Title: Complete genome sequences of African *Salmonella* Enteritidis clinical isolates associated with bloodstream infection

Blanca M. Perez-Sepulvedaa, Alexander V. Predeusa, Wai Yee Fonga, Christopher M. Parrya, John Cheesbroughb\*, Paul Wigleya, Nicholas A. Feaseyc,d, Jay C. D. Hintona#

aInstitute of Infection, Veterinary & Ecological Sciences, University of Liverpool, Liverpool, UK

b Institute of Infection and Global Health, University of Liverpool, Liverpool, UK

cLiverpool School of Tropical Medicine, Pembroke Place, Liverpool, UK

dMalawi-Liverpool-Wellcome Programme, Blantyre, Malawi

# Running title: African *S.* Enteritidis Genome announcement

#Corresponding author

E-mail: [jay.hinton@liverpool.ac.uk](mailto:jay.hinton@liverpool.ac.uk) Tel. (+44) 1517 954573

\*Retired

# Abstract

We report the complete genome sequencing and annotation of four *Salmonella* Enteritidis isolates; two that are representative of the Central/Eastern African clade (CP255 and D7795) and two of the Global Epidemic clade (A1636 and P125109).

*Salmonella enterica* serovar Enteritidis (*S.* Enteritidis) typically causes gastroenteritis and is responsible for a global epidemic linked to poultry and egg production. Over recent decades, *S.* Enteritidis has become a leading cause of invasive non-typhoidal *Salmonella* (iNTS) disease in sub-Saharan Africa (1, 2), and novel clades of this serovar have been isolated from individuals with bloodstream infection (3). In contrast to the Global Epidemic clade, these novel African clades are associated with high mortality in immunocompromised individuals and are typically multidrug-resistant (MDR), representing an important public health challenge (4).

We used long-read sequencing to investigate the genome of two representative *S.* Enteritidis strains of the Central/Eastern African clade (CP255 and D7795) and two of the Global Epidemic clade (P125109 and A1636). CP255 was isolated in the Democratic Republic of Congo (the Zaire) in 1991, from the blood of a child at the Institut-Médical-Evangélique, Kimpese, and is phenotypically MDR (amoxicillin, tetracycline, chloramphenicol and streptomycin-resistant) (5, 6). D7795 (paediatric patient/MDR) and A1636 (adult patient/fully susceptible) were isolated in 1998 and 2000, respectively in Blantyre (Malawi) (3). *S.* Enteritidis P125109, a UK PT4 isolate from 1988, was used as a reference (7–10).

A single colony of each isolate was grown for 16 h in 5 mL of Lennox medium at 37˚C. Total DNA was extracted using the Quick-DNA™ Universal Kit (Zymo, D4069). DNA integrity was verified by 0.5% agarose gel electrophoresis at 90 V for 1.5 h. DNA purity/concentration were measured with a DeNovix DS-11FX spectrophotometer/fluorometer and Qubit dsDNA HS assay Kit. Long-read sequencing was performed by the Centre for Genomic Research (University of Liverpool, UK) in a PacBio SMRT (P6/C4 chemistry) using SMRTbell Template 1.0 (Pacific Biosciences, 100-259-100) library preparation with g-Tube (Covaris) fragmentation and size-selection of 15-50 kb with 0.75% agarose cassette (BluePippin, BMF7510). Illumina HiSeq sequencing was performed as part of the 10KSG project (11) and by MicrobesNG (UK) using Nextera XT Library Prep Kit (Illumina, USA) with modifications (2 ng DNA and 1 min PCR elongation) and 250 bp paired‑end protocol. Reads were adapter-trimmed using Trimmomatic v0.30, with a sliding window quality cut-off of Q15 (12).

Using Filtlong v0.2.0 (<https://github.com/rrwick/Filtlong>) and Illumina reads as a reference, we selected a subset of raw PacBio reads with best quality and length to yield an approximate 100x coverage for each genome. Selected long- and short-reads were assembled using Unicycler v0.4.4 in hybrid mode (13). Genomes were initially annotated using Prokka v1.13.7 (14) through Bacpipe v0.6 (<https://github.com/apredeus/multi-bacpipe>), and automatically re-annotated by GenBank with PGAP (15), and rotated to the origin at the *thrLABC* operon. Variant calling was done with Snippy v4.3.6 (<https://github.com/tseemann/snippy>) in contig mode.

Genome comparison revealed genomic degradation and differences in accessory genome (Table 1). P125109 and A1636 carried the virulence plasmid pSENV, whereas D7795 carried pSEN-BT (3) and CP255 pSEN-DRC, which both had a pSENV backbone with MDR-encoding genes. Other plasmids were identified in A1636, D7795, and CP255. The prophage repertoire of A1636 resembled P125109 (7). Both D7795 and CP255 lacked ΦSE20, and instead carried P88-like and Fels2-like prophages.

# Data Availability

The annotated complete genome assemblies of *S.* Enteritidis CP255, D7795, A1636 and P125109 have been deposited in NCBI. The BioProject accession number is [PRJNA671837](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA671837/), and the individual accession numbers are [SAMN16552338](https://www.ncbi.nlm.nih.gov/assembly/GCA_015241115.1) (A1636), [SAMN16552337](https://www.ncbi.nlm.nih.gov/assembly/GCA_015240995.1) (CP255), [SAMN16552336](https://www.ncbi.nlm.nih.gov/assembly/GCA_015240855.1) (D7795) and [SAMN16552335](https://www.ncbi.nlm.nih.gov/assembly/GCA_015240635.1) (P125109). The Bacpipe annotation pipeline can be accessed at <https://github.com/apredeus/multi-bacpipe>.

# Acknowledgements

We are grateful to present and former members of the Hinton laboratory for helpful discussions, particularly Rocío Canals and Siân Owen for their expert advice. We are indebted to Paul Barrow for kindly sharing the *S.* Enteritidis P125109 strain. This project was supported by the Wellcome Trust Senior Investigator Award (106914/Z/15/Z) to Jay C. D. Hinton. The MLW bacteraemia service is funded by Wellcome Africa and Asia Programme Grant 206545/Z/17/Z.

# References

1. Reddy EA, Shaw A V., Crump JA. 2010. Community-acquired bloodstream infections in Africa: a systematic review and meta-analysis. Lancet Infect Dis 10:417–432.

2. Feasey NA, Masesa C, Jassi C, Faragher EB, Mallewa J, Mallewa M, Maclennan CA, Msefula C, Heyderman RS, Gordon MA. 2015. Three epidemics of invasive multidrug-resistant *Salmonella* bloodstream infection in Blantyre, Malawi, 1998-2014. Clin Infect Dis 61:S363–S371.

3. Feasey NA, Hadfield J, Keddy KH, Dallman TJ, Jacobs J, Deng X, Wigley P, Barquist Barquist L, Langridge GC, Feltwell T, Harris SR, Mather AE, Fookes M, Aslett M, Msefula C, Kariuki S, Maclennan CA, Onsare RS, Weill F-X, Le Hello S, Smith AM, Mcclelland M, Desai P, Parry CM, Cheesbrough JS, French N, Campos J, Chabalgoity JA, Betancor L, Hopkins KL, Nair S, Humphrey TJ, Lunguya O, Cogan TA, Tapia MD, Sow SO, Tennant SM, Bornstein K, Levine MM, Lacharme-Lora L, Everett DB, Kingsley RA, Parkhill J, Heyderman RS, Dougan G, Gordon MA, Thomson NR, Barquist L, Gemma C, Hello S Le, Melita A. 2016. Distinct *Salmonella* Enteritidis lineages associated with enterocolitis in high-income settings and invasive disease in low-income settings. Nat Genet 48:1211–7.

4. Uche IV, MacLennan CA, Saul A. 2017. A Systematic Review of the Incidence, Risk Factors and Case Fatality Rates of Invasive Nontyphoidal *Salmonella* (iNTS) Disease in Africa (1966 to 2014). PLoS Negl Trop Dis 11:e0005118.

5. Green SDR, Cheesbrough JS. 1993. *Salmonella* bacteraemia among young children at a rural hospital in western Zaire. Ann Trop Paediatr 13:45–53.

6. Cheesbrough JS, Taxman BC, Green SDR, Mewa FI, Numbi A. 1997. Clinical definition for invasive *Salmonella* infection in African children. Pediatr Infect Dis J 16:277–283.

7. Thomson NR, Clayton DJ, Windhorst D, Vernikos G, Davidson S, Churcher C, Quail MA, Stevens M, Jones MA, Watson M, Barron A, Layton A, Pickard D, Kingsley RA, Bignell A, Clark L, Harris B, Ormond D, Abdellah Z, Brooks K, Cherevach I, Chillingworth T, Woodward J, Norberczak H, Lord A, Arrowsmith C, Jagels K, Moule S, Mungall K, Sanders M, Whitehead S, Chabalgoity JA, Maskell D, Humphrey T, Roberts M, Barrow PA, Dougan G, Parkhill J. 2008. Comparative genome analysis of *Salmonella* Enteritidis PT4 and *Salmonella* Gallinarum 287/91 provides insights into evolutionary and host adaptation pathways. Genome Res 18:1624–1637.

8. Silva CA, Blondel CJ, Quezada CP, Porwollik S, Andrews-Polymenis HL, Toro CS, Zaldívar M, Contreras I, Mcclelland M, Santiviago CA. 2012. Infection of mice by *Salmonella enterica* serovar Enteritidis involves additional genes that are absent in the genome of serovar Typhimurium. Infect Immun 80:839–849.

9. Barrow PA. 1991. Experimental infection of chickens with *Salmonella* enteritidis. Avian Pathol 20:145–153.

10. Barrow PA, Lovell MA. 1991. Experimental infection of egg‐laying hens with *Salmonella* enteritidis phage type 4. Avian Pathol 20:335–348.

11. Perez-Sepulveda B, Heavens D, Pulford C, Predeus A, Low R, Webster H, Schudoma C, Rowe W, Lipscombe J, Watkins C, Kumwenda B, Shearer N, Costigan K, Baker K, Feasey NA, Hinton J, Hall N. 2020. An accessible, efficient and global approach for the large-scale sequencing of bacterial genomes. bioRxiv 2020.07.22.200840.

12. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120.

13. Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol 13:e1005595.

14. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068–2069.

15. 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.

16. Kröger C, Dillon SCDCDC, Cameron ADS, Papenfort K, Sivasankaran SK, Hokamp K, Chao Y, Sittka A, Hebrard M, Handler K, Colgan AM, Leekitcharoenphon P, Langridge GC, Lohan AJ, Loftus B, Lucchini S, Ussery DW, Dorman CJ, Thomson NR, Vogel J, Hinton JCD. 2012. The transcriptional landscape and small RNAs of *Salmonella enterica* serovar Typhimurium. Proc Natl Acad Sci 109:E1277–E1286.

17. Jørgensen TS, Xu Z, Hansen MA, Sørensen SJ, Hansen LH. 2014. Hundreds of Circular Novel Plasmids and DNA Elements Identified in a Rat Cecum Metamobilome. PLoS One 9:e87924.

**Table 1: Characteristics and accession numbers of genomes of four *S.* Enteritidis isolates**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Isolate | Year | Assembly accession no. | Raw reads accession no.a | No. reads  Illumina (2x250 bp) | Illumina coverage | No. reads  PacBio | N50 PacBio reads | G+C content | Genome size (bp) | Total no. of genes | Virulence plasmidb | Other plasmidsc | SNPc P125109 | SNPc D7795 |
| **CP255** | 1991 | [GCA\_015240995](https://www.ncbi.nlm.nih.gov/assembly/GCA_015240995.1) | [SRR12953596](https://trace.ncbi.nlm.nih.gov/Traces/sra/?run=SRR12953596) [SRR12953597](https://trace.ncbi.nlm.nih.gov/Traces/sra/?run=SRR12953597) | 296,730 | 30.6 | 40,000 | 20,451 | 52.30% | 4,840,946 | 4,729 | pSEN-DRC (96 kb) | pRSF1010 (8.7 kb)(16) | 972 | 61 |
| **D7795** | 2000 | [GCA\_015240855](https://www.ncbi.nlm.nih.gov/assembly/GCA_015240855.1) | [SRR12953602](https://trace.ncbi.nlm.nih.gov/Traces/sra/?run=SRR12953602) [SRR12953603](https://trace.ncbi.nlm.nih.gov/Traces/sra/?run=SRR12953603) | 813,469 | 83.5 | 103,762 | 16,718 | 52.30% | 4,869,504 | 4,777 | pSEN-BT (116 kb)(3) | pRGI00316 (4.9 kb)(17) | 1,017 | 0 |
| **A1636** | 1998 | [GCA\_015241115](https://www.ncbi.nlm.nih.gov/assembly/GCA_015241115.1) | [SRR12953598](https://trace.ncbi.nlm.nih.gov/Traces/sra/?run=SRR12953598) [SRR12953599](https://trace.ncbi.nlm.nih.gov/Traces/sra/?run=SRR12953599) | 3,859,080 | 406.4 | 97,435 | 16,873 | 52.20% | 4,748,456 | 4,608 | pSENV (59 kb) | pSE-GC (3.2 kb) | 46 | 1,022 |
| **P125109** | 1988 | [GCA\_015240635](https://www.ncbi.nlm.nih.gov/assembly/GCA_015240635.1) | [SRR12953600](https://trace.ncbi.nlm.nih.gov/Traces/sra/?run=SRR12953600) [SRR12953601](https://trace.ncbi.nlm.nih.gov/Traces/sra/?run=SRR12953601) | 587,289 | 61.9 | 92,761 | 17,272 | 52.20% | 4,745,224 | 4,606 | pSENV (59 kb) |  | 0 | 1,022 |

aAccession numbers for raw short (Illumina) and long (PacBio) reads

bNumber in parenthesis indicates plasmid size in kb and reference

cSNP (single nucleotide polymorphisms) identified in whole genome-based sequence comparison against either *S.* Enteritidis P125109 or D7795