Draft Genome Sequences of 23 *Salmonella enterica* Strains Isolated from Cattle in Ibadan, Nigeria, Representing 21 *Salmonella* Serovars

Maria Sanchez Leon,[a] Kayode Fashae,[b] George Kastanis,[a] Marc Allard[a]

Center for Food Safety and Applied Nutrition, Division of Microbiology, Office of Regulatory Science, U.S. Food and Drug Administration, College Park, Maryland, USA[a]; Department of Microbiology, University of Ibadan, Ibadan, Nigeria[b]

ABSTRACT   To provide a better understanding of the diversity of *Salmonella enterica*, we report the assembled genome sequences of 23 *Salmonella enterica* strains isolated from fecal samples of cattle in Nigeria comprising 21 different *Salmonella* serovars.

The taxonomic classification of *Salmonella enterica* includes more than 2,600 serovars, which subsequently can be divided into typhoidal and nontyphoidal *Salmonella* (NTS) serovars (1). Typhoidal serovars (*S. enterica* subsp. *enterica* serovars Typhi, Sendai, Paratyphi A, Paratyphi B, and Paratyphi C) primarily infect humans, while NTS serovars have a broad spectrum of hosts, such as mammals (including humans), birds, and reptiles (2). Globalization has increased the availability of diverse strains of *S. enterica* across the world. Consequently, different serovars have crossed borders and emerging antimicrobial resistance (AMR) phenotypes have been described (3), thus presenting the need to study the genotypic diversity of *S. enterica* serovars. We announce here 23 draft genomes from *S. enterica* strains isolated from fecal samples of cattle in Nigeria, including genomes from serovars that are not currently available at GenBank, e.g., *S. enterica* subsp. *enterica* serovars Agoueve, Marseille, Sundsvall, Takoradi, Ekotedo, Tees, Plymouth, Hato, 47:z4,z23:-, Altendorf, Essen, and Glostrup.

Bacterial DNA was extracted from overnight cultures by using the DNeasy blood and tissue kit (Qiagen, Valencia, CA, USA) according to the manufacturer's instructions. Sequencing libraries were constructed with 0.2 ng/μl of prepared DNA using the Nextera XT DNA library prep kit (Illumina, San Diego, CA, USA).

Sequencing was performed on the MiSeq Illumina instrument with the 500-cycle MiSeq reagent V2 kit (2 $\times$ 250 bp) following the manufacturer's guidelines. Raw data were subjected to genome assembly using SPAdes version 3.8 (4), and genome annotation was performed with the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) (5). *Salmonella* serotyping by whole-genome sequencing was predicted with the SeqSero version 1.0 software tool (6). The NCBI pathogen detection website (https://www.ncbi.nlm.nih.gov/pathogens) and annotation pipeline (5) were queried to extract information regarding AMR genes present in these genomes. *In silico* multilocus sequence typing (MLST) data were extracted from the Enterobase online tool (http://enterobase.warwick.ac.uk). *Salmonella* pathogenicity islands (SPIs) were identified using the SPIFinder version 1.0 online tool (https://cge.cbs.dtu.dk/services/SPIFinder).

The total lengths of these genomes range from 4.6 to 5.2 Mb, with an average GC content of 52%, and the numbers of contigs range from 25 to 102. The numbers of predicted genes, noncoding RNAs (ncRNAs), tRNAs, and clustered regularly interspaced short palindromic repeat (CRISPR) arrays range from 4,353 to 5,075, 10 to 18, 71 to 82, and 1 to 4, respectively, as previously described for *Salmonella* genomes (Table 1). *In*

TABLE 1 Sequencing metadata and MLST results for 23 *Salmonella enterica* strains, isolated from cattle in Nigeria

GenBank accession no.	No. of coding genes	No. of ncRNAs	No. of tRNAs	No. of CRISPR arrays	No. of contigs	Total length (bp)	GC content %	Predicted subspecies	Predicted serotype	MLST	Predicted AMR genes[b]
NPMJ00000000	4,485	13	77	2	48	4,616,478	52.16	*enterica*	Johannesburg	ST515	−
NPMI00000000	4,565	15	75	3	41	4,703,003	52.2	*enterica*	Stanleyville	ST2562	+
NPMH00000000	4,552	15	75	3	34	4,697,783	52.2	*enterica*	Stanleyville	ST2562	+
NPMG00000000	4,353	13	76		25	4,535,642	52.2	*enterica*	Eastbourne	ST414	−
NPMF00000000	4,538	12	74	2	54	4,743,981	52.24	*enterica*	Agoueve	ST3963[a]	−
NPME00000000	4,452	10	75	1	25	4,616,387	52.21	*enterica*	Give	ST3969[a]	−
NPMD00000000	4,761	18	79	2	43	4,915,736	52.2	*enterica*	Hadar	ST473	−
NPMC00000000	4,517	13	75		46	4,696,577	52.15	*enterica*	Marseille	ST3966[a]	−
NPMB00000000	4,390	11	78	2	31	4,562,440	52.2	*enterica*	Sundsvall	ST488	−
NPMA00000000	4,457	10	75	3	26	4,667,942	52.12	*enterica*	Takoradi	ST531	−
NPLZ00000000	4,443	14	75	2	42	4,637,247	52.18	*enterica*	Ekotedo	ST3968[a]	−
NPLY00000000	4,492	12	77	2	42	4,689,632	52.14	*enterica*	Tees	ST3955[a]	−
NPLX00000000	4,510	13	75	2	46	4,658,074	52.22	*enterica*	Bergen	ST1279	+
NPLW00000000	4,799	12	72	2	53	4,937,236	51.96	*enterica*	Plymouth	ST565	−
NPLV00000000	4,488	11	74	2	31	4,694,335	52.19	*enterica*	Hato	ST3997[a]	−
NPLU00000000	4,508	11	74	2	36	4,729,809	52.16	*arizonae*	47:z4,z23:-	ST3967[a]	−
NPLT00000000	5,075	11	82	4	102	5,266,492	51.79	*enterica*	Altendorf	ST3974[a]	+
NPLS00000000	4,730	14	71	1	33	4,876,578	52.09	*enterica*	Dublin	ST10	−
NPOH00000000	4,687	12	74	2	48	4,857,624	52.19	*enterica*	Essen	ST3964[a]	+
NPLR00000000	4,483	13	74	2	39	4,677,388	52.15	*enterica*	Ealing	ST2013	−
NPLQ00000000	4,570	18	77	2	33	4,739,515	52.09	*enterica*	Glostrup	ST3961[a]	−
NPLP00000000	4,442	11	73	2	30	4,637,671	52.18	*enterica*	Ekotedo	ST3972[a]	−
NPLO00000000	4,554	17	75	2	26	4,762,386	52.04	*enterica*	Corvallis	ST3962[a]	+

[a]MLST analysis indicates these genomes belong to new STs.
[b]−, absence of AMR genes; +, occurrence of AMR genes.

silico MLST analysis shows that each genome belongs to different sequencing types (STs), and 12 of them have not been previously reported. AMR genes were predicted in 6 out of 23 strains, including those encoding chloramphenicol, quinolone, tetracycline, and fosfomycin resistance. Up to seven SPIs were detected in these genomes, including SPI-1, SPI-2, and SPI-4, which encode a predicted type III and type I secretion system.

Analysis of these genomes will offer a better understanding of the genomic events responsible for AMR, disease transmission, and pathogenicity of *Salmonella enterica*.

Accession number(s). The draft genome sequences reported here have been deposited in DDBJ/EMBL/GenBank under BioProject PRJNA186035. GenBank accession numbers are listed in Table 1.

REFERENCES

1. Gal-Mor O, Boyle EC, Grassl GA. 2014. Same species, different diseases: how and why typhoidal and non-typhoidal *Salmonella enterica* serovars differ. Front Microbiol 5:391. https://doi.org/10.3389/fmicb.2014.00391.
2. Matthews TD, Schmieder R, Silva GG, Busch J, Cassman N, Dutilh BE, Green D, Matlock B, Heffernan B, Olsen GJ, Farris Hanna L, Schifferli DM, Maloy S, Dinsdale EA, Edwards RA. 2015. Genomic comparison of the closely related *Salmonella enterica* serovars Enteritidis, Dublin and Gallinarum. PLoS One 10:e0126883. https://doi.org/10.1371/journal.pone.0126883.
3. MacPherson DW, Gushulak BD, Baine WB, Bala S, Gubbins PO, Holtom P, Segarra-Newnham M. 2009. Population mobility, globalization, and antimicrobial drug resistance. Emerg Infect Dis 15:1727–1732.
4. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski 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.
5. 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.
6. Zhang S, Yin Y, Jones MB, Zhang Z, Deatherage Kaiser BL, Dinsmore BA, Fitzgerald C, Fields PI, Deng X. 2015. *Salmonella* serotype determination utilizing high-throughput genome sequencing data. J Clin Microbiol 53: 1685–1692. https://doi.org/10.1128/JCM.00323-15.