Genomic diversity and antimicrobial resistance among non-typhoidal *Salmonella* associated with human disease in The Gambia

Authors:

Saffiatou Darboe¹, Richard S. Bradbury², Jody Phelan³, Abdoulie Kanteh¹, Abdul-Khalie Muhammad¹, Archibald Worwui¹, Shangxin Yang⁴, Davis Nwakanma¹, Blanca Perez-Sepulveda⁵, Samuel Kariuki⁶, Brenda Kwambana-Adams^{1,7*} and Martin Antonio^{1*}

***These authors contributed equally to this article**

Author affiliations

1. Medical Research Council Unit, The Gambia at London School of Hygiene and Tropical Medicine
2. Federation University, Melbourne, Australia
3. London School of Hygiene and Tropical Medicine, London, United Kingdom
4. University of California, Los Angeles, California, USA
5. University of Liverpool, United Kingdom
6. Kenyan Medical Research Institute, Nairobi, Kenya
7. University College London, London, United Kingdom

Corresponding author:

Saffiatou Darboe, Laboratory Services, Medical Research Council Unit The Gambia at London School of Hygiene and Tropical Medicine, P.O Box 273 Banjul, The Gambia; email: Saffiatou.Darboe@lshtm.ac.uk

Article summary line:

This study revealed high serovar diversity in non-typhoidal *Salmonella* serovars in The Gambia with *Salmonella* Typhimurim ST19 and *Salmonella* Enteritidis ST11 the most prevalent serovars causing bacteraemia. Serovars responsible for gastroenteritidiss were much more diverse. Importantly, high genetic diversity was noted for *Salmonella* Enteritidis with the presence of the virulent multidrug resistant *S. Enteritidis* West African clade present.

Running title:

Genetic diversity and antibiogram of non-typhoidal *Salmonella* in The Gambia

Key words:

Non-typhoidal *Salmonella* (NTS), invasive NTS (iNTS), *Salmonella* Typhimurium, *Salmonella* Enteritidis, The Gambia.

Abstract:

Non-typhoidal *Salmonella* associated with multidrug resistance cause invasive disease in sub-Saharan African. Specific lineages of serovars *S. Typhimurium* and *S. Enteritidis* are implicated. We characterised the genomic diversity of 100 clinical Non-typhoidal *Salmonella* collected from 93 patients in 2001 from the eastern and 2006 to 2018 in the western regions of The Gambia respectively. Phenotypic susceptibility applied Kirby Baur disk diffusion and whole genome sequencing utilized Illumina platforms. The predominant serovars were *S. Typhimurium* ST19 (31/100) and *S. Enteritidis* ST11 (18/100) restricted to invasive disease with the notable absence of *S. Typhimurium* ST313. Phylogenetic analysis performed in the context of 495 African strains from the European Nucleotide Archive confirmed the presence of the *S. Enteritidis* virulent epidemic invasive multidrug resistant West African clade. Multidrug resistance including chloramphenicol and azithromycin has emerged among the West African *S. Enteritidis* clade 7/9 (78%) with potential for spread, thus having important implications for patient management warranting systematic surveillance and epidemiologic investigations to inform control.

Data summary:

Sequences are deposited in the NCBI sequence reads archive (SRA) under BioProject ID:PRJEB38968. The genomic assemblies are available for download from the European Nucleotide Archive (ENA) : <http://www.ebi.ac.uk/ena/data/view/>. Accession numbers SAMEA6991082 to SAME6991180

Introduction:

Non-typhoidal *Salmonella* (NTS) serovars are a common cause of foodborne gastroenteritis but can also cause severe disseminated infections dependent on the pathogen's virulence and the host's immune status [1,2]. Globally, there are over 2,800 serovars, some of which are adapted to non-human hosts [3]. The global annual estimate of *Salmonella* gastroenteritis is 93.8 million illnesses with 155,000 deaths [4]. The highest mortality occurs in Africa which accounts for 4,100 deaths annually with an incidence of 320/100,000 population [4]. Most cases of *Salmonella* gastroenteritis in immunocompetent hosts are self-limiting and do not require antimicrobial therapy; however, infections in infants, the elderly and immunocompromised patients do require antimicrobial treatments such as ciprofloxacin, 3rd generation cephalosporins or azithromycin [5].

The clinical characteristics of non-typhoidal salmonellosis emerging in Africa represent a changing disease pattern, from gastroenteritis to invasive disease with a case fatality ratio of 20-25% [1,6–8]. Although NTS in diarrhoea is less well characterised in Africa, it may be predisposition to invasive disease. Invasive NTS (iNTS) disease, mainly bacteraemia. iNTS is globally estimated at 3.4 million illnesses annually, disproportionately affecting those in sub-Saharan Africa (sSA), with over 50% associated with HIV infection, malnutrition, recent malaria and children between 6 months to 3 years of age [6,8,9]. iNTS disease has a markedly different presentation, closer to enteric fever in its clinical form than typical NTS disease [8]. The predominant serovars responsible for invasive disease in sSA are specialised lineages of *Salmonella enterica* serovars Enteritidis and Typhimurium that are distinct from those circulating in other parts of the world [1,10]. Whole genome sequencing (WGS) has provided new insights into the host adapted signatures associated with pathogenicity and metabolism of these *S. Typhimurium* lineages and *S. Enteritidis* clades characterised by genomic degradation and accessory genome [1,11].

The closely related *S. Typhimurium* ST313 lineages I and II evolved independently around 52 and 35 years ago respectively with the acquisition of the chloramphenicol acetyltransferase (*cat*) resistance gene [7]. Both lineages have been shown to carry a *S. Typhimurium* virulence plasmid, commonly known as pSLT, which also encodes genes conferring resistance to common antimicrobials including tetracycline, sulfamethoxazole-trimethoprim and chloramphenicol [7,11]. In addition, two related, but phylogenetically different epidemic clades of *S. Enteritidis* ST11, the West African clade and the Central/Eastern African clade, characterised by the presence of chloramphenicol acetyl resistance genes *catA1* and *catA2*, respectively, plus the incomplete set of *tra* genes, emerged between 1933 and 1945 [1]. The utility of second-line antimicrobials such as fluoroquinolones, azithromycin and extended spectrum cephalosporins is limited in treating these emerging MDR strains [12].

In The Gambia, iNTS remains a leading cause of invasive diseases, [13] unlike *S. Typhi* which causes typhoid fever in many low and middle income countries (LMIC). We previously described regional serovar variation and emerging MDR in The Gambia [14,15]: *S. Typhimurium* was found to be more prevalent in the western region, while *S. Enteritidis*, including MDR strains, were found to be more prevalent in the eastern region. In this context, we performed whole-genome analysis of clinical NTS isolates to determine prevalent genotypes and antimicrobial resistance genes. The resulting analysis can be used to help guide clinical management and control of NTS diseases in The Gambia.

Methods and Materials:

Study setting and population

The study was conducted at the Medical Research Council Unit The Gambia at the London School of Hygiene and Tropical Medicine (MRCG @ LSHTM) using clinical NTS isolates from the eastern (Upper River Region) and western (West Coast Region and Greater Banjul Area) regions of The Gambia (Figure 1). The eastern region, located on the far east side of the river Gambia, is the commercial centre and a busy economic hub, with an estimated population of 200,000 people. It is an important transit point for merchandise and people going into eastern Senegal, Mali and Guinea Conakry. The western region is densely populated, with a population of over 1 million people including the capital city, Banjul (Figure 1) [16]. Malaria declined in recent years but remains endemic with peak transmission occurring from July to November [17]. Malnutrition remains a problem with the prevalence of underweight, stunting and wasting among children under 5 years old estimated at 16.4%, 25.0% and 4.3% respectively [18]; HIV prevalence among adults aged 15 – 49 years is estimated at 2.1% [19].

Sample collection, microbiological procedures and antimicrobial susceptibility testing

The study evaluated 100 clinical NTS from 93 patients admitted to hospital with suspected sepsis, gastroenteritis or other focal infections in the eastern and western regions of The Gambia (Table 1). Seven patients had multiple samples collected during the same infection episode (Table 2). Three patients had concurrent bacteraemia and gastroenteritis, two had bacteraemia with meningitis whilst two had bacteraemia with two sampling episodes. All NTS from the eastern region (20) were isolated in 2001 from 18 patients, and those from the western region (80) were isolated between 2006 to 2018 from 75 patients (Table 1). All isolates were stored in 15% (v/v) glycerol broth at -70°C. The isolates were grown on MacConkey agar overnight at 37°C in the Clinical Microbiology Laboratory. The laboratory is accredited to

Good Clinical Laboratory Practice (GCLP; 2010) and ISO15189 (2015) as previously described [14]. Antimicrobial susceptibility for amoxicillin-clavulanate, ampicillin, cefotaxime, cefoxitin, ceftazidime, ceftriaxone, cefuroxime, chloramphenicol, ciprofloxacin, gentamicin, nalidixic acid, sulfamethoxazole-trimethoprim and tetracycline, were tested on Mueller-Hinton agar (MHA) using the Kirby-Bauer disk diffusion method. Interpretation was done according to the 2017 Clinical Laboratory Standard Institute (CLSI) guidelines [20]. Antimicrobial agents were from BD Oxoid (Basingstoke, United Kingdom) and *Escherichia coli* (ATCC 25922) was used for quality control.

DNA extraction and whole genome sequencing

Genomic DNA was extracted and sequenced in two locations; the MRCG (n=33), and sample processing (n=67) at the University of Liverpool (UK) with DNA sequencing at the Earlham Institute (UK). The extraction protocol for the MRCG used the QIAamp DNA Mini kit (Qiagen, Germany) to extract the DNA from 1mL of an overnight culture grown in triple soya broth (TSB) from BD Oxoid (Basingstoke, United Kingdom) at 37°C, according to manufacturer's instructions, and quantified using a Qubit fluorometer (ThermoFisher, Qubit dsDNA HS Assay). Libraries were prepared using the Nextera XT kit using the Illumina MiSeq system. The DNA extraction and sequencing of samples processed at the University of Liverpool were carried out using an optimised method for large-scale sequencing [21], including the bespoke LITE (Low Input, Transposase Enabled) pipeline for library construction, and Illumina HiSeq sequencing technology. Both sites used the 2x150 bp read protocol.

Genome assembly and *in silico* analysis

The quality of the raw reads was assessed using FASTQC [22] where on average, all reads had a quality Phred score (Qscore) above 30. Paired-end reads were trimmed using Trimomatic

[23] and assembled into contigs using SPades, with default settings [24]. *In silico* serotyping of the core genome MLST (cgMLST) and serovar was predicted using the *Salmonella in Silico* Typing resource (SISTR) platform [25]. eBurst Groups (eBGs) were assigned using the Enterobase platform which is based on the allelic identity that accounts for homologous recombination, defined by Achtman as closely related natural genetic clusters/populations of two or more sequence types connected by pair-wise identity or single locus variants [26].

Phylogenetic analysis

Assembled contigs were annotated using Prokka (v1.14.6). The pan-genome was determined using Roary (v3.13.0) [27], taking the GFF files from Prokka as input with default settings. The pan-genome was aligned using Mafft (v7.464) to generate a high-quality sequence alignment. The alignment was used to create a maximum likelihood phylogeny using IQ-TREE (v1.6.12). The phylogenetic tree was visualised and annotated using the interactive Tree Of Life (iTOL). iTOL annotations input files for the tree were generated using custom python scripts (<https://github.com/jodyphelan/itol-config-generators>). Antimicrobial resistance (AMR), plasmids and virulence genes were detected by ResFinder [28], PlasmidFinder [29] and virulence factor gene database (VFDB) with a minimum coverage and nucleotide identity of 90% as the cut-off. Publicly available data was downloaded from the European Nucleotide Archive (ENA) to compare the strains from this study against other African strains. All sequence reads from the African samples belonging to the taxid 149539 (NCBI:txid149539) were downloaded and assembled/annotated using the same methods as detailed above. The pangenome phylogeny was constructed using the same methods as outlined above.

Statistical analysis:

The independent variables were serovars and the dependent variable was disease category. The relationships between the serovar and disease were analysed using logistic regression with

176 measures of association expressed in odds ratios. No power calculations were performed and
 177 an alpha value of 0.1 was considered statistically significant. All statistical analyses were
 178 performed in Stata, Version 13.1 (StataCorp. 2013. Stata Statistical Software: Release 13.
 179 College Station, TX: StataCorp LP.)

180 Ethical Review and Approval

181 The study received ethical approval from the Joint MRC/Gambia Government Ethics
 182 Committee (SCC1498).

Results:

Isolate source, associated disease syndrome and regional serovars differences

One hundred isolates recovered from clinical samples from eastern (n=20) and western (n=80) regions of The Gambia were analyzed (Table 1). Isolates were recovered from patients of all ages, with a median age range of 5-14 years. Isolates from the eastern region were predominantly from invasive disease (17 blood and 2 CSF) with only 1 gastroenteritis (stool) source cases. Isolates from the western region were associated with invasive disease (48 blood and 1 CSF), gastroenteritis (25 stool), and other focal non-invasive infections (5 abscesses/pus and 1 urine) (Table 3). Overall, high serovar diversity was noted. *Salmonella* serovars other than *S. Enteritidis*, *S. Typhimurium* and *S. Virchow* were primarily responsible for gastroenteritis (17/26; 65.4%), whilst *S. Typhimurium* was the leading cause of invasive disease (Table 4). *S. Typhimurium* and *S. Enteritidis* were 5 times and twice as likely to cause invasive disease than gastroenteritis respectively.

Sequence types and eBurst groups

All *S. Typhimurium* were in eBG1 and assigned to a single sequence type, ST19 with one or two allelic variants (Table 5). All *S. Enteritidis* belonged to eBG4 assigned to ST11 and ST1925, including two isolates having single locus allelic variants. *S. Virchow* was in eBG9 and assigned to three different STs as follows: ST181, ST755 and ST841. The four *S. Hull* isolates belonged to eBG330 and assigned ST1996 with single locus variant. *S. Stanleyville* eBG79 (ST339), *S. Poona* eBG46 (ST308) and *S. Give* eBG67 all belonged to ST516.

AMR genes, AMR phenotypes, plasmid replicons and virulence genes

Antimicrobial resistance genes belonging to eight classes of antimicrobials were detected, plus a biocide tolerance genetic determinant. Interestingly, genes *aac(6')-Iaa_1* and *mdf(A)_1* encoding an aminoglycoside modifying enzyme and a multidrug transporter, respectively, were present in all strains, but did not have any detectable phenotypic effects on our isolates (Figure

2). Other AMR genes were harboured by 16/100 (16.0%) isolates and confer resistance to aminoglycosides (*aph_3_Ib* and *aph_6_Id*; n=12), tetracyclines (*tet_A* and *tet_B*; n=4), trimethoprim (*dfrA14*, *dfrA7* and *dfrA8*; n=8), sulfamethoxazole (*sul2* and *sul1*; n=7), ampicillin (*blaTEM-1B*; n=8), fosfomycin (*fosA7_1*; n=8), azithromycin (*mph_A*; n=3) and chloramphenicol (*catA1_1*; n=2) (Figure 2). Possession of three or more AMR genes was found in 9/100 (9.0%) isolates, 7/9 (77.8%) of which were found in *S. Enteritidis*.

Phenotypic resistance was observed for tetracycline, ampicillin, sulfamethoxazole-trimethoprim and chloramphenicol in 9/100 (9%) isolates, correlating with the presence of resistance genes, except for streptomycin, fosfomycin and azithromycin that were not phenotypically tested (Figure 2). The odds of resistance to ampicillin, sulfamethoxazole-trimethoprim and tetracycline were 51, 20 and 25 times more likely for *S. Enteritidis* than all other serovars combined (Table 6). No resistance to gentamicin was observed phenotypically despite the presence of two aminoglycoside resistance genes (*aph_3_Ib* and *aph_6_Id*), which only confer resistance to streptomycin.

Nineteen different plasmid replicons were detected in 66/100 (66%) isolates; 7 isolates harboured one plasmid replicon, 39 harboured two, 16 harboured three and 4 harboured four plasmid replicons (Supplementary table 2). The most common plasmid types were *IncFII* (n=55) and *IncFIB* (n=50), harboured by all *S. Typhimurium* and all but the two chloramphenicol-resistant *S. Enteritidis* (Table 7). The *IncN_1* plasmid was associated with MDR including azithromycin resistance and was only found in *S. Enteritidis* from the eastern region (Figure 4a). The *Inc* plasmid is reported to be associated with beta-lactam, streptomycin and sulphonamide resistance. Interestingly, the *IncI1_Alpha* was harboured by the two chloramphenicol MDR *S. Enteritidis* strains from the western region and the susceptible strains from the eastern region (Figure 4b). Notably, no plasmid replicons were detected in *S. Bradford*, *S. Hull*, *S. Stanleyville*, *S. Rubislaw*, *S. Vinohrady* and 1,4,12,27:g,m:1,2 serovars.

A total of 115 virulence genes were found with notable absence of *entA*, *entB*, *entE*, *faeC*, *faeD*, *faeE*, *fepC*, *fepG* in all Gambian NTS serovars (Figure 3). We also identified virulence genes encoding toxins, fimbriae and flagella facilitate invasion, adhesion, type III secretion and survival within host, among other functions increasing bacterial virulence (Figure 3). The highest average number of virulence genes was seen for *S. Typhimurium* (mean: 111/115; 96.52%) with the notable absence of the *cdtB*, *ssPH1* and *shdA* genes. In contrast, the number of virulence genes *S. Enteritidis* isolates was smaller (mean: 105/115, 91%), with most isolates missing the *cdtB*, *gogB*, *grvA*, *shdA*, *sinH*, *slrP*, *sseK2* and *ssPH1* genes.

Phylogenetic analysis

The SNP analysis showed the isolates are clustered into respective serovars and geographic location (Figure 3). *S. Typhimurium* ST19 demonstrated clonality whilst *S. Enteritidis* was much diverse (Supplementary figures 1 and 2). To put our data within the wider regional context, we compared our *S. Enteritidis* strains to 495 available African *S. Enteritidis* genomes from the European Nucleotide Archive (ENA). Our analysis revealed considerable genetic diversity which fell into three clades: the North American poultry-associated clade [30] and the global epidemic clade known to cause human gastroenteritis in addition to the West African clade known to cause invasive diseases carrying the *catA1* gene [1] (Figure 4a). All *S. Enteritidis* strains from the eastern region (n=9) and 3 from the western region fell within the West African clade, clustering closely with *S. Enteritidis* strains from Ghana, Guinea and Mali. Among these, 7/12 (58.3%) were MDR (5 from the eastern region and two from the western region) and the remaining were pan susceptible strains (4 from the eastern region plus one western region). Among all strains within the West African clade from the subregion in the study, azithromycin resistance *mph_A_2* gene was only harboured by strains from the eastern region (Figure 4b). Strains belonging to the North American clade were isolated from blood (n=1) and stool (n=2), whilst those within the global epidemic strains were isolated from blood (n=2) and urine (n=1).

Discussion:

We used phylogenetic analysis to confirm the circulation of a diverse range of NTS serovars including, the epidemic *S. Enteritidis* West African clade mainly in the eastern region of The Gambia, as far back as 2001. This clade is associated with high mortality, harbouring MDR and exhibit genome degradation facilitating an invasive lifestyle, warranting further epidemiological investigations and surveillance as it has important implication in treatment [1,31]. Interestingly, not all *S. Enteritidis* in this study within the clade harboured resistance genes. The *S. Enteritidis* exhibited diversity with other global lineages present. Remarkably, this study found a closely related clonal lineage of *S. Typhimurium* ST19 causing invasive NTS disease as opposed ST313 virulent lineage, which was reported to be circulating in other parts of sSA [7,32]. A possible explanation is that the sequence type ST313 is mainly associated with HIV infection which has a low prevalence in The Gambia [19]. Nevertheless, Panzenhagen *et al.*, reported *S. Typhimurium* ST19-lineage that has evolved in Brazil similar to ST313 restricted to invasive diseases [33]. This unique pathogenesis warrants further comparative genomic and epidemiological investigations into the *S. Typhimurium* ST19-lineage [13].

Two-thirds of serovars responsible for gastroenteritis in this setting were serovars other than *S. Typhimurium* and *S. Enteritidis* as opposed to other parts of the world where these two serovars account for the highest burden [34]. The great diversity in serovars causing gastroenteritis as opposed to iNTS may suggest that NTS gastroenteritis may not be a predisposition to iNTS in The Gambia. In addition, NTS was not a major cause of gastroenteritis in this setting [35]. Although a huge gap exist regarding transmission dynamics of NTS in sSA [36], more insight is needed in understanding the relationship between NTS gastroenteritis and iNTS disease. In addition, the past two decades has revealed geographical serovar diversity between the two regions, indicating a possible regional specific

epidemiological pattern of NTS in The Gambia. However, the time difference in the sampling between the two regions may confound the difference in location and warrant further investigation. Nonetheless, previous phenotypic studies have highlighted these serovar differences [14,15]. The changing disease pattern of NTS in sSA, associated with specific lineages of *S. Typhimurium* and *S. Enteritidis* remain a major concern and warrant surveillance [1,9,10,37]. Although a recent decline has been reported for iNTS, it remains a leading cause of bacteraemia in The Gambia [38,39]. Three cases of *Salmonella* bacterial meningitis were included in this study, all of which were found in paediatric patients under 10 years old. Although rare, NTS meningitis has been reported elsewhere in Africa, and is often associated with high case fatality [40,41]. Therefore, NTS needs to be considered in the differential diagnosis of bacterial meningitis following post-vaccine declines in the prevalence of Hib, *Neisseria meningitidis* and pneumococcal meningitis [38,42].

Antimicrobial resistance was found to be correlated with serovar, plasmid replicon and geographical location. MDR was confined within *S. Enteritidis* noted for first-line antibiotics such as ampicillin, sulfamethoxazole-trimethoprim, tetracycline and chloramphenicol. Although no fluoroquinolone or cephalosporin resistance was identified, implying these drugs might still be effective in The Gambia, the emergence of azithromycin resistance gene *mph_A* requires further monitoring as a recommended drug of choice for iNTS [43]. Nevertheless, the aminoglycoside resistance gene *aac(6')-Iaa* and a multi-drug transporter gene *mdf(A)* were present in all serovars including pan-susceptible isolates. This highlights the potential of using genomic-based AMR prediction to monitor AMR determinants for emerging resistance. Our study did not phenotypically test streptomycin susceptibility which lacks clinical breakpoints and it is not used in the treatment of infections. In addition, the streptomycin resistance genes were frequently found to lack expression [44]. While the development of AMR has been mainly attributed to antibiotic misuse in humans and animals, evidence has shown that environmental

factors such as poor sanitation, hygiene and access to clean water may be equally responsible for driving resistance in LMIC [45,46].

Although the differences in sampling timepoint was a major limitation in the study, the diversity of AMR between serovars and geographic regions highlights the need for real-time surveillance as well as region-appropriate interventions to effectively combat AMR. Geographic differences seen in AMR may suggest differences in selective pressure and ecological factors thus suggesting need for location specific control measures. The study by Carroll *et al.* underscored distinct factors such as use of antimicrobials in food producing animals as contributing to emergence and dispersal of AMR in humans [47]. Our findings are consistent with other studies that show NTS serovar differences in geographical locations within the same country [47,48]. A correlation between phenotypic and genotypic resistance was also observed (Figure 2). The *IncN* type plasmid was strongly associated with resistance and was found only in MDR *S. Enteritidis*, thus requiring closer surveillance. This plasmid is associated with dissemination of antimicrobial resistance with high potential of spread [49].

We found many virulence genes, however, the identification of virulence factors coding for specific phenotypic traits can be challenging, due to differences in specific traits among serovars [50]. Notwithstanding, the pathogenic success of NTS serovars is directly linked to their plethora of virulence factors aided by host susceptibility, serovar fitness, infectious dose and antimicrobial resistance (AMR) [51]. Further studies are needed to understand the clinical implications of these virulence genes.

There are several limitations in this study. First, the isolates were collected at different time points, with a lag of up to 18 years between the two different regions, which may lead to missing temporal differences. However, the higher AMR prevalence of *S. Enteritidis* in the eastern region as early as 2001 compared to the more recent western region proves the point that AMR emerged a lot earlier and more prevalent in the eastern region. Second, relatively

few isolates were analysed from only two regions due to limited microbiology capacity and therefore our results may not reflect the entirety of strains and lineages of the NTS in The Gambia.

In conclusion, this study has revealed great serovar diversity in serovars responsible for gastroenteritis and iNTS and provides evidence for the emergence of MDR *S. Enteritidis* epidemic West African clade in The Gambia. These findings have important implications for antimicrobial prescription policies and regional surveillance of NTS disease. We have demonstrated that a robust genomic epidemiological surveillance of NTS by WGS can be instrumental in generating the critical knowledge and timely information for better disease management and prevention.

Acknowledgement:

We thank the following for their support in the study: Jay C.D Hinton, Senior Investigator Award grant recipient of the Liverpool for leading the 10k Salmonella project, Nurudeen Ikumapayi of MRCG@LSHTM for support in providing the isolates from the eastern regions. We are also grateful to Jarrai Manneh and Abdul Karim Sesay for providing support and training with the Illumina sequencing in the MRCG@LSHTM, Buntung Ceesay and Mamadou Jallow of MRCG@LSHTM for contributing to the microbiology laboratory analysis. We thank the patients, Clinical Laboratories, and Services of the Medical Research Council Unit.

Author notes: Supporting data and protocols are provided as supplementary material and available.

356 **Conflict of interest:**

357 We declare no conflict of interest

358 **Author contributions:**

359 Conceptualization: SD, RSB, MA and BKA.

360 Laboratory analysis and sequencing of isolates in MRCG: SD

361 Data transfer of sequence reads into analysis pipeline AW, AK and JP

362 Processing and sequencing of isolates in the UK, including data transfer. BPS

363 Data curation: SD, AW, AK

364 Formal analysis: SD and AKM

365 Writing original draft: SD

366 Review and editing: SD, RSB, SY, DN, BPS, BKA, MA

367 Review of final draft: all authors

368 Supervision: SK, BKA and MA

369 Project Administration: SD

370

371 **Funding information**

372 This project was jointly supported by MRCG @ LSHTM; the Global Challenges Research

373 Fund (GCRF) data & resources grant BBS/OS/GC/000009D and the BBSRC Core Capability

374 Grant to the Earlham Institute BB/CCG1720/1. This project was also partly supported by the

375 Wellcome Trust Senior Investigator Award (106914/Z/15/Z) to Jay C. D. Hinton. The source

376 of funding has no role in the design, decision to publish and writing of the manuscript.

377

378

References

1. Feasey NA, Hadfield J, Keddy KH, Dallman TJ, Jacobs J, Deng X, et al. Distinct Salmonella Enteritidis lineages associated with enterocolitis in high-income settings and invasive disease in low-income settings. *Nat Genet.* 2016;48(10):1211–7.
2. Jones TF, Ingram AL, Cieslak PR, Vugia DJ, Tobin-D’Angelo M, Hurd S, et al. Salmonellosis Outcomes Differ Substantially by Serotype. *J Infect Dis.* 2008;198(1):109–14.
3. Langridge GC, Fookesa M, Connor TR, Feltwell T, Feasey N, Parsons BN, et al. Patterns of genome evolution that have accompanied host adaptation in Salmonella. *Proc Natl Acad Sci U S A.* 2015;112(3):863–8.
4. Majowicz SE, Musto J, Scallan E, Angulo FJ, Kirk M, O’Brien SJ, et al. The Global Burden of Nontyphoidal Salmonella Gastroenteritis. *Clin Infect Dis.* 2010;50(6):882–9.
5. Crump JA, Sjölund-Karlsson M, Gordon MA, Parry CM, et al. Epidemiology, clinical presentation, laboratory diagnosis, antimicrobial resistance, and antimicrobial management of invasive Salmonella infections. Vol. 28, *Clinical Microbiology Reviews.* 2015. p. 901–37.
6. Ao TT, Feasey NA, Gordon MA, Keddy KH, Angulo FJ, Crump JA. Global Burden of Invasive Nontyphoidal Salmonella Disease, 2010. *Emerg Infect Dis.* 2015 Jun;21(6):941–9.
7. Okoro CK, Kingsley RA, Connor TR, Harris SR, Parry CM, Al-Mashhadani MN, et al. Intracontinental spread of human invasive Salmonella Typhimurium pathovariants in sub-Saharan Africa. *Nat Genet.* 2012;44(11):1215–21.
8. Feasey NA, Dougan G, Kingsley RA, Heyderman RS, Gordon MA. Invasive non-typhoidal salmonella disease: An emerging and neglected tropical disease in Africa.

- Vol. 379(9835), The Lancet. 2012. p. 2489–99.
9. Uche I V, MacLennan CA, Saul A. 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. 2017;11(1):e0005118.
10. Kingsley RA, Msefula CL, Thomson NR, Kariuki S, Holt KE, Gordon MA, et al. Epidemic multiple drug resistant Salmonella Typhimurium causing invasive disease in sub-Saharan Africa have a distinct genotype. Genome Res. 2009;19(12):2279–87.
11. Okoro CK, Barquist L, Connor TR, Harris, Simon R, Clare S, Stevens MP, et al. Signatures of Adaptation in Human Invasive Salmonella Typhimurium ST313 Populations from Sub-Saharan Africa. PLoS Negl Trop Dis. 2015;9(6):e0003848.
12. Lunguya O, Lejon V, Phoba M-F, Bertrand S, Vanhoof R, Glupczynski Y, et al. Antimicrobial Resistance in Invasive Non-typhoid Salmonella from the Democratic Republic of the Congo: Emergence of Decreased Fluoroquinolone Susceptibility and Extended-spectrum Beta Lactamases. PLoS Negl Trop Dis. 2013;7(3):e2103.
13. Darboe S, Okomo U, Muhammad A, Ceesay B, Jallow M, Usuf E, et al. Community-acquired Invasive Bacterial Disease in Urban Gambia , 2005 – 2015 : A Hospital-based Surveillance. Clin Infect Dis. 2019;69(Suppl 2):105–13.
14. Kwambana-Adams B, Darboe S, Nabwera H, Forster-Nyarko E, Ikumapayi UN, Secka O, et al. Salmonella infections in The Gambia, 2005-2015. Clin Infect Dis. 2015;61(Suppl 4):S354-362.
15. Ikumapayi UN, Antonio M, Sonne-Hansen J, Biney E, Enwere G, Okoko B, et al. Molecular epidemiology of community-acquired invasive non-typhoidal Salmonella among children aged 2-29 months in rural Gambia and discovery of a new serovar, Salmonella enterica Dingiri. J Med Microbiol. 2007;56(11):1479–84.
16. The Gambia Bureau Statistics. The Gambia Demographic and Health Survey.

- 2014;(April):2014–5.
17. Ceesay SJ, Casals-Pascual C, Nwakanma DC, Walther M, Gomez-Escobar N, Fulford AJ., et al. Continued decline of malaria in The Gambia with implications for elimination. PLoS One. 2010;5(8):e12242.
18. UNICEF, WHO WBJ child malnutrition estimates (J). Prevalence of underweight, weight for age [Internet]. 2013 [cited 2020 Jul 7]. Available from: <https://data.worldbank.org/indicator/SH.STA.MALN.ZS?locations=GM>
19. National AIDS Secretariat. The Gambia Global AIDS Response Progress Report [Internet]. 2015. Available from: https://www.unaids.org/sites/default/files/country/documents/GMB_narrative_report_2015.pdf
20. CLSI. Performance standards for antimicrobial susceptibility testing. Performance Standards for Antimicrobial Susceptibility Testing; Twenty-seventh Informational Supplement. CLSI Document M100-S27. 2017.
21. Perez-Sepulveda BM, Heavens D, Pulford C V, Predeus A V., Low R, Webster H, et al. Title : An accessible , efficient and global approach for the large-scale sequencing of bacterial genomes. bioRxiv [Internet]. 2020 [cited 2020 Jul 25];1–33. Available from: <https://www.biorxiv.org/content/10.1101/2020.07.22.200840v1>
22. Wingett SW and AS. FASTQ Screen: A tool for multi-genome mapping and quality control. F1000Research. 2018;7:1338.
23. Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114–20.
24. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012;19(5):455–77.

25. Yoshida C, Brumwell SL, Lingohr EJ, Ahmad A, Blimkie TM, Kogan BA, et al. Draft whole-genome sequences of 25 *Salmonella enterica* strains representing 24 serovars. *Genome Announc.* 2016;4(2):e01718-15.
26. Achtman M, Wain J, Weill FX, Nair S, Zhou Z, Sangal V, et al. Multilocus sequence typing as a replacement for serotyping in *Salmonella enterica*. *PLoS Pathog.* 2012;8(6):e1002776.
27. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MTG, et al. Roary: Rapid large-scale prokaryote pan genome analysis. *Bioinformatics.* 2015 May 25;31(22):3691–3.
28. Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, et al. Identification of acquired antimicrobial resistance genes. *J Antimicrob Chemother.* 2012;67(11):2640–4.
29. Carattoli A, Zankari E, García-Fernández A, Larsen M V, Lund O, Villa L, et al. In Silico detection and typing of plasmids using plasmidfinder and plasmid multilocus sequence typing. *Antimicrob Agents Chemother.* 2014;58(7):3895–903.
30. Deng X, Desai PT, den Bakker HC, Mikoleit M, Tolar B, Trees E, et al. Genomic epidemiology of *Salmonella enterica* serotype Enteritidis based on population structure of prevalent lineages. *Emerg Infect Dis.* 2014;20(9):1481–9.
31. Aldrich C, Hartman H, Feasey N, Chattaway MA, Dekker D, Al-Emran HM, et al. Emergence of phylogenetically diverse and fluoroquinolone resistant *salmonella enteritidis* as a cause of invasive nontyphoidal *salmonella* disease in Ghana. *PLoS Negl Trop Dis.* 2019;13(6):e0007485.
32. Branchu P, Bawn M, Kingsley RA. Genome variation and molecular epidemiology of *Salmonella enterica* serovar Typhimurium pathovariants. *Infect Immun.* 2018;86(8):1–17.

33. Panzenhagen PHN, Paul NC, Conte CA, Costa RG, Rodrigues DP, Shah DH. Genetically distinct lineages of Salmonella Typhimurium ST313 and ST19 are present in Brazil. *Int J Med Microbiol*. 2018 Mar 1;308(2):306–16.
34. Hendriksen RS, Vieira AR, Karlsmose S, Lo Fo Wong DMA, Jensen AB, Wegener HC, et al. Global monitoring of salmonella serovar distribution from the world health organization global foodborne infections network country data bank: Results of quality assured laboratories from 2001 to 2007. *Foodborne Pathog Dis*. 2011 Aug 1;8(8):887–900.
35. Kotloff KL, Nataro JP, Muhse K. 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*. 2013.
36. Gordon M a. Europe PMC Funders Group Invasive Non-typhoidal Salmonella Disease – epidemiology , pathogenesis and diagnosis. *Curr Opin Infect Dis*. 2012;24(5):484–9.
37. Mahon BE, Fields PI. Invasive Infections with Nontyphoidal Salmonella in Sub-Saharan Africa. *Microbiol Spectr*. 2016;4(3).
38. Zaman SMA, Howie SRC, Ochoge M, Secka O, Bah A, Baldeh I, et al. Impact of routine vaccination against Haemophilus influenzae type b in The Gambia: 20 years after its introduction. *J Glob Health*. 2020;10(1):010416.
39. Mackenzie G, Ceesay SJ, Hill PC, Walther M, Bojang KA, Satoguina J, et al. A decline in the incidence of invasive non-typhoidal salmonella infection in the gambia temporally associated with a decline in malaria infection. *PLoS One*. 2010;5(5):e10568.
40. Molyneux EM, Mankhambo LA, Ajib P, Graham SM, Forsyth H, Amos P, et al. The outcome of non-typhoidal salmonella meningitis in Malawian children, 1997-2006. *Ann Trop Paediatr*. 2009 Mar;29(1):13–22.

41. Keddy KH, Sooka A, Musekiwa A, Smith AM, Ismail H, Tau NP, et al. Clinical and microbiological features of salmonella meningitis in a South African Population, 2003-2013. Clin Infect Dis. 2015;61(Suppl 4):S272-82.
42. Mackenzie GA, Hill PC, Jeffries DJ, Hossain I, Uchendu U, Ameh D, et al. Effect of the introduction of pneumococcal conjugate vaccination on invasive pneumococcal disease in The Gambia: a population-based surveillance study. Lancet Infect Dis. 2016;16(6):703–11.
43. Gomes C, Martínez-Puchol S, Palma N, Horna G, Ruiz-Roldán L, Pons MJ, et al. Macrolide resistance mechanisms in Enterobacteriaceae: Focus on azithromycin. Vol. 43, Critical Reviews in Microbiology. Taylor and Francis Ltd; 2017. p. 1–30.
44. Springer B, Kidan YG, Prammananan T, Ellrott K, Böttger EC, Sander P. Mechanisms of streptomycin resistance: Selection of mutations in the 16S rRNA gene conferring resistance. Antimicrob Agents Chemother. 2001;45(10):2877–84.
45. Afema JA, Byarugaba DK, Shah DH, Atukwase E, Nambi M, Sischo WM. Potential sources and transmission of salmonella and antimicrobial resistance in Kampala, Uganda. PLoS One. 2016;11(3):e0152130.
46. Laxminarayan R, Duse A, Wattal C, Zaidi AK., Wertheim HF., Sumpradit N, et al. Antibiotic resistance-the need for global solutions. Vol. 13, The Lancet Infectious Diseases. 2013. p. 1057–98.
47. Carroll LM, Wiedmann M, den Bakker H, Siler J, Warchocki S, Kent D, et al. Whole-genome sequencing of drug resistant Salmonella enterica isolates from dairy cattle and humans in New York and Washington States reveals source and geographic associations. Appl Environ Microbiol. 2017;83(12).
48. Kariuki S, Oundo JO, Muyodi J, Lowe B, Threlfall E, Hart CA. Genotypes of multidrug-resistant Salmonella enterica serotype typhimurium from two regions of

529 Kenya. FEMS Immunol Med Microbiol. 2000;29(1).

530 49. García-Fernández A, Villa L, Moodley A, Hasman H, Miriagou V, Guardabassi L, et
531 al. Multilocus sequence typing of IncN plasmids. J Antimicrob Chemother. 2011
532 Sep;66(9):1987–91.

533 50. Van Asten AJAM, Van Dijk JE. Distribution of “classic” virulence factors among
534 Salmonella spp. FEMS Immunol Med Microbiol. 2005;44(3):251–9.

535 51. Cheng RA, Eade CR, Wiedmann M. Embracing diversity: Differences in virulence
536 mechanisms, disease severity, and host adaptations contribute to the success of
537 nontyphoidal salmonella as a foodborne pathogen. Front Microbiol. 2019;
538
539

Tree scale: 0.001

Colored ranges

- N. American Poultry associated clade
- Global epidemic clade
- West African clade

Study samples

- Sample

region

- Central
- East
- Gambia - East
- Gambia - West
- North
- South
- West

Table 1. Baseline characteristics of Gambian non-typhoidal *Salmonella* disease patients from whom isolates were cultured for use in this study

		<i>N (%)</i>	<i>eastern region</i>	<i>western region</i>
<i>Age range</i>	0-4 years	42 (45.2)	4 (25.0)	38 (50.7)
	5-14 years	21 (22.6)	12 (65.0)	9 (12.0)
	≥15 years	26 (27.9)	2 (10.0)	24 (32.0)
	Unknown	4 (4.3)	0	4 (5.3)
<i>Gender</i>	Male	51(54.3)	10 (55.6)	41 (54.7)
	Female	38(41.5)	7 (38.9)	31 (41.3)
	Unknown	4 (4.2)	1 (5.5)	3 (4.0)
<i>Source</i>	Invasive disease	68 (68.0)	19 (95.0)	49 (61.2)
	Gastroenteritis	26 (26.0)	1 (5.0)	25 (31.3)
	Other	6 (6)	0	6 (7.5)
<i>Serovars</i>	<i>S. Enteritidis</i>	18 (18.0)	9 (45.0)	9 (11.2)
	<i>S. Typhimurium</i>	31 (31.0)	8 (40.0)	23 (28.8)
	<i>S. Virchow</i>	8 (8.0)	0	8 (10.0)
	Other serovars*	43 (43.0)	3 (15.0)	40 (50.0)

*Shown in figure supplementary table

Table 2. Patients with multisite NTS simultaneous infections

Patient	Sex	Age range (yrs)	Infection source	Serovar	Region	Date isolated
1	Male	0-4	Bacteraemia	<i>S. Enteritidis</i>	western	19/9/2017
			Gastroenteritis	<i>S. Enteritidis</i>		14/9/2017
2	Female	35-39	Bacteraemia	<i>S. Typhimurium</i>	western	1/1/2014
			Gastroenteritis	<i>S. Typhimurium</i>		1/1/2014
3	Male	0-4	Bacteraemia	<i>S. Virchow</i>	western	1/3/2009
			Meningitis	<i>S. Virchow</i>		1/3/2009
4	Male	30-34	Bacteraemia	<i>S. Typhimurium</i>	western	1/11/2006
			Gastroenteritis	<i>S. Typhimurium</i>		1/11/2006
5	Female	0-4	Bacteraemia	1,6,14,25:y:1,5	western	1/10/2008
			Bacteraemia	1,6,14,25:y:1,5		1/10/2008
6	Male	5-9	Bacteraemia	<i>S. Enteritidis</i>	eastern	1/10/2001
			Bacteraemia	<i>S. Enteritidis</i>		1/10/2001
7	Male	0-4	Bacteraemia	<i>S. Typhimurium</i>	eastern	1/12/2001
			Meningitis	<i>S. Typhimurium</i>		1/12/2001

Table 3. Gambian non-typhoidal *Salmonella* serovar distribution by infection in this study

	eastern region (2001) n=20					western region (2006-2018) n=80				
	Infection source									
	Invasive disease		Gastroenteritis	Other	Invasive disease		Gastroenteritis	Other*		
	<u>Total</u>	19	1	0	<u>Total</u>	49	25	6		
	Bacteraemia	Meningitis		Bacteraemia	Meningitis					
	17	2		48	1					
<i>S. Enteritidis</i>	9	8 (47.1)	0	1	0	9	6 (12.5)	0	2 (8.0)	1 (16.6)
<i>S. Typhimurium</i>	8	6 (35.3))	2	0	0	23	19 (39.6)	0	3 (12.0)	1 (16.6)
<i>S. Virchow</i>	0	0	0	0	0	8	4 (8.3)	1	3 (12.0)	0
<i>Other serovars</i>	3	3 (17.6)	0	0	0	40	19 (39.6)	0	17 (68.0)	4 (66.8)

*Other = Abscess/pus and urine isolates

Table 4. Gambian non-typhoidal *Salmonella* serovar distribution and disease prevalence

Serovar	Total	Invasive N (%)	Gastroenteritis N (%)	Others N (%)	Odds of Invasive vs Gastroenteritis	95% CI	p value
	100	68	26	6			
<i>S. Typhimurium</i>	31	27 (39.7)	3 (11.5)	1 (16.7)	5.05	1.38; 18.48	0.014
<i>S. Enteritidis</i>	18	14 (20.6)	3 (11.5)	1 (16.7)	1.99	.52; 7.58	0.315
<i>S. Virchow</i>	8	5 (7.4)	3 (11.5)	0 (0)	0.61	0.13, 2.75	0.519
Other serovars	43	22 (32.4)	17 (65.4)	4 (66.7)	0.25	0.10; 0.66	0.005

Table 5. eBurst groups of the major Gambian NTS serovars responsible for clinical disease

Serovars	Sequence type	eBG	Number
<i>S. Enteritidis</i>	11	4	15
	1925	4	1
	11	4	2*
<i>S. Typhimurium</i>	19	1	29
	19	1	2*
<i>S. Virchow</i>	181	9	2
	755	9	1
	841	9	4
	841	9	1*

*One or two allelic variants of ST.

Table 6. Odds of *S. Enteritidis* resistance against other NTS serovars causing disease in The Gambia

Antimicrobials	<i>S. Enteritidis</i> n=18	Other serovars n=72	Odds of <i>S.</i> Enteritidis vs all serovars	95% CI	P-value
<i>Ampicillin</i>	7	1	51.5	5.78; 459.60	<0.001
sulfamethoxazole- trimethoprim	6	2	20.0	3.61; 110.74	0.001
Tetracycline	7	2	25.5	4.68; 138.39	<0.001
Chloramphenicol	2	0	1	NA	NA

Table 7. Summary of plasmid replicons and serovar in The Gambian non-typhoidal *Salmonella* isolates

Plasmid Type	Total	<i>S. Enteritidis</i> n=18	<i>S. Typhimurium</i> n=31	<i>S. Virchow</i> n=8	Other serovars n=43
<i>IncFII_S__1</i>	55	16	31	0	8
<i>IncFIB_S__1</i>	50	16	31	0	3
<i>IncII_1_Alpha</i>	6	5	0	0	1
<i>IncN_1</i>	5	5	0	0	0
<i>IncXI_1</i>	4	0	2	0	2
<i>IncFIB_pKPHS1__1_pKPHS1</i>	4	0	0	0	4
<i>IncFII_SARC14__1_SARC14</i>	3	0	0	0	3
<i>IncFII_p14__1_p14</i>	3	0	0	0	3
<i>IncL/M_pOXA-48__1_pOXA-48</i>	3	0	1	0	2
<i>pSL483_1</i>	3	0	3	0	0
<i>ColRNAI_1</i>	2	0	1	0	1
<i>Col_MG828__1</i>	2	0	1	1	0
<i>IncFIB_pB171__1_pB171</i>	2	0	0	0	2
<i>IncI2_1_Delta</i>	2	0	1	0	1
<i>IncXI_4</i>	1	1	0	0	0
<i>IncX3_1</i>	1	1	0	0	0
<i>repUS21__rep_pWBG764</i>	1	0	0	0	1
<i>IncFII_pRSB107__1_pRSB107</i>	1	0	0	0	1
<i>pENTAS02_1</i>	1	0	0	0	1