Genome Sequences of Four *Shigella boydii* Strains Representative of the Major *S. boydii* Clades

Michael J. Sikorski, Tracy H. Hazen, Gopi Vyas, Jane M. Michalski, David A. Rasko

Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, Maryland, USA

Department of Microbiology and Immunology, University of Maryland School of Medicine, Baltimore, Maryland, USA

Center for Vaccine Development and Global Health, University of Maryland School of Medicine, Baltimore, Maryland, USA

**ABSTRACT** There are four bacterial species in the genus *Shigella* that cause shigellosis or dysentery. *Shigella boydii* is one of the least studied *Shigella* species but has been shown to be separated into three phylogenomic clades. Here, we report four complete reference sequences of the *S. boydii* phylogenomic clades.

Shigellosis is a diarrheal disease caused by the bacterial genus *Shigella*, consisting of four species (1), including *Shigella boydii*. The aim of this submission is to describe the complete reference genomes of four *S. boydii* isolates which are representatives of the previously identified *S. boydii* phylogenomic clades (2).

The four *S. boydii* strains were isolated using standard culture methods (3) from the stool of children in Bangladesh under the age of five as part of the Global Enteric Multicenter Study (GEMS) (3, 4). The isolates were grown in lysogeny broth overnight at 37°C, and DNA was purified by alkaline lysis extraction as previously described (5), with the exception that after the phenol:chloroform treatment the upper aqueous phase was added to a heavy phase lock tube (5 Prime, Inc., Gaithersburg, MD), and the extraction was repeated using chloroform-isooamyl alcohol (24:1 [vol/vol]). The upper aqueous phase was collected, and at least 5 volumes of isopropanol was used for precipitation of DNA on ice for 15 min, followed by centrifugation at 12,000 × g for 10 min, ethanol washes, and resuspension in water. Library preparation was conducted using a Kappa kit for 150-bp paired-end reads, and the genomes were sequenced on the Illumina HiSeq 2000 platform using paired-end libraries with 300-bp inserts, as previously described (2). The same genomic DNA preparations of each isolate were also used to generate a 20-kb sequencing library for the Pacific Biosciences (PacBio) RS II platform with P6C4 chemistry in a single flow cell per isolate using standard methods (6). The PacBio raw data of *S. boydii* isolates 600080, 600690, and 602068 were assembled using the Hierarchical Genome Assembly Process v.3 (HGAP3) pipeline and SMRTAnalysis v.2.3.0 software (7), while isolate 600657 was assembled using Canu v.1.4 (8). Assemblies for each isolate were circularized using Minimus2 (9) and polished with the generated Illumina reads using Quiver (7) to close the genomes. All software was run with default values. The total number of reads generated for each isolate on each sequencing platform and the relevant statistics for each genome assembly are listed in Table 1. The assemblies contained between two and four molecules, with each assembly having a single manually circularized chromosome and one or two circularized plasmids that ranged in size from 6.9 kb to 216 kb. Complete and circularized molecules were manually edited to remove overlaps and not rotated. The genomes were annotated with PGAP v.4.12 (12).

Phylogenomic analysis was performed on these four complete *S. boydii* genomes in comparison with complete *Shigella* genomes available from GenBank (n = 102) and a panel of diverse *E. coli* genomes (n = 30) as previously described (2). A maximum-
<table>
<thead>
<tr>
<th>Isolate</th>
<th>E. coli/ S. boydii phylogroup</th>
<th>No. of contigs</th>
<th>Genome size (bp)</th>
<th>Total genome GC %</th>
<th>Molecule Completion level</th>
<th>Molecule length (bp)</th>
<th>GC % by molecule</th>
<th>Plasmid incompatibility type</th>
<th>No. of Illumina reads</th>
<th>No. of PacBio reads</th>
<th>PacBio mean read length (bp)</th>
<th>Illumina sequence coverage (%)</th>
<th>PacBio sequence coverage (%)</th>
<th>Shigella virulence gene(s)</th>
<th>Antimicrobial resistance gene(s)</th>
<th>Genome accession no.</th>
<th>SRA accession no.</th>
</tr>
</thead>
<tbody>
<tr>
<td>600080</td>
<td>B1</td>
<td>1a</td>
<td>2</td>
<td>4,628,191</td>
<td>50.94</td>
<td>Chromosome Circular</td>
<td>4,559,517</td>
<td>50.92</td>
<td>NA##</td>
<td>9,383,921</td>
<td>11,921</td>
<td>8,734</td>
<td>304.13</td>
<td>22.5</td>
<td></td>
<td>CP049600</td>
<td>SRX8156206, SRX8156207</td>
</tr>
<tr>
<td>600690</td>
<td>B1</td>
<td>1b</td>
<td>3</td>
<td>5,043,708</td>
<td>50.45</td>
<td>Chromosome Circular</td>
<td>4,820,649</td>
<td>50.68</td>
<td>NA##</td>
<td>6,758,931</td>
<td>16,566</td>
<td>7,813</td>
<td>201.01</td>
<td>25.7</td>
<td></td>
<td>CP049607</td>
<td>SRX8156204, SRX8156205</td>
</tr>
<tr>
<td>600657</td>
<td>B1</td>
<td>2</td>
<td>4</td>
<td>4,833,494</td>
<td>51.06</td>
<td>Chromosome Circular</td>
<td>4,589,517</td>
<td>51.29</td>
<td>NA##</td>
<td>3,611,833</td>
<td>15,166</td>
<td>8,621</td>
<td>112.09</td>
<td>27.1</td>
<td></td>
<td>CP049280</td>
<td>SRX8163693, SRX8163694</td>
</tr>
<tr>
<td>602068</td>
<td>B1</td>
<td>3</td>
<td>2</td>
<td>4,758,031</td>
<td>50.92</td>
<td>Chromosome Circular</td>
<td>4,572,715</td>
<td>51.15</td>
<td>NA##</td>
<td>3,763,064</td>
<td>12,550</td>
<td>11,536</td>
<td>118.63</td>
<td>30.5</td>
<td></td>
<td>CP049283</td>
<td>SRX8156202, SRX8156203</td>
</tr>
</tbody>
</table>

##Virulence genes listed in the Virulence Factor Database (10) where >50% of the genes of a particular region were detected with a BLAST score ratio (BSR) of >0.8 are reported. Bold type indicates that 100% of genes were detected with a BSR of >0.8.

###Antimicrobial resistance (AMR) genes were predicted using CARD RGI (11), and only *"perfect"* hits are reported here. Highly conserved genes associated with antibiotic resistance phenotypes, such as efflux pumps and select transcriptional regulators, are not included.

##NA, not applicable.

###ND, not detected.

---

Sikorski et al. | Volume 9 Issue 41 e00881-20 | mra.asm.org | 2 | Downloaded from mpra.ub.uni-muenchen.de on February 5, 2021 by guest
likelihood phylogeny tree (Fig. 1) was constructed with 100 bootstrap replicates using RAxML v.8.2.10 (13) and visualized using iTOL v.5.5.1 (10).

The presence or absence of *Shigella* virulence genes (14) was examined using the large-scale BLAST score ratio (11) (Table 1). Antibiotic resistance genes were identified in each assembly using the Resistance Gene Identifier (RGI) v.5.1.0 software against the Comprehensive Antibiotic Resistance Database (CARD) v.3.0.8 with perfect detection criteria (15), and plasmid incompatibility types were predicted using PlasmidFinder (Table 1) (16). Three of the assemblies (600690, 600657, and 602068) contained the *Shigella* virulence plasmid, while the fourth assembly (600080) lacked the *Shigella* virulence plasmid but contained a 68-kb plasmid with the predicted antibiotic resistance genes *qnrS1* and *blaTEM-1* (Table 1).

Given the paucity of *S. boydii* reference isolates, these four complete phylogenomi-
cally distinct genomes presented may serve future studies as representative references from each clade of *S. boydii* (2).

**Data availability.** All data have been released, and the accession numbers are listed in Table 1.

**ACKNOWLEDGMENT**

This project was funded in part by federal funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services, under grant number U19AI110820 and T32 training grant T32DK067872 (M.J.S.).

**REFERENCES**


