



High-Quality Whole-Genome Sequences for 59 Historical *Shigella* Strains Generated with PacBio Sequencing

Justin Kim,^{a,b} Rebecca L. Lindsey,^a Lisle Garcia-Toledo,^{a,b} Vladimir N. Loparev,^a Lori A. Rowe,^a Dhvani Batra,^a Phalasy Juieng,^a Devon Stoneburg,^a Haley Martin,^a Kristen Knipe,^a Peyton Smith,^{a,b} Nancy Strockbine^a

^aCenters for Disease Control and Prevention, Atlanta, Georgia, USA

^bOak Ridge Institute for Science and Education, Oak Ridge, Tennessee, USA

ABSTRACT *Shigella* spp. are enteric pathogens that cause shigellosis. We report here the high-quality whole-genome sequences of 59 historical *Shigella* strains that represent the four species and a variety of serotypes.

Shigellosis is an acute enteric disease; symptoms may include severe diarrhea with mucoid bloody stools, abdominal cramps, and tenesmus (1). *Shigella* species are endemic to most developing countries and are estimated to cause over 80 million cases of bloody diarrhea and 700,000 deaths each year, the majority of which occur among children less than 5 years of age (2). Here, we report the availability of high-quality genome sequences for 59 *Shigella* strains generated via PacBio sequencing.

The DNA of the isolates was extracted using the ArchivePure DNA extraction kit standard protocol (5 Prime, Gaithersburg, MD). Library construction and PacBio sequencing were conducted as previously described (3), and DNA was sheared to an approximate length of 20 kb via needle shearing. BluePippin (Sage Scientific, Beverly, MA) was used to size select the sheared DNA. Sheared DNA and standard library protocols from a DNA template preparation kit (Pacific Biosciences, Menlo Park, CA) were used to generate large SMRTbell libraries. All strains were sequenced on one single-molecule real-time (SMRT) cell. Completed libraries were then bound to proprietary P6 v2 polymerase and sequenced using C4 chemistry on a PacBio RS II sequencer for 360-min movies (3). The sequence reads underwent *de novo* assembly using the PacBio Hierarchical Genome Assembly Process version 3 (4). For 49 strains (designated in Table 1), the sequence order in the resulting PacBio assembly for the chromosomes was verified using restriction enzyme NcoI and AflIII whole-genome mapping (OpGen, Gaithersburg, MD).

Table 1 lists the strain numbers, species, serotypes, accession numbers, genome and plasmid sizes, and the availability of an optical map that could be aligned to the PacBio genome sequence. One chromosomal sequence was obtained for all but seven genomes. The average genome coverage was 133.7×. The average G+C content was 51.1%. The assemblies of 14 isolates failed to be circularized, likely due to unresolved or collapsed repeat regions. Noncircular assemblies in Table 1 are noted.

Accession number(s). This whole-genome shotgun project has been deposited in DDBJ/ENA/GenBank under the accession numbers listed in Table 1. The versions described in this paper are the first versions.

ACKNOWLEDGMENTS

This work was funded by federal appropriations to the Centers for Disease Control and Prevention through the Advanced Molecular Detection Initiative line item.

The findings and conclusions of this article are those of the authors and do not necessarily represent the views of the Centers for Disease Control and Prevention. The

Received 5 March 2018 Accepted 19 March 2018 Published 12 April 2018

Citation Kim J, Lindsey RL, Garcia-Toledo L, Loparev VN, Rowe LA, Batra D, Juieng P, Stoneburg D, Martin H, Knipe K, Smith P, Strockbine N. 2018. High-quality whole-genome sequences for 59 historical *Shigella* strains generated with PacBio sequencing. *Genome Announc* 6:e00282-18. <https://doi.org/10.1128/genomeA.00282-18>.

This is a work of the U.S. Government and is not subject to copyright protection in the United States. Foreign copyrights may apply. Address correspondence to Rebecca L. Lindsey, rebecca.lindsey@cdc.hhs.gov.

TABLE 1 Strain information and GenBank accession numbers for 59 *Shigella* isolates

Strain (reference no.)	Species	Serotype	GenBank accession no.	Chromosome size(s) (bp)	Associated plasmid size(s) (bp)	Optical map
65-6310	<i>S. boydii</i>	1	PTE500000000	2,603,357, ^a 2,046,275 ^a	164,492 ^a	No
ATCC 8700	<i>S. boydii</i>	2	CP026731–CP026733	4,575,738	59,816, ^a 51,078	Yes
NCTC 9850	<i>S. boydii</i>	3	CP026762 and CP026763	4,575,797	150,497	Yes
NCTC 9770	<i>S. boydii</i>	4	PSSU000000000	3,106,825, ^a 1,499,125 ^a	202,816 ^a	Yes
NCTC 9733	<i>S. boydii</i>	5	CP026844 and CP026845	4,932,934	230,094	Yes
ATCC 12027	<i>S. boydii</i>	6	CP026865 and CP026866	4,680,168 ^a	135,321 ^a	Yes
NCTC 9734	<i>S. boydii</i>	7	CP026874 and CP026875	5,163,976 ^a	193,652 ^a	Yes
NCTC 9353	<i>S. boydii</i>	8	CP026797 and CP026798	4,488,529	146,337	Yes
ATCC 49812	<i>S. boydii</i>	9	CP026836 and CP026837	5,106,737	213,410	Yes
ATCC 12030 (SH135)	<i>S. boydii</i>	10	CP026867–CP026870	4,607,824 ^a	145,169, 52,696, ^a 36,641 ^a	Yes
59-2708	<i>S. boydii</i>	11	CP026846	4,856,857	None	Yes
59-248	<i>S. boydii</i>	14	CP026766 and CP026767	4,678,058	209,166 ^a	Yes
ATCC 12034	<i>S. boydii</i>	15	PSSP000000000	5,100,100, ^a 24,879, ^a 44,022 ^a	251,107 ^a	Yes
ATCC 35964	<i>S. boydii</i>	16	CP026876 and CP026877	5,129,092	166,502	Yes
ATCC 35965	<i>S. boydii</i>	17	PSSR000000000	3,275,068, ^a 1,127,682, ^a 662,313 ^a	182,849 ^a	Yes
ATCC 35966	<i>S. boydii</i>	18	PSSS000000000	4,267,999, ^a 342,076 ^a	128,923 ^a	Yes
83-578	<i>S. boydii</i>	19	CP026813 and CP026814	4,580,582	122,029	Yes
ATCC BAA-1247	<i>S. boydii</i>	20	CP026795 and CP026796	4,575,738	226,559 ^a	Yes
54-1621 (5)	<i>S. boydii</i>	Provisional 54-1621	CP026810	4,531,304 ^a	None	Yes
ATCC 13313	<i>S. dysenteriae</i>	1	CP026774 and CP026775	4,395,762	182,697	Yes
BU53M1 ^b (6)	<i>S. dysenteriae</i>	1	CP024466–CP024469	4,409,083	54,993, 115,922, 184,894	Yes
69-3818	<i>S. dysenteriae</i>	1	CP026777–CP026779	4,390,268	115,128, ^a 202,417	Yes
07-3308	<i>S. dysenteriae</i>	1	CP026878 and CP026879	4,382,687	69,799	No
53-3937	<i>S. dysenteriae</i>	1	CP026780 and CP026781	4,382,743	67,064 ^a	Yes
08-3380	<i>S. dysenteriae</i>	1	CP026782 and CP026783	4,464,195	183,210	No
80-547	<i>S. dysenteriae</i>	1	CP026784 and CP026785	4,391,331	204,250 ^a	Yes
NCTC 9718	<i>S. dysenteriae</i>	1	CP026786 and CP026787	4,371,869	189,871 ^a	No
ATCC 9750 (Sd44)	<i>S. dysenteriae</i>	2	CP026824	4,971,516	None	Yes
ATCC 9751	<i>S. dysenteriae</i>	3	CP026825	4,699,491	None	Yes
ATCC 9753	<i>S. dysenteriae</i>	4	CP026840 and CP026841	4,716,399	215,925	Yes
ATCC 9764	<i>S. dysenteriae</i>	5	CP026872 and CP026873	4,711,098 ^a	166,878 ^a	Yes
ATCC 9754	<i>S. dysenteriae</i>	6	CP026842 and CP026843	4,588,477	106,437	Yes
ATCC 9752	<i>S. dysenteriae</i>	7	CP026838 and CP026839	4,273,636	103,367	Yes
ATCC 12021 (Sd41)	<i>S. dysenteriae</i>	8	CP026826 and CP026827	4,558,619	224,352 ^a	Yes
ATCC 12037	<i>S. dysenteriae</i>	9	CP026828 and CP026829	4,642,274	216,328	Yes
ATCC 12039	<i>S. dysenteriae</i>	10	CP026830 and CP026831	4,880,735	181,100	Yes
ATCC 12038	<i>S. dysenteriae</i>	11	PSSQ000000000	4,292,356, ^a 99,306, ^a 257,965 ^a	165,104 ^a	Yes
ATCC 49550	<i>S. dysenteriae</i>	12	PSST000000000	3,270,959, ^a 1,465,326 ^a	230,788 ^a	Yes
ATCC 49346	<i>S. dysenteriae</i>	14 ^c	CP026832 and CP026833	4,619,326	224,419	Yes
ATCC 49347	<i>S. dysenteriae</i>	15 ^d	CP026834 and CP026835	4,684,535	99,066	Yes
2017C-4522	<i>S. dysenteriae</i>	Provisional 2009C-3478	CP026805 and CP026806	4,609,265	69,966	Yes
96-3162 (7)	<i>S. dysenteriae</i>	Provisional 96-3162	CP026821–CP026823	4,804,763	79,586, 215,820	Yes
204/96 (8)	<i>S. dysenteriae</i>	Provisional 204/96	CP026807–CP026809	4,800,156	79,695, 226,481	Yes
93-119 (9)	<i>S. dysenteriae</i>	Provisional 93-119	CP026815–CP026817	4,799,750	79,553, 226,441	Yes
96-265 ^e (10)	<i>S. dysenteriae</i>	Provisional 96-265	CP026818–CP026820	4,744,419	80,213, 195,489	Yes
E670/74 (11)	<i>S. dysenteriae</i>	Provisional E670/74	CP027027 and CP027028	5,036,586	149,179	Yes
73-5612	<i>S. flexneri</i>	1b	CP026871	4,490,153 ^a	None	No
ATCC 29903	<i>S. flexneri</i>	2a	CP026788–CP026790	4,659,463	113,130, 165,702	Yes
61-4982	<i>S. flexneri</i>	4b	CP026791 and CP026792	4,631,337	59,834	Yes
74-1170	<i>S. flexneri</i>	5a	CP026793 and CP026794	4,733,503	251,323 ^a	No
NCTC 9728	<i>S. flexneri</i>	5b	CP026799 and CP026800	4,511,010	135,368	Yes
64-5500	<i>S. flexneri</i>	6	CP026811 and CP026812	4,659,714	181,479	Yes
98-3193	<i>S. flexneri</i>	7 ^f	CP026776	4,508,802	None	No
04-3145	<i>S. flexneri</i>	7 ^f	CP026764 and CP026765	4,571,921	68,319	No

(Continued on next page)

TABLE 1 (Continued)

Strain (reference no.)	Species	Serotype	GenBank accession no.	Chromosome size(s) (bp)	Associated plasmid size(s) (bp)	Optical map
95-3008	<i>S. flexneri</i>	7 ^f	CP026772 and CP026773	4,516,380 ^a	52,051 ^a	Yes
94-3007 ^b (6)	<i>S. flexneri</i>	7 ^f	CP024473–CP024476	4,533,699	69,554, 82,833, 220,282	Yes
93-3063	<i>S. flexneri</i>	Y	CP026768–CP026771	4,628,330	27,054, ^a 71,925, ^a 220,759	Yes
89-141 (12)	<i>S. flexneri</i>	Provisional 89-141	CP026803 and CP026804	4,481,548	245,004 ^a	No
ATCC 29930	<i>S. sonnei</i>	NA ^g	CP026801 and CP026802	4,975,028	18,973 ^a	Yes

^aIndicates genome/plasmid could not be circularized.

^bTwo strains were previously sequenced but included here because they were part of the originally selected historical *Shigella* strain collection.

^cProposed serotype designation for provisional serovar is E22383 (9).

^dProposed serotype designation for provisional serovar is E23507 (9).

^eFrank Rogers, National Laboratory for Enteric Pathogens, Health Canada, Winnipeg, Canada, personal communication.

^fProposed serotype designation for provisional serotype is 88-893 (94-3007).

^gNA, not applicable.

use of trade names is for identification only and does not imply endorsement by the Centers for Disease Control and Prevention or by the U.S. Department of Health and Human Services.

REFERENCES

- Niyogi SK. 2005. Shigellosis. *J Microbiol* 43:133–143.
- Yang F, Yang J, Zhang X, Chen L, Jiang Y, Yan Y, Tang X, Wang J, Xiong Z, Dong J, Xue Y, Zhu Y, Xu X, Sun L, Chen S, Nie H, Peng J, Xu J, Wang Y, Yuan Z, Wen Y, Yao Z, Shen Y, Qiang B, Hou Y, Yu J, Jin Q. 2005. Genome dynamics and diversity of *Shigella* species, the etiologic agents of bacillary dysentery. *Nucleic Acids Res* 33:6445–6458. <https://doi.org/10.1093/nar/gki954>.
- Smith P, Lindsey RL, Rowe L, Batra A, Stripling D, Garcia-Toledo D, Drapeau L, Knipe DK, Strockbine N. 2017. High-quality whole genome sequences for 21 enterotoxigenic *Escherichia coli* strains generated with PacBio sequencing. *Genome Announc* 6:e01311-17. <https://doi.org/10.1128/genomeA.01311-17>.
- 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>.
- Takasaka M, Kono A, Kono M, Nakamura A, Sakakibara I, Honjo S. 1983. Isolation and pathogenicity of provisional serovar 1621-54 of *Shigella* from imported cynomolgus monkeys. *Jpn J Med Sci Biol* 36:27–37. <https://doi.org/10.7883/yoken1952.36.27>.
- Schroeder MR, Juieng P, Batra D, Knipe K, Rowe LA, Sheth M, Smith P, Garcia-Toledo L, Loparev VN, Lindsey RL. 2018. High-quality complete and draft genome sequences for three *Escherichia* spp. and three *Shigella* spp. generated with Pacific Biosciences and Illumina sequencing and optical mapping. *Genome Announc* 6:e01384-17. <https://doi.org/10.1128/genomeA.01384-17>.
- Kuijper EJ, van Eeden A, de Wever B, van Ketel R, Dankert J. 1997. Nonserotypeable *Shigella dysenteriae* isolated from a Dutch patient returning from India. *Eur J Clin Microbiol Infect Dis* 16:553–554. <https://doi.org/10.1007/BF01708247>.
- Matsushita S, Noguchi Y, Yanagawa Y, Igarashi H, Ueda Y, Hashimoto S, Yano S, Morita K, Kanamori M, Kudoh Y. 1998. *Shigella dysenteriae* strains possessing a new serovar (204/96) isolated from imported diarrheal cases in Japan. *Kansenshogaku Zasshi* 72:499–503. <https://doi.org/10.11150/kansenshogakuzasshi1970.72.499>.
- Matsushita S, Noguchi Y, Yanagawa Y, Kobayashi K, Nakaya H, Igarashi H, Kudoh Y. 1997. *Shigella dysenteriae* strains possessing a new serovar isolated from imported diarrheal cases in Japan. *Kansenshogaku Zasshi* 71:412–416. <https://doi.org/10.11150/kansenshogakuzasshi1970.71.412>.
- Garrity G, Brenner DJ, Krieg NR, Staley JR. 2005. *Bergey's manual of systematic bacteriology: vol 2: the Proteobacteria, part B: the Gamma-proteobacteria*, Springer, New York, NY.
- Gross RJ, Thomas LV, Cheasty T, Rowe B, Lindberg AA. 1989. Four new provisional serovars of *Shigella*. *J Clin Microbiol* 27:829–831.
- Matsushita S, Yamada S, Kudoh Y. 1992. *Shigella flexneri* strains having a new type antigen 89-141. *Kansenshogaku Zasshi* 66:1628–1633. <https://doi.org/10.11150/kansenshogakuzasshi1970.66.1628>.