Genomic Portrait of the Evolution and Epidemic Spread of a Recently Emerged Multidrug-Resistant Shigella flexneri Clone in China


Published Ahead of Print 22 January 2014.

Updated information and services can be found at:
http://jcm.asm.org/content/52/4/1119

These include:

SUPPLEMENTAL MATERIAL
Supplemental material

REFERENCES
This article cites 40 articles, 21 of which can be accessed free at: http://jcm.asm.org/content/52/4/1119#ref-list-1

CONTENT ALERTS
Receive: RSS Feeds, eTOCs, free email alerts (when new articles cite this article), more»
Genomic Portrait of the Evolution and Epidemic Spread of a Recently Emerged Multidrug-Resistant Shigella flexneri Clone in China

Nan Zhang,a Ruiting Lan,b Qiangzheng Sun,a Jianping Wang,a Yiting Wang,a Jin Zhang,c Deshan Yu,d Wanfu Hu,e Shoukui Hu,e Hang Dai,a Pengcheng Du,a Haiyin Wang,a Jianguo Xu a

State Key Laboratory for Infectious Disease Prevention and Control, Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, National Institute for Communicable Disease Prevention and Control, Beijing, China; aBranch for Enteric Disease Control and Prevention, Institute for Infectious Disease Control and Prevention, Henan Center for Disease Control and Prevention, Zhengzhou, China; bGansu Center for Disease Control and Prevention, Lanzhou, China; cAnhui Center for Disease Control and Prevention, Hefei, China

Shigella flexneri is the major cause of shigellosis in developing countries. A new S. flexneri serotype, Xv, appeared in 2000 and replaced serotype 2a as the most prevalent serotype in China. Serotype Xv is a variant of serotype X, with phosphoethanolamine modification of its O antigen mediated by a plasmid that contained the opt gene. Serotype Xv isolates belong to sequence type 91 (ST91). In this study, whole-genome sequencing of 59 S. flexneri isolates of 14 serotypes (serotypes 1 to 4, Y, Yv, X, and Xv) indicated that ST91 arose around 1993 by acquiring multidrug resistance (MDR) and spread across China within a decade. A comparative analysis of the chromosome and opt-carrying plasmid pSFXv_2 revealed independent origins of 3 serotype Xv clusters in China, with different divergence times. Using 18 cluster-dividing single-nucleotide polymorphisms (SNPs), SNP typing divided 380 isolates from 3 provinces (Henan, Gansu, and Anhui) into 5 SNP genotypes (SGs). One SG predominated in each province, but substantial interregional spread of SGs was also evident. These findings suggest that MDR is the key selective pressure for the emergence of the S. flexneri epidemic clone and that Shigella epidemics in China were caused by a combination of local expansion and interregional spread of serotype Xv.

Among the four species of Shigella, Shigella flexneri is the most common cause of bacillary dysentery in developing countries (1). Worldwide, there are 164.7 million cases of shigellosis annually with 1.1 million deaths, mostly of children less than 5 years of age (1). S. flexneri is divided into at least 19 serotypes (serotypes 1a, 1b, 1c, 1d, 2a, 2b, 3a, 3b, 4a, 4v, 4b, 5a, 5b, X, Xv, Y, Yv, F6, and 7b), based on the combinations of antigenic determinants present on the O antigen of the cell envelope lipopolysaccharide (LPS) (2–14). Serotypes 1c, 4v, 7b, 1d, Yv, and Xv are new serotypes reported in recent years (3–14), among which serotypes 1c and Xv have caused epidemic-level disease (5, 7, 9). Serotype Xv first appeared in 2000 in Henan province and subsequently replaced serotype 2a to become the predominant serotype in Henan between 2002 and 2006. Serotype Xv was also the most prevalent serotype in Gansu (67%) and Anhui (54%) in 2007 (9). A novel plasmid (pSFXv_2) carrying the opt gene encoding the O antigen phosphoethanolamine (PEtN) transferase is responsible for serotype conversion from X to Xv (10).

Our previous study indicated that S. flexneri epidemics in China have been caused by a sequence type 91 (ST91) clone, and the majority of the ST91 isolates belong to serotype Xv (9). Genome sequencing of a representative ST91 strain, 2002017, revealed that it gained 2 unique multidrug resistance (MDR) islands, i.e., the Shigella resistance locus (SRL) and Tn7 (9). The SRL, which was first identified in the Japanese S. flexneri 2a strain YSH6000, carries genes for resistance against streptomycin, ampicillin, chloramphenicol, and tetracycline and also carries an iron acquisition system (9, 15–17). The SRL in 2002017 carries an additional set of tetracycline resistance genes but does not contain the iron acquisition system present in YSH6000. Tn7 is a composite transposon conferring resistance to trimethoprim, streptomycin, and streptomycin/spectinomycin, all of which were recommended antibiotics used to treat dysentery in the early 1990s (9).

To elucidate the temporal and geographic dynamics of S. flexneri epidemics in China, we sequenced 59 strains, predominantly of ST91 and serotype Xv, across a 10-year time period and found that acquisition of multidrug resistance was the key factor for the expansion of ST91 across China in the early 1990s; we also found that the novel serotype Xv emerged more than once in different regions independently.

MATERIALS AND METHODS

Bacterial strains and DNA extraction. Fifty-nine strains, which were isolated in the Henan (38 strains), Gansu (8 strains), Anhui (7 strains), Shanxi (3 strains), Guizhou (2 strains), and Sichuan (1 strain) provinces of China between 1997 and 2006, were selected for Illumina paired-end sequencing (18). The 59 isolates covered 14 serotypes, including Xv (25 isolates), X (9 isolates), 1a (7 isolates), 1b (1 isolate), 1d (1 isolate), 2a (5 isolates), 2b (1 isolate), 3a (3 isolates), 3b (1 isolate), 4a (1 isolate), 4av (1 isolate), 4v (1 isolate), 1d (1 isolate), and Yv (1 isolate). The majority of the strains belong to ST91, within which serotype Xv strains dominate; therefore, more serotype Xv isolates were selected. The selection also reflected pulsed-
field gel electrophoresis (PFGE) diversity, representing 30 pulotypes. The strain details are presented in Table S1 in the supplemental material. Multilocus sequence typing (MLST) using 15 housekeeping genes was performed based on the protocol obtained from the EcMLST website (http://www.shigatox.net/ecmlst).

Bacterial isolates were grown on brain heart infusion agar at 37°C. Genomic DNA was extracted using a Wizard genomic DNA purification kit (Promega Corp., Madison, WI), as described in the manufacturer’s manual. The plasmid purification minikit from Qiagen (Hilden, Germany) was used to isolate the plasmids, according to the manufacturer’s instructions.

**Sequencing and bioinformatics analysis.** The whole-genome sequencing of the 59 isolates was performed using an Illumina Genome Analyzer IIx system (San Diego, CA), with mate-pair and pair-end sequencing technology (18). The Illumina reads were mapped to the 2002017 reference genome using Burrows-Wheeler alignment (BWA) with default parameters (19). Repeat regions and phage sequences were excluded from single-nucleotide polymorphism (SNP) analysis. SAMtools (20) was used to calculate per-position coverage and to determine the nucleotide for each position. High-quality potential SNPs were identified where a position was covered by ≥10 reads with >80% covering reads exhibiting the same SNP. All SNPs were recorded by position in reference to the 2002017 genome. Potential genome-wide SNPs were obtained.

Recombination was detected using the method of Feng et al. (21). The raw sequence data were also assembled de novo using SOAPdenovo (release 1.04) (22). Genomic indels were analyzed by methods described previously (23). Identification of insertional sequence (IS) sites was performed using the method of Smith (24).

Detection of plasmid SNPs was performed by mapping sequences reads to the reference plasmid sequences of pSFyv_2 from 2002017 and pSFyv_2 from HN006 (10). The opt gene was detected by PCR amplification, as described by Sun et al. (10).

**Phylogenetic analysis.** Phylogenetic trees were constructed by the maximum likelihood (ML) method (25). Determination of the time of divergence of a branch and substitution rates was performed using the BEAST v1.7.5 package (26). The best-fit evolutionary model for the data set was found to be the general time reversible (GTR) model with a gamma-distribution of among-site rate heterogeneity and a proportion of invariable sites.

**Antibiotic resistance analysis.** Antimicrobial susceptibility testing was performed using the disk diffusion method, following the standard protocol recommended by the Clinical and Laboratory Standards Institute (CLSI) (27). Resistance genes commonly present in Enterobacteriaceae, including the SRL, Tn7, CMY, TEM, SHV, and CMY, were used as references (9, 28–31) (see Table S1 in the supplemental material). The antibiotic resistance genes were searched using BLASTn, with an E value of 1 × 10−15 as the cutoff value for a significant match, based on the data generated by whole-genome sequencing.

**Detection of SNPs by Sequenom MassArray.** SNP typing was performed with a Sequenom MassArray detection system (San Diego, CA), as described by Jaremko et al. (32). Eighteen cluster-specific SNPs (see results below) obtained from whole-genome sequencing were used to type a large set of isolates. The distribution and nature of the 18 SNPs and primers used are listed in Table 1 and Table S2 in the supplemental material. The initial PCR amplification was carried out with 10 ng of genomic DNA in a 5-μl PCR mixture, and base-extension reactions were performed in a 9-μl PCR extension reaction mixture, according to the manufacturer’s instructions. Excess deoxynucleoside triphosphates (dNTPs) were removed by incubation with shrimp alkaline phosphatase for 40 min at 37°C. The final base-extension products were treated with SpectroCLEAN resin (Sequenom, San Diego, CA) to remove salts. The final reaction so-

### TABLE 1 Cluster-specific SNPs

<table>
<thead>
<tr>
<th>Cluster</th>
<th>SNP genotype</th>
<th>SNP Location</th>
<th>Reference nucleotide</th>
<th>SNP</th>
<th>Nature of SNP</th>
<th>Amino acid change</th>
<th>Gene (locus tag)</th>
<th>Gene function</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 SG2</td>
<td>1</td>
<td>660,521</td>
<td>C</td>
<td>A</td>
<td>ns</td>
<td>E/D</td>
<td>nagE</td>
<td>Pts N-acetylglucosamine-specific EIIcBA</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>2,093,579</td>
<td>T</td>
<td>A</td>
<td>ns</td>
<td>L/Q</td>
<td>hchA</td>
<td>Chaperone protein hchA</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>3,385,278</td>
<td>T</td>
<td>C</td>
<td>nc</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>3,400,397</td>
<td>A</td>
<td>C</td>
<td>s</td>
<td>A/A</td>
<td>yhdP</td>
<td>Putative membrane protein</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>4,408,641</td>
<td>T</td>
<td>C</td>
<td>s</td>
<td>Y/Y</td>
<td>ypeE</td>
<td>Putative LysR-type transcriptional regulator</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>1,409,095</td>
<td>T</td>
<td>G</td>
<td>s</td>
<td>Q/Q</td>
<td>bamA</td>
<td>Outer membrane protein assembly factor YaeT precursor</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>1,026,507</td>
<td>T</td>
<td>G</td>
<td>ns</td>
<td>T/P</td>
<td>yewW</td>
<td>Putative SAM-dependent methyltransferase</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>1,083,507</td>
<td>G</td>
<td>A</td>
<td>ns</td>
<td>R/H</td>
<td>SFxv_1115</td>
<td>Putative P4-family integrase</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>3,107,817</td>
<td>A</td>
<td>A</td>
<td>ns</td>
<td>V/F</td>
<td>ychF</td>
<td>GTP-dependent nucleic acid-binding protein EngD</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>2,077,539</td>
<td>A</td>
<td>T</td>
<td>ns</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>2,470,089</td>
<td>C</td>
<td>T</td>
<td>s</td>
<td>A/A</td>
<td>SFxv_2626</td>
<td>Hypothetical protein</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>2,667,345</td>
<td>C</td>
<td>G</td>
<td>ns</td>
<td>V/V</td>
<td>htsS</td>
<td>Histidyl-tRNA synthetase</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>1,317,434</td>
<td>G</td>
<td>A</td>
<td>nc</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>3,301,086</td>
<td>A</td>
<td>G</td>
<td>nc</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>3,504,262</td>
<td>G</td>
<td>A</td>
<td>ns</td>
<td>G/S</td>
<td>cysG</td>
<td>Uroporphyrinogen III methylase</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>3,804,107</td>
<td>T</td>
<td>A</td>
<td>ns</td>
<td>T/S</td>
<td>SFxv_3993</td>
<td>Lipopolysaccharide 1,2-galactosyltransferase</td>
</tr>
<tr>
<td></td>
<td>17</td>
<td>3,817,753</td>
<td>T</td>
<td>C</td>
<td>ns</td>
<td>K/R</td>
<td>pyrE</td>
<td>Orotate phosphoribosyltransferase</td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>2,041,242</td>
<td>G</td>
<td>A</td>
<td>ns</td>
<td>C/Y</td>
<td>pfPB</td>
<td>Glycyl-radical enzyme-activating protein family</td>
</tr>
<tr>
<td></td>
<td>19</td>
<td>2,402,145</td>
<td>A</td>
<td>G</td>
<td>ns</td>
<td>V/I</td>
<td>frwD</td>
<td>PTS fructose-like IIb component 2</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>4,456,749</td>
<td>T</td>
<td>C</td>
<td>nc</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 SG3</td>
<td>21</td>
<td>2,821,789</td>
<td>C</td>
<td>A</td>
<td>nc</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>22</td>
<td>3,058,327</td>
<td>A</td>
<td>G</td>
<td>ns</td>
<td>S/F</td>
<td>tktA</td>
<td>Transketolase 1 isozyme</td>
</tr>
<tr>
<td>3 (extended)</td>
<td>23</td>
<td>267,157</td>
<td>G</td>
<td>T</td>
<td>ns</td>
<td>L/F</td>
<td>SFxv_0262</td>
<td>Hypothetical protein</td>
</tr>
<tr>
<td></td>
<td>24</td>
<td>4,109,093</td>
<td>T</td>
<td>C</td>
<td>nc</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>2,969,268</td>
<td>T</td>
<td>A</td>
<td>ns</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>26</td>
<td>4,511,008</td>
<td>C</td>
<td>T</td>
<td>s</td>
<td>T/T</td>
<td>dlpZ</td>
<td>Thiol-disulfide interchange protein</td>
</tr>
<tr>
<td></td>
<td>27</td>
<td>4,600,805</td>
<td>T</td>
<td>C</td>
<td>ns</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* References: (1) strain 2002017; (2) nonsynonymous SNP (nsSNP); (3) synonymous SNP (sSNP); (4) noncoding SNP (ncSNP).

a. See Table S4 in the supplemental material.

b. Reference strain 2002017.

c. ns, nonsynonymous SNP (nsSNP); s, synonymous SNP (sSNP); nc, noncoding SNP (ncSNP).

d. Pts, phosphotransferase system; SAM, S-adenosylmethionine.
**RESULTS AND DISCUSSION**

Whole-genome sequence analysis of 59 S. flexneri isolates. In order to elucidate the genomic basis of the epidemic spread of S. flexneri serotype Xv and the endemcity of shigellosis in China, we sequenced 59 S. flexneri isolates, of which 50 were ST91 and 25 were serotype Xv, from different regions of China, covering 10 years from 1997 to 2006 (Fig. 1A). The genomes were sequenced using Illumina 100-bp paired-end sequencing technology, with an average of 121-fold coverage. Reads were mapped to the reference sequence 2002017 to produce 4.48 Mb per genome, on average. Five published complete genome sequences (9, 33–36). A final set of 1,790 chromosomal SNPs in the core genome regions, excluding SNPs in repetitive or mobile sequences, were identified from a total of 64 genomes. Fifty-two homologous recombination regions with high frequencies of sub- stitutions that were assumed to be introduced by horizontal transfer were detected. The recombinant regions containing 108 homoplastic SNPs were removed from the phylogenetic analysis described below.

We analyzed the 220-kb virulence plasmid pINV for SNPs. The presence of SNPs in the IS regions was noted, but no SNPs were present in other regions of the plasmid genome of the 59 isolates. We also searched for potential new IS sites and found only one novel IS site, located in gene SFxv_1148 (encoding TnpA transposase), in 15 of the 59 isolates examined (see Table S1 in the supplemental material).

The genomes were assembled de novo and examined for any strain-specific genes. With exclusion of serotype-specific phage sequences, the average number of strain-specific genes was 12, with values ranging from 65 strain-specific genes for the most diverse isolate S15757 to none for 4 Xv isolates (see Table S1 in the supplemental material).

Phylogenetic relationship of the 64 isolates. The 59 genomes together with 5 published genome sequences of S. flexneri were used to construct a strain phylogeny using the ML method. The topology shows that the 64 strains can be divided into 6 lineages, with 2 unique SNPs and a 544-bp deletion at base position 339,247. Cluster 1 contained 17 isolates, 12 of which were grouped into 3 distinct phylogenetic clusters (Fig. 1B). Only 2 isolates (2001014 and Shi06AH091) fell outside the 3 main clusters. These Xv isolates form distinctive and extremely tight clusters on branches separate from those of other serotype isolates, suggesting multiple origins of the Xv serotype. The 3 serotype Xv clusters are supported by multiple SNPs, with 5, 15, and 2 SNPs supporting clusters 1, 2 and 3, respectively (Table 1). Four SNPs are shared by clusters 1 and 2, showing a possible common origin. However, cluster 3 seems to have arisen independently, with 2 unique SNPs and a 544-bp deletion at base position 339,247. Cluster 1 contained 17 isolates, 12 of which were from Henan (including the reference strain 2002017), 3 from Shanxi, 1 from Gansu, and 1 from Anhui, covering 14 serotype Xv isolates. Thus, cluster 1 is predominantly a Henan cluster. Cluster 2 contained 6 serotype Xv isolates, 5 of which were from Gansu and 1 from Henan. Cluster 3 contained 4 serotype Xv isolates, all of which were from Anhui. Note that the 3 cluster 3 isolates from 2006 were not known to be epidemiologically linked. Thus, there seems to be a regional prevalence of different serotype Xv clusters.

Substitution rate and divergence time estimates for lineage I and serotype Xv clusters. The 58 isolates with known isolation years from 1954 to 2006 were used for BEAST analysis. With a total of 1,790 SNPs, a BEAST tree was constructed to visualize the overall relationships of the strains, revealing a relationship between root-to-tip branch length and the known isolation dates for the sequenced isolates. We estimated a substitution rate of $8.66 \times 10^{-4}$ substitutions per site per year at 1,790 chromosomal SNP loci (95% highest posterior density [HPD], 4.17 $\times 10^{-4}$ to 1.34 $\times 10^{-3}$), corresponding to the accumulation of approximately 1.5 SNPs per chromosome per year. The estimated time to the most recent common ancestor (MRCA) for lineage I is 12.25 years (95%
Coevolution between chromosome and plasmid pSFXv_2. The opt gene, on a 6.85-kb plasmid, encodes the O antigen PEtN transferase, which is responsible for attaching a PEtN at the second (RhaII) or third (RhaIII) rhamnose, or both, of the tetrasaccharide repeat unit on the O antigen for the variant serotypes, Xv, Yv, and 4av strains (10, 14). There are 2 known opt alleles, optII and optIII, which are carried by plasmids pSFXv_2 and pSFyv_2, respectively, and preferentially mediate modifications of RhaII and RhaIII, respectively (10, 14). We determined the presence of opt by mapping illumina reads to the pSFXv_2 sequence and further confirmed the sequence by sequencing the amplified opt PCR products. All Xv strains carried optII, and 3 variants, differing from the optII prototype by 1 or 2 bases, were observed. A single Yv strain carried an optIII form with a single base change from the prototype optIII. All cluster 1 Xv strains carried an opt form identical to the prototype optII. However, the optII from 3 isolates from cluster 2 (Shi06GS55, Shi06GS07, and Shi06GS48), which are grouped together on the SNP tree, contained a single base change. The opt alleles of the other 2 Gansu isolates (Shi06GS37 and Shi06GS43) and 1 Henan isolate (Shi06HN016) in cluster 2 were identical to the optII prototype. The optII allele in cluster 3 differed from the prototype optII by 2 bases. Of the 2 Xv strains outside the main clusters, the optII allele of Shi06AH091, isolated in Anhui, shared 1 base change with cluster 3, while the optII allele of 2001044, isolated in Henan, was identical to the prototype optII. Three non-functional opt alleles were also observed, with each in 3 serotype X strains. The 3 serotype X isolates (2002141, Shi06HN344, and 2005184) were found to carry the pSFXv_2 plasmid. A single-base deletion rendered the opt gene nonfunctional in all 3 cases. The 2002141 optII contained a single-base deletion at base 560, while Shi06HN344 and 2005184 contained a single-base deletion at base 295, resulting in a stop codon that abolished the function of gene. The single-base deletions in Shi06HN344 and 2005184 occurred independently, since the 2 isolates were not clustered together and 2005184 shared 2 base changes with cluster 3 isolates (Fig. 1B).

We further extended our SNP analysis to the whole pSFXv_2 plasmid (see Table S3 in the supplemental material). pSFXv_2 is 6,850 bp in length and encodes a mobilization protein, a replication initiation protein, a lipoprotein, and 7 proteins of unknown function in addition to the O antigen PEtN transferase (10). By mapping raw reads to the reference pSFXv_2 plasmid, we found that all serotype Xv isolates and 3 serotype X isolates carried pSFXv_2, while all other strains did not contain the plasmid. The pSFXv_2 plasmids of cluster 1 and cluster 2 isolates are highly homologous. The 12 serotype Xv isolates and 1 serotype X isolate (2002141) in cluster 1 share 7 base changes with cluster 2. One serotype Xv isolate (Shi06HN347) and 1 serotype X isolate (Shi06HN344) in cluster 1 share only 5 of these 7 bases. There is an additional base change in the opt gene in the 3 isolates from cluster 2. Cluster 3 was dissimilar from the other 2 clusters, with 15 unique base changes. There was good concordance between pSFXv_2 variants and chromosomal clusters, consistent with coevolution of the plasmid and chromosome. The pSFXv_2 variation further confirms that cluster 3 arose independently.

Acquisition of drug resistance genetic elements accompanying lineage 1 expansion. Our previous analysis of serotype Xv isolates showed that all were resistant to ampicillin and nalidixic acid, 96% were also resistant to chloramphenicol and tetracycline, and 89.7% were resistant to trimethoprim-sulfamethoxazole (9).

In serotype Xv isolate 2002017, the RSR island resistance carries MDR genes for resistance to tetracycline (tetA, tetR, tetC, and tetD), streptomyein (aadA2), ampicillin (oxa-1), and chloramphenicol (cat), while Tn7 harbors MDR elements for resistance to trimethoprim (dfrA1, encoding dihydrofolate reductase), streptothricin (sat1, encoding streptothricin acetyltransferase), and streptomycin/spectinomycin (aadA1, encoding aminoglycoside adenyltransferase) (9). All lineage 1 isolates carried the full RSR resistance island, except that 2 isolates had partial deletions. Shi06GS48 did not contain the tetD gene, whereas 2005AH264 lacked the oxa-1 and cat genes. All lineage I isolates carried Tn7. There was no sequence variation in the SRL or Tn7 among the lineage I isolates. No other lineages carried the SRL except for a lineage IV isolate, 2002091, which also carried a SRL and Tn7 identical to those of 2002017.

Mutations in gyrA are known to confer quinolone resistance (39). All lineage I isolates harbored 2 gyrA point mutations (248 harbored C→T, resulting in Ser83→Leu, and 631 C→T, resulting in His211→Tyr). Ser83→Leu is a known quinolone resistance mutation (40), whereas His211→Tyr is a novel mutation, located in the quinolone resistance-determining region. The 2 nonsynonymous mutations are markers for lineage I and possibly conferred an advantage to the lineage with quinolone resistance, although it is yet to be determined experimentally whether the His211→Tyr mutation confers quinolone resistance. Additionally, 4 ST91 isolates (Shi06HN378, Shi06SX53, 2001044, and 2002021) harbored an Asp87→Gly mutation in gyrA, 1 ST91 isolate (2001004) harbored an Asp87→Tyr mutation, and 1 ST91 isolate (2005002) had an Asp87→Asn mutation. Only 1 nonlin-

FIG 1 Genomic relationships of S. flexneri isolates. (A) Phylogenetic tree of 64 S. flexneri isolates based on 1,790 SNPs constructed by the maximum likelihood method. Maximum likelihood bootstrap values that support major lineages are indicated below the branches. The 6 lineages are marked on the right. The distribution of antibiotic resistance genomic contents is shown for the Shigella resistance locus (SRL) island, Tn7, gyrA mutations, and dfrA5. +, presence of a given gene or mutation; −, absence of a given gene or mutation. Isolate details (year, region, serotype, ST, and opt type) are shown. opt* indicates that the isolate carries a defective opt. The median and estimated 95% highest posterior density (HPD) of divergence time are given for the major lineages. Blue, cluster 1; purple, cluster 2; red, cluster 3; TET, tetracycline; STR, streptomycin; AMP, ampicillin; GHH, chloramphenicol; TMP, trimethoprim; NAL, nalidixic acid. (B) Phylogenetic tree of lineage I, to show opt gene variations. The phylogenetic tree was from A. The 3 clusters of Xv isolates are marked with boxed cluster names and divergence times. Blue, cluster 1; purple, cluster 2; red, cluster 3. The number above the branch that defines a cluster is the number of cluster-specific SNPs. ns, nonsynonymous SNP (nsSNP); s, synonymous SNP (sSNP); nc, noncoding SNP (ncSNP). Isolate details (year, region, serotype, ST, and opt type) are shown. opt* indicates that the isolate carries a defective opt. The opt SNPs are listed on the right, with the consensus bases in the top row. Dots, bases identical to the consensus bases. +, single-base deletion, resulting in a stop codon at the position shown.
eage I isolate (1997005) had an Asp87→Asn mutation, which clearly is a mutation parallel to that in the lineage I isolate.

There are other signals of antibiotic resistance-driven selection. dfrA5 confers resistance to trimethoprim. Cotrimoxazole, a combination of trimethoprim and sulfamethoxazole, is known to have been used widely in China in the 1990s, although it is no longer recommended for empirical treatment (41). All lineage I isolates (except 2005AH264) carried dfrA5, which is identical to that in 2002017. Outside lineage I, 1 lineage IV isolate (2002091) and 1 lineage III isolate (51581) also carried dfrA5. Additionally, 3 isolates (Shi06AH130, Shi06HN244, and 2005AH264) carried bla<sub>CTX-M-14</sub>, another 3 isolates (Shi06HN250, Shi06HN118, and 2001042) carried bla<sub>TEM-1</sub>, and Shi06HN250 carried bla<sub>CTX-M-3</sub>. Strains 2001042 and Shi06HN250 also carried qnrS. Strains 2001042 and Shi06HN250 also carried qepA and aac(6′)-Ib-cr, respectively. Acquisition of these genes occurred sporadically and independently. However, acquisition of the SRL, Tn7, and two gyrA mutations are likely to have occurred once in lineage I, as they are present in all lineage I isolates and are associated with lineage I expansion during the 1990s.

**Regional dynamics of serotype Xv clusters.** The overwhelming regional clustering of serotype Xv isolates as 3 genome clusters based on the whole-genome sequences prompted us to determine the regional dynamics of the 3 clusters using a much larger sample, including 3 from Henan and 7 from Anhui, belonged to this lineage. We used the 5 unique SNPs present in the genome-sequenced strains is confirmed with this large set of isolates. However, there is also a substantial interregional flow of strains. Henan isolates included 79.1% cluster 1 (SG2), 7.3% cluster 2 (SG3 and SG4), and 9% cluster 3 (SG5) isolates. Gansu isolates included 58.4% cluster 2 (SG3 and SG4), 23.2% cluster 1 (SG2), and 1.6% cluster 3 (SG5) isolates, while Anhui isolates included 57.7% cluster 3 (SG5), 23.1% cluster 1 (SG2), and 9% cluster 2 (SG3 and SG4) isolates (Fig. 2; also see Table S4 in the supplemental material).

Since strain 2005184 carries a defective optII (Fig. 1), serotype Xv in cluster 3 must have arisen before 2005184 diverged. This led us to hypothesize that some of the unclustered SG1 isolates may belong to this lineage. We used the 5 unique SNPs present in the branch to 2005184 (Table 1; also see Table S2 in the supplemental material) to type the 46 SG1 isolates, and we found that 10 isolates, including 3 from Henan and 7 from Anhui, belonged to this lineage. Thus, cluster 3 could be extended to include 2005184 and could be defined using the 5 unique SNPs.

**Conclusions.** Whole-genome sequencing of 59 S. flexneri isolates revealed that, by acquiring MDR islands (SRL and Tn7) and quinolone resistance gyrA mutations, ST91 arose in the early 1990s, rapidly expanded, and spread across China within a decade. The acquisition of a novel plasmid by S. flexneri, leading to the emergence of a new serotype, occurred more than once. Both the

---

**FIG 2** Frequency and geographic distribution of the SNP genotypes of the 380 S. flexneri isolates in China, typed using cluster-specific SNPs. (A) Frequencies of the 5 SGs in different provinces, shown in pie charts. The numbers in parentheses after the province names are the numbers of isolates from the given province. SG3 and SG4 were grouped together to be consistent with genome clusters. (B) Geographic distribution of different SGs (genome clusters). The frequencies of the 5 SGs in different geographic regions (provinces) are shown in pie charts. For consistency with genome clusters, SG3 and SG4 are combined in one group. The genome clusters are indicated in parentheses after the SGs. The numbers in parentheses are the numbers of isolates from the given SG(s).
pSFXv_2 plasmid and the opt gene variation suggest independent gain of the plasmid by different clusters, with clear evidence that cluster 3 arose independently as an Xv serotype. There were 3 main clusters of serotype Xv isolates, which were localized in their respective regions, further supporting local origins of serotype Xv. However, there was also a substantial interregional flow of the clusters. In particular, cluster 1, which was predominant in Henan and emerged earliest, spread to other regions with much greater proportions. Therefore, S. flexneri epidemics represent a dynamic play of local strains and interregional spread of different strains. Our results provide significant insights into the evolution of S. flexneri.

ACKNOWLEDGMENTS

This work was supported by grants from the National Natural Science Foundation of China (grant no. 81271788, 81290340, and 81290345), the National Basic Research Priorities Program (grant no. 2011CB504901), the Project of State Key Laboratory for Infectious Disease Prevention and Control (grant no. 2011SKLID203 and 2008SKLID106), and the National Key Program for Infectious Diseases (grant no. 2013ZX10004221 and 2013ZX10004216-001-002).

REFERENCES


