

Draft Genome Sequence of 24570, the Type Strain of *Shigella flexneri*

Kate S. Baker,^a  Julian Parkhill,^a Nicholas R. Thomson^{a,b}

Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, United Kingdom^a; London School of Hygiene and Tropical Medicine, London, United Kingdom^b

***Shigella flexneri* is a diarrheal pathogen that causes a large disease burden worldwide. We sequenced the genome of the publicly available type strain (*S. flexneri* 2a strain 24570) of this bacterial species to increase its utility as a reference. We present genome assembly results and comparisons with other reference strains.**

Received 30 March 2015 Accepted 30 April 2015 Published 28 May 2015

Citation Baker KS, Parkhill J, Thomson NR. 2015. Draft genome sequence of 24570, the type strain of *Shigella flexneri*. *Genome Announc* 3(3):e00393-15. doi:10.1128/genomeA.00393-15.

Copyright © 2015 Baker et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 3.0 Unported license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Nicholas R. Thomson, nrt@sanger.ac.uk.

Shigella flexneri is a highly contagious bacterial pathogen responsible for a large global disease burden. Every year, approximately 750,000 children under the age of 5 years die from diarrheal disease in developing nations (1). Within that demographic group, *Shigella* is one of the top four attributable causes of diarrheal disease (2), with *Shigella flexneri* causing the majority of the shigellosis burden in these areas (3). Bacterial type strains serve as a benchmarking tool for diagnostic and research laboratories around the world (4) and the type strain of the *Shigella flexneri* strain (strain 24570) is available at the American Type Culture Collection.

Genomic DNA from the *Shigella flexneri* type strain 24570 (ATCC 29903) was obtained and sequenced using 150 bp paired-end Illumina (MiSeq) sequencing according to in-house protocols (5, 6), with an approximately 500 bp insert size. Reads ($n = 2,037,336$) were assembled into 327 contiguous sequence scaffolds using Velvet Optimizer (7) and these were annotated using Prokka (8), to produce a high-quality draft genome (9). BLAST comparisons were made to determine the presence/absence of genes in the annotated assembly, and phylogenetic analysis was performed as in reference (10).

The reads of the strain 24570 draft genome mapped against reference genome 2457T (11) with approximately 70× coverage. The assembly had an N_{50} of 39,568 bp and a total length of 4.7 MB. Automated annotation predicted the presence of 4,653 coding sequences, including the serotype conferring *gtrII*-gene. Phylogenetic sequence analysis (performed as in reference 10) confirmed that the type strain of *Shigella flexneri* fits within a serotype 2a lineage (not shown). When compared with complete *Shigella flexneri* serotype 2a genomes, 24570 had 1,103 single nucleotide polymorphisms (SNPs) relative to first completed reference genome strain 301 (12), 203 SNPs relative to the common laboratory strain 2457T (11), and 414 SNPs relative to the World War 1 strain, NCTC1 (10). The availability of a draft genome for this publicly available type strain increases its utility as a resource for diagnostic and research laboratories.

Nucleotide sequence accession numbers. This whole-genome shotgun project has been deposited in the European Nucleotide Archive under Bioproject number PRJEB2976, sample ERS574920, run ERR738429. Contiguous sequences of

the *de novo* assembly have also been deposited at the European Nucleotide Archive under the accession numbers CELV01000001 to CELV01000327.

ACKNOWLEDGMENT

This study was funded by Wellcome Trust grant 098051.

REFERENCES

- Liu L, Johnson HL, Cousens S, Perin J, Scott S, Lawn JE, Rudan I, Campbell H, Cibulskis R, Li M, Mathers C, Black RE, Child Health Epidemiology Reference Group of WHO and UNICEF. 2012. Global, regional, and national causes of child mortality: an updated systematic analysis for 2010 with time trends since 2000. *Lancet* 379:2151–2161. [http://dx.doi.org/10.1016/S0140-6736\(12\)60560-1](http://dx.doi.org/10.1016/S0140-6736(12)60560-1).
- Kotloff KL, Nataro JP, Blackwelder WC, Nasrin D, Farag TH, Panchalingam S, Wu Y, Sow SO, Sur D, Breiman RF, Faruque AS, Zaidi AK, Saha D, Alonso PL, Tamboura B, Sanogo D, Onwuchekwa U, Manna B, Ramamurthy T, Kanungo S, Ochieng JB, Omore 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, Acacio 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 Multicenter Study, GEMS): a prospective, case-control study. *Lancet* 382:209–222. [http://dx.doi.org/10.1016/S0140-6736\(13\)60844-2](http://dx.doi.org/10.1016/S0140-6736(13)60844-2).
- Livio S, Strockbine NA, Panchalingam S, Tennant SM, Barry EM, Marohn ME, Antonio M, Hossain A, Mandomando I, Ochieng JB, Oundo JO, Qureshi S, Ramamurthy T, Tamboura B, Adegbola RA, Hossain MJ, Saha D, Sen S, Faruque AS, Alonso PL, Breiman RF, Zaidi AK, Sur D, Sow SO, Berkeley LY, O'Reilly C, Mintz ED, Biswas K, Cohen D, Farag TH, Nasrin D, Wu Y, Blackwelder WC, Kotloff KL, Nataro JP, Levine MM. 2014. *Shigella* isolates from the Global Enteric Multicenter Study inform vaccine development. *Clin Infect Dis* 59: 933–941. <http://dx.doi.org/10.1093/cid/ciu468>.
- Rowe B, Gross RJ. 1984. *Bergey's manual of systematic bacteriology*, 1st ed, vol 1. Lippincott Williams & Wilkins, Baltimore, MD.
- Quail MA, Kozarewa I, Smith F, Scally A, Stephens PJ, Durbin R, Swerdlow H, Turner DJ. 2008. A large genome center's improvements to the Illumina sequencing system. *Nat Methods* 5:1005–1010. <http://dx.doi.org/10.1038/nmeth.1270>.
- Quail MA, Otto TD, Gu Y, Harris SR, Skelly TF, McQuillan JA, Swerdlow HP, Oyola SO. 2012. Optimal enzymes for amplifying sequencing libraries. *Nat Methods* 9:10–11. <http://dx.doi.org/10.1038/nmeth.1814>.
- Zerbino DR, Birney E. 2008. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Res* 18:821–829. <http://dx.doi.org/10.1101/gr.074492.107>.

8. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30:2068–2069. <http://dx.doi.org/10.1093/bioinformatics/btu153>.
9. Chain PS, Grafham DV, Fulton RS, Fitzgerald MG, Hostettler J, Muzny D, Ali J, Birren B, Bruce DC, Buhay C, Cole JR, Ding Y, Dugan S, Field D, Garrity GM, Gibbs R, Graves T, Han CS, Harrison SH, Highlander S, Hugenholtz P, Khouri HM, Kodira CD, Kolker E, Kyrpides NC, Lang D, Lapidus A, Malfatti SA, Markowitz V, Metha T, Nelson KE, Parkhill J, Pitluck S, Qin X, Read TD, Schmutz J, Sozhamannan S, Sterk P, Strausberg RL, Sutton G, Thomson NR, Tiedje JM, Weinstock G, Wollam A, Genomic Standards Consortium Human Microbiome Project Jumpstart, Detter JC. 2009. Genome project standards in a new era of sequencing. *Science* 326:236–237. <http://dx.doi.org/10.1126/science.1180614>.
10. Baker KS, Mather AE, McGregor H, Coupland P, Langridge GC, Day M, Deheer-Graham A, Parkhill J, Russell JE, Thomson NR. 2014. The extant World War I dysentery bacillus NCTC1: a genomic analysis. *Lancet* 384:1691–1697. [http://dx.doi.org/10.1016/S0140-6736\(14\)61789-X](http://dx.doi.org/10.1016/S0140-6736(14)61789-X).
11. Wei J, Goldberg MB, Burland V, Venkatesan MM, Deng W, Fournier G, Mayhew GF, Plunkett G III, Rose DJ, Darling A, Mau B, Perna NT, Payne SM, Runyen-Janecky LJ, Zhou S, Schwartz DC, Blattner FR. 2003. Complete genome sequence and comparative genomics of *Shigella flexneri* serotype 2a strain 2457T. *Infect Immun* 71:2775–2786. <http://dx.doi.org/10.1128/IAI.71.5.2775-2786.2003>.
12. Jin Q, Yuan Z, Xu J, Wang Y, Shen Y, Lu W, Wang J, Liu H, Yang J, Yang F, Zhang X, Zhang J, Yang G, Wu H, Qu D, Dong J, Sun L, Xue Y, Zhao A, Gao Y, Zhu J, Kan B, Ding K, Chen S, Cheng H, Yao Z, He B, Chen R, Ma D, Qiang B, Wen Y, Hou Y, Yu J. 2002. Genome sequence of *Shigella flexneri* 2a: insights into pathogenicity through comparison with genomes of *Escherichia coli* K12 and O157. *Nucleic Acids Res* 30:4432–4441. <http://dx.doi.org/10.1093/nar/gkf566>.