



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

 Michael J. Sikorski,^{a,b,c} Tracy H. Hazen,^{a,b} Gopi Vyas,^a Jane M. Michalski,^{a,b}  David A. Rasko^{a,b}

^aInstitute for Genome Sciences, University of Maryland School of Medicine, Baltimore, Maryland, USA

^bDepartment of Microbiology and Immunology, University of Maryland School of Medicine, Baltimore, Maryland, USA

^cCenter 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-isoamyl 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-

Citation Sikorski MJ, Hazen TH, Vyas G, Michalski JM, Rasko DA. 2020. Genome sequences of four *Shigella boydii* strains representative of the major *S. boydii* clades. *Microbiol Resour Announc* 9:e00881-20. <https://doi.org/10.1128/MRA.00881-20>.

Editor Steven R. Gill, University of Rochester School of Medicine and Dentistry

Copyright © 2020 Sikorski et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to David A. Rasko, drasko@som.umaryland.edu.

Received 28 July 2020

Accepted 21 August 2020

Published 8 October 2020

TABLE 1 Sequencing statistics

Isolate	<i>E. coli/Shigella</i> phylogroup	<i>S. boydii</i> clade	No. of contigs	Genome size (bp)	Total genome GC %	Molecule	Completion level	Molecule length (bp)	GC % by molecule	Plasmid incompatibility type	No. of Illumina reads	PacBio mean read length (bp)	Illumina sequence coverage (x)	PacBio sequence coverage (x)	Shigella virulence genes ^a	Antimicrobial resistance gene(s) ^b	Genome accession no.	SRA accession no.
600080	B1	1a	2	4,628,191	50.94	Chromosome	Circular	4,559,517	50.92	NA ^c	9,383,921	8,734	304.13	22.5	Aerobactin iron uptake (<i>iucABCD</i> , <i>iutA</i>), SHI-2, SHI-3	ND ^d	CP049606	SRX8156206, SRX8156207
600690	B1	1b	3	5,043,708	50.45	p600080_68 Chromosome	Circular	68,674 4,820,649	52.72 50.68	IncFIB(K) NA	6,758,931	7,813	201.01	25.7	None Aerobactin iron uptake (<i>iucABCD</i> , <i>iutA</i>), T6SS, SHI-2, SHI-3	<i>qnrS1</i> , <i>bla_{TEM-1}</i> <i>sat-2</i>	CP049607 CP049278	SRX8156204, SRX8156205
600657	B1	2	4	4,833,494	51.06	p600690_216 Chromosome	Circular	216,073	45.47	IncFII	3,611,833	8,621	112.09	27.1	<i>icsA</i> (<i>virG</i>), <i>icsP</i> (<i>sopA</i>), <i>mxi-1</i> , <i>mxi-2</i> , <i>SHET2</i> *, SHI-3	ND	CP049279	
602068	B1	3	2	4,758,031	50.92	p600657_26 p600657_8 Chromosome	Not circular Circular	26,762 8,379	51.52 55.48	NA None	3,763,064	11,556	118.63	30.5	None Aerobactin iron uptake (<i>iucABCD</i> , <i>iutA</i>), <i>sigA</i> , T2SS, SHI-2, SHI-3	ND <i>ampH</i> , <i>pmrF</i> , <i>sat-2</i>	CP049282 CP049283 CP049284 CP049285	SRX8156202, SRX8156203
602068	B1	3	2	4,758,031	50.92	p602068_185 Chromosome	Circular	185,316	45.12	IncFII	3,763,064	11,556	118.63	30.5	<i>icsA</i> (<i>virG</i>), <i>icsP</i> (<i>sopA</i>), <i>mxi-1</i> , <i>mxi-2</i> , <i>SHET2</i>	ND	CP049286	

^aVirulence 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.

^bAntimicrobial 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.

^cNA, not applicable.

^dND, not detected.

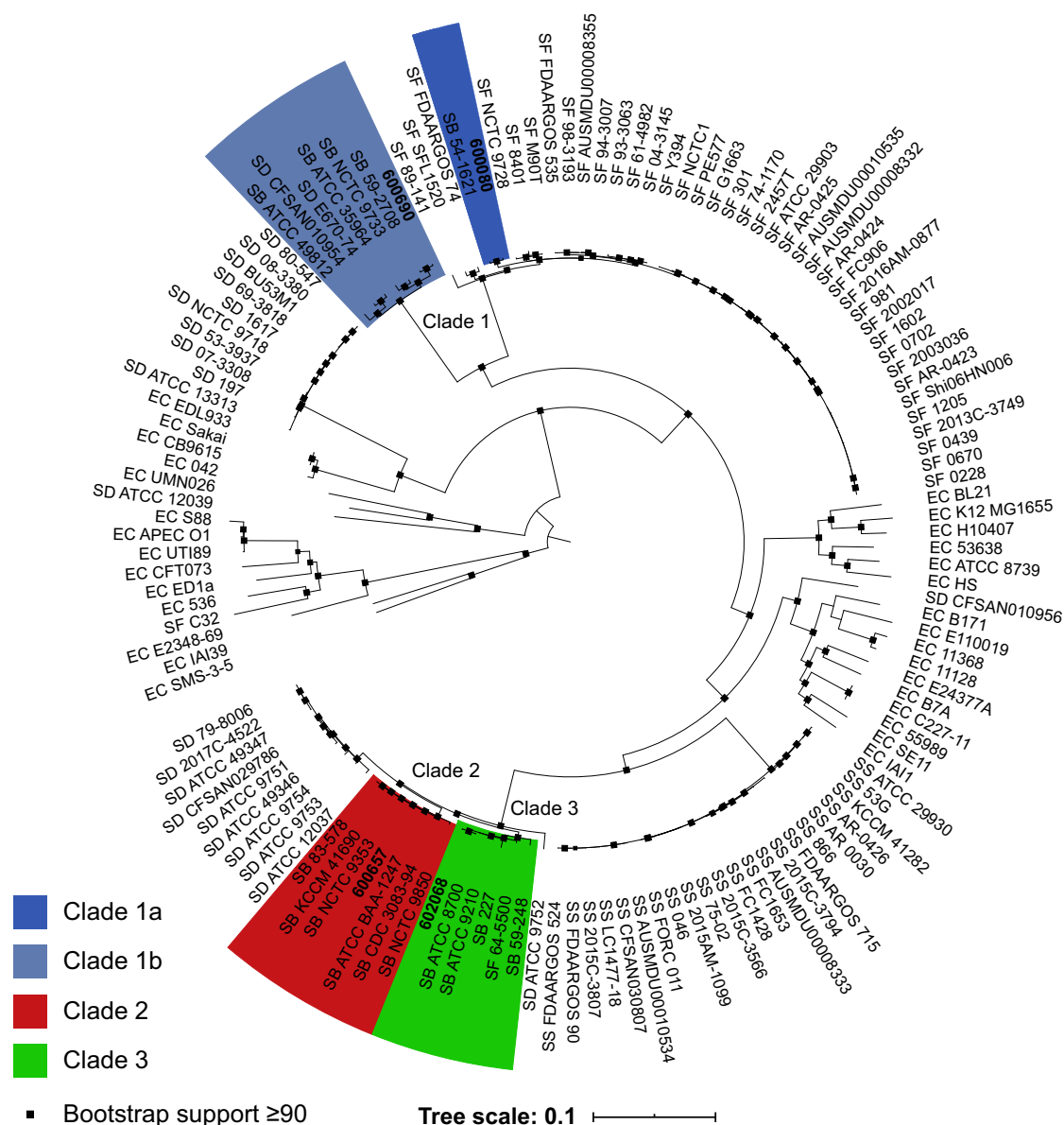


FIG 1 Phylogenomic tree containing four reference *S. boydii* isolates (in bold), reference *Shigella* isolates currently available in the public databases ($n = 102$), and a collection of diverse reference *E. coli* isolates ($n = 30$). Three *S. boydii* clades are identified by color, with clade 1 in blue (separated into two smaller subclades denoted by two shades of blue), clade 2 in red, and clade 3 in green. The tree was inferred with Interactive Tree Of Life (iTOL) v.5.5.1 with bootstrap support values greater than or equal to 90% shown at the black square nodes. Distance for the number of nucleotide changes is shown to be 0.1 with the corresponding tree scale bar length.

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 *bla*_{TEM-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

1. Sansonetti PJ. 2010. Shigella. In Mahy BW, Meulen VT, Borriello SP, Murray PR, Funke G, Kaufmann SH, Steward MW, Merz WG, Hay RJ, Cox F, Wakelin D, Gillespie SH, Despommier DD (ed), Topley and Wilson's microbiology and microbial infections, 10th ed. John Wiley & Sons, Chichester, United Kingdom.
2. Kania DA, Hazen TH, Hossain A, Nataro JP, Rasko DA. 2016. Genome diversity of Shigella boydii. Pathog Dis 74:ftw027. <https://doi.org/10.1093/femspd/ftw027>.
3. Panchalingam S, Antonio M, Hossain A, Mandomando I, Ochieng B, Oundo J, Ramamurthy T, Tamboura B, Zaidi AKM, Petri W, Houghton E, Murray P, Prado V, Vidal R, Steele D, Strockbine N, Sansonetti P, Glass RI, Robins-Browne RM, Tauschek M, Svennerholm A-M, Berkeley LY, Kotloff K, Levine MM, Nataro JP. 2012. Diagnostic microbiologic methods in the GEMS-1 case/control study. Clin Infect Dis 55:S294–S302. <https://doi.org/10.1093/cid/cis754>.
4. Kotloff KL, Nataro JP, Blackwelder WC, Nasrin D, Farag TH, Panchalingam S, Wu Y, Sow SO, Sur D, Breiman RF, Faruque ASG, Zaidi AKM, Saha D, Alonso PL, Tamboura B, Sanogo D, Onwuchekwa U, Manna B, Ramamurthy T, Kanungo S, Ochieng JB, Omere R, Oundo JO, Hossain A, Das SK, Ahmed S, Qureshi S, Quadri F, Adegbola RA, Antonio M, Hossain MJ, Akinsola A, Mandomando I, Nhampossa T, Acácio S, Biswas K, O'Reilly CE, Mintz ED, Berkeley LY, Muhsen K, Sommerfelt H, Robins-Browne RM, Levine MM. 2013. Burden and aetiology of diarrhoeal disease in infants and young children in developing countries (the Global Enteric Multi-center Study, GEMS): a prospective, case-control study. Lancet 382: 209–222. [https://doi.org/10.1016/S0140-6736\(13\)60844-2](https://doi.org/10.1016/S0140-6736(13)60844-2).
5. Sambrook JF, Fritsch EF, Maniatis T. 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
6. Hazen TH, Michalski J, Nagaraj S, Okeke IN, Rasko DA. 2017. Characterization of a large antibiotic resistance plasmid found in enteropathogenic Escherichia coli strain B171 and its relatedness to plasmids of diverse E. coli and Shigella strains. Antimicrob Agents Chemother 61: e00995-17. <https://doi.org/10.1128/AAC.00995-17>.
7. Chin C-S, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, Turner SW, Korlach J. 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat Methods 10:563–569. <https://doi.org/10.1038/nmeth.2474>.
8. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res 27:722–736. <https://doi.org/10.1101/gr.215087.116>.
9. Sommer DD, Delcher AL, Salzberg SL, Pop M. 2007. Minimus: a fast, lightweight genome assembler. BMC Bioinformatics 8:64. <https://doi.org/10.1186/1471-2105-8-64>.
10. Letunic I, Bork P. 2019. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res 47:W256–W259. <https://doi.org/10.1093/nar/gkz239>.
11. Sahl JW, Caporaso JG, Rasko DA, Keim P. 2014. The large-scale BLAST score ratio (LS-BSR) pipeline: a method to rapidly compare genetic content between bacterial genomes. PeerJ 2:e332. <https://doi.org/10.7717/peerj.332>.
12. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI Prokaryotic Genome Annotation Pipeline. Nucleic Acids Res 44: 6614–6624. <https://doi.org/10.1093/nar/gkw569>.
13. Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313. <https://doi.org/10.1093/bioinformatics/btu033>.
14. Chen L, Yang J, Yu J, Yao Z, Sun L, Shen Y, Jin Q. 2005. VFDB: a reference database for bacterial virulence factors. Nucleic Acids Res 33: D325–D328. <https://doi.org/10.1093/nar/gki008>.
15. Alcock BP, Raphenya AR, Lau TTY, Tsang KK, Bouchard M, Edalatmand A, Huynh W, Nguyen A-LV, Cheng AA, Liu S, Min SY, Miroshnichenko A, Tran H-K, Werfalli RE, Nasir JA, Oloni M, Speicher DJ, Florescu A, Singh B, Faltyn M, Hernandez-Koutoucheva A, Sharma AN, Bordeleau E, Pawlowski AC, Zubyk HL, Dooley D, Griffiths E, Maguire F, Winsor GL, Beiko RG, Brinkman FSL, Hsiao WWL, Domselaar GV, McArthur AG. 2020. CARD 2020: antibiotic resistance surveillance with the comprehensive antibiotic resistance database. Nucleic Acids Res 48:D517–D525. <https://doi.org/10.1093/nar/gkz935>.
16. Carattoli A, Hasman H. 2020. PlasmidFinder and in silico pMLST: identification and typing of plasmid replicons in whole-genome sequencing (WGS). Methods Mol Biol 2075:285–294. https://doi.org/10.1007/978-1-4939-9877-7_20.