



Draft Genome Sequences of 72 Isolates from All Four Species of *Shigella*

Jayanthi Gangiredla,^a Mark K. Mammel,^a David W. Lacher,^a Carmen Tartera,^a Scott A. Jackson,^{a*} Isha R. Patel,^a Christopher A. Elkins,^{a*} Amit Mukherjee^a

^aDivision of Molecular Biology, Office of Applied Research and Safety Assessment, Center for Food Safety and Applied Nutrition, U.S. Food and Drug Administration, Laurel, Maryland, USA

ABSTRACT *Shigella* is a genus of Gram-negative enteric pathogenic bacteria which has four species, *Shigella dysenteriae*, *S. flexneri*, *S. boydii*, and *S. sonnei*. *Shigella* species are clinically important bacteria because they cause shigellosis or dysentery. Here we report the genome sequences of 72 *Shigella* isolates from these four species.

Bacteria of the genus *Shigella*, which have been studied for more than 100 years, are foodborne and waterborne enteric pathogens that cause shigellosis or dysentery and are still a global health problem, particularly in developing countries (1–3). The four species of *Shigella*, namely, *Shigella dysenteriae*, *S. flexneri*, *S. boydii*, and *S. sonnei*, were identified by serotyping and biochemical differences (4, 5). Elucidation of the evolutionary relationships among the *Shigella* species was made possible by DNA sequencing of eight housekeeping genes that led to the finding that most *Shigella* serotypes belong to one of three main phylogenetic groups, while *S. sonnei* and some *S. dysenteriae* serotypes lie outside these three groups (6). Subsequent studies led to molecular serotyping based on DNA sequences of genes coding various O antigens (7). In an earlier study, we used phenotypic microarray analysis to identify a distinguishing phenotype in *Escherichia coli* O157:H7 from the 2006 spinach outbreak (8). In continuation of that effort to find new metabolic phenotypes in the genus *Shigella*, which could distinguish the four species, we sequenced the genomes of 72 *Shigella* isolates in order to identify the genetic basis of such phenotypes; the genomic sequences are reported here.

The 72 *Shigella* isolates, which include members of the species *S. dysenteriae* ($n = 16$), *S. flexneri* ($n = 22$), *S. boydii* ($n = 17$), and *S. sonnei* ($n = 17$), are listed in Table 1 with their molecular serotypes and phylogenetic groups. Single-colony cultures of these isolates were grown overnight with shaking in LB liquid medium at 37°C, and 1 ml of these cultures was used to extract total genomic DNA using the DNeasy blood and tissue kit (Qiagen, Hilden, Germany) and the Qiagen QIAcube instrument following the manufacturer's Gram-negative bacteria protocol. Sequencing libraries were prepared with 1 ng of DNA using the Nextera XT DNA sample prep kit (Illumina, San Diego, CA, USA) and sequenced on the Illumina MiSeq platform. The resulting 2 × 250-bp paired-end reads were quality controlled using FastQC, trimmed and adapter sequences removed by Trimmomatic, and *de novo* assembled using SPAdes 3.8.2 (9, 10). Additional adapter trimming after assembly was performed on a few contigs after identification by NCBI.

The average genome coverages ranged between 25× and 489×, with the genome sizes ranging from 4,024,324 bp to 4,917,633 bp, and the overall average G+C content was 50.7%. The number of contigs ranged from 362 to 968, while the N_{50} values ranged from 10,575 bp to 35,986 bp. Molecular serotyping was performed by BLAST analysis

Received 9 October 2018 Accepted 7 November 2018 Published 29 November 2018

Citation Gangiredla J, Mammel MK, Lacher DW, Tartera C, Jackson SA, Patel IR, Elkins CA, Mukherjee A. 2018. Draft genome sequences of 72 isolates from all four species of *Shigella*. *Microbiol Resour Announc* 7:e01399-18. <https://doi.org/10.1128/MRA.01399-18>.

Editor John J. Dennehy, Queens College
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 Amit Mukherjee, amit.mukherjee@fda.hhs.gov.

* Present address: Scott A. Jackson, National Institute of Standards and Technology, Gaithersburg, Maryland, USA; Christopher A. Elkins, Centers for Disease Control and Prevention, Atlanta, Georgia, USA.

TABLE 1 Accession numbers, molecular serotypes, and assembly metrics for 72 *Shigella* strains

Species	Group	Molecular serotype	Strain	GenBank accession no.	No. of reads	Avg coverage (x)	No. of contigs	Genome size (bp)	N ₅₀ value (bp)	G+C content (%)
<i>S. boydii</i>	1	SB1:H45	DMB SH129	QWSD00000000	4,677,396	260	572	4,382,335	16,731	50.9
<i>S. boydii</i>	1	SB1:H45	DMB SH132	QWSG00000000	4,747,996	312	553	4,287,257	17,020	50.7
<i>S. boydii</i>	1	SB1:H45	DMB SH20008	QWSK00000000	1,218,348	46	594	4,287,598	17,088	50.7
<i>S. boydii</i>	1	SB1:H45	DMB SH20139	QWTH00000000	1,115,244	43	582	4,236,879	16,676	50.7
<i>S. boydii</i>	1	SB2:H45	DMB SH112	QWRU00000000	3,080,314	187	472	4,342,486	19,887	50.7
<i>S. boydii</i>	1	SB2:H45	DMB SH121	QWRX00000000	1,166,222	42	520	4,327,919	19,513	50.7
<i>S. boydii</i>	1	SB2:H45	DMB SH130	QWSE00000000	763,174	31	626	4,106,788	14,520	50.9
<i>S. boydii</i>	1	SB2:H45	DMB SH136	QWSI00000000	4,113,256	248	503	4,447,944	21,333	50.8
<i>S. boydii</i>	1	SB3:H45	DMB SH131	QWSF00000000	6,317,096	489	545	4,218,811	17,032	50.9
<i>S. boydii</i>	1	SB10:H45	DMB SH135	QXHH00000000	915,640	34	731	4,275,334	13,860	51.0
<i>S. boydii</i>	1	SB10:H45	DMB SH20138	QWTG00000000	1,140,324	44	643	4,265,386	15,923	50.7
<i>S. boydii</i>	1	SB10:H45	DMB SH20140	QWTI00000000	1,033,272	40	633	4,264,430	16,283	50.7
<i>S. boydii</i>	1	SB10:H45	DMB SH20247	QWTW00000000	1,511,832	110	562	4,252,934	18,050	50.7
<i>S. boydii</i>	1	SB10:H45	DMB SH20248	QWTX00000000	2,458,940	169	586	4,295,019	17,602	50.7
<i>S. boydii</i>	2	SB7:H14	DMB SH134	QWSH00000000	1,482,596	46	516	4,772,209	26,716	50.6
<i>S. boydii</i>	2	SB9:H14	DMB SH20032	QWSV00000000	3,243,250	181	456	4,917,633	35,986	50.4
<i>S. boydii</i>	2	SB16:H9	DMB SH137	QWSJ00000000	889,534	28	743	4,879,160	17,442	50.5
<i>S. dysenteriae</i>	1	SD3:H45	DMB SH116	QXHX00000000	1,305,836	45	649	4,348,151	16,810	50.7
<i>S. dysenteriae</i>	1	SD3:H45	DMB SH20208	QXIH00000000	1,369,306	45	644	4,351,163	17,978	50.7
<i>S. dysenteriae</i>	1	SD4:H45	DMB SH109	QXHW00000000	1,083,012	37	726	4,382,812	14,288	50.6
<i>S. dysenteriae</i>	1	SD4:H45	DMB SH20168	QXIC00000000	1,687,918	61	674	4,457,263	19,461	50.7
<i>S. dysenteriae</i>	1	SD4:H45	DMB SH20196	QXID00000000	1,310,606	97	531	4,326,989	21,317	50.7
<i>S. dysenteriae</i>	1	SD6:H45	DMB SH20198	QXIE00000000	1,456,558	119	603	4,328,352	19,812	50.7
<i>S. dysenteriae</i>	1	SD8:H26	DMB SH20207	QWTU00000000	1,711,382	63	531	4,327,664	23,909	50.7
<i>S. dysenteriae</i>	1	SD11:H45	DMB SH20201	QXIG00000000	1,273,758	48	515	4,073,274	19,045	50.9
<i>S. dysenteriae</i>	1	ONT:H45	DMB SH20200	QXIF00000000	1,410,470	47	596	4,210,534	18,645	50.9
<i>S. dysenteriae</i>	2	SD2:H14	DMB SH120	QWRW00000000	1,707,606	58	501	4,539,019	24,700	50.6
<i>S. dysenteriae</i>	2	SD2:H14	DMB SH127	QWSB00000000	1,992,700	66	520	4,545,392	23,447	50.5
<i>S. dysenteriae</i>	2	SD2:H14	DMB SH128	QWSC00000000	4,897,112	267	515	4,641,380	25,093	50.5
<i>S. dysenteriae</i>	2	SD2:H14	DMB SH20169	QWTO00000000	899,848	29	933	4,628,618	10,691	50.5
<i>S. dysenteriae</i>	SD1	SD1:H18	DMB SH20028	QWSS00000000	1,391,440	51	682	4,138,748	13,147	50.8
<i>S. dysenteriae</i>	SD1	SD1:H18	DMB SH20174	QWTS00000000	1,605,964	56	657	4,024,324	13,197	51.0
<i>S. dysenteriae</i>	SD1	SD1:H18	DMB SH20191	QWTT00000000	1,027,776	37	758	4,120,378	11,203	50.8
<i>S. flexneri</i>	1	SF6:H45	DMB SH20107	QXHY00000000	1,650,448	123	475	4,320,907	20,376	50.7
<i>S. flexneri</i>	1	SF6:H45	DMB SH20108	QXHZ00000000	2,526,490	170	493	4,315,004	20,142	50.7
<i>S. flexneri</i>	1	SF6:H45	DMB SH20109	QXIA00000000	4,486,192	286	455	4,256,784	20,366	50.8
<i>S. flexneri</i>	1	SF6:H45	DMB SH20134	QXIB00000000	1,580,886	55	562	4,389,422	19,425	50.7
<i>S. flexneri</i>	3	SF1a:H14	DMB SH20092	QWTB00000000	651,772	25	727	4,305,976	12,910	50.8
<i>S. flexneri</i>	3	SF1b:H14	DMB SH125	QWSA00000000	643,812	27	652	4,378,060	16,015	50.5
<i>S. flexneri</i>	3	SF1b:H14	DMB SH20017	QWSM00000000	1,069,082	43	391	4,164,268	26,816	50.8
<i>S. flexneri</i>	3	SF1b:H14	DMB SH20026	QWSR00000000	725,528	31	542	4,214,513	18,735	50.8
<i>S. flexneri</i>	3	SF2a:H14	DMB SH107	QWRR00000000	1,595,260	56	444	4,415,938	28,786	50.5
<i>S. flexneri</i>	3	SF2a:H14	DMB SH123	QWRZ00000000	739,976	30	544	4,403,207	20,805	50.5
<i>S. flexneri</i>	3	SF2a:H14	DMB SH20018	QWSN00000000	691,076	25	812	4,356,278	10,765	50.7
<i>S. flexneri</i>	3	SF2a:H14	DMB SH20173	QWTR00000000	1,817,964	131	430	4,610,016	31,303	50.3
<i>S. flexneri</i>	3	SF3a:H14	DMB SH108	QWRS00000000	1,781,898	61	499	4,356,861	25,857	50.5
<i>S. flexneri</i>	3	SF3a:H14	DMB SH20100	QWTC00000000	875,648	27	968	4,575,761	10,575	50.7
<i>S. flexneri</i>	3	SF3a:H14	DMB SH20170	QWTP00000000	3,641,616	309	442	4,383,781	27,268	50.4
<i>S. flexneri</i>	3	SF4b:H14	DMB SH20101	QWTD00000000	3,938,282	266	385	4,340,521	31,440	50.8
<i>S. flexneri</i>	3	SF4b:H14	DMB SH20103	QWTF00000000	1,543,768	103	403	4,473,264	30,069	50.7
<i>S. flexneri</i>	3	SF4bv:H14	DMB SH20102	QWTE00000000	842,058	36	519	4,250,168	19,953	50.8
<i>S. flexneri</i>	3	SF5a:H14	DMB SH20012	QWSL00000000	3,331,184	220	383	4,404,256	30,069	50.5
<i>S. flexneri</i>	3	SF5a:H14	DMB SH20025	QWSQ00000000	6,204,462	373	362	4,341,244	29,579	50.7
<i>S. flexneri</i>	3	SF5a:H14	DMB SH20030	QWST00000000	688,084	25	815	4,412,366	10,680	50.5
<i>S. flexneri</i>	3	SFY:H14	DMB SH20024	QWSP00000000	2,118,110	42	471	4,292,243	26,569	50.7
<i>S. sonnei</i>	SS	SS:H16	DMB SH111	QWRT00000000	5,161,528	314	502	4,565,219	22,247	50.5
<i>S. sonnei</i>	SS	SS:H16	DMB SH20033	QWSW00000000	810,324	30	719	4,749,503	18,660	50.6
<i>S. sonnei</i>	SS	SS:H16	DMB SH20034	QWSX00000000	2,968,762	186	490	4,769,276	25,218	50.6
<i>S. sonnei</i>	SS	SS:H16	DMB SH20158	QWTK00000000	1,092,978	42	581	4,559,368	22,109	50.5
<i>S. sonnei</i>	SS	SS:H16	DMB SH20171	QWTQ00000000	2,958,016	36	554	4,599,591	23,352	50.5
<i>S. sonnei</i>	SS	O--:H16	DMB SH114	QWRV00000000	988,926	34	561	4,428,377	17,676	50.8
<i>S. sonnei</i>	SS	O--:H16	DMB SH122	QWRY00000000	883,062	32	652	4,455,949	15,399	50.8
<i>S. sonnei</i>	SS	O--:H16	DMB SH20023	QWSO00000000	1,580,420	94	463	4,576,543	25,245	50.8
<i>S. sonnei</i>	SS	O--:H16	DMB SH20031	QWSU00000000	3,708,078	231	446	4,566,720	25,218	50.8
<i>S. sonnei</i>	SS	O--:H16	DMB SH20036	QWSY00000000	3,225,642	210	458	4,638,655	25,245	50.8

(Continued on next page)

TABLE 1 (Continued)

Species	Group	Molecular serotype	Strain	GenBank accession no.	No. of reads	Avg coverage (×)	No. of contigs	Genome size (bp)	N_{50} value (bp)	G+C content (%)
<i>S. sonnei</i>	SS	O–:H16	DMB SH20046	QWSZ00000000	958,702	32	636	4,577,160	17,407	50.8
<i>S. sonnei</i>	SS	O–:H16	DMB SH20090	QWTA00000000	3,757,052	41	486	4,687,163	25,118	50.9
<i>S. sonnei</i>	SS	O–:H16	DMB SH20142	QWTJ00000000	1,150,584	34	636	4,484,359	17,533	50.8
<i>S. sonnei</i>	SS	O–:H16	DMB SH20160	QWTL00000000	681,874	26	815	4,759,065	15,244	50.8
<i>S. sonnei</i>	SS	O–:H16	DMB SH20163	QWTM00000000	783,574	30	752	4,633,102	17,526	50.9
<i>S. sonnei</i>	SS	O–:H16	DMB SH20165	QWTN00000000	961,914	33	640	4,667,361	20,357	50.8
<i>S. sonnei</i>	SS	O–:H16	DMB SH20230	QWTV00000000	947,578	33	563	4,542,501	19,521	50.8

using the *wzx*, *wzy*, and *fliC* loci. Strains classified as *S. flexneri* were subtyped by BLAST using the *gtr1*, *gtr2*, *gtr4*, *gtr5*, *gtrX*, *oac*, and *opt* loci.

Data availability. The draft genome assemblies were deposited in DDBJ/ENA/GenBank through the FDA's GenomeTrakr pipeline under BioProject accession number [PRJNA273284](https://www.ncbi.nlm.nih.gov/BioProject/ accession/PRJNA273284) with the accession numbers shown in Table 1. The versions described in this announcement are the first ones. All annotations were done by using the NCBI PGAP pipeline (11).

ACKNOWLEDGMENTS

The views expressed in this article are those of the authors and do not necessarily reflect the official policy of the Department of Health and Human Services, the U.S. Food and Drug Administration (FDA), or the U.S. Government. Reference to any commercial materials, equipment, or process does not in any way constitute approval, endorsement, or recommendation by the FDA.

REFERENCES

- Niyogi SK. 2005. Shigellosis. *J Microbiol* 43:133–143.
- Lampel KA, Formal SB, Maurelli AT. 2018. A brief history of Shigella. *EcoSal Plus* 8. <https://doi.org/10.1128/ecosalplus.ESP-0006-2017>.
- Scallan E, Hoekstra RM, Angulo FJ, Tauxe RV, Widdowson MA, Roy SL, Jones JL, Griffin PM. 2011. Foodborne illness acquired in the United States—major pathogens. *Emerg Infect Dis* 17:7–15. https://wwwnc.cdc.gov/eid/article/17/1/P1-1101_article.
- Ewing WH. 1949. *Shigella* nomenclature. *J Bacteriol* 57:633–638.
- Lan R, Reeves PR. 2002. *Escherichia coli* in disguise: molecular origins of Shigella. *Microbes Infect* 4:1125–1132. [https://doi.org/10.1016/S1286-4579\(02\)01637-4](https://doi.org/10.1016/S1286-4579(02)01637-4).
- Pupo GM, Lan R, Reeves PR. 2000. Multiple independent origins of Shigella clones of *Escherichia coli* and convergent evolution of many of their characteristics. *Proc Natl Acad Sci U S A* 97:10567–10572. <https://doi.org/10.1073/pnas.180094797>.
- Liu B, Knirel YA, Feng L, Perepelov AV, Senchenkova SN, Wang Q, Reeves PR, Wang L. 2008. Structure and genetics of *Shigella* O antigens. *FEMS Microbiol Rev* 32:627–653. <https://doi.org/10.1111/j.1574-6976.2008.00114.x>.
- Mukherjee A, Mammel MK, Leclerc JE, Cebula TA. 2008. Altered utilization of *N*-acetyl-D-galactosamine by *Escherichia coli* O157:H7 from the 2006 spinach outbreak. *J Bacteriol* 190:1710–1717. <https://doi.org/10.1128/JB.01737-07>.
- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>.
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Pribelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19:455–477. <https://doi.org/10.1089/cmb.2012.0021>.
- 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>.