

Draft Whole-Genome Sequences of 25 *Salmonella enterica* Strains Representing 24 Serovars

Catherine Yoshida,^a Stephanie L. Brumwell,^a Erika J. Lingohr,^a Aaminah Ahmad,^a Travis M. Blimkie,^a Benjamin A. Kogan,^a Jessica Pilsworth,^a Muhammad A. Rehman,^a Krista L. Schleicher,^a Jenitta Shanmugaraj,^a Andrew M. Kropinski,^b John H. E. Nash^{a,b}

National Microbiology Laboratory at Guelph, Infectious Disease Prevention and Control Branch, Public Health Agency of Canada, Government of Canada, Guelph, Ontario, Canada^a; Department of Pathobiology, Ontario Veterinary College, University of Guelph, Guelph, Ontario, Canada^b

We report the draft genome sequences of 25 *Salmonella enterica* strains representing 24 different serotypes, many of which were not available in public repositories during our selection process. These draft genomes will provide useful reference for the genetic variation between serotypes and aid in the development of molecular typing tools.

Received 15 December 2015 Accepted 22 January 2016 Published 3 March 2016

Citation Yoshida C, Brumwell SL, Lingohr EJ, Ahmad A, Blimkie TM, Kogan BA, Pilsworth J, Rehman MA, Schleicher KL, Shanmugaraj J, Kropinski AM, Nash JHE. 2016. Draft whole-genome sequences of 25 *Salmonella enterica* strains representing 24 serovars. *Genome Announc* 4(2):e01718-15. doi:10.1128/genomeA.01718-15.

Copyright © 2016 Yoshida 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 Catherine Yoshida, Catherine.Yoshida@phac-aspc.gc.ca.

Salmonella spp. are Gram-negative bacteria commonly responsible for human diarrheal illness through the contamination of foods including meats and produce. There are an estimated 1.7 million incidents of salmonellosis each year in North America alone, with over 80 million occurring globally (1). Having the ability to attribute illness to specific serotypes and subtypes can aid outbreak detection and source tracking. There is a momentum toward the acceptance and integration of new molecular methods for pathogen detection and classification. This places great impor-

tance on the availability of unique whole-genome sequences to accurately differentiate serotypes and subtypes and to validate methodologies. Genome sequences facilitate advancements in the detection and characterization of *Salmonella* serotypes, and those presented here contributed to the development of the *Salmonella* genoserotyping array (SGSA), a molecular-based serotyping assay (2), as well as a web-based platform for *in silico* serotyping of draft genome assemblies of *Salmonella* spp. (<https://lfz.corefacility.ca/sistr-app>; 3). The SGSA is able to provide the antigenic formula

TABLE 1 *Salmonella* strains sequenced in this study

Serovar	Antigenic formula	Isolate no.	Sequencing method(s) ^a	Accession no.
Blegdam	9,12:g,m,q: -	SA20065575	1	LHSP00000000
Enteritidis	1,9,12:g,m,p: -	SA20093032	1, 2	LHSQ00000000
Enteritidis	1,9,12:f,g,m,t: -	SA20103550	1, 2	LHSR00000000
Fresno	9,46:z ₃₈ : -	ST224	1, 2	LHSS00000000
Gallinarum	1,9,12: -: -	ST572	1, 2	LHST00000000
Gaminara	16:d: 1,7	SA20063285	1, 2	LHSU00000000
Hadar	6,8:z ₁₀ :e,n,x	SA20026260	1, 2	LHSV00000000
Hillingdon	9,46:g,m: -	S01-0588	1	LHSX00000000
Hvittingfoss	16:b:e,n,x	SA20014981	1, 2	LHSW00000000
Itami	9,12: 1,z ₁₃ : 1,5	SA20014991	1, 3	LHSY00000000
Johannesburg	40:b:e,n,x	SA20025782	1	LHSZ00000000
Kentucky	8,20:i:z ₆	SA20030505	1, 2	LHTA00000000
Manhattan	6,8:d: 1,5	SA20034532	1, 2	LHTB00000000
Milwaukee	43:f,g,[t]: -	SA19950795	1, 2	LHTC00000000
Moscow	9,12:g,q: -	SA20061414	1	LIXO00000000
Newport	6,8,20:e,h: 1,2	L0167	1	LIXP00000000
Panama	1,9,12:l:v: 1,5	SA20030878	1, 3	LHTD00000000
Paratyphi_A	2,12:a: -	SA19950809	1, 2	LHTE00000000
Pullorum	1,9,12: -: -	S4037-07	1, 2	LHTF00000000
Rubislaw	11:r:e,n,x	SA20030553	1, 2	LHTG00000000
subsp. II 56:z ₁₀ :e,n,x	56:z ₁₀ :e,n,x	1369-1373	1, 2	LHTH00000000
subsp. IIIa 62:z ₃₆ : -	62:z ₃₆ : -	5335/86	1, 2	LHTK00000000
subsp. IIIa 63:g,z ₅₁ : -	63:g,z ₅₁ : -	So 20/20	1, 2	LHTL00000000
Virginia	8:d: 1,2	SA19971529	1, 2	LHTI00000000
Weslaco	42:z ₃₆ : -	247K	1, 3	LHTJ00000000

^a 1 = Roche 454 GS-FLX titanium; 2 = Illumina MiSeq; 3 = Illumina HiSeq.

and serovar, preserving the nomenclature of legacy data gathered using traditional serotyping methods.

When this study began, there were numerous *Salmonella* serogroups and serovars not represented in GenBank, and isolates were selected to create a more complete public data set and expand the identification capacity of the SGSA. While this manuscript was being readied for publication, several of these serogroups and/or serovars became publicly available; however these will contribute to the accumulation of multiple representative genomes for the development of molecular typing tools. As of 10 July 2015, this is the first assembly and publication of whole-genome sequences for three *Salmonella* serogroups, O:43(U), O:56, and O:63, and 10 serovars: Blegdam, Fresno, Hillingdon, Itami, Milwaukee, Moscow, Virginia, Weslaco, subsp. *salamae* (II) 56:z₁₀:e,n,x, and subsp. *arizonae* (IIIa) 63:g,z₅₁:-.

The genomic DNA required for sequencing was extracted from *Salmonella* isolates grown overnight at 37°C on Luria-Bertani agar (BD Canada, Mississauga, ON, Canada) using an EZ1 DNA tissue kit (Qiagen Ltd., Mississauga, ON, Canada). Three technologies were used for sequencing: Roche's 454 GS-FLX Titanium, obtaining an average coverage of 40-fold; Illumina HiSeq 2000 with TruSeq sample preparation of 2 × 100 paired-end runs, obtaining an average coverage of 90-fold; and Illumina MiSeq with TruSeq sample preparation of 2 × 251 paired-end runs, obtaining an average coverage of 90-fold. Of the 25 draft genomes, 20 isolates were sequenced using both Illumina and 454 technologies, while five were sequenced by 454 only. Mira assembler version 4.0 (4) was used to assemble the reads into contigs, which were then proofread and corrected in the program Gap5 of the Staden software package (5). The genomes were annotated using the NCBI Prokaryotic Genome Annotation Pipeline (http://www.ncbi.nlm.nih.gov/genome/annotation_prok).

These draft genomes can provide useful information for the development and validation of molecular diagnostic tools, as well as other research activities.

Nucleotide sequence accession numbers. The draft genome sequences for these 25 *Salmonella* isolates have been deposited in DDBJ/ENA/GenBank under Bioproject no. PRJNA294295. The GenBank accession numbers are listed in Table 1. The raw sequence data are available in the Sequence Read Archive (SRA).

ACKNOWLEDGMENTS

This work was funded in part by the Genome Research and Development Initiative through the Office of Biotechnology and Science, Health Canada.

We sincerely thank the following for providing isolates: Roger Johnson, Linda Cole, Betty Wilkie, Ann Perets, Ketna Mistry, OIE Reference Laboratory for Salmonellosis, National Microbiology Laboratory, Public Health Agency of Canada, Guelph, ON, Canada; Vanessa Allen, Anne Maki, Anayn Peralta, Enteric Section, Public Health Ontario Laboratory, Toronto, ON, Canada; Danielle Daignault, National Microbiology Laboratory, Public Health Agency of Canada, St. Hyacinthe, QC, Canada (ASHQ); Martin Cormican, School of Medicine, National University of Ireland Galway, Galway, Ireland; Niall Delappe, National *Salmonella*, *Shigella* and *Listeria* Reference Laboratory, Galway University Hospital, Galway, Ireland; Muna Anjum, Sarah North, Victoria Barrett, Gail Wise, Sam Chappell, Animal and Plant Health Agency, Weybridge, New Haw, Addlestone, Surrey, United Kingdom; Durda Slavic, Animal Health Laboratory, University of Guelph, Guelph, ON, Canada; John Devenish, Canadian Food Inspection Agency, Ottawa, ON, Canada; S.M. Laboratory Services Inc., Longueuil, QC, Canada; Helen Tabor, National Microbiology Laboratory, Public Health Agency of Canada, Winnipeg, MB, Canada; Sean Byrne, Animal Health Centre, Abbotsford, BC, Canada; Francois-Xavier Weill, Institut Pasteur, Paris, France.

We also thank the NCBI PGAP team for annotation services and the National Microbiology Laboratory, Public Health Agency of Canada, Winnipeg, MB, Canada, and McGill University and G enome Qu ebec Innovation Centre, Montr eal, QC, Canada, for genome sequencing.

FUNDING INFORMATION

This work, including the efforts of Catherine Yoshida, Andrew M. Kropinski, and John H. E. Nash, was funded by Government of Canada (Gouvernement du Canada).

REFERENCES

1. Majowicz SE, Musto J, Scallan E, Angulo FJ, Kirk M, O'Brien SJ, Jones TF, Fazil A, Hoekstra RM. 2010. The global burden of nontyphoidal *Salmonella* gastroenteritis. *Food Safety* 50:882–889. <http://dx.doi.org/10.1086/650733>.
2. Franklin K, Lingohr EJ, Yoshida C, Anjum M, Bodrossy L, Clark CG, Kropinski AM, Karmali MA. 2011. Rapid genoserotyping tool for classification of *Salmonella* serovars. *J Clin Microbiol* 49:2954–2965. <http://dx.doi.org/10.1128/JCM.02347-10>.
3. Yoshida CE, Kruczkiewicz P, Laing CR, Lingohr EJ, Gannon VP, Nash JH, Taboada EN. 2016. The *Salmonella In Silico* Typing Resource (SISTR): An Open Web-Accessible Tool for Rapidly Typing and Subtyping Draft *Salmonella* Genome Assemblies. *PLoS One* 11(1):e0147101. <http://dx.doi.org/10.1371/journal.pone.0147101>.
4. Chevreur B, Wetter T, Suhai S. 1999. Genome sequence assembly using trace signals and additional sequence information, p 45–56. *In* Computer science and biology. Proceedings of the German Conference on Bioinformatics, GCB '99. GCB, Hannover, Germany.
5. Bonfield JK, Whitwham A. 2010. Gap5—editing the billion fragment sequence assembly. *Bioinformatics* 26:1699–1703. <http://dx.doi.org/10.1093/bioinformatics/btq268>.