneri* strains used for *oac* gene analysis in this study were shown in Table 1. All strains were identified biochemically using the Dade Behring MicroScan WalkAway 40 (Dade Behring, Hessen, Germany). Serotypes were confirmed using two commercial serotyping antiserum kits: the antisera made by Denka Seiken (Tokyo, Japan) and the monoclonal antibodies against *S. flexneri* (Reagensia AB, Stockholm, Sweden). Ampicillin sensitive *S. flexneri* strains 03XZ014 (serotype Y), NCTC9725 (serotype 4a), 05004 (serotype 1a) and 04SH03 (serotype X) were used as hosts for *oac* gene functional analysis. *S. flexneri* strains 03HL12 (serotype 3a) and 019 (serotype 1a)[13] were used as hosts for phage Sf6 and SfI induction, respectively. Four serotype X strains (014, 51580, 04SH03 and 062), 11 serotype Y strains (036, 035, 51581, 017, 03XZ014, 038, 043, 064, 065, 025 and 026), 12 serotype 1a strains (51571, 019, GS30, HB31, 080, SX25, QH20, SX12, 05004, HN184, QH37 and AH93), 4 serotype 4a strains (NCTC9725, NCTC8296, NCTC7885 and 004) and 2 serotype 3b strains (110 and 061) were used for Sf6 infection experiments. All the Chinese strains were isolated from diarrheal patients. Other strains were obtained from National Collection of Type Cultures (NCTC). All strains were generally grown at 37 ℃ in Luria broth (LB) with agitation, or on LB agar.

Oligonucleotide primers, PCR and DNA sequencing

Primers used in this study were listed in Table 2. All primers were synthesized by Sangon Biotech (Shanghai, China). Unless otherwise stated, PCR amplification was performed using a standard protocol with the following thermal cycling profile: 94 ℃ for 5 min followed by 30 cycles of 94 ℃ for 30 s, 55 ℃ for 50 s and 72 ℃ for 5 min, on a SensoQuest LabCycler (SensoQuest, Germany). Walking PCR was performed using the Genome Walking PCR Kit (TaKaRa, Kyoto, Japan) according to the manufacturer's protocol. PCR products were either sequenced directly or cloned into the pMD20T TA cloning vector (TaKaRa, Japan) for sequencing.

oac gene sequencing and functional analysis

The full length of the *oac* gene was amplified using primer pair O1 (Table 2). DNA and deduced protein sequence comparison was performed using BLASTn or BLASTp at NCBI

Table 2 Primers used in this study

Primer pairs	Primer sequence (5′-3′)	Target gene or fragments	Amplicon size (bp)
O1	O1U: TCA ATC CAG GGA TAA TTT AGG CG O1L: ATG CAT AAG AGC AAC TGC TTT GA	*oac* or oac_{1b}	1002
O2	O2U: TAG AAA CAG AAG CCA CTG GAG CAC C O2L: GCT GCG TGG AAA AGA ACT CCA CCT T	oac_{1b}, promoter and terminator	1260
O3	O3U: CCG CCA GGA TGG TGA AAA AGA G O3L: AGA ACG CCA GTC CAC GCA AAG G	*yfdC-oac*	3034
O4	O4U: ATG CAT AAG AGC AAC TGC TTT GA O4L: GGT TTA TGG CTG GGT ATT TGA T	*oac-tsp*	2075
O5	O5U: CCG CCA GGA TGG TGA AAA AGA GC O5L: GGT TTA TGG CTG GGT ATT TG	*yfdC-tsp*	4489

(http://www.ncbi.nlm.nih.gov/). Additionally, the fragment (nt 2540–3799 of accession NO. JF450728) carrying the *oac* gene, together with the putative promoter and terminator regions (nts 2615–2643 and 3683–3698 of accession NO. JF450728, respectively) was amplified using primer pair O2 (Table 2). Purified products were cloned into pMD20T (Amp^r) vector and transformed into *S. flexneri* strains 03XZ014 (serotype Y), NCTC9725 (serotype 4a), 05004 (serotype 1a) and 04SH03 (serotype X), respectively. Transformation was performed by electroporation using a Bio-Rad Gene Pulsar (BioRad, Hercules, CA, USA). Transformants were identified by PCR amplification of *oac* gene and serotyping by slide agglutination with both monovalent antisera (Denka Seiken, Japan) and monoclonal antibodies against *S. flexneri* (Reagensia AB, Sweden).

Characterization of the regions flanking the oac_{1b} gene

Based on known arrangements of Sf6 genome in its host strains,[14] primer pairs O3, O4 and O5 (Table 2), which are complementary to sequences of *yfdC*, *oac* and *tsp* genes, respectively, were designed for PCR identification of the chromosomal regions flanking *oac* gene in serotypes 3a, 3b and 4b strains. In order to identify the regions flanking oac_{1b} in serotype 1b, we first performed PCR walking starting from gene oac_{1b}, and then used Illumina Solexa sequencing on the whole genome of strain 1997020. Genomic DNA was extracted from broth culture using a Wizard Genomic DNA Purification kit (Promega, Madison, WI, USA) with methods as described in their manual. A paired-end library was constructed, and the average length of insert was about 500 bp. Reads were generated with Illumina Solexa GA IIx (Illumina, San Diego, CA, USA) and re-assembled into scaffolds using SOAPdenovo (Release 1.04). Fragments carrying the gene oac_{1b} and flanking sequence were extracted. Open reading frames (ORFs) were determined using the ORF Finder program, which is accessible through the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/gorf/gorf.html), and conformed to the codon usage table for *Escherichia coli*. Searches for homologous DNA and protein sequences were conducted with the BLAST software against the non-redundant GenBank database (http://www.ncbi.nlm.nih.gov/blast/blast/). Based on the DNA sequence of the oac_{1b} carrying fragment in 1997020 (accession NO. JN377795), a series of primers were designed and used for overlapping PCR to confirm the genetic structures in the other 19 serotype 1b strains.

Phages conversion assay

Phages Sf6 and SfI used in this study were induced, isolated and purified from *S. flexneri* strains 03HL12 (serotype 3a) and 019 (serotype 1a), respectively, using the methods as described previously.[13] Phage infection and identification of lysogens were performed essentially according to the methods for phage λ with the following modifications.[15] Firstly, *S. flexneri* host strains were inoculated in LB and incubated for 3 h at 37 ℃ with shaking. Cells were collected by centrifugation when OD_{600} reached 1.2 and resuspended in $MgSO_4$ (10 mM). Then 100 μl purified phage particles were added into 200 μl competent cells at a ratio of 1 phage to 100 cells. After further incubation at 37 ℃ for 20 min, 3 ml of semisolid agar (LB with 0.7% (w/v) agar) were added and mixture was laid on the Brain-Heart solid medium, and then incubated at 37 ℃. The area of turbid growth was streaked for single colony isolation and serotype identification.

Nucleotide sequence accession number

The nucleotide sequences obtained in this study have been published in GenBank (accession NOs. JF450698–JF450729 and JN377795).

RESULTS

A new oac_{1b} gene was identified from *S. flexneri* serotype 1b strains, which was divergent from the *oac* genes in serotypes 3a, 3b and 4b strains

Serotypes 1b, 3a, 3b and 4 are known to contain an *O*-acetyl group and thus carry an *oac* gene. We sequenced the *oac* gene from 36 *S. flexneri* strains to determine its diversity. The *oac* genes in serotype 3a, 3b and 4b strains were highly homologous; 6 strains (HB05, 51575, NCTC8522, NCTC8598, NCTC8336 and 51577) had an identical sequence, whereas the remaining 10 strains were nearly identical, but differed by one base (334, A→G) from the other six strains (Table 1). The *oac* gene from phage Sf6,[9] which was derived from an *S. flexneri* serotype 3a strain was identical to the former group of strains.

Sequences of the *oac* genes amplified from all 20 *S. flexneri* serotype 1b strains were also highly homologous to each other. All except three strains were identical with the remaining three strains (07GS73, 09GS70 and 09GS119) differing by one base (783, C→T) (Table 1). Surprisingly, the *oac* gene in serotype 1b was significantly different from that in serotype 3a, 3b and 4b strains (Table 1). There are 110 base changes in the 1002 bp gene, of which, 72 and 38 are synonymous and non-synonymous changes respectively. The distribution of these changes along the gene is shown in Figure 1A. We calculated the ratio of synonymous and non-synonymous substitution rates (K_a/K_s) using a sliding window. We found that the ratios in 2 regions are much higher than the rest of the gene (Figure 1A), suggesting that these regions may have been subjected to diversifying selection. The three major regions (amino acid residues 40–57, 69–81 and 138–156) conserved among the inner membrane *trans*-acylase family proteins[10] show very low K_a/K_s ratio (Figure 1A and 1B). The residues R73 and R75R76, known to be critical for Oac function,[12] are also conserved (Figure 1A and 1B).

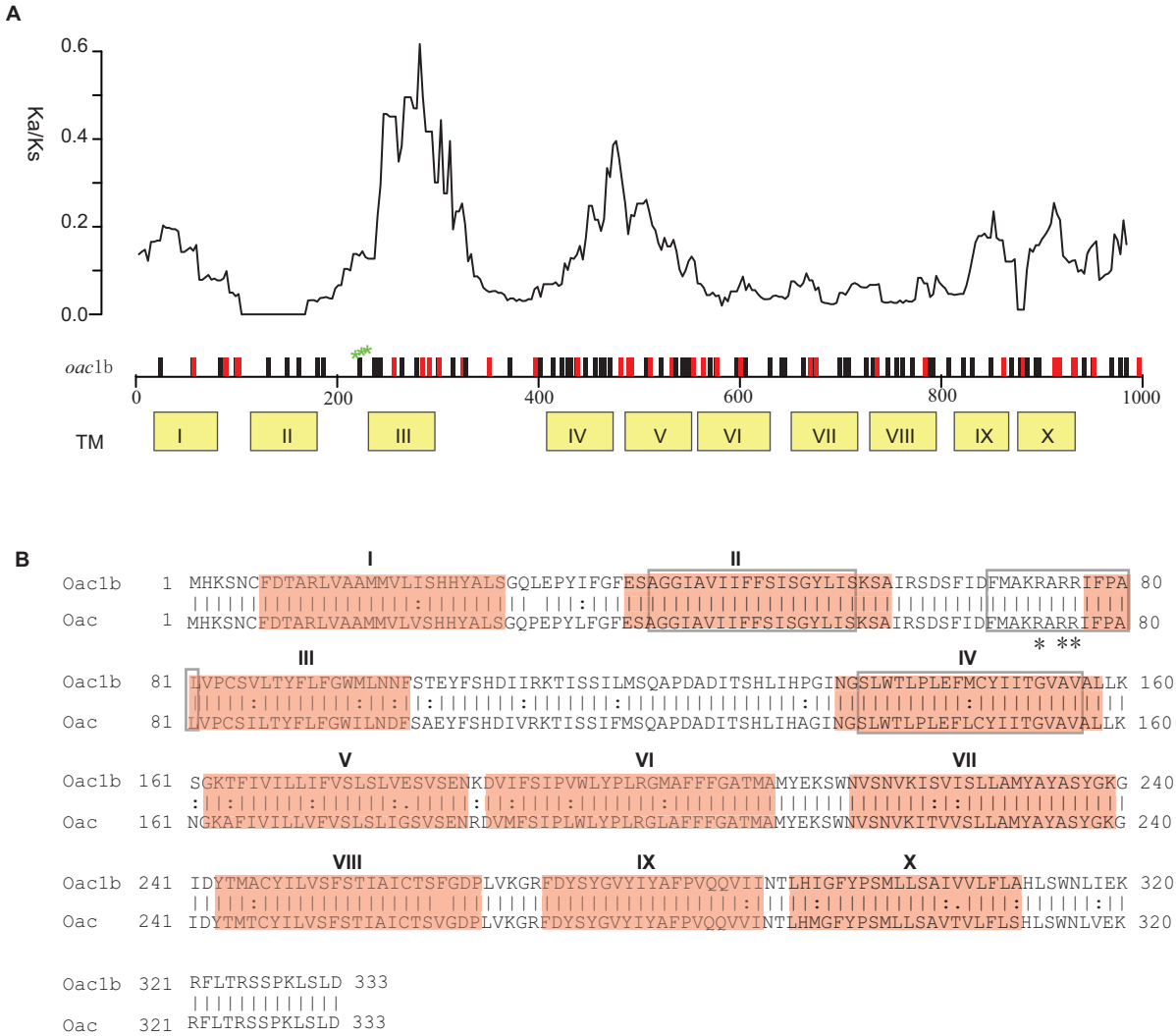

Figure 1 Comparison of protein Oac_{1b} and Oac. (**A**) Plot of variation between Oac_{1b} and Oac. Nucleotide differences were plotted across the horizontal axis with synonymous and non-synonymous mutations plotted as black and red vertical lines. Ticks and numbers across the horizontal axis are base positions. Transmembrance (TM) segments as predicted by Verma *et al.* are shown in yellow boxes below the gene. The critical residues for Oac function are indicated by green stars. The ratio of non-synonymous and synonymous substitution rate (K_a/K_s) across the gene using a sliding window of 90 bases (30 codons) and overlap of 3 bases (1 codon) is shown above the gene. (**B**) Pair-wise list the amino acid sequences of protein Oac_{1b} and Oac. Lines represent amino acid residues that are identical, whereas dots represent amino acids that are similar. TM segments are indicated in red color. The three major regions conserved among the inner membrane *trans*-acylase family proteins are boxed. The three critical residues for Oac function are marked by asterisk.

The serotype 1b strains used in this study were isolated from different countries, different regions, at different times (Table 1), and belonged to at least two different sequence types based on multi-locus sequence typing of 15 house keeping genes (Sun *et al*, unpublished data), while the other serotypes were also from diverse sources. Thus, we conclude that the divergent *oac* gene is specific to serotype 1b and hence, we named it oac_{1b}.

Oac_{1b} has the same enzymatic function as Oac

To determine whether Oac_{1b} performs the same *O*-acetylation function as the Oac, oac_{1b} from strain 1997020, together with its promoter and terminator elements were amplified and cloned into TA vector pMD20T and transformed into 03XZ014 (serotype Y), NCTC9725 (serotype 4a), 05004 (serotype 1a) and 04SH03 (serotype X), respectively. All the transformants were found to be capable of agglutinating with *S. flexneri* group 6 antisera, and these serotype Y, X, 4a and 1a strains were converted into serotypes 3b, 3a, 4b and 1b respectively. The agglutination intensity of the transformants showed no difference from that of serotype 3a, 3b, 4b and 1b clinical isolates. These results suggest that Oac_{1b} possesses Oac activity, mediating the acetylation of the O-antigen in these serotypes.

The oac_{1b} gene was located in a chromosomal region characteristic of phage origin but different from the Sf6 phage genome

In order to determine whether the oac_{1b} gene is also carried by a Sf6 phage, we performed PCR on all the 20 serotype 1b strains using primer pairs O3 and O4 (Table 2) designed on the Sf6 genome sequence, and all PCR amplifications were negative, suggesting the region is very different. We then used genome walking PCR starting from the oac_{1b} gene and obtained two fragments about 3.1 kb and 2.3 kb up- and down-stream of the oac_{1b} gene respectively from serotype 1b strain 1997020. The 5451 bp sequence encodes six complete

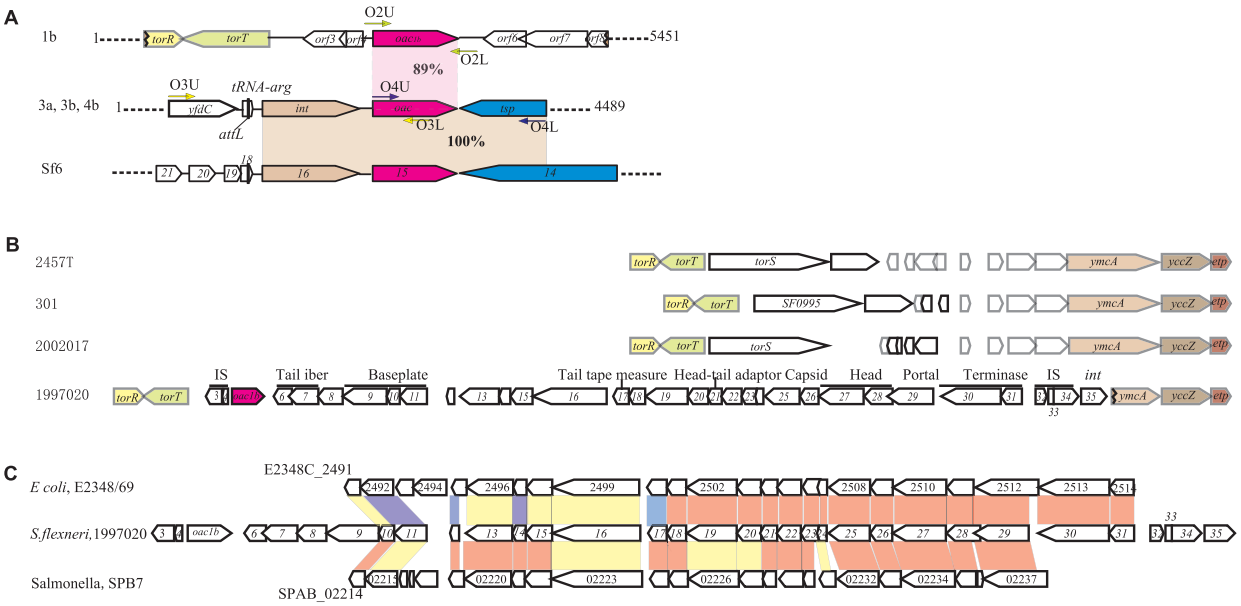

Figure 2 Genomic structure of the oac_{1b} region. (**A**) Comparison of chromosomal regions flanking *O*-acetyltransferase gene oac_{1b} in serotype 1b strains, and *oac* in serotype 3a, 3b and 4b strains and serotype-converting phage Sf6. Regions sharing >85% sequence identity are indicated by shaded boxes. Genes coding the same function are shown in the same color. Key primers used in this study are marked by arrows. (**B**) Genomic structure of regions flanking *O*-acetyltransferase gene oac_{1b} in serotype 1b strain 1997020 and comparison with relevant regions of sequenced *S. flexneri* strain 301, 2457t and 2002017. The details of ORFs in strain 1997020 are listed in Supplementary Table S1. Genes share high homologies are shown in the same color. (**C**) Comparison of the genomic structure of oac_{1b}-carrying prophage in serotype 1b strain 1997020 with prophage genomes in *Salmonella enterica* serovar Paratyphi B strain SPB7 and *E. coli* O127:H6 strain E2348/69. Genes sharing >40% identity at amino acid level between the strains are marked by color, red, >80% identity; yellow, 60%–80%; blue, 40%–60%.

ORFs and two incomplete ORFs (Figure 2A). Upstream of oac_{1b} are two housekeeping genes (*torR and torT*) and an IS element (IS1), whereas downstream of oac_{1b} are three genes of high identity to phage or prophage genes, which are non-homologous to any of the genes in phage Sf6 (Figure 2A), suggesting the presence of a novel prophage. In order to obtain the entire sequence of the prophage, the whole genome of strain 1997020 was sequenced using Illumina Solexa sequencing technology. A total of 4 637 796 reads were generated to reach about 110-fold coverage and these were assembled *de novo* into 280 contigs (>1000 bp). A contig of 30 kb carrying oac_{1b} and flanking regions was identified and a total of 38 ORFs (including one pseudogene (ORF36)) were able to be predicted by the ORF finder (Supplementary Table S1 and Figure 2B) (accession NO. JN377795).

DNA and protein level analyses of the 38 ORFs found that ORF1 and ORF2 subsume genes *torR* and *torT*, while ORF36, ORF37 and ORF38 are genes *ycmA*, *yccZ* and *ept*, respectively, which are present in the genomes of sequenced *S. flexneri* strains 301, 2457t and 2002017. The sequence between ORF3 to ORF35 has non-homologus sequences among *torT* and *ycmA* in the genomes of 301, 2457t and 2002017 (Figure 2B). The insertion of ORF3 to ORF35 has been accompanied by deletion of *torS* to part of the ORF36 (*ycmA*), which has its 5′ 935-bp region truncated, thereby resulting in a pseudogene (Figure 2B).

BLASTp analysis found that most of the proteins encoded by the 33 ORFs (ORF3–ORF35) are similar to bacteriophage proteins except for 9 ORFs, whose functions are unknown. These ORFs are ORF6 (tail fiber), ORF7 (tail protein), ORF9 (baseplate protein), ORF11 (baseplate assembly protein), ORF17 (tail tape measure protein), ORF18 (structural protein), ORF21 (head-tail adaptor), ORF23 (structural protein), ORF25 (capsid), ORF27–ORF28 (head), ORF29 (portal), ORF30 (terminase large subunit) and ORF31 (terminase small subunit) (Supplementary Table S1 and Figure 2B). Two putative insertion sequences, IS1 (ORF3 and ORF4) and IS911 (ORF32–ORF34), were located downstream *torT* and ORF31, respectively (Supplementary Table S1 and Figure 2B). The sequences of IS1 and IS911 are identical to the IS sequences found in the *S. flexneri* genomes. However, there are many copies of IS1 and IS911 in the genome, making it difficult to draw any inference of their origins. It should be noted that an IS911 is also present in the Sf6 genome. However, that IS was located in the Nin region but not the virion head domain as was found here. ORF35 is an integrase sharing 99% amino acid identity with an integrase of phage HK022 (Supplementary Table S1) and may have played a role in the integration of the bacteriophage. These data clearly indicate that this segment of DNA carrying oac_{1b} originated from a phage. Since no phage genes for recombination, immunity, replication and lysis were found, this sequence represents an incomplete prophage genome. Attempts to induce the phage from all 20 serotype 1b strains available in our collection were unsuccessful using conditions described by Mavris *et al.*[16] We also performed overlapping PCR amplification to show that the genomic organization of this prophage region is similar among the serotype 1b strains.

Apart from two regions of homology—the gene *oac* and IS911 as described above, the DNA sequence and gene organization of oac_{1b}-carrying prophage is entirely different from that of phage Sf6 and contains no remnants of Sf6 phage genes (Figure 2A and 2B). Thus, we can conclude that the oac_{1b}-carrying prophage in serotype 1b strains had a non-Sf6 phage origin. Sequence analysis also indicates that this prophage remnant is not homologous to any of other known bacteriophages. However, continuous similarity of genes was found with prophage regions of *Salmonella enterica* serovar Paratyphi B

strain SPB7 and *E. coli* O127:H6 strain E2348/69, respectively, as shown in Figure 2C.

Current studies show that all the serotype-converting phages integrate into host chromosome at two conserved positions, *tRNA-thrW* for SfI, SfII, SfIV, SfV and SfX[17] and *tRNA-argW* for Sf6.[14] Once integrated, the *int* and O-antigen modification genes are located at opposition ends of the phage DNA, ending with *attL* and *attR* sites.[17] To confirm that the Sf6 was inserted into the *tRNA-argW* in non-serotype 1b strains, a fragment of 4489 bp (accession NO. JF450729) including the *oac* gene was obtained from strain 03HL12 (serotype 3a) by PCR using primer pairs O5 (Table 2). The sequence and gene structure downstream *tRNA-argW* were identical to that of Sf6 and as expected from the structural organization of Sf6 (Figure 2A). The insertion site of the presumable oac_{1b}-carrying phage is unusual. It has apparently inserted between gene *torT* and *ycmA*, resulting in the deletion of genome region between gene *torT* and *ycmA*, including part of *ycmA* (1–935 bp) (Figure 2B).

Phage Sf6 is unable to infect and convert serotype 1a strains

We induced the Sf6 phage from serotype 3a strain 03HL12 and used it to infect the four serotypes (X, Y, la and 4a) that are expected to be convertible based on O-antigen structure. All four serotype X strains and 11 serotype Y strains tested can be converted into 3a and 3b, respectively, as shown previously.[7–9] Interestingly, four serotype 4a strains were also converted to serotype 4b, which is contrary to previous reports that serotype 4a cannot infected by Sf6.[7] However, all 12 serotype 1a strains tested cannot be infected and converted into serotype 1b by phage Sf6 (Figure 3). Similar phenomenon was also observed by Clark *et al.*[7] Serotype 3b strains contain only an *O*-acetyl group connected to the O-2 position in the rhamnose III of the tetrasaccharide of the O-antigen. Theoretically, serotype 3b can be converted to serotype 1b by adding a glucosyl group to the *N*-acetylglucosamine of tetrasaccharide, which is mediated by the *gtrI* genes carried by phage SfI. We tested two serotype 3b strains by infecting them with phage SfI, but no serotype 1b convertants were found (Figure 3). The non-conversion cannot be attributed to the phage since the same stock was previously used successfully to infect serotype X strains which was converted to serotype 1d.[13]

DISCUSSION

The divergent oac_{1b} is more likely to have been obtained from outside *S. flexneri* than to have undergone rapid divergence from the *oac* gene in the other serotypes (3a, 3b and 4b) within *S. flexneri*. Previous studies have shown that *S. flexneri* strains of serotypes 1 to 5 arose as an independent lineage from within *E. coli* recently and there is very low level of variation in house keeping genes.[18] The virulence plasmid carried by these strains also showed high levels of similarity among the serotype 1–5 strains.[19] Thus, it seems less probable that oac_{1b} was evolved from the *oac* gene within *S. flexneri* given the high level of divergence.

Clark *et al.*[7] had previously showed that the *oac* gene cloned from Sf6 was capable of converting Y, X, 4a and 1a to 3b, 3a, 4b and 1b, respectively. Thus the *oac* and oac_{1b} genes are functionally interchangeable despite the high level of sequence variation.

The host range for Sf6 was proposed to be restricted to strains with a group 3;4 antigen of the *O*-polysaccharide chain which is presumably recognized by the phage tail protein TSP.[20] Hydrolysis by TSP of the 1,3-α-linkage between rhamnose II and rhamnose III exposes the host cell membrane to phage DNA to allow entry into host and complete lysogenic conversion.[21,22] Since all four serotypes (X, Y, la and 4a) carry group 3;4 antigen, there must be additional antigenic difference that render serotype 1a resistant to Sf6 infection.

The putative insertion site of the likely oac_{1b}-carrying phage appeared to be unusual. It had apparently inserted between gene *torT* and *ycmA*, resulting in the deletion of genome region between gene *torT* and *ycmA* including 936 bp of the *ycmA gene.* The *tor* operon which encodes the trimethylamine N-oxide respiratory system is apparently nonfunctional in *S. flexneri* as *torD*, *torA* and *torS* are known pseudogenes in *S. flexneri* 2a strains Sf301, 2457T and 2002017. *ymcA*, which encode a putative outer membrane lipoprotein, which is highly conserved among *Shigella* and *E. coli*. The effect of inactivation of the *ycmA* gene in 1b strains is not clear. It should be noted that neither the *tRNA* genes, nor the *att* site sequence, was found in this region of 1b strains.

Our data clearly indicated that the DNA carrying oac_{1b} originated from a phage; however, since no phage genes for recombination, immunity, replication and lysis were found, this sequence appears to represent an incomplete prophage genome. Our attempts to induce the phage, using conditions described above, from all of the 20 serotype 1b strains available in our collection, proved to be unsuccessful. Additionally overlapping PCR amplification showed that the genomic organization of this prophage region is similar among all the serotype 1b strains.

Of significance, we found that the oac_{1b} gene mediating the O-antigen acetylation in the *S. flexneri* serotype 1b strains was highly divergent different from the *oac* gene in serotype 3a, 3b and 4b strains and phage Sf6. We have shown that oac_{1b} was likely part of a prophage and had a non-Sf6 phage origin. In comparison to the

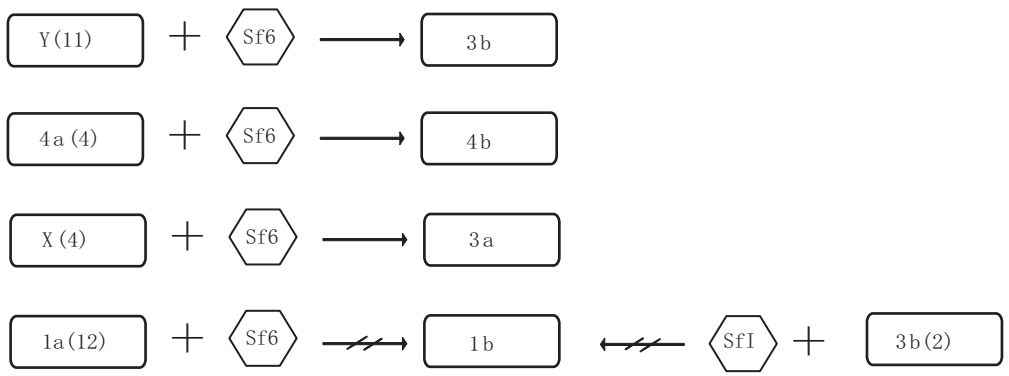

Figure 3 Schematic diagram of sensitivity of *S. flexneri* serotypes (1a, X, Y and 4a) to serotype-converting phage Sf6 and serotype 3b to serotype-converting phage SfI. Numbers of strains tested were showed in parentheses.

Sf6-like genomic structure in serotype 3a, 3b and 4b strains, the organization of the prophage carrying oac_{1b} in serotype 1b chromosome is rather unique. The ancestral phage was apparently inserted between genes *torT* and *ymcA*, however, with no typical phage attachment sites found. Sf6 infection experiments showed that the LPS of serotype 1a must contain additional changes to render it resistant to Sf6 infection. This study extends the current understanding of serotype conversion in *S. flexneri* and helps identify mechanisms that give rise to *Shigella* serotype evolution and diversity.

ACKNOWLEDGMENTS

This work was supported by grants (2011CB504901, 2011SKLID203, 2008SKLID106 and YB20098450101) from the Ministry of Science and Technology, and State Key Laboratory for Infectious Disease Prevention and Control, China.

1 Kotloff KL, Winickoff JP, Ivanoff B *et al*. Global burden of *Shigella* infections: implications for vaccine development and implementation of control strategies. *Bull World Health Organ* 1999; **77**: 651–666.
2 Bardhan P, Faruque AS, Naheed A, Sack DA. Decrease in shigellosis-related deaths without *Shigella* spp.-specific interventions, Asia. *Emerg Infect Dis* 2010; **16**: 1718–1723.
3 von Seidlein L, Kim DR, Ali M *et al*. A multicentre study of *Shigella* diarrhoea in six Asian countries: disease burden, clinical manifestations, and microbiology. *PLoS Med* 2006; **3**: e353.
4 Wang XY, Tao F, Xiao D *et al*. Trend and disease burden of bacillary dysentery in China (1991–2000). *Bull World Health Organ* 2006; **84**: 561–568.
5 Ye C, Lan R, Xia S *et al*. Emergence of a new multidrug-resistant serotype X variant in an epidemic clone of *Shigella flexneri*. *J Clin Microbiol* 2010; **48**: 419–426.
6 Clemens JD, Kotloff KL, Kay B. *Generic protocol to estimate the burden of* Shigella *diarrhoea and dysenteric mortality*. Geneva: World Health Organization, 1999.
7 Clark CA, Beltrame J, Manning PA. The *oac* gene encoding a lipopolysaccharide O-antigen acetylase maps adjacent to the integrase-encoding gene on the genome of *Shigella flexneri* bacteriophage Sf6. *Gene* 1991; **107**: 43–52.
8 Verma NK, Brandt JM, Verma DJ, Lindberg AA. Molecular characterization of the O-acetyl transferase gene of converting bacteriophage SF6 that adds group antigen 6 to *Shigella flexneri*. *Mol Microbiol* 1991; **5**:71–75.
9 Gemski P Jr, Koeltzow DE, Formal SB. Phage conversion of *Shigella flexneri* group antigens. *Infect Immun* 1975; **11**: 685–691.
10 Slauch JM, Lee AA, Mahan MJ, Mekalanos JJ. Molecular characterization of the *oafA* locus responsible for acetylation of *Salmonella typhimurium* O-antigen: *oafA* is a member of a family of integral membrane *trans*-acylases. *J Bacteriol* 1996; **178**: 5904–5909.
11 Cheon HG, Hanna PE. Effect of group-selective modification reagents on arylamine *N*-acetyltransferase activities. *Biochem Pharmacol* 1992; **43**: 2255–2268.
12 Thanweer F, Tahiliani V, Korres H, Verma NK. Topology and identification of critical residues of the O-acetyltransferase of serotype-converting bacteriophage, SF6, of *Shigella flexneri*. *Biochem Biophys Res Commun* 2008; **375**: 581–585.
13 Sun Q, Lan R, Wang Y *et al*. Genesis of a novel *Shigella flexneri* serotype by sequential infection of serotype-converting bacteriophages SfX and SfI. *BMC Microbiol* 2011; **11**: 269–274.
14 Casjens S, Winn-Stapley DA, Gilcrease EB *et al*. The chromosome of *Shigella flexneri* bacteriophage Sf6: complete nucleotide sequence, genetic mosaicism, and DNA packaging. *J Mol Biol* 2004; **339**: 379–394.
15 Sambrook J, Fritsch EF, Maniatis T. *Molecular cloning: a laboratory manual*. 2nd ed. New York: Cold Spring Harbor Laboratory, 1989.
16 Mavris M, Manning PA, Morona R. Mechanism of bacteriophage SfII-mediated serotype conversion in *Shigella flexneri*. *Mol Microbiol* 1997; **26**: 939–950.
17 Allison GE, Verma NK. Serotype-converting bacteriophages and O-antigen modification in *Shigella flexneri*. *Trends Microbiol* 2000; **8**: 17–23.
18 Pupo GM, Lan R, Reeves PR. Multiple independent origins of *Shigella* clones of *Escherichia coli* and convergent evolution of many of their characteristics. *Proc Natl Acad Sci U S A* 2000; **97**: 10567–10572.
19 Lan R, Lumb B, Ryan D, Reeves PR. Molecular evolution of large virulence plasmid in *Shigella* clones and enteroinvasive *Escherichia coli*. *Infect Immun* 2001; **69**: 6303–6309.
20 Lindberg AA, Wollin R, Gemski P, Wohlhieter JA. Interaction between bacteriophage Sf6 and *Shigella flexner*. *J Virol* 1978; **27**: 38–44.
21 Chua JE, Manning PA, Morona R. The *Shigella flexneri* bacteriophage Sf6 tailspike protein (TSP)/endorhamnosidase is related to the bacteriophage P22 TSP and has a motif common to exo- and endoglycanases, and C-5 epimerases. *Microbiology* 1999; **145**: 1649–1659.
22 Freiberg A, Morona R, van den Bosch L *et al*. The tailspike protein of *Shigella* phage Sf6. A structural homolog of *Salmonella* phage P22 tailspike protein without sequence similarity in the beta-helix domain. *J Biol Chem* 2003; **278**: 1542–1548.

Supplementary Information for this article can be found on *Emerging Microbes and Infections*' website (http://www.nature.com/EMI/)