Original Article

Genomics of an emerging clone of *Salmonella* serovar Typhimurium ST313 from Nigeria and the Democratic Republic of Congo

Pimlapas Leekitcharoenphon[1,2], Carsten Friis[1], Ea Zankari[1], Christina A. Svendsen[1], Lance B. Price[3], Maral Rahmani[1], Ana Herrero-Fresno[1], Kayode Fashae[4], Olivier Vandenberg[5, 6], Frank M. Aarestrup[1], Rene S. Hendriksen[1]

[1]*Division of Bacterial Genomics and Epidemiology, WHO Collaborating Center for Antimicrobial Resistance in Food borne Pathogens and European Union Reference Laboratory for Antimicrobial Resistance, National Food Institute, Technical University of Denmark, Kemitorvet, Denmark*
[2]*Center for Biological Sequence Analysis, Department of System Biology, Technical University of Denmark, Kemitorvet, Denmark*
[3]*Division of Pathogen Genomics, Translational Genomics Research Institute (TGen), Flagstaff, USA*
[4]*Department of Microbiology, University of Ibadan, Ibadan, Nigeria.*
[5]*Infectious Diseases Epidemiological Unit, Public Health School, Université Libre de Bruxelles, Brussels, Belgium.*
[6]*Department of Microbiology, Saint-Pierre University Hospital, Brussels, Belgium.*

Abstract

Introduction: *Salmonella enterica* serovar Typhimurium ST313 is an invasive and phylogenetically distinct lineage present in sub-Saharan Africa. We report the presence of *S.* Typhimurium ST313 from patients in the Democratic Republic of Congo and Nigeria.

Methodology: Eighteen *S.* Typhimurium ST313 isolates were characterized by antimicrobial susceptibility testing, pulsed-field gel electrophoresis (PFGE), and multilocus sequence typing (MLST). Additionally, six of the isolates were characterized by whole genome sequence typing (WGST). The presence of a putative virulence determinant was examined in 177 *Salmonella* isolates belonging to 57 different serovars.

Results: All *S.* Typhimurium ST313 isolates harbored resistant genes encoded by bla_{TEM1b}, *cat*A1, *str*A/B, s*ul*1, and *dfr*A1. Additionally, *aac(6')1aa* gene was detected. Phylogenetic analyses revealed close genetic relationships among Congolese and Nigerian isolates from both blood and stool. Comparative genomic analyses identified a putative virulence fragment (ST313-TD) unique to *S.* Typhimurium ST313 and *S.* Dublin.

Conclusion: We showed in a limited number of isolates that *S.* Typhimurium ST313 is a prevalent sequence-type causing gastrointestinal diseases and septicemia in patients from Nigeria and DRC. We found three distinct phylogenetic clusters based on the origin of isolation suggesting some spatial evolution. Comparative genomics showed an interesting putative virulence fragment (ST313-TD) unique to *S.* Typhimurium ST313 and invasive *S.* Dublin.

Key words: *Salmonella* serovar Typhimurium ST313; Sub-Saharan Africa; MLST; whole genome sequence typing; SNP analysis; virulence genes

J Infect Dev Ctries 2013; 7(10):696-706. doi:10.3855/jidc.3328

(Received 22 January 2013 – Accepted 30 April 2013)

Introduction

Globally, non-typhoidal salmonellae are one of the most important food-borne pathogens [1-4] causing an estimated 93.8 million cases of gastroenteritis each year, including 155,000 deaths [5]. *Salmonella* infections represent a considerable burden in both developing and developed countries leading to high medical expense, lost productivity, and mortality [5-7].

Most human infections are self-limiting; however, approximately 5% of all patients infected with non-typhoidal *Salmonella* will develop bacteremia. The very young, the elderly, the malnourished, or people with underlying diseases such as malaria or HIV are at a significantly higher risk of developing bacteremia compared to otherwise healthy individuals [8]. Bacteremic patients have higher rates of hospitalization, often have prolonged courses of illness, and have higher case fatality rates [1, 9].

While severe infections with non-typhoidal *Salmonella* are relatively rare in Europe and North America, several studies have shown that invasive

non-typhoidal *Salmonella* is endemic in sub-Saharan Africa [10-13]. In some of those countries the mortality in children caused by non-typhoidal *Salmonella* bacteremia exceeds the burden of malaria [12]. In the Democratic Republic of Congo (DRC), a study conducted from 2002 to 2006 in a rural children's hospital showed that 62.1% of all bloodstream infections in children were caused by non-typhoidal *Salmonella* with *Salmonella enterica* serovars Typhimurium and Enteritidis accounting for 60.5% and 22.3% of the cases, respectively [13].

A retrospective study of *S.* Typhimurium describing invasive diseases from 1997 to 2004, identified 31 isolates from Malawi and 13 out of 20 from Kenya to be of a novel multilocus sequence type (MLST) ST313 [11]. One *S.* Typhimurium ST313 isolate was completely sequenced and found to be phylogenetically distinct from other *S.* Typhimurium isolated in sub-Saharan Africa. It was suggested that *S.* Typhimurium ST313 is strongly associated with invasive disease due to adaptation to human host as a result of genome degradation, similar to the evolutionary history of *S.* Typhi [11].

However, there is limited data describing the isolation of ST313 from stool and thus it is unknown whether the presence of *S.* Typhimurium ST313 in invasive diseases is the consequence of the high prevalence of a regionally dominant clone, possibly causing gastroenteritis in a susceptible human population, or due to a more invasive clone which is just one of many strains circulating in the local population [14].

The objective of this study was to investigate the genetic characteristics of 18 *S.* Typhimurium ST313 isolates in a spatial and temporal context. The isolates were obtained from both human stool and blood samples in patient from two sub-Saharan African countries; Nigeria and the DRC isolated in 2002 and 2005 in Nigeria and from 2002 to 2006 in DRC.

Methodology

Bacterial isolates and molecular typing

During 2004 and 2005, blood samples were collected from seven major hospitals and a private diagnostic laboratory in the province of Ibadan, Nigeria. A total of 16 *Salmonella* isolates were recovered from 223 samples; of which seven were *S.* Typhimurium [15].

From January 2002 to December 2006, blood cultures (n= 1567) of febrile children (n = 1528) either out-patients (n = 829) or hospitalized (n = 699); as well as unknown number of stool cultures were investigated at the Children's Hospital of Lwiro, in DRC. Of the blood cultures, 241 yielded *Salmonella* of which 133 were *S.* Typhimurium [13]. Unfortunately, it was not possible due to inadequate reporting to obtain an exact number of the stool, which yielded *S.* Typhimurium.

The procedures for isolation, identification, serotyping, antimicrobial susceptibility testing and PFGE included in this study have been described previously [16]. Multilocus sequence typing (MLST) was performed on all of the isolates as described previously [17].

Whole genome sequencing

Publicly available genomic sequences were obtained from GenBank (accessed 10/10/2011). Forty-three whole genomic sequences for *Salmonella* serovar Typhimurium ST313 from Malawi were downloaded from the European Nucleotide Archive, [18] (accessed 14/02/2012) [19]. In addition, six strains were sequenced by Illumina GAIIx genome analyzer (Illumina, Inc., San Diego, CA). Raw sequence data have been submitted to the European Nucleotide Archive [18] under accession no. ERP002011. The raw Illumina data were assembled using the pipeline available on the Center for Genomic Epidemiology [20] which is based on Velvet, algorithms for de novo short reads assembly [21]. For a complete list of genomic sequence data refer to the supplementary Table 1.

Identification of Single Nucleotide Polymorphisms

Paired-end reads were aligned against the reference genome using Burrows-Wheeler Aligner (BWA) [22]; SNPs were identified using SAMtools [22] and bedtools [23]. The informative SNPs required a minimum coverage of 20X and a minimum distance of 20 bps between each SNP. We used the *S.* Typhimurium ST313 D23580 genome as a reference (National Center for Biotechnology Information, accession: FN424405) in the analysis. Additionally, we included in the analysis only 16 out of 43 previously sequenced and deposited genomes [37 originating from Malawian patients due to a potential low quality of the raw reads in the remaining genomes]. The analysis, excluding indels, resulted in 44 SNPs among the six whole genome sequences and a total of 92 from all genomes including those originating from Malawian patients. The informative SNPs from each isolate were concatenated to a single alignment according to the position of the reference

Table 1: Epidemiological and clinical information for the 18 *Salmonella* serovar Typhimurium ST313 from DRC and Nigeria.

ID	Country	Year	Isolation site (City)	Age (Years)	Sex (F / M)	Symptoms	Specimen	Hospitalization (Days)	Outcome
02-03/008	Republic of Congo [d]	2002-2003	Lwiro	4	F	Diarrhea / stool submitted 4 days prior to onset of septicemia, kwashiorkor, [b]	Stool[a]	4	Fatal
02-03/002	Republic of Congo [d]	2002-2003	Lwiro	-	-	-	Blood	-	-
05/157	Republic of Congo [d]	2005	Lwiro	1	F	Septicemia, meningococcemia,no malnutrition, [b]	Stool	5	Fatal
05/159	Republic of Congo [d]	2005	Lwiro	1	M	Diarrhea / stool submitted 5 days prior to onset septicemia, kwashiorkor, [b]	Stool[a]	16	Absconded [c]
06/001	Republic of Congo [d]	2006	Lwiro	3	M	Pneumonia, no malnutrition	Blood	5	Recovered
06/013	Republic of Congo [d]	2006	Lwiro	1	M	Diarrhea, septicemia, no malnutrition, [b]	Stool[a]	10	Recovered
07/015	Republic of Congo [d]	2007	Lwiro	1	M	Diarrhea, septicemia, no malnutrition, [b]	Stool[a]	10	Recovered
07/022	Republic of Congo [d]	2007	Lwiro	1	F	Pneumococcal meningitis, diarrhea/ stool submitted 3 days prior to blood sample,no malnutrition, [b]	Stool	20	Recovered
02-03/028	Republic of Congo [d]	2002-2003	Lwiro	1	M	Malaria, diarrhea, no malnutrition, [b]	Stool[a]	21	Recovered
05/102	Republic of Congo [d]	2005	Lwiro	-	-	-	Blood	-	-
07/006	Republic of Congo [d]	2007	Lwiro	-	-	-	-	-	-
BL25	Nigeria	2004-2005	Ibadan	1	M	Diarrhea / fever	Blood	-	-
BL39	Nigeria	2004-2005	Ibadan	1	M	Gastroenteritis secondary to malaria	Blood	-	-
BL67	Nigeria	2004-2005	Ibadan	1.5	M	Convulsion / Severe malaria	Blood	-	-
BL	Nigeria	2004-2005	Ibadan	2.5	M	Severe malaria / anemia	Blood	-	-
BL95	Nigeria	2004-2005	Ibadan	1	F	Septicemia	Blood	-	-
BL51	Nigeria	2004-2005	Ibadan	2.5	M	Septicemia / malaria / anemia / malnutrition	Blood	-	-
F86	Nigeria	2004-2005	Ibadan	7	F	Typhoid fever	Blood	-	-

(-) no data; (a) Only stool specimen was submitted to the hospital laboratory; (b) Antimalaria treatment with quinine prior to submitting blood samples; (c) left hospital due to war related insecurities; (d) isolates included [13]; (e) isolates included [15]

genome by Perl script. Subsequently, multiple alignments were performed using MUSCLE from MEGA5 [24]. A parsimony tree was generated based on MEGA5 via maximum parsimony method [24]. The tree was evaluated for the support of the notes by bootstrap analyses with 10,000 replicates. The approximation of dN/dS ratio has been calculated by dividing the sum of non-synonymous SNPs with the sum of synonymous SNPs on the protein-coding sequences of the reference genome [25].

Resistance gene database

The web-server ResFinder [26, 27] was used to identify acquired antimicrobial resistance genes with a selected threshold equal to 95% identity and results were compared with phenotypic antimicrobial susceptibility testing results.

Identification of core genes

A set of 119 genomes (Supplementary Table 1) from NCBI and the collection of assembled genomes were subjected to gene finding using Prodigal [28]. All predicted genes were aligned all-against-all at the amino acid level using BLASTP [29] and further grouped into gene clusters using MCL [30].

Defining gene islands

Identification of variable gene islands was initially performed only from the complete genomes but with plasmids excluded. Islands were defined as containing at least ten non-core genes within a region of DNA no larger than 5,000 bp. This produced 1,305 individual islands, many of which contained the same genes present in different bacterial isolates. Homology reduction was therefore performed using BLASTN alignments of the islands, all-against-all [29].

BLASTN tends to break up long imperfect matches. To avoid this, an overall identity score for every sequence pair was calculated by summarizing the identity from each individual, non-overlapping, hit. Any island pair having more than an overall identity of 70% was considered a variation of the same island and was therefore pruned using the second algorithm of Hobohm *et al*. This eliminates homology but preservs the maximum size of the data set [31]. The resulting set of 205 unique islands was then aligned with BLASTN against all genomes. At this stage we also included the draft genomes. Similar to the alignment for homology reduction, the resulting identity for each island against a given genome was calculated as the sum of the identity for each non-overlapping hit. This provided a distance matrix expressing a percent identity for each island against each genome. This matrix was pruned by removing islands found only in a single genome, or in all examples of subspecies enterica.

This produced a total set of 145 gene islands. A pruned matrix consisting of 145 gene islands in 118 bacterial genomes was clustered in both dimensions and rendered in a heatmap using the R software [32].

Analysis of the putative virulent determinant (ST313-TD)

PCR amplification of a 924 bp long putative virulent determinant (ST313-TD) was performed on a global collection of 177 *Salmonella* isolates. The isolates belonged to *Salmonella enterica* subsp. enterica (n = 170), *Salmonella enterica* subsp. salamae (n = 1), *Salmonella enterica* subsp. *arizonae* (n = 1), *Salmonella enterica* subsp. *diarizonae* (n = 1), and *Salmonella enterica* subsp. *houtenae* (n = 4). Among the *Salmonella enterica* subsp. *enterica* 57 different serovars were tested belonging to serogroup O:4 (n = 43), O:7 (n = 44), O:8 (n = 34), O:9 (n = 27), O:3,10 (n = 12), O:1,3,19 (n = 2), O:13 (n = 6), O:6,14 (n = 1), and O:28 (n = 1).

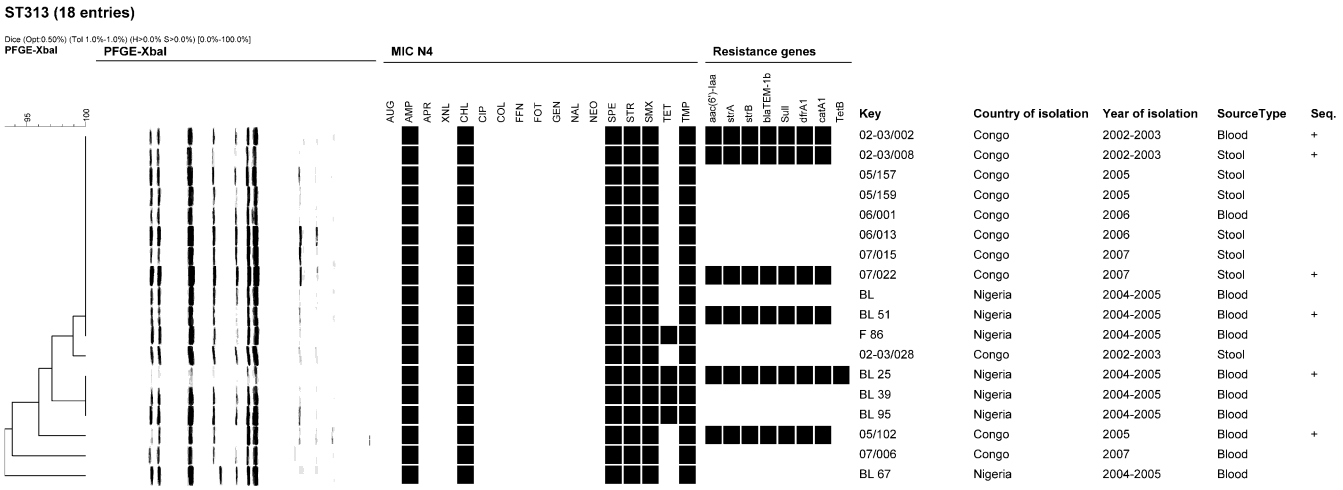
The PCR assay was designed to target 518 bp out of the 924 bp corresponding to the ST313-TD determinant. The amplifications were performed with buffer supplied by the manufacturer, 20 pmol/µl of each primer (forward primer: 5'GAA CAG TTT TAG GGC CCA A3' were paired with the reverse primer: 5'GGG AGT TCT TAA CGA TGG AA3') and 0.5 U of Amplicon Taq Polymerase (Ampliqon, Pennsylvania, United States) in a final reaction volume of 50 µl. The following amplification conditions were used: 5 min at 94°C; 25 cycles of 1 min at 94°C, 1 min at 52°C, and 1 min at 72°C; ending with one cycle of 10 min at 72°C. *S.* Typhimurium ST313 06/004 (DRC) providing an amplicon size of 518 bp was used as positive control.

Results

Epidemiological information of cases from Nigeria and DRC

The seven isolates originating from blood samples were collected in 2004 and 2005 from patient admitted to two hospitals within the city of Ibadan, Nigeria. All originated from infants or children suffering from severe illness such as malaria (laboratory confirmed), typhoid fever (diagnosed as typhoid fever, blood culture revealed *S.* Typhimurium), septicemia, diarrhea, anemia, and malnutrition (Table).

Figure 1. Dendrogram showing the genotypic relatedness of 18 *S.* Typhimurium ST 313 isolates based on *Xba*I-PFGE fingerprints and results of the antimicrobial susceptibility testing and resistance gene content.



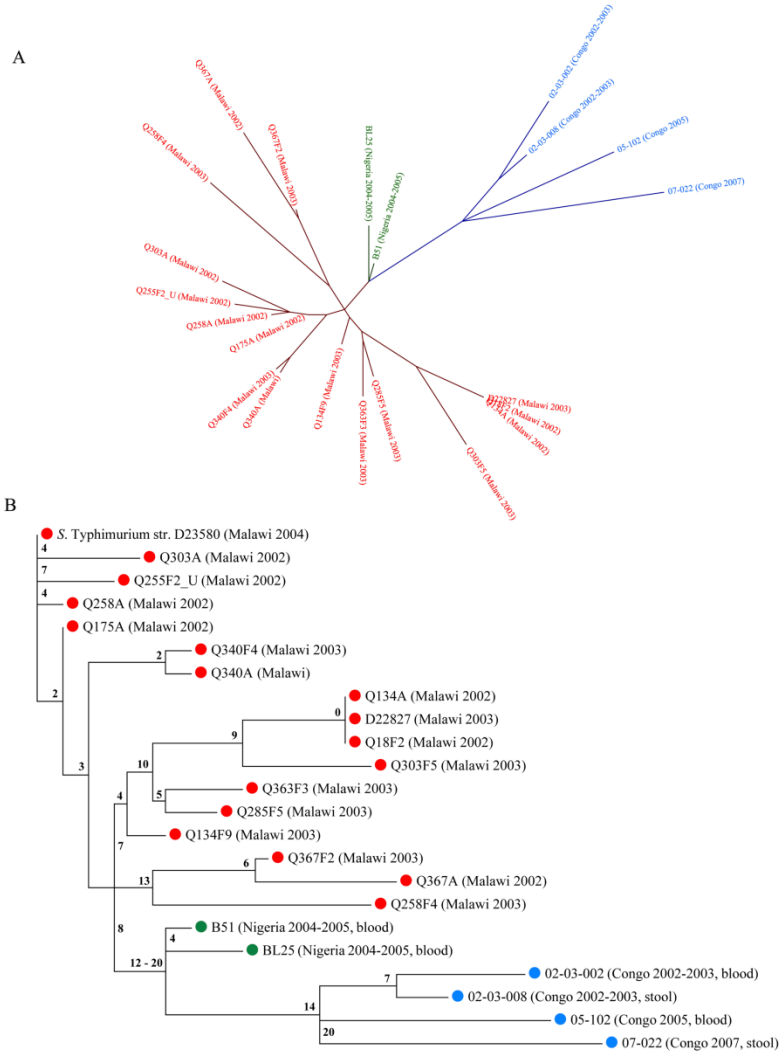
Black squares represent the presence of resistance gene or resistance to the antibiotics listed. Seq., whole genome sequencing was performed; AMP, ampicillin; AUG, amoxicillin + clavulanic acid; APR apramycin; FOT, cefotaxime; XNL, ceftiofur; CHL, chloramphenicol; CIP, ciprofloxacin; COL, colistin; FFN, florfenicol; GEN, gentamicin; NAL, nalidixic acid; NEO, neomycin; SPT, spectinomycin; STR, streptomycin; SMX, sulphamethoxazole; TET, tetracycline; and TMP, trimethoprim.

Between 2002 to 2006, 11 *Salmonella* isolates, including three from blood samples, seven from stool samples and one of unknown origin, were collected from children admitted to the Children's Hospital of Lwiro, DRC. Infections included one hospital-acquired infection, eight community-acquired infections, and two of unknown origin. Three patients had primary symptoms of uncomplicated diarrhea and five had diarrhea and septicemia. Four patients also suffered from meningococcemia, pneumonia, pneumococcal meningitis, and malaria; and two suffered from acute protein-energy malnutrition (kwashiorkor) in contrast to the six others where no sign of malnutrition was observed. Neither primary nor secondary symptoms were known for three of the patients. Seven of the patients were treated with anti-malaria medication (quinine) prior to blood sample submission. The patients were hospitalized from four to 21 days with a median of ten days (exact hospitalization times were not known for two patients). Five patients were discharged from the hospital as recovered, two died during treatment, one left the hospital prematurely due to war related insecurities, and the outcomes for three of the patients were unknown (Table 1).

Antimicrobial susceptibility testing and resistance genes

The determination of the minimum inhibitory concentration (MIC) of a series of antibacterial agents for 18 *S.* Typhimurium ST313 isolates revealed two antimicrobial resistance profiles. The most common profile (n = 14) exhibited resistance to six core antimicrobials: ampicillin, chloramphenicol, spectinomycin, streptomycin, sulfamethoxazole, and trimethoprim whereas the other profile (n = 4) showed resistance to tetracycline in addition to first resistance profile (Figure 1). All isolates were susceptible to amoxicillin + clavulanic acid, apramycin, cefotaxime, ceftiofur, ciprofloxacin, colestin, florfenicol, gentamicin, nalidixic acid, and neomycin. The genomic sequence of the six isolates were submitted to the ResFinder based on which the following resistance genes were determined: *aac(6')-Iaa* (kanamycin), *str*A, *str*B (spectinomycin and streptomycin), $bla_{\text{TEM-1b}}$ (ampicillin), *sul*1 (sulfamethoxazole), *dfr*A1 (trimethoprim), and *cat*A1 (chloramphenicol). In addition to those resistance genes, one isolate resistant to tetracycline also harbored the *tet*(B) gene (Figure 1).

Figure 2. Phylogeny of *S.* Typhimurium ST313 from Nigeria (green), DRC (blue), and Malawi (red).



The parsimony tree was generated based on 92 informative SNPs from twenty-two whole-genome sequences. (A) radial tree layout shows the overview distribution of branches. (B) rectangular tree layout implicates the number of SNPs differences marked in black numbers.

Multi Locus Sequence Typing and Pulsed-Field Gel Electrophoresis

All 18 isolates belonged to ST313 and exhibited six unique *Xba*I PFGE patterns. Two PFGE clusters with indistinguishable profiles were observed containing three and eleven isolates, respectively (Figure 1). The cluster containing the three isolates all conferred resistance to the same antimicrobials including tetracycline, were all from Nigeria between 2004 and 2005 and all came from blood samples. Ten of the 11 isolates belonging to the second cluster conferred resistance to the same antimicrobials. However, one of the eleven clustered isolates was also resistant to tetracycline (Figure 1). These 11 closely related isolates originated from Congolese and Nigerian patients from where both stool and blood samples were submitted in a period from 2002 to 2007 (Figure 1). The four isolates not in cluster one or two were all very similar (between 93% and 99% similarity to the remaining isolates). These four isolates were all cultured from both stool and blood samples originating from Congolese and Nigerian patients in 2002 to 2007.

Parsimony SNP tree

The genetic evolution of *S.* Typhimurium ST313 was examined using analysis of Next Generation Sequence (NGS) data from twenty-two isolates from

the DRC (n = 4), Nigeria (n = 2), and Malawi (n = 16). High-quality SNPs were identified by mapping Illumina paired-end reads to the completed genome of *S.* Typhimurium ST313 str. D23580. A parsimony tree was generated based on 92 informative SNPs (Figure 2).

All of the *S.* Typhimurium ST313 isolates from DRC, Nigeria, and Malawi were grouped within three distinct phylogenetic clusters based on the origin of isolation showing the clear overview of spatial evolution (Figure 2A). The recent study of Okoro *et al*. suggested that the population of *S.* Typhimurium ST313 is divided into two distinct lineages and Malawi and DRC served as potentially early transmission hubs in each lineage [33]; the isolates from Nigeria and DRC were very closely related, differing by only 12 – 20 SNPs despite the large geographically distance between the countries (Figure 2B). In this study, we observed from zero to six SNPs among the highly clonal isolates from Malawi [19]. A similarly low number of SNPs was observed within the Nigerian branch, where only four SNPs separated the isolates; BL25 and B51. The isolates from DRC were subdivided according to the isolation time; 2002 – 2003 and 2005 – 2007, into two closely related clusters differing by 14 SNPs (Figure 2B). Interestingly, there were only seven SNPs separating the Congolese isolates; 02-03-002 and 02-03-008, despite these being isolated from blood and stool samples, respectively.

The non-synonymous SNP/synonymous SNP ratio (dN/dS) is a measurement of stabilizing selection [34]. A ratio of 1 is expected in the absence of selection, a low ratio (dN/dS<1) indicates stabilizing selection, while a high ratio (dN/dS>1) indicates positive selection [35]. The genetic evolution among the Nigerian and Congolese isolates seems to be under positive selection as the dN/dS ratio is 1.7 (Supplementary Table 2). Nonetheless, the importance of these finding needs to be confirmed by additional analysis on the larger set of *S.* Typhimurium ST313 genomes.

Comparative genomic analysis

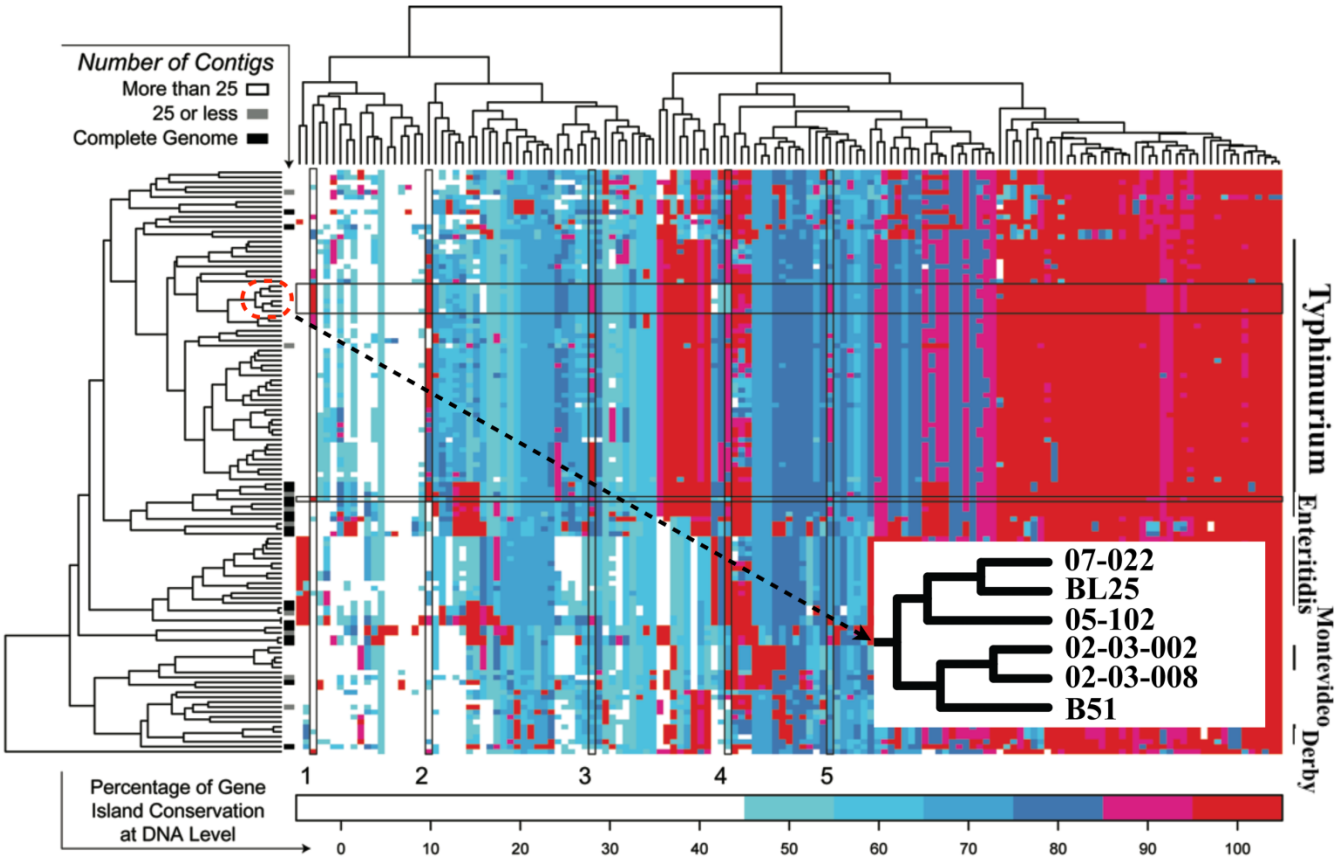
We identified gene islands in all completely assembled genomes, defined as chromosomal regions containing at least 10 non-core genes in close proximity. These islands were aligned against all genomes (including the fragmented genomes) and alignment identities were clustered into a heatmap (Figure 3 and supplementary figure for the complete isolate name). The heatmap S1 clearly illustrates that several gene islands are ubiquitous within certain serotypes but rarely found in isolates of other serotypes. This is especially evident for *S.* Typhimurium due to the amount of published data for this serovar, but can also be seen for *Salmonella* serovars Enteritidis, Derby, and Montevideo.

The *S.* Typhimurium ST313 genomes are marked by horizontal rectangles in Figure 3 and all clustered together with the exception of one genome, *S.* Typhimurium ST313 D23580. The genome of D23580 is instead clustered with a group of complete (or near-complete) genomes. The reason is that in general, complete or near-complete genomes in Figure 3 typically share certain islands, which are rarely found in the draft sequences. This is an indication that certain regions of the genome are harder to assemble, and thus only fragments of these were found in the draft genomes. As such, this suggests that an assembly bias exists in Figure 3, and one should take care to reference the number of contigs when interpreting the illustration.

In Figure 3, *S.* Typhimurium ST313 genomes stand out compared to the other *Salmonella* genomes in five columns marked with numbered vertical rectangles. These regions were extracted and analyzed for functional profiles through Gene Ontology (GO) terms [36] using the Blast2GO suite [37] (data not shown). Region 1 contained 47 genes, of which the vast majority of genes were prophage related. Region 2 contained 45 genes and was dominated by many hypothetical or putative proteins. A few phage-related genes were also found, particularly encoding bacteriophage tail proteins. Region 3 contained 157 genes, including a disproportionate number related to nuclease activities DNA binding, DNA metabolism and recombination. Region 3 also included several known virulence factors, such as type VI systems, and many more putative virulence genes. Region 4 was different from the others in that it was found conserved to about 90-100% in most of the *S.* Typhimurium investigated except for the *S.* Typhimurium ST313 genomes, where the conservation was 50% or less. The region contained only 20 genes, but several of those were known virulence factors, which were interestingly not found in *S.* Typhimurium ST313.

Region 5 contains 123 genes including a disproportionate number related to enzymatic and metabolic processes as well as fsr a gene (locus_tag SeD_A0541), which encodes resistance to fosmidomycin, not previously known virulence factors were found. The region 5 carried a presumptive

Figure 3. Heatmap showing the distribution of gene islands across *Salmonella* genomes.



The heatmap color indicates the percentage of conservation at the DNA level of 145 gene islands (horizontal axis) in each of the 118 strains under investigation (vertical axis). The location of the ST313 strains are indicated by two black horizontal boxes; the singleton is the D23580 strain from GenBank, while the strains sequenced as part of this work all cluster together. Likewise, the five island regions discussed in the text are marked with vertical boxes. The ST313-GI harboring ST313-TD gene is located in region 5. Also indicated is whether each genome was sequenced to a complete contigous genome, nearly complete (<25 contigs) or should be considered a draft assembly (>25 contigs). Certain islands are typically only found conserved in complete or near-complete genomes, which suggests that the genome quality provides a degree of bias. The representation diplays genomic DNA only, although it cannot be ruled out that some of the draft genomes might contain plasmid DNA not annotated as such.

uncharacterized 17.7 kb gene island, (ST313-GI), present in the *S.* Typhimurium ST313 sequenced strains including D23580. The gene island ST313-GI is integrated between a tRNA gene at the 5´end encoding a tRNA for threonine and two tRNAs genes at the 3´end encoding tRNAs for arsenic and threonine. ST313-GI, encodes several predicted prophage proteins and other phage-related factors such as an integrase, an excisionase, replication and regulatory proteins, an anti-restriction, repressor, anti-repressor and anti-terminator proteins and a transcriptional activator. A putative virulence gene of 924 bp tentatively named ST313-TD was found within ST313-GI which showed high similarity (100% similarity) to a gene annotated in the genomes of *S.* Typhimurium ST313 D23580 (accession no. FN424405 gene position 377825..378748) and *S.* Dublin CT_02021853 (accession no. CP001143) and 3246 (accession no. CM001151) from NCBI, exclusively. In the *S.* Dublin strains CT_02021853 and 3246, the determinant was contained in a region of app. 6.8 kb showing high similarity (99%) to a part of the ST313-GI island in *S.* Typhimurium ST313.

Distribution of the putative virulence determinant (ST313-TD) among different Salmonella serovars

The presence of ST313-TD was detected in nine isolates assigned to the following serovars; six (n = 6; 100%) *S.* Dublin originating from Denmark, Taiwan, Thailand, Columbia, and Nigeria; one (n = 2, 50%)

Salmonella serovar Bredeney of unknown origin; one (n = 6, 17%) *Salmonella* serovar Saintpaul from Italy, and one (n = 9, 11%) *Salmonella* serovars Kentucky from Nigeria. None of the five non-ST313 *S.* Typhimurium isolates included in the collection in this study were positive.

Discussion

The high incidence of invasive salmonellosis in sub-Saharan Africa is a major human health problem [10, 11, 19]. It has been suggested that this is mainly due to a large susceptible population suffering from a range of maladies such as: malaria, diarrhea, septicemia, anemia, HIV, malnutrition, and fever. In particular, malaria has been suspected to increase the risk of invasive non-typhoidal *Salmonella* (NTS) infections and may contribute to the seasonality of NTS disease [12]. Recently, it has also been suggested that a specific highly invasive sub-type of *Salmonella*; *S.* Typhimurium ST313 might be causing the majority of the infections [11]. The study from Okoro *et al*. found that *S.* Typhimurium ST313 could be split into two lineages and provided evidence that these arose independently around 1960 and 1977 from Malawi and DRC, respectively [33]. Nonetheless, it remains unknown whether *S.* Typhimurium ST313 causes mainly invasive diseases in humans or whether it is also a major cause of gastroenteritis.

A recent study of *S.* Typhimurium ST313 from Malawi revealed that isolates were highly clonal, differing only by a maximum of six SNPs [19]. In the study reported here, we also found that isolates were highly clonal despite large temporal and spatial distances. This confirms previous speculations that *S.* Typhimurium ST313 is widespread in sub-Saharan African countries and is most likely the consequence of a recent emergence, possibly disseminated by the migration of animals and humans or by continental trade.

Until this report, no evidence of *S.* Typhimurium ST313 being associated with gastroenteritis in sub-Saharan Africa has been published [11]. In this study, we found a high degree of similarity between isolates originating from blood and stool samples in DRC; this was supported by the clinical data. These data indicate that the *S.* Typhimurium ST313 isolates from DRC, and mostly likely all *S.* Typhimurium ST313, can cause gastroenteritis in humans like other NTS serotypes. However, the dominance of ST313 amongst isolates from blood is puzzling.

Interestingly, we found a putative virulence determinant (ST313-TD) unique to *S.* Typhimurium ST313, but also present in global isolates of *S.* Dublin. It is well known that *S.* Dublin is highly invasive in cattle and humans when compared to other NTS [38]. In humans, *S.* Dublin often results in bacteraemia with severe disease and high mortality [39]. In *S.* Typhimurium ST313, the hypothetical virulence determinant (ST313-TD) is included within a 17.7 kb region that may represent a novel pathogenicity island (ST313-GI). We also observed ST313-GI integrated between tRNA genes. Insertion of phages and pathogenicity islands within or in close proximity to tRNAs is very common [1, 8, 15, 40]. It is well known that most of the phages described in *Salmonella*, are involved in virulence [41] and several other *Salmonella* gene islands also encode pathogenic functions, e.g. *Salmonella* pathogenicity islands (SPIs) that encode multiple effectors required for *Salmonella* virulence [42].

The putative virulent determinant (ST313-TD) was present in a few isolates other than *S.* Typhimurium ST313 and *S.* Dublin. Unfortunately, we only have sparse information on those isolates complicating any speculation on why those few isolates represent serotypes which also harbored the genetic determinant. We know that the *S.* Kentucky isolate from Nigeria harbored the determinant whereas the *S.* Kentucky isolate from poultry isolated in the United States (US) did not. Two different lineages exist of *S.* Kentucky; one is present in the US affecting only poultry [43] where the other one is a highly pathogenic lineage causing human infections in Africa and Europe [44]. This could explained why only the Nigerian isolate harbored the determinant as we speculate if this isolate belonged to the pathogenic lineage in contrast to the isolates from the US which might belong to the poultry lineage. Due to the low frequency of this putative virulence determinant in other serovars, our data suggest that the impact on acquisition of the gene island needs to be further investigated.

In this study, the 18 *S.* Typhimurium ST313 isolates from DRC and Nigeria conferred resistance to the exact same panel of antimicrobials; ampicillin, chloramphenicol, kanamycin, streptomycin, sulfamethoxazole, and trimethoprim as observed in isolates from Malawi and Kenya [11]. In addition to the antimicrobial resistance pattern observed in Malawi and Kenya, we also revealed that some of the isolates from Nigeria conferred resistance to tetracycline harbored by the *tet*B gene. The antimicrobial resistance patterns and the corresponding genes are not unique to *S.* Typhimurium

ST313 but frequently found among other NTS and gram negative *Enterobacteriaceae* [45, 46].

This study is limited in terms of the low number of isolates investigated and the lack of clinical details such as HIV status of the patients, co-infections and local population data however, the authors believe that the data raise important questions about the epidemiology of *S.* Typhimurium ST313.

Conclusions

In this study, we showed in a limited number of isolates that *S.* Typhimurium ST313 is a prevalent sequence-type causing gastrointestinal diseases and septicemia in patients from Nigeria and DRC. We found three distinct phylogenetic clusters based on the origin of isolation suggesting some spatial evolution. Comparative genomics showed an interesting putative virulent fragment (ST313-TD) unique to *S.* Typhimurium ST313 and invasive *S.* Dublin. Further studies will have to be conducted on a large number of isolates and to determine the reservoirs and transmission routes of *S.* Typhimurium ST313 in sub-Saharan Africa.

Acknowledgements

This work is dedicated to the nursing staff of the Children's Hospital of Lwiro. We are grateful to Zozo Nyarukweba for outstanding technical assistance. This work was supported by the Center for Genomic Epidemiology [20] at the Technical University of Denmark funded by grant 09-067103/DSF from the Danish Council for Strategic Research, as well as by grant 3304-FVFP-08 from the Danish Food Industry Agency and by the World Health Organization Global Foodborne Infections Network [47] and the Centre Scientifique et Médical de 1'Université Libre de Bruxelles pour ses Activités de Coopération (CEMUBAC).

References

1. Hohmann EL (2001) Nontyphoidal salmonellosis. Clin Infect Dis 32: 263-269
2. Menezes GA, Khan MA, Harish BN, Parija SC, Goessens W, Vidyalakshmi K, Baliga S, Hays JP (2010) Molecular characterization of antimicrobial resistance in non-typhoidal salmonellae associated with systemic manifestations from India. J Med Microbiol 59: 1477-1483
3. Schlundt J, Toyofuku H, Jansen J, Herbst SA (2004) Emerging food-borne zoonoses. Rev Sci Tech 23: 513-533
4. Voetsch AC, Van Gilder TJ, Angulo FJ, Farley MM, Shallow S, Marcus R, Cieslak PR, Deneen VC, Tauxe RV (2004) FoodNet estimate of the burden of illness caused by nontyphoidal Salmonella infections in the United States. Clin Infect Dis 38 Suppl 3: S127-S134
5. 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. Clin Infect Dis 50: 882-889
6. Mead PS, Slutsker L, Dietz V, McCaig LF, Bresee JS, Shapiro C, Griffin PM, Tauxe RV (1999) Food-related illness and death in the United States. Emerg Infect Dis 5: 607-625
7. Flint JA, Van Duynhoven YT, Angulo FJ, DeLong SM, Braun P, Kirk M, Scallan E, Fitzgerald M, Adak GK, Sockett P, Ellis A, Hall G, Gargouri N, Walke H, Braam P (2005) Estimating the burden of acute gastroenteritis, foodborne disease, and pathogens commonly transmitted by food: an international review. Clin Infect Dis 41: 698-704
8. Brent AJ, Oundo JO, Mwangi I, Ochola L, Lowe B, Berkley JA (2006) Salmonella bacteremia in Kenyan children. Pediatr Infect Dis J 25: 230-236
9. Jones TF, Ingram LA, Cieslak PR, Vugia DJ, Tobin-D'Angelo M, Hurd S, Medus C, Cronquist A, Angulo FJ (2008) Salmonellosis outcomes differ substantially by serotype. J Infect Dis 198: 109-114
10. Graham SM (2010) Nontyphoidal salmonellosis in Africa. Curr Opin Infect Dis 23:409-414
11. Kingsley RA, Msefula CL, Thomson NR, Kariuki S, Holt KE, Gordon MA, Harris D, Clarke L, Whitehead S, Sangal V, Marsh K, Achtman M, Molyneux ME, Cormican M, Parkhill J, Maclennan CA, Heyderman RS, Dougan G (2009) Epidemic multiple drug resistant Salmonella Typhimurium causing invasive disease in sub-Saharan Africa have a distinct genotype. Genome Res 19: 2279-2287
12. Morpeth SC, Ramadhani HO, Crump JA (2009) Invasive non-Typhi Salmonella disease in Africa. Clin Infect Dis 49: 606-611
13. Vandenberg O, Nyarukweba DZ, Ndeba PM, Hendriksen RS, Barzilay EJ, Schirvel C, Bisimwa BB, Collard JM, Aidara KA, Aarestrup FM (2010) Microbiologic and clinical features of Salmonella species isolated from bacteremic children in eastern Democratic Republic of Congo. Pediatr Infect Dis J 29: 504-510
14. Wain J, Keddy KH, Hendriksen RS, Rubino S (2013) Using next generation sequencing to tackle non-typhoidal Salmonella infections. J Infect Dev Ctries 7: 1-5
15. Fashae K, Ogunsola F, Aarestrup FM, Hendriksen RS (2010) Antimicrobial susceptibility and serovars of Salmonella from chickens and humans in Ibadan, Nigeria. J Infect Dev Ctries 4: 484-494
16. Hendriksen RS, Le HS, Bortolaia V, Pulsrikarn C, Nielsen EM, Pornruangmong S, Chaichana P, Svendsen CA, Weill FX, Aarestrup FM (2012) Characterization of isolates of Salmonella enterica serovar Stanley, a serovar endemic to Asia and associated with travel. J Clin Microbiol 50: 709-720
17. Achtman M, Wain J, Weill FX, Nair S, Zhou Z, Sangal V, Krauland MG, Hale JL, Harbottle H, Uesbeck A, Dougan G, Harrison LH, Brisse S (2012) Multilocus sequence typing as a replacement for serotyping in Salmonella enterica. PLoS Pathog 8:e1002776
18. European Nucleotide Archive (2012) [http://www.ebi.ac.uk/ena/]. European Bioinformatics Institute
19. Okoro CK, Kingsley RA, Quail MA, Kankwatira AM, Feasey NA, Parkhill J, Dougan G, Gordon MA (2012) High-resolution single nucleotide polymorphism analysis distinguishes recrudescence and reinfection in recurrent invasive nontyphoidal Salmonella Typhimurium disease. Clin Infect Dis 54: 955-963

20. Center for Genomic Epidemiology (2012) [www.genomicepidemiology.org]. DTU
21. Zerbino DR, Birney E (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res 18: 821-829
22. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics 25: 2078-2079
23. Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26: 841-842
24. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28: 2731-2739
25. Yi S (2007) Synonymous and Nonsynonymous Rates. eLS
26. ResFinder (2012) [www.cbs.dtu.dk/services/ResFinder]. DTU
27. Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, Aarestrup FM, Larsen MV (2012) Identification of acquired antimicrobial resistance genes. J Antimicrob Chemother 67: 2640-2644
28. Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ (2010) Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11: 119
29. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403-410
30. Enright AJ, Van DS, Ouzounis CA (2002) An efficient algorithm for large-scale detection of protein families. Nucleic Acids Res 30: 1575-1584
31. Hobohm U, Scharf M, Schneider R, Sander C (1992) Selection of representative protein data sets. Protein Sci 1: 409-417
32. R Development Core Team (2008) R: A language and enviroment for statistical computing. R Foundation for Statistical Computing
33. Okoro CK, Kingsley RA, Connor TR, Harris SR, Parry CM, Al-Mashhadani MN, Kariuki S, Msefula CL, Gordon MA, De PE, Wain J, Heyderman RS, Obaro S, Alonso PL, Mandomando I, Maclennan CA, Tapia MD, Levine MM, Tennant SM, Parkhill J, Dougan G (2012) Intracontinental spread of human invasive Salmonella Typhimurium pathovariants in sub-Saharan Africa. Nat Genet 10
34. Holt KE, Parkhill J, Mazzoni CJ, Roumagnac P, Weill FX, Goodhead I, Rance R, Baker S, Maskell DJ, Wain J, Dolecek C, Achtman M, Dougan G (2008) High-throughput sequencing provides insights into genome variation and evolution in Salmonella Typhi. Nat Genet 40:987-993
35. Rocha EP, Smith JM, Hurst LD, Holden MT, Cooper JE, Smith NH, Feil EJ (2006) Comparisons of dN/dS are time dependent for closely related bacterial genomes. J Theor Biol 239: 226-235
36. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 25: 25-29
37. Gotz S, Garcia-Gomez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talon M, Dopazo J, Conesa A (2008) High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Res 36: 3420-3435
38. Helms M, Vastrup P, Gerner-Smidt P, Molbak K (2003) Short and long term mortality associated with foodborne bacterial gastrointestinal infections: registry based study. BMJ 326:357
39. Betancor L, Yim L, Martinez A, Fookes M, Sasias S, Schelotto F, Thomson N, Maskell D, Chabalgoity JA (2012) Genomic Comparison of the Closely Related Salmonella enterica serovars Enteritidis and Dublin. Open Microbiol J 6:5-13. Epub;%2012 Feb 10: 5-13
40. Mmolawa PT, Schmieger H, Heuzenroeder MW (2003) Bacteriophage ST64B, a genetic mosaic of genes from diverse sources isolated from Salmonella enterica serovar typhimurium DT 64. J Bacteriol 185: 6481-6485
41. Lemire S, Figueroa-Bossi N, Bossi L (2011) Bacteriophage crosstalk: coordination of prophage induction by trans-acting antirepressors. PLoS Genet 7:e1002149
42. Srikanth CV, Mercado-Lubo R, Hallstrom K, McCormick BA (2011) Salmonella effector proteins and host-cell responses. Cell Mol Life Sci 68: 3687-3697
43. Fricke WF, McDermott PF, Mammel MK, Zhao S, Johnson TJ, Rasko DA, Fedorka-Cray PJ, Pedroso A, Whichard JM, Leclerc JE, White DG, Cebula TA, Ravel J (2009) Antimicrobial resistance-conferring plasmids with similarity to virulence plasmids from avian pathogenic Escherichia coli strains in Salmonella enterica serovar Kentucky isolates from poultry. Appl Environ Microbiol 75: 5963-5971
44. Le HS, Hendriksen RS, Doublet B, Fisher I, Nielsen EM, Whichard JM, Bouchrif B, Fashae K, Granier SA, Jourdan-Da SN, Cloeckaert A, Threlfall EJ, Angulo FJ, Aarestrup FM, Wain J, Weill FX (2011) International spread of an epidemic population of Salmonella enterica serotype Kentucky ST198 resistant to ciprofloxacin. J Infect Dis 204: 675-684
45. Sirichote P, Hasman H, Pulsrikarn C, Schonheyder HC, Samulioniene J, Pornruangmong S, Bangtrakulnonth A, Aarestrup FM, Hendriksen RS (2010) Molecular Characterization of Extended-Spectrum Cephalosporinase-Producing Salmonella enterica Serovar Choleraesuis Isolates from Patients in Thailand and Denmark. J Clin Microbiol 48: 883-888
46. Hendriksen RS, Bangtrakulnonth A, Pulsrikarn C, Pornreongwong S, Hasman H, Song SW, Aarestrup FM (2008) Antimicrobial resistance and molecular epidemiology of Salmonella Rissen from animals, food products, and patients in Thailand and Denmark. Foodborne Pathog Dis 5: 605-619
47. World Health Organization Global Foodborne Infections Network (2012) [www.who.int/gfn]. WHO

Corresponding author

Rene S. Hendriksen
Division of Bacterial Genomics and Epidemiology, WHO Collaborating Center for Antimicrobial Resistance in Food borne Pathogens and European Union Reference Laboratory for Antimicrobial Resistance, National Food Institute, Technical University of Denmark, Kemitorvet, Building 204, DK-2800 Kgs. Lyngby, Denmark
Email: rshe@food.dtu.dk

Conflict of interests: No conflict of interests is declared.

Supplementary items

Supplementary Table 1: The set of 119 genomes from NCBI and the collection of assembled genomes.

Strain	# Genes	Accession
S. Agona str. SL483	4508	CP001137,CP001138
S. arizonae str. RSK2980	4278	CP000880
S. Choleraesuis str. A50	4596	CM001062,CM001063
S. Choleraesuis str. SC-B67	4789	AY509003,AY509004,AE017220
S. Derby 0807H16988	4707	-
S. Derby 0807T13477	4627	-
S. Derby 0810F45685	4426	-
S. Derby 0810W40256	4705	-
S. Derby 0903F3864	4768	-
S. Dublin str. 3246	4696	CM001151,CM001152
S. Dublin str. CT_02021853	4682	CP001143,CP001144
S. Enteritidis 0905R11565	4439	-
S. Enteritidis 0905R11609	4437	-
S. Enteritidis 0905R11615	4424	-
S. Enteritidis 0907R11860	4513	-
S. Enteritidis 0909R12018	4451	-
S. Enteritidis 0910R12234	4465	-
S. Enteritidis 0910R12287	4395	-
S. Enteritidis 14405	4628	-
S. Enteritidis 7310	4440	-
S. Enteritidis 909R12091	4459	-
S. Enteritidis D10	4671	-
S. Enteritidis HN-GSS-2007-013	4496	-
S. Enteritidis str. P125109	4363	AM933172
S. Enteritidis TZH-237	4590	-
S. Gallinarum str. 287/91	4466	AM933173
S. Gallinarum str. 9	4579	CM001153,CM001154
S. Hadar 9973011-2	4369	-
S. Hadar str. RI_05P066	4487	ABFG01000001-ABFG01000050
S. Heidelberg 05126-20353	4696	-
S. Heidelberg str. SL476	4679	CP001118,CP001119,CP001120
S. Heidelberg str. SL486	4432	ABEL01000001-ABEL01000048
S. Infantis 9961191-4	4456	-
S. Javiana str. GA_MM04042433	4221	ABEH02000001-ABEH02000019
S. Kentucky 2007-51-7625-1	4365	-
S. Kentucky str. CDC 191	4383	ABEI01000001-ABEI01000053
S. Kentucky str. CVM29188	4442	ABAK02000001
S. Montevideo str. 19N	4364	AESV01000001-AESV01000060
S. Montevideo str. 315996572	4364	AESH01000001-AESH01000050
S. Montevideo str. 495297-3	4342	AESJ01000001-AESJ01000039
S. Montevideo str. 515920-2	4342	AESM01000001-AESM01000034
S. Montevideo str. CASC_09SCPH15965	4512	AESQ01000001-AESQ01000069
S. Newport 98-24475-1	4678	-
S. Newport str. SL254	4710	CP000604,CP001112,CP001113
S. Newport str. SL317	4720	ABEW01000001-ABEW01000063
S. Paratyphi_A str. AKU_12601	4351	FM200053
S. Paratyphi_A str. ATCC 9150	4348	CP000026
S. Paratyphi_B str. SPB7	4561	CP000886
S. Paratyphi_C str. RKS4594	4690	CP000857,CP000858
S. Rissen 7622232-1	4614	-
S. Saintpaul 98-24527-3	4570	-
S. Saintpaul str. SARA23	4350	ABAM02000001,ABAM02000002
S. Saintpaul str. SARA29	4757	ABAN01000001-ABAN01000182
S. Schwarzengrund 99-22292-3	4346	-
S. Schwarzengrund str. CVM19633	4552	CP001125,CP001126,CP001127
S. Schwarzengrund str. SL480	4547	ABEJ01000001-ABEJ01000067
S. Tennessee str. CDC07-0191	4546	NZ_ACBF01000001-NZ_ACBF01000094
S. Typhi str. CT18	5065	AL513382,AL513383,AL513384
S. Typhi str. Ty2	4632	AE014613
S. Typhimurium 000419417	4713	-
S. Typhimurium 0110T17035	4675	-

Supplementary Table 1 (*continued*): The set of 119 genomes from NCBI and the collection of assembled genomes.

Strain	# Genes	Accession
S. Typhimurium 0112F33212	4745	
S. Typhimurium 02-03-002	4713	-
S. Typhimurium 02-03-008	4734	-
S. Typhimurium 0207T6419	4715	-
S. Typhimurium 0210H31581	4800	-
S. Typhimurium 0406R5753	4829	-
S. Typhimurium 0407M287	4749	-
S. Typhimurium 0407W47858	4732	-
S. Typhimurium 0408R5930	4792	-
S. Typhimurium 0408R5960	4790	-
S. Typhimurium 0409R5985	4731	-
S. Typhimurium 04-1046	4855	-
S. Typhimurium 0505F37633	4468	-
S. Typhimurium 0507R6701	4776	-
S. Typhimurium 0508R6706	4651	-
S. Typhimurium 0508R6707	4838	-
S. Typhimurium 0508R6762	4783	-
S. Typhimurium 0508R6811	4701	-
S. Typhimurium 05-102	4735	-
S. Typhimurium 0510R6956	4776	-
S. Typhimurium 06-004	6160	-
S. Typhimurium 07-022	4738	-
S. Typhimurium 0803T57157	4707	-
S. Typhimurium 0804R9234	4610	-
S. Typhimurium 0808F31478	4635	-
S. Typhimurium 0808R10031	7533	-
S. Typhimurium 0808S61603	4686	-
S. Typhimurium 0810R10649	4660	-
S. Typhimurium 0811R10987	4712	-
S. Typhimurium 0901M16079	4691	-
S. Typhimurium 0902R11254	4694	-
S. Typhimurium 0903R11327	4687	-
S. Typhimurium 0905W16624	4626	-
S. Typhimurium 0909R12120	4886	-
S. Typhimurium 1005R12913	4709	-
S. Typhimurium 1006R12965	4686	-
S. Typhimurium 1269	4765	-
S. Typhimurium 4,[5],12:i:- str. CVM23701	4694	ABAO01000001-ABAO01000113
S. Typhimurium 7632	4771	-
S. Typhimurium 7717	4773	-
S. Typhimurium 8035	4823	-
S. Typhimurium 8110	4560	-
S. Typhimurium 98-22876-3	4845	-
S. Typhimurium B51	4848	-
S. Typhimurium BL25	4624	-
S. Typhimurium SH1467-03	4826	-
S. Typhimurium str. 14028S	4655	CP001362,CP001363
S. Typhimurium str. D23580	4707	FN424405,FN432031
S. Typhimurium str. LT2	4635	AE006468,AE006471
S. Typhimurium str. SL1344 NCTC13347	4549	FQ312003
S. Typhimurium str. T000240	4652	AP011957
S. Typhimurium str. TN061786	4626	AERV01000001-AERV01000024
S. Typhimurium str. UK-1	4469	CP002614
S. Typhimurium VGC-192	4847	-
S. Virchow 9922243-1	4559	-
S. Virchow str. SL491	4585	ABFH02000001-ABFH02000003
S. Weltevreden KA5-5	4854	-
S. Weltevreden str. HI_N05-537	4784	ABFF01000001-ABFF01000081

Supplementary Table 2: the dN/dS ratio of the six Whole-genome sequences.

ref_genome position	ref_genome_sequence	Salmonella-spp-02-03-002	Salmonella-spp-02-03-008	Salmonella-spp-05-102	Salmonella-spp-07-022	Salmonella-spp-BL25	Salmonella.sp-B51	SNPs
2442872	T	C [I/V]	C [I/V]	C [I/V]	C [I/V]	-	-	nSNPs
3965328	A	G	G	G	G	G	G	sSNPs
2978391	C	T	-	-	-	-	-	sSNPs
605180	G	A	A	-	-	-	-	sSNPs
3219461	T	C [E/G]	C [E/G]	C [E/G]	C [E/G]	-	-	nSNPs
75492	T	C*	C*	C*	C*	C*	C*	-
4355466	G	A	A	-	-	-	-	sSNPs
4413544	A	G [I/V]	-	-	-	-	-	nSNPs
1450047	C	T [P/S]	T [P/S]	T [P/S]	T [P/S]	T [P/S]	T [P/S]	nSNPs
271609	A	G	G	G	G	G	G	sSNPs
1109671	A	C [Y/S]	C [Y/S]	C [Y/S]	C [Y/S]	-	-	nSNPs
2776052	C	T [G/E]	T [G/E]	-	-	-	-	nSNPs
2357228	G	A [W/stop]	-	-	-	-	-	nSNPs
574134	T	C*	-	-	-	-	-	-
1681200	A	G	G	G	G	-	-	sSNPs
296601	T	C [Q/R]	C [Q/R]	C [Q/R]	C [Q/R]	-	-	nSNPs
1006786	T	C [H/R]	C [H/R]	C [H/R]	C [H/R]	C [H/R]	C [H/R]	nSNPs
2881210	A	G*	-	G*	G*	-	-	-
1630836	G	A	-	-	-	-	-	sSNPs
110852	G	-	A	-	-	-	-	sSNPs
2646601	G	-	A [V/I]	-	-	-	-	nSNPs
3081367	T	-	-	A [N/I]	-	-	-	nSNPs
4131211	G	-	-	A [A/V]	-	-	-	nSNPs
2917205	T	-	-	C [N/S]	-	-	-	nSNPs
4423375	G	-	-	A [R/C]	-	-	-	nSNPs
1669678	G	-	-	A [A/V]	-	-	-	nSNPs
2495775	G	-	-	C [A/G]	-	-	-	nSNPs
2415867	T	-	-	C	-	-	-	sSNPs
3336011	C	-	-	T	-	-	-	sSNPs
1031932	T	-	-	-	C [N/S]	-	-	nSNPs
1108469	T	-	-	-	C*	-	-	-
1563260	G	-	-	-	T	-	-	sSNPs
860050	C	-	-	-	A	-	-	sSNPs
4080695	G	-	-	-	A [T/I]	-	-	nSNPs
288816	G	-	-	-	T*	-	-	-
3319690	C	-	-	-	T*	-	-	-
4576554	C	-	-	-	A [T/N]	-	-	nSNPs
320350	G	-	-	-	A	-	-	sSNPs
1595468	G	-	-	-	A [A/V]	-	-	nSNPs
2960970	G	-	-	-	A	-	-	sSNPs
216551	A	-	-	-	-	C [I/S]	-	nSNPs
4564003	G	-	-	-	-	A [L/F]	-	nSNPs
2194196	A	-	-	-	-	T [F/I]	-	nSNPs
1223632	T	-	-	-	-	-	C [F/S]	nSNPs

* = SNPs in stop codon

sSNPs = 14
nSNPs = 24
nSNPs/sSNPs = 1.7

Fact box – Online easy to use genomic analysis tools

The Center for Genomic Epidemiology aims to develop algorithms for rapid analyses of whole genome DNA-sequences, tools for analyses and extraction of information from the sequence data in plain language reports.

- Identification; a fast and crude prediction of bacterial taxonomy by using the "Reads2Type" tool.
- Multi Locus Sequence Typing (MLST) (multiple schemes) from an assembled genome or from set of reads. Read more: "Multilocus sequence typing of total-genome-sequenced bacteria" by Larsen *et al*., *JCM* 2012.
- Identification of acquired antibiotic resistance genes from a file with sequence reads using the ResFinder tool. Read more: "Identification of acquired antimicrobial resistance genes" by Zankari *et al*., *JAC* 2012.
- Identification of plasmid replicons and typing of those by PlasmidFinder and pMLST tool.
- Creation of a SNPs phylogenetic tree from assembled genomes or sets of reads.

Read more: "snpTree--a web-server to identify and construct SNP trees from whole genome sequence data" by Leekitcharoenphon *et al., BMC Genomics 2012*.

The tools are available free of charge and online http://cge.cbs.dtu.dk/services/all.php