AMERICAN SOCIETY FOR MICROBIOLOGY   genomeAnnouncements™

Complete Genome and Methylome Sequences of *Salmonella enterica* subsp. *enterica* Serovars Typhimurium, Saintpaul, and Stanleyville from the SARA/SARB Collection

Kuan Yao,[a,b] Richard J. Roberts,[c] Marc W. Allard,[a] Maria Hoffmann[a]

Division of Microbiology, Office of Regulatory Science, Center for Food Safety and Nutrition, U.S. Food and Drug Administration, College Park, Maryland, USA[a]; School of Systems Biology, George Mason University, Manassas, Virginia, USA[b]; New England Biolabs, Inc., Ipswich, Massachusetts, USA[c]

ABSTRACT In this announcement, we report the complete genome and methylome sequences of three *Salmonella enterica* strains from the SARA and SARB collection: *S. enterica* subsp. *enterica* serovar Typhimurium (SARA13), *S. enterica* subsp. *enterica* serovar Saintpaul (SARA26), and *S. enterica* subsp. *enterica* serovar Stanleyville (SARB61).

The three *Salmonella* strains included in this study—*S. enterica* subsp. *enterica* serovars Typhimurium, Saintpaul, and Stanleyville—belong to the Salmonella Reference Collection SARA/SARB (1, 2). The reference collection was established based on the genetic structure of *S. enterica* characterized by multilocus enzyme electrophoresis (MEE) (3). *S.* Typhimurium is the most common cause of food poisoning in the United States, and outbreaks have been linked to poultry, beef products, and rodents. The prevalence of this serotype has increased from 9% to 33% since 1990 (4). *S.* Saintpaul is closely related to *S.* Typhimurium (5), with infections having resulted from consumption of several fresh produce commodities such as cucumbers, jalapeño peppers, and tomatoes (6). *S.* Stanleyville (SARB61) belongs to the SARB collection, which encompasses a more genetically diverse group of *S. enterica* that are commonly found in the environment, animals, and humans (3). Multiple cases of *S.* Stanleyville were reported in Cameroon, Mali, and Uganda (7, 8). Between 2003 and 2013, there were 65 cases of multistate *S.* Stanleyville infections reported in the United States (9).

S. Typhimurium, Saintpaul, and Stanleyville isolates were sequenced based on previously published procedures (10, 11). The continuous long-read data were *de novo* assembled using the PacBio Hierarchical Genome Assembly Process (HGAP) version 3.0 (12). The assembled sequences were annotated using the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) and deposited at DDBJ/EMBL/GenBank.

The complete *S.* Saintpaul (SARA26) genome was sequenced with 125× coverage. The complete genome size was 4,686,793 bp with a G/C content of 52.02%. *S.* Saintpaul consisted of 4,491 genes. The PHAge Search Tool (PHAST) analysis for prophage sequence detection did not identify any intact phages (13). The *S.* Typhimurium genome was sequenced with 171× coverage. The closed genome for the chromosome was 4,819,807 bp and 93,826 bp for the plasmid. The genome consisted of 4,770 genes with a G+C content of 52.21% for the chromosome and 53.1% for the plasmid. Using PHAST analysis, prophages Gifsy-1, Gifsy-2, and Salmon-ST64B were identified. The *S.* Stanleyville genome was sequenced with 144× coverage. The genome consisted of 4,888,463 bp for the chromosome and three plasmids with sizes of 106,449 bp, 58,302 bp, and 49,762 bp. The complete genome contained 4,991 genes. The G+C

TABLE 1 Summary of active methylases and their recognition sequences

Strain	Assignment	Methyltransferase specificity	Methylation type	Restriction modification type
S. Saintpaul	M.SenSARA26Dam	GATC	m6A	Orphan alpha
	M.SenSARA26I	CAGAG	m6A	III beta
	M.SenSARA26II	ATGCAT	m6A	II beta
	SenSARA26III	ACRCAG	m6A	II G,S, alpha
S. Typhimurium	M.Sen13Dcm	CCWGG	m5C	Orphan
	M.Sen13I	GAGNNNNNNRTAYG	m6A	I gamma
	M.Sen13II	CAGAG	m6A	III beta
	M.Sen13IV	ATGCAT	m6A	II beta
	Sen13III	GATCAG	m6A	II G,S, alpha
S. Stanleyville	M.Sen624I	CAGAG	m6A	III beta
	M.Sen624II	GAGNNNNNNRTAYG	m6A	I gamma
	M.Sen624III	ATGCAT	m6A	II beta

content for the chromosome was 52.13%; the G+C content ranging from the largest to the smallest plasmid was 51.0%, 52.1%, and 52.0%, respectively. PHAGE analysis indicated the presence of prophage Salmon-SPN1S.

The DNA methyltransferase activities were analyzed based on the kinetic variations of the nucleotide incorporation rate of the PacBio RSII sequencing platform (14). The single-molecule real-time data of the methylomes are summarized in Table 1. They are also deposited in REBASE (15) as follows: *S.* Saintpaul, http://rebase.neb.com/cgi-bin/pacbioget?20626; *S.* Typhimurium, http://rebase.neb.com/cgi-bin/pacbioget?20625; and *S.* Stanleyville, http://rebase.neb.com/cgi-bin/pacbioget?20799. While most of the motifs have been found in other *Salmonella* strains, the motif ACRCAG found as the recognition sequence of a type IIG restriction/modification enzyme is new and unique.

Accession number(s). Sequences have been deposited in GenBank under the following accession numbers: *S.* Saintpaul, CP017727; *S.* Typhimurium, CP017728 and CP017729; and *S.* Stanleyville, CP017723, CP017724, CP017725, and CP017726.

REFERENCES

1. Beltran P, Plock SA, Smith NH, Whittam TS, Old DC, Selander RK. 1991. Reference collection of strains of the *Salmonella typhimurium* complex from natural populations. J Gen Microbiol 137:601–606. https://doi.org/10.1099/00221287-137-3-601.
2. Boyd EF, Wang FS, Beltran P, Plock SA, Nelson K, Selander RK. 1993. Salmonella reference collection B (SARB): strains of 37 serovars of subspecies I. J Gen Microbiol 139:1125–1132. https://doi.org/10.1099/00221287-139-6-1125.
3. Achtman M, Hale J, Murphy RA, Boyd EF, Porwollik S. 2013. Population structures in the SARA and SARB reference collections of *Salmonella enterica* according to MLST, MLEE and microarray hybridization. Infect Genet Evol 16:314–325. https://doi.org/10.1016/j.meegid.2013.03.003.
4. Switt AI, Soyer Y, Warnick LD, Wiedmann M. 2009. Emergence, distribution, and molecular and phenotypic characteristics of *Salmonella enterica* serotype 4,5,12:i. Foodborne Pathog Dis 6:407–415. https://doi.org/10.1089/fpd.2008.0213.
5. Timme RE, Pettengill JB, Allard MW, Strain E, Barrangou R, Wehnes C, Van Kessel JS, Karns JS, Musser SM, Brown EW. 2013. Phylogenetic diversity of the enteric pathogen *Salmonella enterica* subsp. *enterica* inferred from genome-wide reference-free SNP characters. Genome Biol Evol 5:2109–2123. https://doi.org/10.1093/gbe/evt159.
6. Centers for Disease Control and Prevention (CDC). 28 August 2008. Multistate outbreak of *Salmonella* Saintpaul infections linked to raw produce. https://www.cdc.gov/salmonella/2008/raw-produce-8-28-2008.html.
7. Ngandjio A, Tchendjou P, Koki Ndombo P, Gonsu Kamga H, Fonkoua MC. 2012. Emergence of multi-drug resistant *Salmonella enterica* serotype Stanleyville infections among children in Yaoundé, Cameroon. J Trop Pediatr 58:161–163. https://doi.org/10.1093/tropej/fmr054.
8. Afema JA, Byarugaba DK, Shah DH, Atukwase E, Nambi M, Sischo WM. 2016. Potential sources and transmission of Salmonella and antimicrobial resistance in Kampala, Uganda. PLoS One 11:e0152130. https://doi.org/10.1371/journal.pone.0152130.
9. Centers for Disease Control and Prevention (CDC). 2013. National enteric disease surveillance: *Salmonella* annual report, 2011. https://www.cdc.gov/ncezid/dfwed/PDFs/salmonella-annual-report-2011-508c.pdf.
10. Yao K, Muruvanda T, Roberts RJ, Payne J, Allard MW, Hoffmann M. 2016. Complete genome and Methylome sequences of *Salmonella enterica* subsp. *enterica* serovar Panama (ATCC 7378) and *Salmonella enterica* subsp. *enterica* serovar Sloterdijk (ATCC 15791). Genome Announc 4(2): e00133-16. https://doi.org/10.1128/genomeA.00133-16.
11. Hoffmann M, Payne J, Roberts RJ, Allard MW, Brown EW, Pettengill JB. 2015. Complete genome sequence of *Salmonella enterica* subsp. *enterica* serovar Agona 460004 2-1, associated with a multistate outbreak in the United States. Genome Announc 3(4):e00690-15. https://doi.org/10.1128/genomeA.00690-15.

12. Chin CS, 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.

13. Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. 2011. PHAST: a fast phage search tool. Nucleic Acids Res 39:W347–W352. https://doi.org/10.1093/nar/gkr485.

14. Flusberg BA, Webster DR, Lee JH, Travers KJ, Olivares EC, Clark TA, Korlach J, Turner SW. 2010. Direct detection of DNA methylation during single-molecule, real-time sequencing. Nat Methods 7:461–465. https://doi.org/10.1038/nmeth.1459.

15. Roberts RJ, Vincze T, Posfai J, Macelis D. 2015. REBASE—a database for DNA restriction and modification: enzymes, genes and genomes. Nucleic Acids Res 43:D298–D299. https://doi.org/10.1093/nar/gku1046.