### Title

Shigella Draft Genome Sequences: Resources for Food Safety and Public Health




### Journal

Microbiology Resource Announcements, 5(16)

### ISSN

2576-098X



### Authors

Weis, Allison M

Gilpin, Brent

Huang, Bihua C

et al.



### Publication Date

2017-04-20

### DOI

10.1128/genomea.00176-17

Peer reviewed

AMERICAN SOCIETY FOR MICROBIOLOGY | genomeAnnouncements™

PROKARYOTES

# *Shigella* Draft Genome Sequences: Resources for Food Safety and Public Health

**Allison M. Weis,[a] Brent Gilpin,[b] Bihua C. Huang,[a] Nguyet Kong,[a] Poyin Chen,[a] Bart C. Weimer[a]**

School of Veterinary Medicine, 100K Pathogen Genome Project, UC Davis, Davis, California, USA[a]; Institute of Environmental Science & Research Ltd., Christchurch, New Zealand[b]

**ABSTRACT** *Shigella* is a major foodborne pathogen that infects humans and nonhuman primates and is the major cause of dysentery and reactive arthritis worldwide. This is the initial public release of 16 *Shigella* genome sequences from four species sequenced as part of the 100K Pathogen Genome Project.

**Received** 15 February 2017 **Accepted** 6 March 2017 **Published** 20 April 2017

**Citation** Weis AM, Gilpin B, Huang BC, Kong N, Chen P, Weimer BC. 2017. *Shigella* draft genome sequences: resources for food safety and public health. Genome Announc 5: e00176-17. https://doi.org/10.1128/genomeA.00176-17.



Address correspondence to Bart C. Weimer, bcweimer@ucdavis.edu.

*Shigella* spp. are Gram-negative enteric pathogens that infect humans and nonhuman primates. They are an important cause of dysentery, affecting more than 80 million people and causing more than 700,000 deaths each year worldwide (1, 2). The burden of disease is carried by children, where 99% of infections occur in children in developing nations, and most cases (70%) and deaths (60%) occur in children age 5 and under (1, 2). Rare cases of shigellosis can lead to reactive arthritis (3). *Shigella* is spread by direct contact with an infected person or by ingesting contaminated food or water (1, 4). The infective dose can be as few as 10 organisms, making *Shigella* a foodborne pathogen of global importance based on wide distribution, water quality concerns, and an important risk for public health (4).

The genus *Shigella* is composed of four species: *S. dysenteriae*, *S. flexneri*, *S. boydii*, and *S. sonnei*, all of which cause acute bloody diarrhea (2, 5). *Shigella* genomics has emerged as an important tool in basic and clinical applications for diagnosis and classification, and will inform treatment plans (5, 6), but the ability to conduct source tracking using whole-genome sequencing remains challenging due to the relatively few publically available genomes. In this release, the 100K Pathogen Genome Project sequenced and assembled the genomes of 16 novel *Shigella* isolates of the four species: two *S. boydii*, three *S. dysenteriae*, nine *S. flexneri*, and two *S. sonnei* isolates (Table 1).

The 100K Pathogen Genome Project (http://www.100kgenomes.org) is a large-scale sequencing effort to inform food safety and public health in genome-based identification and source tracking (7, 8). All *Shigella* isolates were shipped to Bart Weimer's laboratory (UC Davis, Davis, CA). DNA isolation, sequencing, and assembly were done as previously described (7–9). Briefly, isolates were checked for purity (10) prior to extracting genomic DNA (gDNA) from cultures grown on brain heart infusion agar (catalog no. 241830; BD Difco, Franklin Lakes, NJ) for 1 to 2 days at 37°C. Cells were lysed (11), gDNA was purified using the Qiagen QIAamp DNA minikit (catalog no. 51306), and quality was measured using the Agilent 2200 TapeStation system with the Genomic DNA ScreenTape (12). After isolation, gDNA was fragmented using Covaris E220 (13), end-repaired (5′), adenylated (3′), and ligated with double-stranded DNA (dsDNA) adapters NEXTflex-96 DNA barcode (Bioo Scientific, Austin, TX), and gDNA (1 $\mu$g) was used for library construction with the Kapa high-throughput (HTP) library preparation kit (catalog no. KK8234; Kapa Biosystems, Boston, MA), using the Agilent Bravo automated liquid handling platform workstation option B (Santa Clara, CA). The

**TABLE 1** *Shigella* species draft genome sequence information

| GenBank accession no. | Strain ID | Species | Depth (×) | No. of contigs | No. of bases |
|---|---|---|---|---|---|
| MSJS00000000 | BCW_4868 | *S. boydii* | 115 | 243 | 4,863,576 |
| MSJT00000000 | BCW_4869 | *S. boydii* | 108 | 297 | 4,246,029 |
| MSJU00000000 | BCW_4870 | *S. dysenteriae* | 114 | 285 | 4,018,103 |
| MSJV00000000 | BCW_4871 | *S. dysenteriae* | 117 | 299 | 4,078,019 |
| MSJW00000000 | BCW_4872 | *S. dysenteriae* | 72 | 292 | 4,490,659 |
| MSJX00000000 | BCW_4874 | *S. flexneri* | 109 | 269 | 4,252,909 |
| MSJY00000000 | BCW_4875 | *S. flexneri* | 90 | 249 | 4,196,256 |
| MSJZ00000000 | BCW_4876 | *S. flexneri* | 101 | 293 | 4,396,898 |
| MSKA00000000 | BCW_4877 | *S. flexneri* | 100 | 296 | 4,330,224 |
| MSKC00000000 | BCW_4879 | *S. flexneri* | 124 | 287 | 4,167,963 |
| MSKB00000000 | BCW_4880 | *S. flexneri* | 106 | 267 | 4,224,783 |
| MSKD00000000 | BCW_4881 | *S. flexneri* | 170 | 253 | 4,334,622 |
| MSKG00000000 | BCW_4882 | *S. flexneri* | 96 | 297 | 4,099,589 |
| MSKF00000000 | BCW_4883 | *S. flexneri* | 118 | 289 | 4,305,926 |
| MSKE00000000 | BCW_4885 | *S. sonnei* | 101 | 299 | 4,392,417 |
| MSKH00000000 | BCW_4886 | *S. sonnei* | 100 | 286 | 4,530,575 |

libraries were size selected using dual SPRI selection (0.2× to 0.6×) to produce libraries with fragments between 300 and 450 bp. Final library amplification was done with eight cycles using the Kapa HiFi HotStart ReadyMix, followed by a 1× SPRI bead cleanup. Prior to sequencing, the library size was confirmed using the Agilent 2100 Bioanalyzer system with high-sensitivity DNA kit (14, 15), quantified with a quantitative PCR (qPCR)-based Kapa library quantification kit (catalog no. KK4824), pooled with multiplexing up to 96 isolates, and sequenced on the Illumina HiSeq 2000 with PE100 plus index read at BGI@UC Davis (Sacramento, CA). The paired-end reads were assembled using CLC Genomics Workbench version 6.5.1 (Qiagen).

**Accession number(s).** Sequences can be found in the NCBI SRA 100K Project BioProject PRJNA186441 and in GenBank (Table 1).

## ACKNOWLEDGMENTS

We thank the Weimer lab and all their efforts in isolate logistics and technical assistance and all of the collaborators for the 100K Pathogen Genome Project.

This project was funded by the 100K Pathogen Genome Project with initial funding from Agilent Technologies to produce these sequences.

## REFERENCES

1. Kotloff KL, Winickoff JP, Ivanoff B, Clemens JD, Swerdlow DL, Sansonetti PJ, Adak GK, Levine MM. 1999. Global burden of *Shigella* infections: implications for vaccine development and implementation of control strategies. Bull World Health Organ 77:651–666.
2. WHO. 2005. Guidelines for the control of shigellosis, including epidemics due to *Shigella dysenteriae* 1. World Health Organization, Geneva, Switzerland.
3. Gaston JSH, Lillicrap MS. 2003. Arthritis associated with enteric infection. Best Pract Res Clin Rheumatol 17:219–239.
4. DuPont HL, Levine MM, Hornick RB, Formal SB. 1989. Inoculum size in shigellosis and implications for expected mode of transmission. J Infect Dis 159:1126–1128. https://doi.org/10.1093/infdis/159.6.1126.
5. Hale TL. 1991. Genetic basis of virulence in *Shigella* species. Microbiol Rev 55:206–224.
6. Yang F, Yang J, Zhang XB, Chen LH, Jiang Y, Yan YL, Tang XD, Wang J, Xiong ZH, Dong J, Xue Y, Zhu YF, Xu XY, Sun LL, Chen SX, Nie H, Peng JP, Xu JG, Wang Y, Yuan ZH, Wen YM, Yao ZJ, Shen Y, Qiang BQ, Hou YD, 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.
7. Weis AM, Clothier KA, Huang BC, Kong N, Weimer BC. 2016. Draft genome sequences of *Campylobacter jejuni* strains that cause abortion in livestock. Genome Announc 4(6):e01324-16. https://doi.org/10.1128/genomeA.01324-16.
8. Weis AM, Storey DB, Taff CC, Townsend AK, Huang BC, Kong NT, Clothier KA, Spinner A, Byrne BA, Weimer BC. 2016. Genomic comparison of *Campylobacter* spp. and their potential for zoonotic transmission between birds, primates, and livestock. Appl Environ Microbiol 82:7165–7175. https://doi.org/10.1128/AEM.01746-16.
9. Weis AM, Huang BC, Storey DB, Kong N, Chen P, Arabyan N, Gilpin B, Mason C, Townsend AK, Smith WA, Byrne BA, Taff CC, Weimer BC. 2017. Large-scale release of *Campylobacter* draft genomes: resources for food safety and public health from the 100K pathogen genome project. Genome Announc 5(1):e00925-16. https://doi.org/10.1128/genomeA.00925-16.
10. Kong N, Ng W, Lee V, Kelly L, Weimer BC. 2013. Production and analysis of high molecular weight genomic DNA for NGS pipelines using Agilent DNA extraction kit (p/n 200600). Application note. Agilent Technologies, Santa Clara, CA. https://www.agilent.com/cs/library/applications/5991-3722EN.pdf.
11. Jeannotte R, Lee E, Kong N, Ng W, Kelly L, Weimer BC. 2014. High-throughput analysis of foodborne bacterial genomic DNA using Agilent 2200 TapeStation and genomic DNA ScreenTape system. Application note. Agilent Technologies, Santa Clara, CA. https://www.agilent.com/cs/library/applications/5991-4003EN.pdf.
12. Kong N, Ng W, Cai L, Leonardo A, Kelly L, Weimer BC. 2014. Integrating the DNA integrity number (DIN) to assess genomic DNA (gDNA) quality control using the Agilent 2200 TapeStation system. Application note. Agilent Technologies, Santa Clara, CA. http://www.agilent.com/cs/library/applications/5991-5442EN.pdf.

13. Jeannotte R, Lee E, Arabyan N, Kong N, Thao K, Huang BH, Kelly L, Weimer BC. 2014. Optimization of Covaris settings for shearing bacterial genomic DNA by focused ultrasonication and analysis using Agilent 2200 TapeStation. Application note. Agilent Technologies, Santa Clara, CA. http://cn.agilent.com/cs/library/applications/5991-5075EN.pdf.
14. Kong N, Ng W, Foutouhi A, Huang BH, Kelly L, Weimer BC. 2014. Quality control of high-throughput library construction pipeline for KAPA HTP library using an Agilent 2200 TapeStation. Application note. Agilent Technologies, Santa Clara, CA. http://www.agilent.com/cs/library/applications/5991-5141EN.pdf.
15. Kong N, Thao K, Huang C, Appel M, Lappin S, Knapp L, Kelly L, Weimer BC. 2014. Automated library construction using KAPA library preparation kits on the Agilent NGS workstation yields high-quality libraries for whole-genome sequencing on the Illumina platform. Application note. Agilent Technologies, Santa Clara, CA. http://www.agilent.com/cs/library/applications/5991-4296EN.pdf.