Europe PMC Funders Group
Author Manuscript
Nat Microbiol. Author manuscript; available in PMC 2021 June 21.

Published in final edited form as:
Nat Microbiol. 2021 March 01; 6(3): 327–338. doi:10.1038/s41564-020-00836-1.

Stepwise evolution of *Salmonella* Typhimurium ST313 causing bloodstream infection in Africa

Caisey V. Pulford[1], **Blanca M. Perez-Sepulveda**[1], **Rocío Canals**[1], **Jessica A. Bevington**[1], **Rebecca J. Bengtsson**[1], **Nicolas Wenner**[1], **Ella V. Rodwell**[1], **Benjamin Kumwenda**[2], **Xiaojun Zhu**[1], **Rebecca J. Bennett**[1], **George E. Stenhouse**[1], **P. Malaka De Silva**[1], **Hermione J. Webster**[1], **Jose A. Bengoechea**[3], **Amy Dumigan**[3], **Alicia Tran-Dien**[4], **Reenesh Prakash**[5], **Happy C. Banda**[5], **Lovemore Alufandika**[5], **Mike P. Mautanga**[5], **Arthur Bowers-Barnard**[1], **Alexandra Y. Beliavskaia**[1], **Alexander V. Predeus**[1], **Will P. M. Rowe**[1], **Alistair C. Darby**[1], **Neil Hall**[6], **François-Xavier Weill**[4], **Melita A. Gordon**[5], **Nicholas A. Feasey**[5], **Kate S. Baker**[1], **Jay C. D. Hinton**[1,*]

[1]Institute of Infection, Veterinary and Ecological Sciences, University of Liverpool, Liverpool, UK

[2]University of Malawi, College of Medicine, Blantyre, Malawi

[3]Wellcome-Wolfson Institute for Experimental Medicine, Queen's University Belfast, Belfast, UK

[4]Institut Pasteur, Unité des Bactéries Pathogènes Entériques, Paris, 75015, France

[5]Malawi-Liverpool-Wellcome Trust (MLW) Clinical Research Programme, Blantyre, Malawi

[6]Earlham Institute, Norwich Research Park, Norwich, United Kingdom

Introductory Paragraph

Bloodstream infections (BSI) caused by nontyphoidal *Salmonella* (NTS) are a major public health concern in Africa, causing ~49,600 deaths every year. The most common *Salmonella enterica* pathovariant associated with invasive NTS (iNTS) disease is *S.* Typhimurium ST313. It has been

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use: http://www.nature.com/authors/editorial_policies/license.html#terms

Correspondence: Please address all correspondence and requests for materials to Jay C.D Hinton (jay.hinton@liverpool.ac.uk).

Author Contributions

C.V.P. was involved in conceptualising and designing the overall study and performed the majority of phylogenetic, genomic and statistical analyses. B.M.P-S was involved in conceptualising, designing and performing phenotypic characterisation of isolates. R.C contributed to conceptualisation of pseudogene analysis and constructed S. Typhimurium 4/74 Δ*lpxO*::*aph* mutant for mass spectrometry analysis by λ Red recombineering. R.P, H.C.B, L.A, M.P.M and A.B-B assisted in collection of isolates from Malawi-Liverpool-Wellcome Clinical Research Center. A.T-D assisted in the collection of isolates from Institut Pasteur. R.J.Bengtsson and W.M.P.R contributed phylogenetic and genomic analysis and supported troubleshooting. R.J.Bennett and G.E.S contributed to BEAST analysis and bioinformatic troubleshooting. A.Y.B, A.V.P and A.C.P completed long read sequencing and assembly of ST313 L3 reference strain BKQZM9. B.K contributed to analysis of contemporary whole genome sequences. J.A.B performed phenotypic characterisation of antimicrobial resistance profiles. N.W performed phenotypic characterisation of the pAnkS encoded LsoA/LsoB toxin-antitoxin proteins. J.A.B and A.D did lipid A analysis by mass spectrometry. E.V.R performed phenotypic confirmation of plasmid and prophage repertoire. H.J.W and X.Z performed phenotypic characterisation of isolates. P.M.D.S provided support in reviewing and interpreting results. N.H sequenced isolates as part of the 10,000 *Salmonella* genomes project. F-X.W, M.A.G, N.A.F, K.S.B and J.C.D.H were involved in conceptualising and designing the overall study. K.S.B and J.C.D.H additionally supervised the entire study.

Conflicts of interest

RC was employed by the University of Liverpool at the time of the study and is now an employee of the GSK group of companies.

proposed that antimicrobial resistance (AMR) and genome degradation contributed to the success of ST313 lineages in Africa, but the evolutionary trajectory of such changes was unclear.

To define the evolutionary dynamics of ST313 we sub-sampled from two comprehensive collections of *Salmonella* isolates from African patients with BSI, spanning 1966 to 2018. The resulting 680 genome sequences led to the discovery of a pan-susceptible ST313 lineage (ST313 L3) which emerged in Malawi in 2016 and is closely related to ST313 variants that cause gastrointestinal disease in the United Kingdom and Brazil. Genomic analysis revealed degradation events in important virulence genes in ST313 L3, which had not occurred in other ST313 lineages. Despite only arising recently in the clinic, ST313 L3 is a phylogenetic intermediate between ST313 L1 and L2 with a characteristic accessory genome. Our in-depth genotypic and phenotypic characterisation identifies the crucial loss-of-function genetic events that occurred during the stepwise evolution of invasive *S.* Typhimurium across Africa.

Introduction

Over the past decade, bloodstream infections (BSI) caused by nontyphoidal *Salmonella* (NTS) have killed approximately 650,00 people globally[1]. Invasive NTS (iNTS) disease disproportionally affects immunocompromised individuals such as adults with Human Immunodeficiency Virus (HIV) and children under five years of age with malaria, malnutrition or severe anaemia living in lower-to-middle income countries[2]. Recently, a systematic review established that *Salmonella* was the most frequently isolated pathogen in hospitalised patients diagnosed with community-onset bloodstream infections in Africa and Asia (2008 to 2018)[3]. *Salmonella enterica* serovar Typhimurium (*S.* Typhimurium) is responsible for approximately two-thirds of iNTS disease cases in Africa[4], and consequently is a focal point for research.

The majority of *S.* Typhimurium isolated from BSI in Africa belong to sequence type (ST)313[5]. Two distinct and tightly clustered lineages of ST313 exist, differing from each other by 455 single nucleotide polymorphisms (SNPs)[5]. ST313 lineage 1 (L1) and ST313 lineage 2 (L2) are a 700 SNP distance[6] from ST19, which commonly causes gastrointestinal infections globally. The African ST313 L1 and L2 are genetically distinct from ST313 found in the United Kingdom and Latin America[7,8].

It has been proposed that antimicrobial resistance (AMR) contributed to the success of ST313 lineages in Africa, with a large proportion of strains being multidrug resistant (MDR)[9]. Amongst *S.* Typhimurium ST313, chloramphenicol resistance is specifically found in L2, and is thought to have contributed to the clonal replacement of L1 by L2 prior to 2001, when chloramphenicol was the empirical treatment choice for suspected sepsis in Malawi[6]. Subsequently, chloramphenicol was replaced by the oral fluoroquinolone, ciprofloxacin, for treatment of iNTS infections in several parts of Africa including Malawi. The 3rd-generation cephalosporin, ceftriaxone became the empirical therapeutic for suspected sepsis from 2005[9]. More recently, variants of ST313 L2 with resistance to 3rd generation-cephalosporins and ciprofloxacin or azithromycin have been identified in Malawi and the Democratic Republic of Congo, respectively[10,11].

Europe PMC Funders Author Manuscripts
Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

In addition to AMR, the ST313 lineages have accumulated loss-of-function mutations (genome degradation)[12]. Evidence of host adaptation in ST313 involved specific pseudogenes that inactivated gene function and modified metabolic and virulence phenotypes[5,13–18]. The functional relevance of genome degradation in ST313 has been investigated at the transcriptomic, proteomic and phenotypic levels for several key isolates[5,13–17,19]. To understand the evolutionary trajectory and epidemiological relevance of functional degradative events, a study of *Salmonella* genomes from a wide range of contemporary and historical isolates was required.

Here, we study a large and up-to-date collection of ST313, revealing the stepwise evolution of *S.* Typhimurium causing BSI in Africa. We uncovered a pan-susceptible lineage (ST313 L3) which emerged in Malawi in 2016, with a distinct genome-degradation pattern. Despite only appearing recently in clinics, our core and accessory genome analyses revealed ST313 L3 to be a phylogenetic intermediate between ST313 L1 and L2. We performed genotypic and phenotypic characterisation of the key loss-of-function events that occurred along the evolutionary pathway of invasive *S.* Typhimurium across Africa.

Results and Discussion

Assembling an informative collection of *S.* Typhimurium isolates

A combination of historical and contemporary *S.* Typhimurium bloodstream isolates was sampled from two sources (Figure 1). The main dataset consisted of 608 human bloodstream isolates collected between 1996 and 2018 from the Malawi-Liverpool Wellcome Trust Clinical Research Programme (MLW) in Blantyre, Malawi. To provide context, 72 human bloodstream isolates collected between 1966 and 2012 by the Unité des Bactéries Pathogènes Entériques, Institut Pasteur, Paris, France were also included. This contextual collection was derived from 13 African, 1 European and 1 Asian countries. A description of the isolates is summarised in Figure 1, details of sequencing quality control are available in Supplementary Table 1 and the complete metadata, including genome accession numbers, are in Supplementary Table 2.

Population structure of *S.* Typhimurium

To define the population structure of *S.* Typhimurium currently causing BSI in Africa, we explored the relationship of the 680 *S.* Typhimurium genomes from the MLW and contextual collections. The two predominant sequence types causing BSI in Malawi were ST313 (n=549) and ST19 (n=34). A small number of isolates (n=25) were typed as ST313 single-locus variants (ST3257 [n=17], ST2080 [n=3], ST302 [n=3], ST4200 [n=1], ST4274 [n=1]). The contextual dataset contained two sequence types: ST313 (n=43) and ST19 (n=29). Taken together, ST313 isolates originated from nine African countries (Figure 1). Previous genome-based analyses of the African ST313 epidemic included isolates from Eastern and Central Africa[6,11]. Here, by including several Western Africa countries, we have expanded the known geographic range of ST313[6,8,11,20–23] to include Cameroon, Central African Republic, Niger, Senegal, Sudan and Togo.

Europe PMC Funders Author Manuscripts

To investigate population structure, a core-genome maximum likelihood phylogeny was constructed, and cluster designation was performed. Three major ST313 clusters, four ST19 clusters, and a small number of miscellaneous strains were identified (Extended Data 1). Multi-locus sequence type variants formed discrete sublineages. The three ST313 clusters were clonal, with Cluster 1 and 2 corresponding to previously defined lineages (ST313 L1 and L2)[5].

The genomes of contemporary isolates revealed that a third lineage has been circulating in Malawi since 2016 (ST313 L3, Figure 2). Characterisation based on long-read sequencing showed that ST313 L3 strain BKQZM9 had a circular chromosome (4.8 Mb) with 99.98 % sequence identity to the ST313 L2 reference strain D23580[5].

To fully contextualise ST313 L3, we constructed a core-gene maximum-likelihood phylogeny, including previously published ST313 genomes (Supplementary Table 3), alongside the current dataset (Extended Data 2). ST313 L3 forms a monophyletic cluster within a group of ST313 strains isolated in the United Kingdom and Brazil[7,8], raising the possibility of an international transmission event. ST313 L3 isolates are closely related, with a maximum pairwise distance of 16 SNPs. To determine if ST313 L3 represents an outbreak in Malawi, it would be necessary to combine genome-based information with details on time, person and place. Patient-level GPS data were unavailable, meaning the geospatial distribution of ST313 L3 is currently unknown.

Accessory genome of *S.* Typhimurium ST313 lineages

To understand the relationship between ST313 L3 and other ST313 lineages, we examined the accessory genome in the context of population structure (Figure 2). All lineages carried the virulence plasmid pSLT[24], and ST313 L2 additionally contained plasmids pBT1, pBT2 and pBT3, consistent with previous studies[5,17]. ST313 L1 carried a previously unreported 8,274 bp plasmid, sharing 99.27 % sequence identity with pAnkS[25] (GenBank NC_010896.1) which encoded the LsoA/ LsoB toxin-antitoxin proteins. This toxin-antitoxin system functioned as a bacteriophage exclusion system in *Salmonella* (Supplementary Discussion 1)[26]. ST313 L3 strain BKQZM9 contained the 95 kb plasmid (pSLT), a 1,975 bp plasmid (pBT3) and a 2,556 bp plasmid (pBT2, identified using short read data only). Thus, the plasmid profile of ST313 L3 shared more similarity to ST313 L2 than other lineages, but lacked pBT1 (Figures 2, S1).

A pairwise comparison of reference strains BKQZM9 (ST313 L3), D23580 (ST313 L2) and U2 (the representative UK-isolated ST313 strain) (Extended Data 3) revealed chromosomal differences involving the prophage repertoires (Figures 2, S1). The common *S.* Typhimurium prophages Gifsy-1, Gifsy-2[27] and ST64B[28] were identified in all three ST313 African lineage reference strains. ST313 L3 carried a Fels-2-like prophage that was absent from the ST313 L2 and the UK-isolated ST313 representative, and shared 99.99 % sequence identity to the prophage RE-2010[29] found in other extraintestinal *Salmonella* serovars including a global outlier cluster of *Salmonella* Enteritidis[30] and *Salmonella* Panama[31,32] (Extended Data 3). Prophage BTP5 was absent from the ST313 L3 reference strain. BTP5 is associated with ST313 L1 and L2[33], but is generally absent from UK-isolated ST313[20].

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Prophage BTP1 was not found in ST313 L3, although a related P22-like prophage with 74 % fragmented sequence identity to BTP1 occupied the same attachment site (Extended Data 4a). The P22-like prophage had almost identical replication, lysis and capsid genes to BTP1, differing only in the Ea, immunity, tail and *gtr* regions, and lacking the *bstA* [34] *gene*. The ST313 L3 P22-like prophage carried the *sieB* gene, which encodes a superinfection exclusion phage immunity protein[35]. The SieB proteins of ST313 L3 and P22 were 92% identical, with the two proteins sharing 177/192 amino acids. Carriage of a gene that mediates abortive phage infection could confer a selective advantage to ST313 L3, reflecting the functional role of the BTP1 prophage-encoded BstA virus defence protein in ST313 L2[34]. BstA defends bacterial cells that carry BTP1 against exogenous attack by a variety of lytic phages, conferring a beneficial trait to ST313 L2 without sacrificing lytic autonomy[34]. Although BTP1 is absent from the UK reference strain U2, two UK-isolated ST313 strains (U15 and U8[7]) carried a P22-like prophage with 99.96 % sequence identity to that found in ST313 L3 (Extended Data 4b). In summary, the prophage repertoire of ST313 L3 shares similarities with UK-isolated ST313.

Genomic epidemiology of antimicrobial resistance

To obtain epidemiological insights into the emergence of ST313 lineages, the dynamics of AMR were investigated. The draft genome sequences of all 680 isolates were examined for genes and mutations that confer reduced susceptibility to antimicrobials (Supplementary Discussion 2). In total, 65 % (n = 440) of isolates had an MDR genotype[36]. Amongst ST19, only 11 % (7/63) of isolates were resistant to≥ 1 antimicrobial. However, amongst ST313, 94 % of isolates were resistant to ≥ 1 antimicrobial. Chloramphenicol-resistant isolates of ST313 L2 have decreased over time, with 89.19 % (n=66/74) prior to 2005, decreasing to 84.11 % (n=307/365) between 2006-2015 and 65.28 % (n=47/72) between 2016-2018 (Figure 3 and Supplementary Table 2). Interestingly, there has been antimicrobial usage policy changes at the local level in Malawi during this timeframe, including the phased withdrawal of chloramphenicol from clinical practice beginning in 2001.

Overall, ST313 in this study had numerous genetic determinants of antibiotic resistance, with 29 different AMR genotype patterns involving 11 antimicrobials (Figure 3). The AMR phenotype of ST313 L1 and L2 is encoded by resistance cassettes carried on a composite Tn*21*-like transposable element, inserted in the *S.* Typhimurium virulence plasmid pSLT[5–7,14]. The plasmid backbone of pSLT was present in all ST313 strains and we observed lineage-specific variation in the Tn*21* insertion site, consistent with previous reports[5]. The Tn*21*-like region varied between isolates, reflecting AMR profile (Extended Data 5). Plasmid pAnkS$^{ST313\text{-}L1}$ carried the *bla* $_{TEM}$ gene which mediated beta-lactam resistance of ST313 L1.

ST313 L3 was pan-susceptible to all antimicrobials tested, both genotypically and phenotypically (Figure 2), as were the UK-isolated ST313[20]. In contrast to L1 and L2, ST313 L3 did not carry an MDR cassette on the pSLT-BT plasmid (Extended Data 6). Phylogenetic reconstruction of the pSLT plasmid (Extended Data 7) reflected the core genome SNP-based phylogeny, demonstrating four phylogenetic clades (ST313 L1, L2, L3

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

and ST19). There was no evidence of a scar associated with excision of the Tn*21*-like transposable element around location 48,530 on the ST313 L3 pSLT plasmid.

Less than 1 % (*n*=5/180) of African *S.* Typhimurium ST313 L2 carried mutations associated with reduced fluoroquinolone susceptibility (eg. *gyrA* [37]), originating in Malawi (*n*=3), Mali (*n*=1) or Cameroon (*n*=1). The GyrA proteins had either the S83F, D87N, D87Y amino acid substitutions. All five strains were phenotypically-susceptible to ciprofloxacin. Ongoing surveillance will be important for identification of strains with triple mutations which would confer full resistance to this clinically important antibiotic.

The evolutionary trajectory of *S.* Typhimurium ST313

Dates for lineage-divergence were provided by Bayesian phylogenetic inference (Figure 4, Extended Data 8). The most recent common ancestor (MRCA) of all ST19 and ST313 sampled in this study was ~918 AD (95 % Highest Posterior Distribution (HPD) = 570-1080) and the MRCA of all ST313 sampled in this study (Figure 4, event 0) was ~1566 (95 % HPD = 1271-1953). The incorporation of historical and diverse isolates that predate those used in previous phylodynamic analysis[6,7] generates an older MRCA of ST313 lineages than previously described (~1787)[7]. However, we note that the branch linked to the MRCA is deeply rooted in the phylogeny and the 95% HPD values reported here have a substantial breadth. A limitation of Bayesian analysis is that it relies upon a series of user-defined prior probability distributions which influence the estimation of tip dates (Methods).

The MRCA of ST313 L1 dated to ~1794 (95 % HPD = 1738 - 1965), with Malawian ST313 L1 forming a discrete sublineage with an MRCA (Figure 4, event 1) dated to ~1950 (95 % HPD = 1921-1986). The ST313 L2 MRCA (Figure 4, event 2) dated to ~1948 (95 % HPD = 1929-1959). These MRCA estimations overlap with previous 95 % HPD values for ST313 lineage dating[6]. The MRCA of ST313 L3 (Figure 4, event 3) was 2007 (95 % HPD 1998-2012), showing that the clonal expansion of ST313 L3 occurred in recent evolutionary history.

Phylodynamics of pseudogene formation in *S.* Typhimurium ST313

To explore the association of pseudogenisation and invasive salmonellosis, we characterised the stepwise progression of functional gene loss associated with host adaptation that likely facilitated the ST313 epidemic in Africa (Methods). This phenomenon is exemplified by the timeline of pseudogene accumulation in ST313 L2 shown in Figure 4.

Our discovery of ST313 L3, allowed the timeline of gene degradation events to be scrutinised in more detail. For example, mutations in *melR* (F311L), *flhA* (A166T) and *pipD* (283 bp deletion) occurred prior to the divergence of ST313 L3 and ST313 L2, while the mutations in *bcsG* (W247*), *sseI* (IS*26* insertion) and *lpxO* (E198*) preceded the clonal expansion of ST313 L2. A small number of gene degradation events were only associated with ST313 L3, specifically an additional non-synonymous mutation in *ratB* (G820S) and a 109bp deletion in *pipD.* Both mutations occurred in genes which had been pseudogenised earlier in evolutionary history, and so are unlikely to cause specific phenotypic changes. The *macB, sseI* and *lpxO* genes that were non-functional in ST313 L2 were functional in ST313 L3, as reported previously for the UK-ST313 representative (U2)[7]. We recently reported the

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

functional significance of the pseudogenisation of *macB* for ST313 L2[18]. The frameshift event that inactivated the *macB* gene occurred between 1948 and 1953 (95% HPD = 1931-1963), representing the final stage in the evolution of ST313 L2 detected in our analysis (Figure 4).

Function of evolutionarily significant *S*. Typhimurium ST313 genes

To complement our genome-derived evolutionary insights, we linked a number of lineage-specific phenotypes to individual SNPs (summarised in Table 1 and detailed fully in Supplementary Table 4). For example, the KatE catalase protects bacteria from environmental oxidative stress, and shows a maximal activity in stationary phase cultures of *S*. Typhimurium ST19[15]. The E117G KatE mutation caused low level catalase activity in ST313 L1, L2, L2.1 and UK-isolated ST313[7,11,15]. We report a similarly low activity in ST313 L3 (Table 1, Supplementary Table 4). Catalase is required for biofilm formation[38], a phenotype associated with survival outside the host, leading to the suggestion that pseudogenisation of *katE* in ST313 reflects adaptation to a restricted host-range[16].

The RDAR-negative (red, dry and rough) phenotype of ST313 L2[39] is caused by pseudogenisation of *bcsG* [15] and two nucleotide changes in the *csgD* gene promoter region[39]. ST313 L3 has an intermediary RDAR phenotype, with an incomplete wrinkling pattern (Supplementary Table 4). ST313 L3 have a functional *bcsG* gene and had the same *csgD*-associated −189 promotor mutation as ST313 L1 and L2[39]. Our finding that ST313 L3 has lost traits required for stress-resistance and biofilm formation is consistent with a reduced requirement for environmental survival and the mounting evidence for human-human transmission of ST313[40,41].

S. Typhimurium ST19 relies upon carbon metabolism to colonise the mammalian gastrointestinal tract[42], whereas genes required for utilisation of specific carbon sources, such as tartrate and melibiose, are not functional in ST313 lineages[17]. ST313 L3 was also unable to grow with tartrate or melibiose as a sole carbon source, consistent with the pseudogenisation of the *ttdA* and *melR* genes in both ST313 L3 and ST313 L2[5,17] (Figure 4).

Further evidence of genome degradation is provided by the pseudogenisation of the genes *ratB* in all ST313 lineages, *pipD* in ST313 L2 and L3, and *lpxO* in ST313 L2. The PipD effector protein has been implicated in gastrointestinal pathogenesis of ST19, although a causal relationship has not been demonstrated[43,44]. The *lpxO* gene is pseudogenised by a stop codon in ST313 L2[5]. LpxO hydroxylates lipid A, a modification required for virulence of S. Typhimurium ST19[45]. We used mass spectrometry to show that the lack of functional LpxO caused structural modifications of Lipid A in ST313 L2 (Table 1, Supplementary Table 4). The functionality of the *lpxO* gene in ST313 L3, reflects the distinct evolutionary path of this lineage.

Invasiveness potential of ST313 L3

To determine the potential of ST313 L3 to cause bloodstream infections, the invasiveness index of each sample was calculated and compared between different ST313 lineages using the Wilcoxon Mann Whitney test[46]. The “invasiveness index” represents the extent of

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

genome degradation and diversifying-selection specific to invasive serovars, using a proven set of 196 extraintestinal predictor genes[47]. Consistent with previous studies, we observed a significant increase in the invasiveness index of ST313 L2 compared with ST313 L1 (W=7636.5, p<0.001), and ST313 L1 compared with ST19 (W=552.5, p<0.001)[11,47]. ST313 L3 had a particularly high invasiveness index (median=0.187, standard deviation (SD)=0.008), which was significantly greater than that of ST19 (median=0.110, SD=0.017), ST313 L1 (median=0.129, SD=0.008) or ST313 L2 (median=0.134, SD=0.006) (W=0, p<0.001) (Figure 5). ST313 L3 also had a significantly greater invasiveness index than UK-isolated ST313 (median=0.134, SD=0.018), (W=480, p-value<0.001).

In total, 17 of the 196 extraintestinal predictor genes had undergone additional degradation in ST313 L3 (isolate BKQZM9) compared with ST313 L2 (D23580), involving the presence of non-synonymous SNPs, indels or gene loss (Supplementary Table 5). A number of the degraded genes in ST313 L3 are required for colonisation of the gastrointestinal tract by *S.* Typhimurium ST19, including *mrcB*, a gene that allows *Salmonella* Typhi to survive in the presence of bile[48]. In *S.* Typhimurium, a functional *damX* gene is required for resistance to bile[49], and was degraded in seven of the nine ST313 L3 isolates. The *napA* gene had extensive deletions in all but one of the ST313 L3 isolates, and encodes a periplasmic nitrate reductase required for gut colonisation[17,50]. Taken together, our findings suggest that ST313 L3 is in the process of adapting from an intestinal to a systemic lifestyle.

Clearly, this *in silico* approach does not prove that ST313 L3 has adapted to an extraintestinal lifestyle. In the past, infection models have provided understanding of the virulence and systemic spread of *Salmonella* pathovariants that cause gastroenteritis[51]. However, cellular and animal infection models have failed to discriminate between levels of invasiveness of ST313 L1 and L2[14] and an ST313 sublineage identified in the Democratic Republic of Congo[11]. Although the invasiveness index cannot yet be experimentally validated, *Salmonella* isolates with different invasiveness indices produce distinct clinical symptoms in a human population[52]. The development and validation of experimental approaches for the robust measurement of the invasiveness of African *S.* Typhimurium will be an important focus for future research, but goes beyond the scope of this paper. Meanwhile, the “invasiveness index” is a useful tool for monitoring the genetic signatures associated with invasiveness and host adaptation in ST313 lineages.

Perspective

Our data provide an expanded and contemporary insight into the dynamics of *S.* Typhimurium associated BSI in Africa, elucidating the stepwise evolution of ST313. This study provides a snapshot of the large repertoire of genomic changes that shaped the emergence of successful ST313 lineages. The stepwise pseudogenisation of genes reported here is consistent with the host adaptation of *S.* Typhimurium ST313 lineages over evolutionary history[6,40] and supports mounting evidence for a human ST313 reservoir[40]. Our findings reflect similar patterns of evolution found in other *Salmonella* serovars that currently cause BSI disease globally. For example, a third of all iNTS disease cases in Africa are caused by *S.* Enteritidis which also shows signs of genome degradation consistent with niche-adaptation[30,53].

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

We discovered that ST313 L3 emerged in 2016 as a cause of BSI in the Malawian population, and had an MRCA of ~2007. It is possible that the phased removal of chloramphenicol from clinical practice following local policy changes in Malawi between 2002 and 2005 created a window of opportunity for the emergence of fully-susceptible ST313 L3 (Supplementary Discussion 2). ST313 L3 forms a monophyletic cluster within a diverse group of ST313 originally identified in the UK[7]. The combination of our phylogenetic and accessory genome analyses led us to conclude that ST313 L3 has been introduced into the Malawian population, likely via an international transmission event. Importantly, ST313 L3 had an elevated invasiveness index compared to ST313 L2, unlike the UK-isolated ST313 strains that had a lower invasiveness index than ST313 L2[47]. The majority of ST313 L3-unique genes were plasmid or prophage-encoded, including the prophage RE-2010[29] which is found in other extraintestinal *Salmonella* pathovariants[30–32]. Apart from accessory genome composition, other ST313 L3-specific changes involve loss-of-function SNPs in genes that are not required for systemic infection. The findings predict that ST313 L3 will be better adapted to cause extraintestinal infection than other ST313 lineages from both the UK and Malawi, possibly explaining the emergence of ST313 L3.

Methods

Dataset

Salmonella samples were derived from two archived blood culture collections. The main dataset was sourced from the Malawi Liverpool Wellcome Clinical Research Programme (MLW) in Blantyre, Malawi and consists of over 14,000 *Salmonella* bloodstream isolates from patients with iNTS disease at the Queen Elizabeth Central Hospital between 1996 and 2018. Isolates listed as *S.* Typhimurium were extracted from the metadata file for the entire collection and stratified by AMR profile into the following four categories: susceptible, resistant to one first line agent, resistant to two first line agents or MDR. First line agents were considered to be ampicillin, cotrimoxazole and chloramphenicol. Random subsampling was then performed using random number generator (Excel, Microsoft) and selected samples were collected and resuscitated from the freezer archives at MLW. We successfully recovered 647 isolates for whole-genome sequencing.

To complete the MLW data, a contextual dataset was sourced from the Institut Pasteur, Paris. The dataset consisted of 96 *S.* Typhimurium isolates collected from human extraintestinal sites from patients contaminated in Algeria, Burkina Faso, Cameroon, Central African Republic, Republic of the Congo, France, Côte d'Ivoire, Madagascar, Mali, Niger, Senegal, Sudan, Togo and Vietnam between 1966 and 2012. All 96 *S.* Typhimurium samples were selected and sent for whole genome sequencing.

Whole genome sequencing of short reads

Isolates were prepared by inoculating a single bead of frozen *Salmonella* stock into a FluidX 2D Sequencing Tube (FluidX Ltd, UK) containing 100 μL of buffered peptone water (Oxoid, CM0509) and incubating overnight at 37°C. Thermolysates were generated by heat-killing cultures in a 95°C water bath for 20 minutes before DNA extraction (MagAttract kit, Qiagen) and whole-genome sequencing, as part of the 10,000 *Salmonella* Genomes

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Project[54] (https://10k-salmonella-genomes.com/). Illumina Nextera XT DNA Libraries were prepared (Illumina, FC-131-1096) and sequenced (Illumina HiSeq 4000) in multiplex (768) as 150 bp paired-end reads.

Assembly and annotation of short reads

Trimmomatic[55] v0.36 was used to trim adapters and Seqtk v1.2-r94 (https://github.com/lh3/seqtk) was used to trim low-quality regions using the trimfq flag. Fastqc v0.11.5 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and multiqc v1.0 (http://multiqc.info) were used to pass sequence reads according to the following criteria: passed basic quality statistics, per base sequence quality, per base N content, adapter content and an average GC content of between 47 % and 57 %. Only high-quality reads were used in downstream analysis. Unicycler[56] v0.3.0 was used to assemble genomes and QUAST[57] v4.6.3 was used to evaluate assembly quality to standards consistent with Enterobase[58]. Specifically, N50 > 20 kb, 600 or fewer contiguous sequences, total number of bases between 4Mbp and 5.8Mbp. Prokka[59] v1.12 was used to annotate the genomes.

Sequence typing

Serotyping was confirmed using the *Salmonella in Silico* Typing Resource (SISTR)[60] v1.0.2 and compared with the original metadata table. The strains were assigned a Multi Locus Sequence Type (MLST) using MLST[61] v2.10 based on the conservation of seven housekeeping genes.

Reference mapping to D23580

Trimmed sequencing data were mapped against the *S.* Typhimurium reference genome D23580 using BWA mem[62] v0.7.10-r789. The reference genome consists of the D23550 chromosome (accession: LS997973.1), and four plasmids including pSLT-BT (accession: LS997974.1), pBT1 (accession: LS997975.1), pBT2 (accession: LS997976.1) and pBT3 (accession: LS997977.1). Mapped reads were then cleaned and sorted using the SAMtools suite[63] v1.7. Reads were realigned against the reference using GATK[64] v3.7 by creating targets for realignment (RealignerTargetCreator) and performing realignment (IndelRealigner). Removal of optical duplicates was completed using Picard v2.10.1-SNAPSHOT (https://broadinstitute.github.io/picard/). Sequence variants were called using Bcftools v1.9-80 (http://samtools.github.io/bcftools) to generate a reference-based pseudogenome for each sample with greater than 10x depth. QualiMap[65] v2.0 was used to identify a mean sample depth of 35.17x across all isolates.

Phylogenetic reconstruction of dataset

High quality pseudogenomes (MLW n=608, contextual n= 72) were concatenated, plasmids were excluded and Gubbins[66] v2.2 was used to remove recombinant regions and invariable sites. The resultant multiple sequence alignment of reference based pseudogenomes (12,013 variant sites) was used to infer a maximum likelihood phylogeny using RAxML-ng[67] v0.6.0 with 100 bootstrap replicates to assess support. To assign clusters, rhierBAPs[68] was used specifying two cluster levels, 20 initial clusters and infinite extra rounds. Visualisations were performed using interactive Tree Of Life (iTOL)[69] v4.2.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Phylogenetic reconstruction of dataset and publicly available dataset

For contextual interpretation, previously published *S.* Typhimurium ST313 were retrieved from public repositories (Supplementary Table 2) and constructed in a core gene phylogeny with our data. Roary[70] v3.11.0 was used to generate a core gene alignment and SNP-sites[71] v2.3.3 was used to extract SNPs. The resultant multiple sequence alignment (15,240 variant sites) was used to construct a maximum likelihood phylogeny using RAxML-ng[67] v0.6.0 with 100 bootstrap replicates to assess branch support. Visualisations were made using iTOL[69] v4.2.

Temporal phylogenetic reconstruction

To determine the evolutionary history of ST313, a chronogram was produced using Bayesian phylogenetic inference. Because the MLW collection is substantially larger than the contextual collection and to reduce computation time, 150/680 samples were chosen for inclusion in the analysis as described below. Sampling reflected original sample selection for this study and included all contextual isolates which passed quality control (n=72) and a subset of MLW isolates randomly selected using a random number generator (n=78, Microsoft Excel). A reference-mapped multiple sequence alignment (7,231 variant sites) was created as described above. BEAUTI[72] v2.6.1 and BEAST2[72] v2.6.1 were used to create and execute three independent chains of length 250,000,000 with 10% burn in, logging every 1,000 and accounting for invariant sites. We included the prior assumptions of a coalescent Bayesian skyline model for population growth, and a relaxed log normal clock rate to account for rate heterogeneity amongst branches (the full model is described in Supplementary Data 1). Tracer[73] v1.7.1 was used to assess convergence, with all parameter effective sampling sizes being > 200. LogCombiner[72] v2.6.1 was used to combine tree files and DensiTree v2.2.7[74] was used for visualisation. Finally, a maximum clade credibility tree was created using TreeAnnotator[72] v2.6.0.

Determination of invasiveness index

The invasiveness index of each isolate was calculated using previously defined methods[47]. Because isolates used in our study originated from human bloodstream, we used the invasiveness index model that was pre-trained by Wheeler *et al.,* using a mixture of gastrointestinal and extraintestinal salmonellae[47]. Specifically, samples were analysed using 196 top predictor genes for measuring invasiveness in *S. enterica.* The machine-learning approach uses a random forest classifier and delta bitscore functional variant-calling to discriminate between the genomes of invasive and gastrointestinal *Salmonella*. The invasiveness index has been validated using multiple *Salmonella* serovars, and the approach clearly discriminated between the *S.* Enteritidis and *S.* Typhimurium lineages associated with extraintestinal infections in sub-Saharan Africa[11,47]. The distribution of invasiveness index values for each lineage were compared using the Wilcoxon Mann Whitney test[46] implemented through R[75] v3.4.0. A custom-made database of the top 196 invasiveness predictor genes was then created from the multi-fasta file provided in Wheeler *et al.* (2018). SRST2[76] v0.2.0 was used to flag any genes with mutations in ST313 L3 compared with ST313 L2. Exact mutations were investigated manually using MegaX[77] v0.1.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Antimicrobial resistance testing and statistical analysis

Genetic determinants for AMR were identified using staramr v0.5.1 (https://github.com/phac-nml/staramr) against the ResFinder[78] and PointFinder[79] databases. Phenotypic antimicrobial susceptibility testing was performed using the EUCAST disk diffusion method[80] (Antibiotic disks; ampicillin 10 μg, chloramphenicol 30 μg and trimethoprim/sulfamethoxazole 25 μg from Mast Group). To compare the results of genotypic against phenotypic testing, sensitivity and specificity were calculated for first line agents used over the study period using MedCalc's test evaluation calculator with Clopper-Pearson confidence intervals (https://www.medcalc.org/calc/diagnostic_test.php).

Genomic conservation of plasmids, prophages and pseudogenes

A literature search was used to identify genomic regions known to vary between ST19 and ST313, including plasmids[5], prophages[33] and pseudogenes[7,13–17,43,81,82]. We compiled a list of previously identified pseudogenes associated with the emergence of ST313 lineages in Africa (Supplementary Table 4)[14]. The collection included metabolic genes (*melR* [17] and *ttdA* [14]), environmentally-responsive genes (*bcsG* [15] , *katE* [15] and *macB* [13]), genes that modulate the bacterial cell surface (*flhA* [17] and *lpxO* [83]) and virulence genes (*pipD* [43] , *ratB* [14] , *sseI* [16]). Genes of unknown function were not included. We detected the SNPs responsible for pseudogene formation, as well as other non-synonymous mutations. A custom-made database of the identified regions was created, along with known variants. Sequences were downloaded using the online tool SalComD23580[17]. SRST2[76] v0.2.0 was then used to identify alleles present and the non-exact matches were investigated manually using MegaX[77] v0.1. Prophage and plasmid presence/absence were manually mapped onto the phylogeny using iTOL[69] v4.2. The conservation of pseudogene-associated mutations was identified at the population level and integrated into the temporal phylogenetic reconstruction using iTOL[69] v4.2. These analyses allowed inference of branches in which functional gene loss and plasmid and prophage acquisition occurred. Phenotypic testing was conducted to determine the relevance of plasmid and functional gene loss as detailed in Supplementary Methods 1.

Phylogenetic reconstruction of pSLT plasmid

Plasmid sequences were extracted from high quality pseudogenomes and snp-sites[71] was used to extract SNPs. The resultant multiple sequence alignment of length 1,034 sites was used to infer a maximum likelihood phylogeny using RAxML-ng[67] v0.6.0 with 100 bootstrap replicates to assess support. The phylogeny was visualised in iTOL[69] and convenience rooted to display lineages.

Long read sequencing of ST313 L3 reference strain BKQZM9

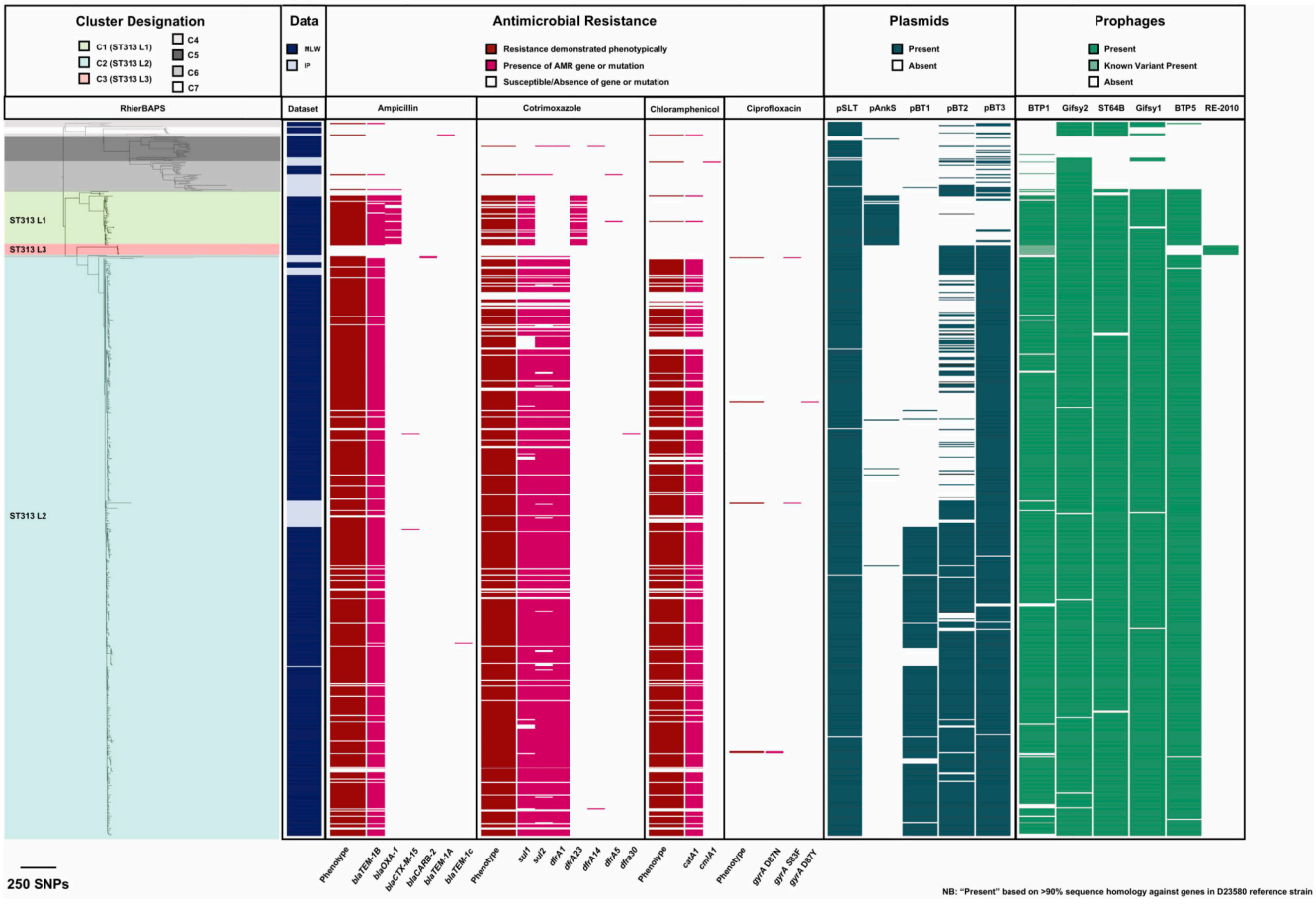
ST313 L3 strain BKQZM9 was selected for long read sequencing as it has the highest overall quality statistics based on short read sequencing. An Oxford Nanopore MinION[84] 9.4.1 flowcell and SQK-RAD004 rapid sequencing kit was used with base calling by Guppy v3.1.5 (https://nanoporetech.com/nanopore-sequencing-data-analysis#). Approximately 120x coverage (600 Mb, 60,164 reads, read N50 33 kb) was generated. Hybrid genome assembly with the Illumina reads was done using Unicycler[56] v0.4.4 which resulted in a 4.8

Mb circular polished chromosome (GenBank Accession: CP060169), a 94 kb circular plasmid (pSLT) (Accession CP060170) and a 1,975 bp circular plasmid (pBT3) (Accession: CP060171. Prokka[59] v1.12 was used for annotation as above.

Genomic comparison of ST313 Lineages

To determine the basic architecture of the ST313 L3 BKQZM9 genome, the identity of the chromosomal and plasmid regions were compared with existing sequences from ST313 L1 (A130), ST313 L2 (D23580) and ST313 UK (U2) using the Basic Local Alignment Search Tool (BLAST)[85] and visualised using the Artemis Comparison Tool[86] v13.0.0. Putative prophage and plasmid sequences were extracted and identified using a BLAST[85] search against known ST313 prophages and plasmids and the non-redundant NCBI nucleotide database. Prophage and plasmid presence/absence were also determined experimentally (Supplementary Discussion 2).

Extended Data



Extended Data Fig. 1. Dataset maximum likelihood core genome SNP phylogeny

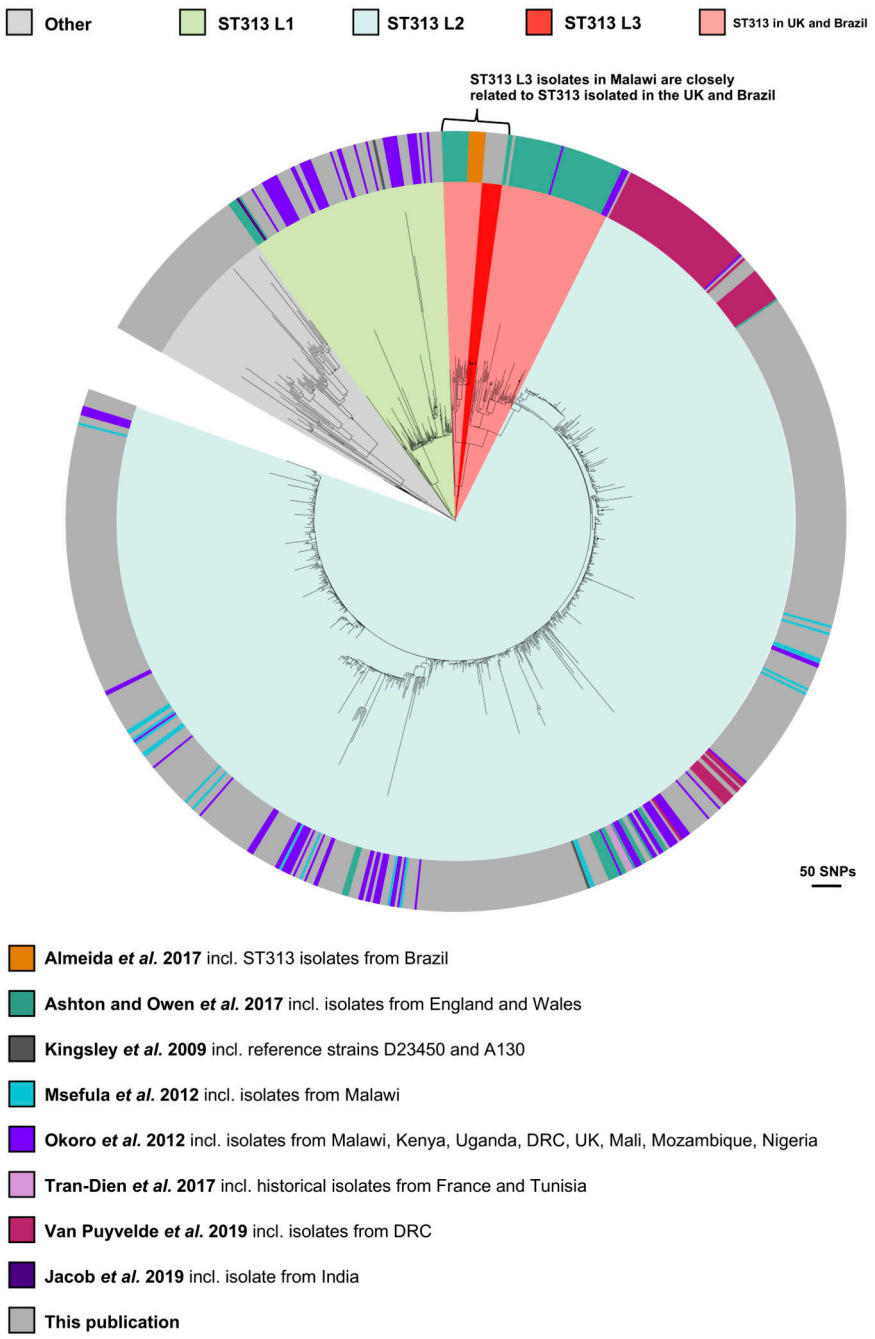
Maximum likelihood phylogeny based on the core genome SNP alignment of strains used in this study. Background shading on phylogeny represents cluster designation (rHierBAPs) and additional metadata is represented as adjacent colour strips. Note that Malawi-Liverpool

Europe PMC Funders Author Manuscripts
Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

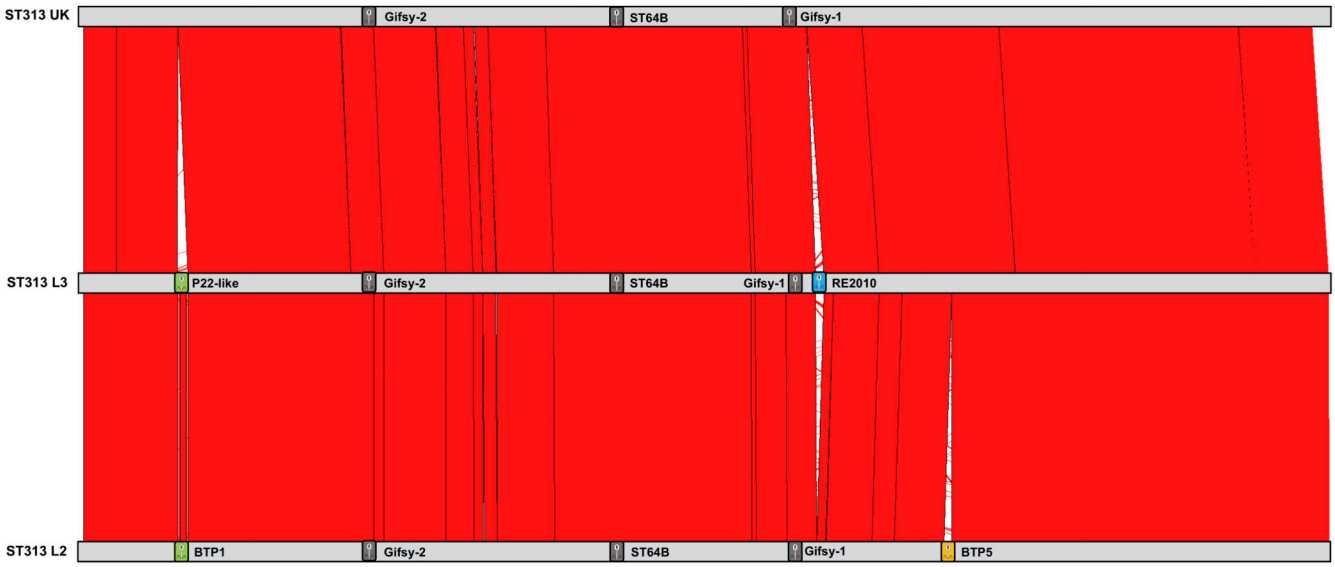
Wellcome Clinical Research Center is abbreviated to MLW and Institute Pasteur Unité des Bactéries Pathogènes Entériques Contextual Collection is abbreviated to IP. Figure visualised using iTOL[69].



Extended Data Fig. 2. Contextual maximum likelihood core gene SNP phylogeny

Maximum likelihood phylogeny based on the core gene SNP alignment of strains in this study in the context of published ST313 genomes[5–8,11,21,22,89]. Background shading on

phylogeny represents cluster designation (rHierBAPs). Outer ring provides details on the original publication. Grey represents this publication. Figure visualised using iTOL[69].



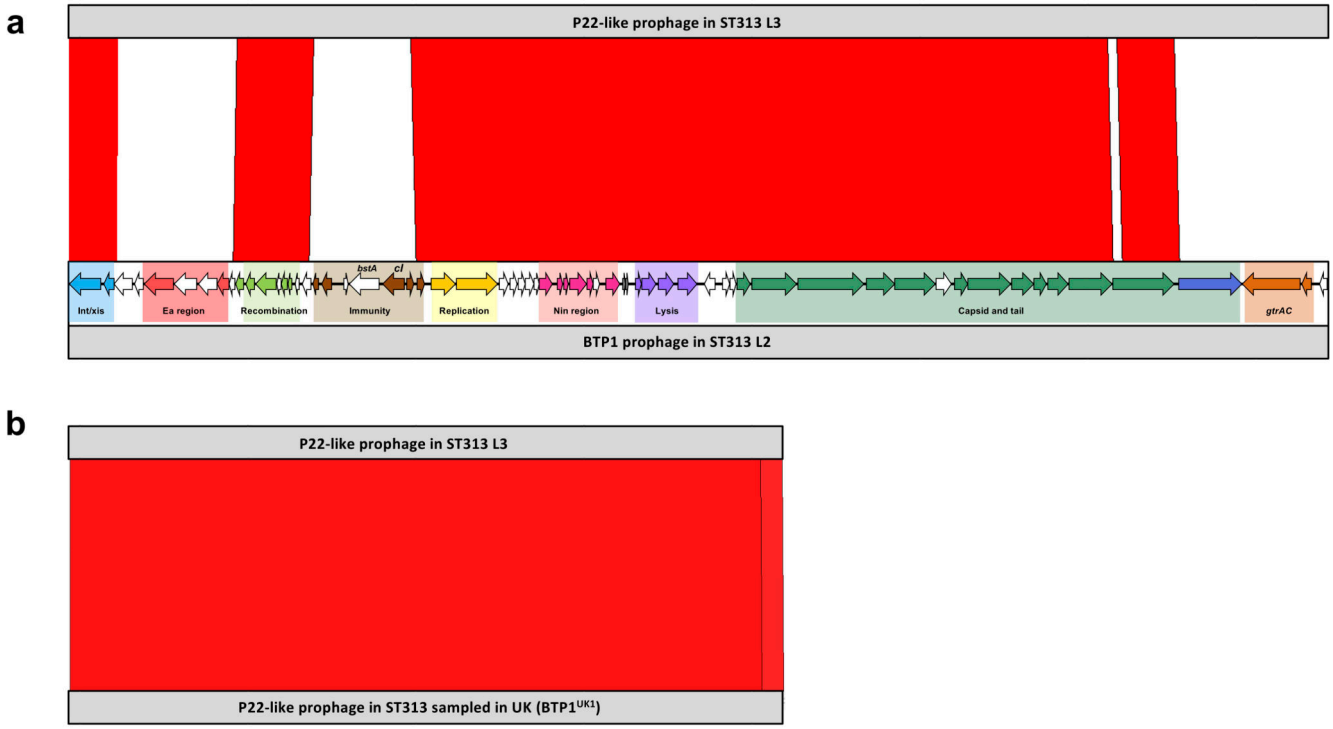
Extended Data Fig. 3. Chromosomal comparison of ST313 lineages

Comparison of the complete genome sequences of ST313 lineage representatives generated using the Artemis Comparison Tool[86]. Reference isolates used; D23580 (ST313 L2), BKQZM9 (ST313 L3) and U2 (UK strains). The number of genes shared between D23580 and BKQZM9 was 4,529. An additional 167 genes were BKQZM9-specific, whereas 309 genes were exclusive to D23580, largely explained by differences in the prophage and plasmid repertoire of the two strains. Red represents sequence similarity and white represents regions absent. Prophage positions are represented as coloured boxes.

Europe PMC Funders Author Manuscripts
Europe PMC Funders Author Manuscripts

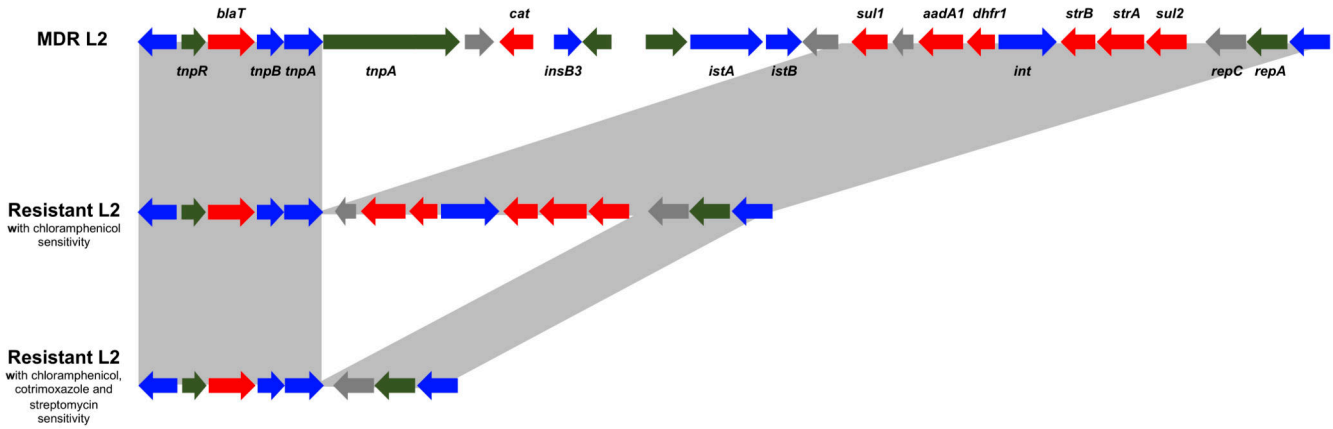
Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts



Extended Data Fig. 4. Prophage BTP1 comparisons

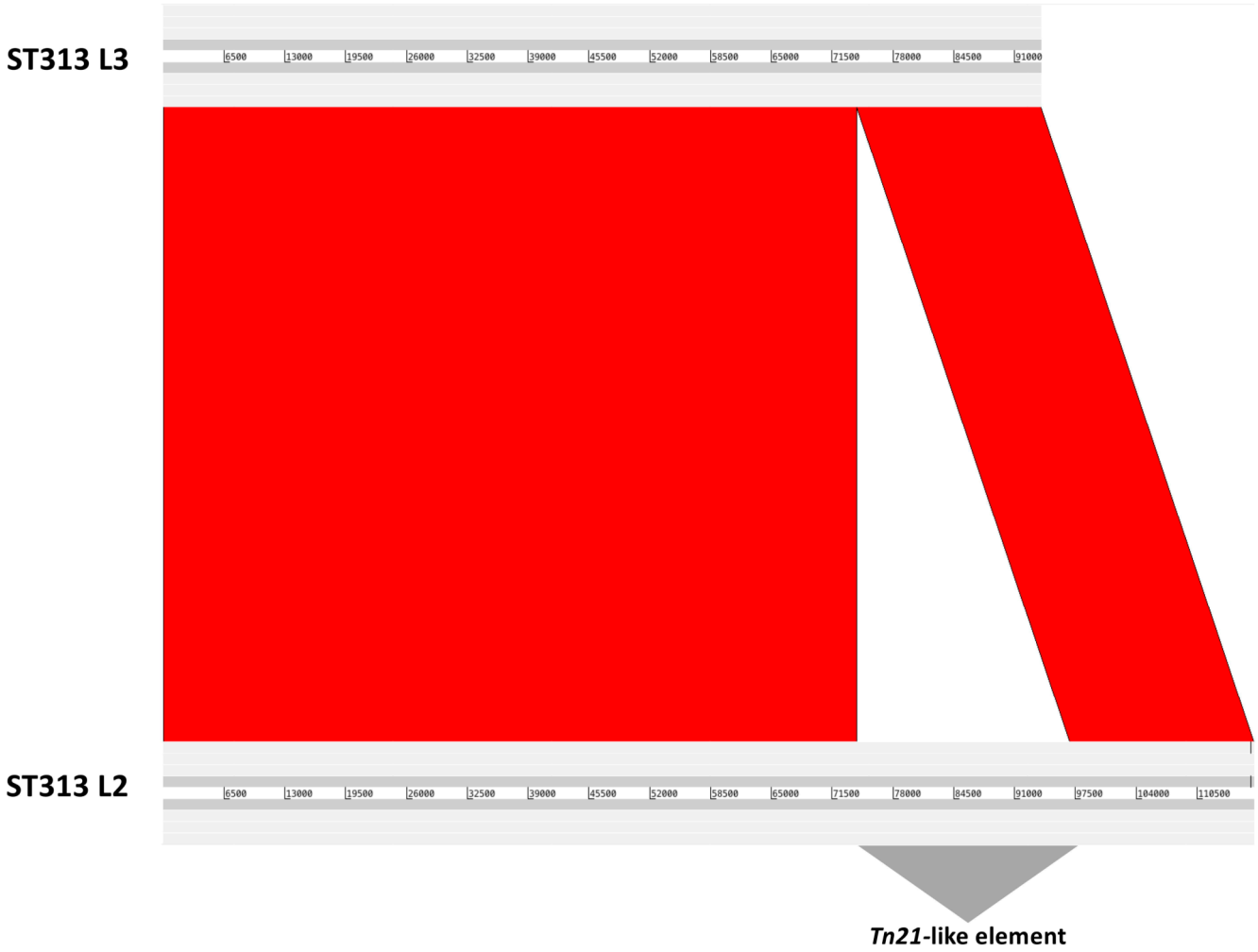
Comparison of P22-like prophage regions identified in ST313 lineages generated using the Artemis Comparison Tool[86]. Red represents sequence similarity and white represents regions absent. **(a)** shows the comparison between P22-like prophage in ST313 L3 (strain BKQZM9) and ST313 L2 (strain D23580). Prophage annotation is adapted from that shown in Owen *et al.,* 2017[33], with different colours highlighting different prophage regions. **(b)** shows the comparison between P22-like prophage in ST313 L3 (strain BKQZM9) and a P22-like prophage in ST313 sampled in the UK.



Extended Data Fig. 5. Examples of variation within the Tn*21*-like element

Annotation of the resistance cassette carried on the Tn*21*-like element integrated into the *Salmonella* virulence plasmid pSLT-BT. Examples of variation in the Tn*21*-like element is

shown for three different resistance profiles in ST313 L2, Grey boxes between annotations represents gene presence. Annotation is adapted from Kingsley *et al.*, 2009[5], and shows antibiotic resistance genes (red), integrase or transposase (blue), pseudogenes (green) and other genes (grey).



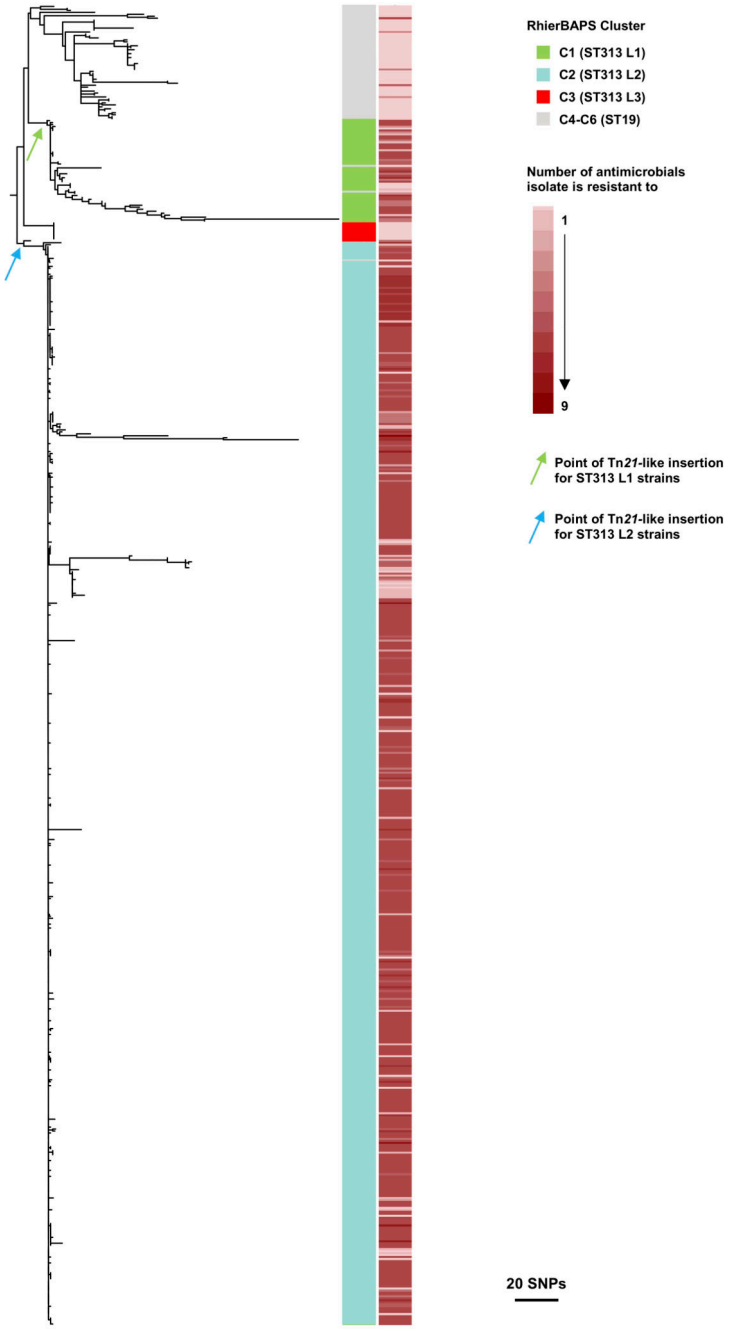
Extended Data Fig. 6. pSLT comparison of D23580 vs BKQZM9

Comparison of the *Salmonella* virulence plasmid pSLT identified in ST313 L3 (BKQZM9) and ST313 L2 (D23580) generated using the Artemis Comparison Tool[86]. Red represents sequence similarity and white represents absent regions. The location of the Tn*21*-like element is indicated.

Europe PMC Funders Author Manuscripts
Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

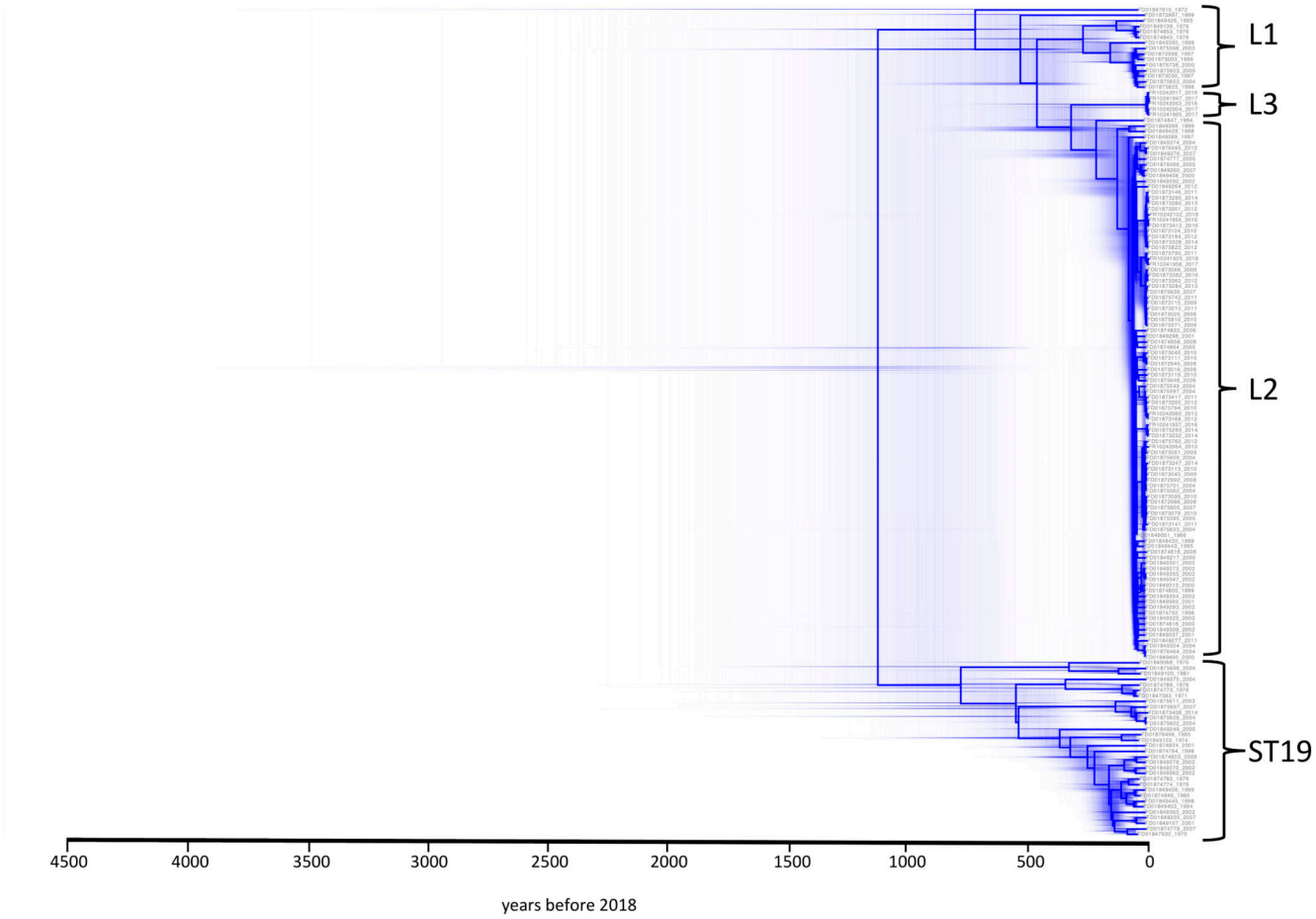


Extended Data Fig. 7. Phylogenetic reconstruction of pSLT plasmid

Maximum likelihood phylogenetic tree showing the population structure of the pSLT plasmid in ST19 and ST313 lineages. Colour strip from right to left; cluster assignment of each isolate and number of antimicrobials each isolate is resistant to. The likely point of insertion of the Tn*21*-like element is indicated for ST313 L1 strains (green arrow) and ST313 L2 strains (blue arrow).

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts



Extended Data Fig. 8. Distribution of possible hierarchies generated in BEAST analysis

Chronograph of 150 *S.* Typhimurium strains isolated from bloodstream of human iNTS disease patients. The figure displays the distribution of possible hierarchies and highlights uncertainty in the timeline. Figure was visualised using DensiTree[74].

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We are extremely grateful to present and former members of Lab H at the Institute of Infection, Veterinary and Ecological Sciences, University of Liverpool, including members of the Hinton and Baker groups for invaluable discussions. In particular we would like to thank Charlotte Chong, Mal Horsburgh, Siân Owen and Karl Costigan for their support, contributions and advice. We would like to express our gratitude to the Malawi Liverpool Wellcome Trust (MLW) clinical research programme and the Institute Pasteur for access to their *Salmonella* archives. We would like to thank Philip Ashton for providing very helpful comments on the manuscript. We would additionally like to thank Ross Low for his assistance in uploading metadata to repositories.

We appreciated funding from several sources. C.V.P is supported by the John Lennon Memorial Scholarship from the University of Liverpool (https://www.liverpool.ac.uk/study/international/tuition-fees-and-scholarships/scholarships-and-awards/john-lennon-memorial-scholarship/) and a Fee Bursary Award from the Institute of Integrative Biology at the University of Liverpool. B.K was funded by an AESA-RISE fellowship from the African Academy of Sciences. N.H is funded by the BBSRC, Core Strategic Programme Grant at the Earlham Institute (BB/

CSP17270/1). F.X.W and A.T.D are supported by the Institut Pasteur, Santé publique, and the French government's Investissement d'Avenir programme, Laboratoire d'Excellence 'Integrative Biology of Emerging Infectious Diseases' (ANR-10-LABX-62-IBEID). K.S.B is funded by a Wellcome Trust Clinical Research Career Development Fellowship (106690/A/14/Z). J.C.D.H is funded by a Wellcome Trust Senior Investigator Award (106914/Z/15/Z).

Genome sequencing was done by the Earlham Institute as part of the 10,000 *Salmonella* genomes project which is supported by the Global Challenges Research Fund data and resources grant (BBS/OS/GC/000009D). Next-generation sequencing and library construction were delivered via the BBSRC National Capability in Genomics and Single Cell (BB/CCG1720/1) at Earlham Institute, by members of the Genomics Pipelines Group.

Data availability

Sequence data that support the findings of this study have been deposited in the Sequence Read Archive (https://www.ncbi.nlm.nih.gov/sra). Accession numbers are available in Supplementary Table 2. The complete genome and plasmid sequence for ST313 L3 strain BKQZM9 can be found under bioproject ID PRJNA656707, specifically, complete chromosome (GenBank accession: CP060169), pSLT (GenBank Accession: CP060170), pBT3 (GenBank Accession: CP060171). The following three *Salmonella enterica* serovar Typhimurium strains are in the process of being deposited in the DSMZ strain collection: A130 (ST313 L1), D23580 (ST313 L2) and BKQZM9 (ST313 L3). Publicly available sequence data were downloaded from one of the following sources: GenBank (https://www.ncbi.nlm.nih.gov/genbank/), Sequence Read Archive (https://www.ncbi.nlm.nih.gov/sra), European Nucleotide Archive (https://www.ebi.ac.uk/ena) or Enterobase (https://enterobase.warwick.ac.uk). Accession numbers are listed in Supplementary Table 3. The data underlying Figures 1, 2, 3, 4, 5, S1 and S2 are available in Supplementary Table 2. The data summarised in Extended Data 2 are available in Supplementary Table 3. The remaining relevant data are within the manuscript and its supporting information files.

References

1. Stanaway JD, et al. The global burden of non-typhoidal salmonella invasive disease: a systematic analysis for the Global Burden of Disease Study 2017. Lancet Infect Dis. 2019; 19:1312–1324. [PubMed: 31562022]
2. Feasey NA, Dougan G, Kingsley RA, Heyderman RS, Gordon MA. Invasive non-typhoidal salmonella disease: an emerging and neglected tropical disease in Africa. Lancet. 2012; 379:2489–2499. [PubMed: 22587967]
3. Marchello CS, Dale AP, Pisharody S, Rubach MP, Crump JA. A Systematic Review and Meta-analysis of the Prevalence of Community-Onset Bloodstream Infections among Hospitalized Patients in Africa and Asia. Antimicrob Agents Chemother. 2019; 64:e01974–19. [PubMed: 31636071]
4. Reddy EA, Shaw AV, Crump JA. Community-acquired bloodstream infections in Africa: a systematic review and meta-analysis. Lancet Infect Dis. 2010; 10:417–432. [PubMed: 20510282]
5. Kingsley RA, et al. Epidemic multiple drug resistant Salmonella Typhimurium causing invasive disease in sub-Saharan Africa have a distinct genotype. Genome Res. 2009; 19:2279–2287. [PubMed: 19901036]
6. Okoro CK, et al. Intracontinental spread of human invasive Salmonella Typhimurium pathovariants in sub-Saharan Africa. Nat Genet. 2012; 44:1215–1221. [PubMed: 23023330]
7. Ashton PM, et al. Public health surveillance in the UK revolutionises our understanding of the invasive Salmonella Typhimurium epidemic in Africa. Genome Med. 2017; 9:92. [PubMed: 29084588]
8. Almeida F, et al. Multilocus sequence typing of Salmonella Typhimurium reveals the presence of the highly invasive ST313 in Brazil. Infect Genet Evol. 2017; 51:41–44. [PubMed: 28288927]

Europe PMC Funders Author Manuscripts
Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

9. Feasey NA, et al. Three Epidemics of Invasive Multidrug-Resistant Salmonella Bloodstream Infection in Blantyre, Malawi, 1998-2014. Clin Infect Dis. 2015; 61(4):S363–71. [PubMed: 26449953]
10. Feasey NA, et al. Drug Resistance in Salmonella enterica ser. Typhimurium Bloodstream Infection, Malawi. Emerg Infect Dis. 2014; 20:1957–1959. [PubMed: 25340988]
11. Van Puyvelde S, et al. An African Salmonella Typhimurium ST313 sublineage with extensive drug-resistance and signatures of host adaptation. Nat Commun. 2019; 10
12. Piccini G, Montomoli E. Pathogenic signature of invasive non-typhoidal Salmonella in Africa: implications for vaccine development. Hum Vaccin Immunother. 2020; :1–16. DOI: 10.1080/21645515.2020.1785791
13. Bogomolnaya LM, et al. The ABC-Type Efflux Pump MacAB Protects Salmonella enterica serovar Typhimurium from Oxidative Stress. MBio. 2013; 4:e00630–13. [PubMed: 24169575]
14. Okoro CK, et al. Signatures of adaptation in human invasive Salmonella Typhimurium ST313 populations from sub-Saharan Africa. PLoS Negl Trop Dis. 2015; 9:e0003611. [PubMed: 25803844]
15. Singletary LA, et al. Loss of Multicellular Behavior in Epidemic African Nontyphoidal Salmonella enterica Serovar Typhimurium ST313 Strain D23580. MBio. 2016; 7:e02265. [PubMed: 26933058]
16. Carden SE, et al. Pseudogenization of the Secreted Effector Gene sseI Confers Rapid Systemic Dissemination of S. Typhimurium ST313 within Migratory Dendritic Cells. Cell Host Microbe. 2017; 21:182–194. [PubMed: 28182950]
17. Canals R, et al. Adding function to the genome of African Salmonella Typhimurium ST313 strain D23580. PLOS Biol. 2019; 17:e3000059. [PubMed: 30645593]
18. Honeycutt JD, et al. Genetic variation in the MacAB-TolC efflux pump influences pathogenesis of invasive Salmonella isolates from Africa. PLOS Pathog. 2020; 16:e1008763. [PubMed: 32834002]
19. Aulicino A, et al. Invasive Salmonella exploits divergent immune evasion strategies in infected and bystander dendritic cell subsets. Nat Commun. 2018; 9
20. Ashton P, et al. Revolutionising Public Health Reference Microbiology using Whole Genome Sequencing: Salmonella as an exemplar. bioRxiv. 2015; doi: 10.1101/033225
21. Jacob JJ, et al. Genomic analysis of human invasive Salmonella enterica serovar Typhimurium ST313 isolate B3589 from India. Infect Genet Evol. 2019; 73:416–424. [PubMed: 31170530]
22. Tran-Dien A, Le Hello S, Bouchier C, Weill F-X. Early transmissible ampicillin resistance in zoonotic Salmonella enterica serotype Typhimurium in the late 1950s: a retrospective, whole-genome sequencing study. Lancet Infect Dis. 2018; 18:207–214. [PubMed: 29198740]
23. Kagambega A, Lienemann T, Frye JG, Barro N, Haukka K. Whole genome sequencing of multidrug-resistant Salmonella enterica serovar Typhimurium isolated from humans and poultry in Burkina Faso. Trop Med Health. 2018; 46:4. [PubMed: 29449781]
24. Rotger R, Casadesus J. The virulence plasmids of Salmonella. Int Microbiol. 1999; 2:177–184. [PubMed: 10943411]
25. Sahin F, Karasartova D, Gerceker D, Aysev AD, Erdem B. [A novel Salmonella Typhimurium plasmid, pAnkS: an example for plasmid evolution in antibiotic resistance]. Mikrobiyol Bul. 2008; 42:383–388. [PubMed: 18822880]
26. Otsuka Y, Yonesaki T. Dmd of bacteriophage T4 functions as an antitoxin against Escherichia coli LsoA and RnlA toxins. Mol Microbiol. 2012; 83:669–681. [PubMed: 22403819]
27. Figueroa-Bossi N, Uzzau S, Maloriol D, Bossi L. Variable assortment of prophages provides a transferable repertoire of pathogenic determinants in Salmonella. Mol Microbiol. 2001; 39:260–271. [PubMed: 11136448]
28. Tucker CP, Heuzenroeder MW. ST64B is a defective bacteriophage in Salmonella enterica serovar Typhimurium DT64 that encodes a functional immunity region capable of mediating phage-type conversion. Int J Med Microbiol. 2004; 294:59–63. [PubMed: 15293455]
29. Hanna LF, Matthews TD, Dinsdale EA, Hasty D, Edwards RA. Characterization of the ELPhiS Prophage from Salmonella enterica Serovar Enteritidis Strain LK5. Appl Environ Microbiol. 2012; 78:1785–1793. [PubMed: 22247173]

Europe PMC Funders Author Manuscripts

30. Feasey NA, et al. Distinct Salmonella Enteritidis lineages associated with enterocolitis in high-income settings and invasive disease in low-income settings. Nat Genet. 2016; 48:1211–1217. [PubMed: 27548315]
31. Pulford CV, et al. Salmonella enterica Serovar Panama, an Understudied Serovar Responsible for Extraintestinal Salmonellosis Worldwide. Infect Immun. 2019; 87:e00273–19. [PubMed: 31262982]
32. Yao K, et al. Complete Genome and Methylome Sequences of Salmonella enterica subsp. enterica Serovar Panama (ATCC 7378) and Salmonella enterica subsp. enterica Serovar Sloterdijk (ATCC 15791). Genome Announc. 2016; 4
33. Owen SV, et al. Characterization of the Prophage Repertoire of African Salmonella Typhimurium ST313 Reveals High Levels of Spontaneous Induction of Novel Phage BTP1. Front Microbiol. 2017; 8:235. [PubMed: 28280485]
34. Owen SV, et al. Prophage-encoded phage defence proteins with cognate self-immunity. bioRxiv. 2020; doi: 10.1101/2020.07.13.199331
35. Susskind MM, Wright A, Botstein D. Superinfection exclusion by P22 prophage in lysogens of Salmonella typhimurium: IV. Genetics and physiology of sieB exclusion. Virology. 1974; 62:367–384. [PubMed: 4610993]
36. Magiorakos A-P, et al. Multidrug-resistant, extensively drug-resistant and pandrug-resistant bacteria: an international expert proposal for interim standard definitions for acquired resistance. Clin Microbiol Infect. 2012; 18:268–281. [PubMed: 21793988]
37. Hooper DC, Jacoby GA. Mechanisms of drug resistance: quinolone resistance. Ann N Y Acad Sci. 2015; 1354:12–31. [PubMed: 26190223]
38. Ma M, Eaton JW. Multicellular oxidant defense in unicellular organisms. Proc Natl Acad Sci U S A. 1992; 89:7924–7928. [PubMed: 1518815]
39. MacKenzie KD, et al. Parallel evolution leading to impaired biofilm formation in invasive Salmonella strains. PLOS Genet. 2019; 15:e1008233. [PubMed: 31233504]
40. Post AS, et al. Supporting evidence for a human reservoir of invasive non-Typhoidal Salmonella from household samples in Burkina Faso. PLoS Negl Trop Dis. 2019; 13:e0007782. [PubMed: 31609964]
41. Crump JA, et al. Investigating the meat pathway as a source of human nontyphoidal Salmonella bloodstream infections and diarrhea in East Africa. Clin Infect Dis. 2020; :ciaa1153.doi: 10.1093/cid/ciaa1153 [PubMed: 32777036]
42. Nguyen BD, et al. Import of Aspartate and Malate by DcuABC Drives H2/Fumarate Respiration to Promote Initial Salmonella Gut-Lumen Colonization in Mice. Cell Host Microbe. 2020; 27:922–936.e6 [PubMed: 32416061]
43. Lawley TD, et al. Genome-Wide Screen for Salmonella Genes Required for Long-Term Systemic Infection of the Mouse. PLOS Pathog. 2006; 2:e11. [PubMed: 16518469]
44. Wood MW, et al. Identification of a pathogenicity island required for Salmonella enteropathogenicity. Mol Microbiol. 1998; 29:883–891. [PubMed: 9723926]
45. Moreira CG, et al. Virulence and stress-related periplasmic protein (VisP) in bacterial/host associations. Proc Natl Acad Sci. 2013; 110:1470 LP–1475. [PubMed: 23302685]
46. McKnight PE, Najab J. Mann-Whitney U Test. The Corsini Encyclopedia of Psychology. 2010; 1doi: 10.1002/9780470479216.corpsy0524
47. Wheeler NE, Gardner PP, Barquist L. Machine learning identifies signatures of host adaptation in the bacterial pathogen Salmonella enterica. PLoS Genet. 2018; 14:e1007333. [PubMed: 29738521]
48. Langridge GC, et al. Simultaneous assay of every Salmonella Typhi gene using one million transposon mutants. Genome Res. 2009; 19:2308–2316. [PubMed: 19826075]
49. Lopez-Garrido J, Cheng N, Garcia-Quintanilla F, Garcia-del Portillo F, Casadesus J. Identification of the Salmonella enterica damX gene product, an inner membrane protein involved in bile resistance. J Bacteriol. 2010; 192:893–895. [PubMed: 19948803]
50. Lopez CA, Rivera-Chavez F, Byndloss MX, Baumler AJ. The Periplasmic Nitrate Reductase NapABC Supports Luminal Growth of Salmonella enterica Serovar Typhimurium during Colitis. Infect Immun. 2015; 83:3470–3478. [PubMed: 26099579]

Europe PMC Funders Author Manuscripts

51. Tsolis RM, Xavier MN, Santos RL, Bäumler AJ. How to become a top model: impact of animal experimentation on human Salmonella disease research. Infect Immun. 2011; 79:1806–1814. [PubMed: 21343352]
52. Tack B, Vanaenrode J, Verbakel JY, Toelen J, Jacobs J. Invasive non-typhoidal Salmonella infections in sub-Saharan Africa: a systematic review on antimicrobial resistance and treatment. BMC Med. 2020; 18:212. [PubMed: 32677939]
53. Langridge GC, et al. Patterns of genome evolution that have accompanied host adaptation in Salmonella. Proc Natl Acad Sci U S A. 2015; 112:863–868. [PubMed: 25535353]
54. Perez-Sepulveda BM, et al. An accessible, efficient and global approach for the large-scale sequencing of bacterial genomes. bioRxiv. 2020; doi: 10.1101/2020.07.22.200840
55. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014; 30:2114–2120. [PubMed: 24695404]
56. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol. 2017; 13:e1005595. [PubMed: 28594827]
57. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. Bioinformatics. 2013; 29:1072–1075. [PubMed: 23422339]
58. Alikhan N-F, Zhou Z, Sergeant MJ, Achtman M. A genomic overview of the population structure of Salmonella. PLOS Genet. 2018; 14:e1007261. [PubMed: 29621240]
59. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014; 30:2068–2069. [PubMed: 24642063]
60. Yoshida CE, et al. The Salmonella In Silico Typing Resource (SISTR): An Open Web-Accessible Tool for Rapidly Typing and Subtyping Draft Salmonella Genome Assemblies. PLoS One. 2016; 11:e0147101. [PubMed: 26800248]
61. Larsen MV, et al. Multilocus sequence typing of total-genome-sequenced bacteria. J Clin Microbiol. 2012; 50:1355–1361. [PubMed: 22238442]
62. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009; 25:1754–1760. [PubMed: 19451168]
63. Li H, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009; 25:2078–2079. [PubMed: 19505943]
64. McKenna A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010; 20:1297–1303. [PubMed: 20644199]
65. García-Alcalde F, et al. Qualimap: evaluating next-generation sequencing alignment data. Bioinformatics. 2012; 28:2678–2679. [PubMed: 22914218]
66. Croucher NJ, et al. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. Nucleic Acids Res. 2014; 43:e15–e15. [PubMed: 25414349]
67. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014; 30:1312–1313. [PubMed: 24451623]
68. Tonkin-Hill G, Lees JA, Bentley SD, Frost SDW, Corander J. RhierBAPS: An R implementation of the population clustering algorithm hierBAPS. Wellcome Open Res. 2018; 3:93. [PubMed: 30345380]
69. Letunic I, Bork P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res. 2016; 44:W242–5. [PubMed: 27095192]
70. Page AJ, et al. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics. 2015; 31:3691–3693. [PubMed: 26198102]
71. Page AJ, et al. SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. Microb genomics. 2016; 2:e000056.
72. Bouckaert R, et al. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. PLOS Comput Biol. 2019; 15:e1006650. [PubMed: 30958812]
73. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. Syst Biol. 2018; 67:901–904. [PubMed: 29718447]
74. Bouckaert RR. DensiTree: making sense of sets of phylogenetic trees. Bioinformatics. 2010; 26:1372–1373. [PubMed: 20228129]
75. R Core Team. R: A language and environment for statistical computing. 2017

Europe PMC Funders Author Manuscripts

76. Inouye M, et al. SRST2: Rapid genomic surveillance for public health and hospital microbiology labs. Genome Med. 2014; 6:90. [PubMed: 25422674]
77. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Mol Biol Evol. 2018; 35:1547–1549. [PubMed: 29722887]
78. Zankari E, et al. Identification of acquired antimicrobial resistance genes. J Antimicrob Chemother. 2012; 67:2640–2644. [PubMed: 22782487]
79. Zankari E, et al. PointFinder: a novel web tool for WGS-based detection of antimicrobial resistance associated with chromosomal point mutations in bacterial pathogens. J Antimicrob Chemother. 2017; 72:2764–2768. [PubMed: 29091202]
80. Matuschek E, Brown DFJ, Kahlmeter G. Development of the EUCAST disk diffusion antimicrobial susceptibility testing method and its implementation in routine microbiology laboratories. Clin Microbiol Infect. 2014; 20:O255–66. [PubMed: 24131428]
81. Nishino K, Latifi T, Groisman EA. Virulence and drug resistance roles of multidrug efflux systems of Salmonella enterica serovar Typhimurium. Mol Microbiol. 2006; 59:126–141. [PubMed: 16359323]
82. Yang J, et al. Characterization of the Invasive, Multidrug Resistant Non-typhoidal Salmonella Strain D23580 in a Murine Model of Infection. PLoS Negl Trop Dis. 2015; 9:e0003839. [PubMed: 26091096]
83. Gibbons HS, Lin S, Cotter RJ, Raetz CR. Oxygen requirement for the biosynthesis of the S-2-hydroxymyristate moiety in Salmonella typhimurium lipid A. Function of LpxO, A new Fe2+/alpha-ketoglutarate-dependent dioxygenase homologue. J Biol Chem. 2000; 275:32940–32949. [PubMed: 10903325]
84. Jain M, Olsen HE, Paten B, Akeson M. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. Genome Biol. 2016; 17:239. [PubMed: 27887629]
85. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990; 215:403–410. [PubMed: 2231712]
86. Carver TJ, et al. ACT: the Artemis Comparison Tool. Bioinformatics. 2005; 21:3422–3423. [PubMed: 15976072]
87. Conway JR, Lex A, Gehlenborg N. UpSetR: an R package for the visualization of intersecting sets and their properties. Bioinformatics. 2017; 33:2938–2940. [PubMed: 28645171]
88. Wickham, H. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag; 2016.
89. Msefula CL, et al. Genotypic Homogeneity of Multidrug Resistant S. Typhimurium Infecting Distinct Adult and Childhood Susceptibility Groups in Blantyre, Malawi. PLoS One. 2012; 7:e42085. [PubMed: 22848711]

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

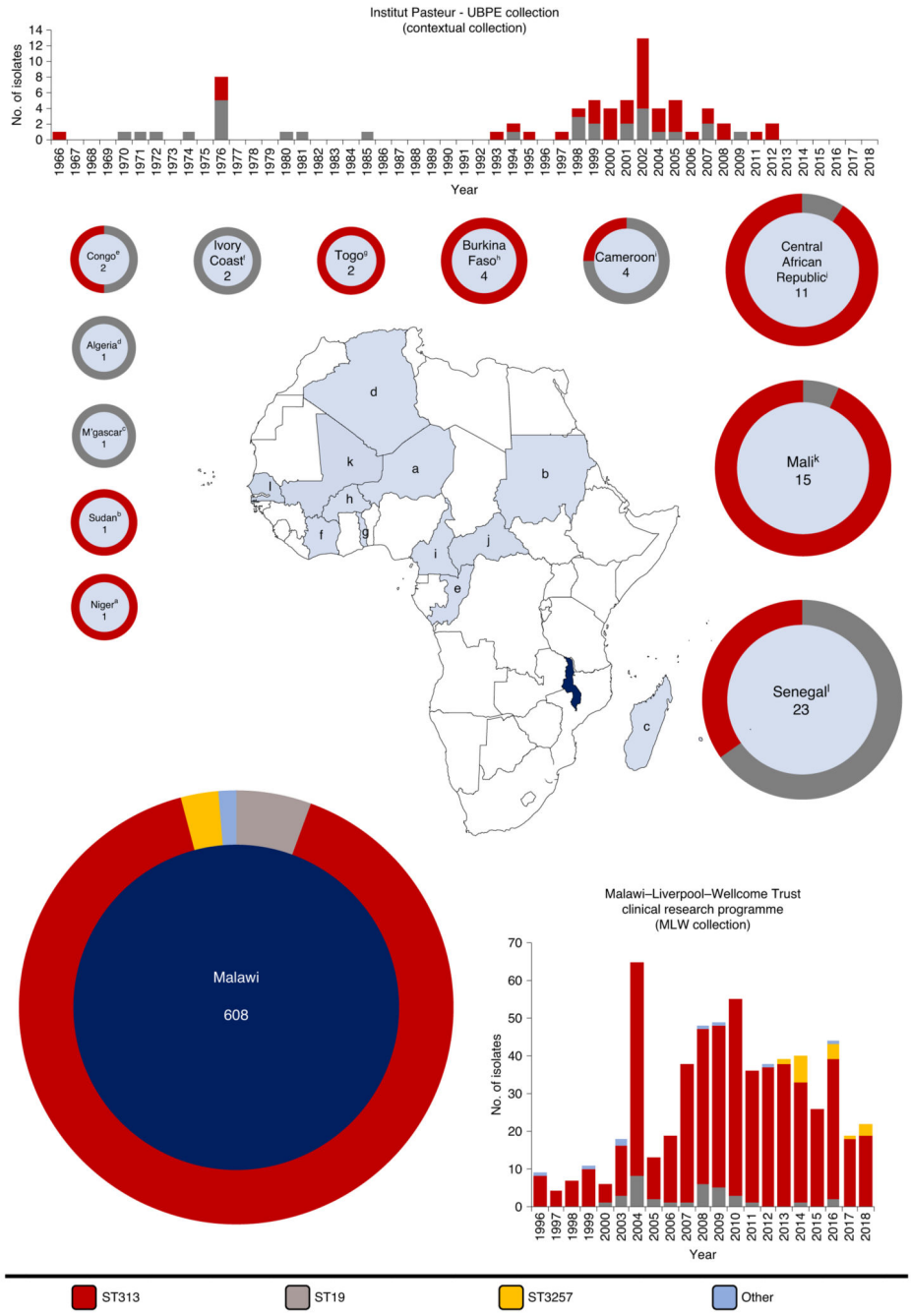


Figure 1. Bloodstream isolates of *Salmonella* Typhimurium used in this study.

Isolates were collected from either the Unité des Bactéries pathogènes entériques (UBPE) of the Institut Pasteur Centres ($n = 72$, light blue) or the Malawi-Liverpool Wellcome Trust Clinical Research Center ($n = 608$, dark blue). Bar graphs show the number of isolates of different sequence types collected by centre per year. Donut charts indicate the proportion of sequence types collected per country and show the total number of isolates from each location in the centre. Note that only isolates from African countries are shown in donut plots. Madagascar is written as M'gascar and Republic of the Congo is written as Congo.

Letters appearing in superscript relate to location on map. The colour code is shown at the bottom of the figure, with red indicating *S.* Typhimurium ST313.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

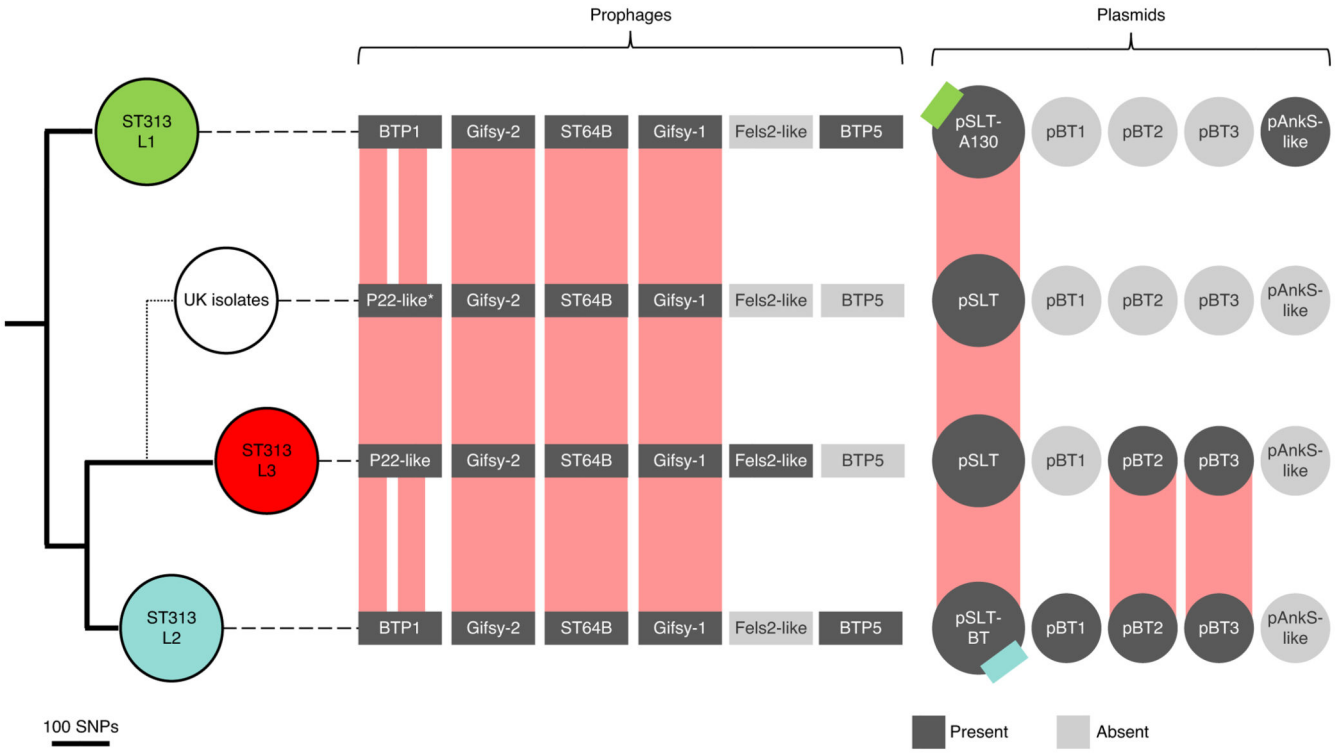


Figure 2. The four major ST313 clusters have different prophage and plasmid repertoires

Maximum likelihood phylogeny demonstrating the population structure of ST313 L1, L2 and L3. Approximate location of UK lineages is indicated schematically due to the diversity of isolates (see Extended Data 2). Prophage and plasmid repertoire of the reference strain for each lineage are shown (dark grey indicates presence; light grey represents absence). Reference isolates used were A130 (ST313 L1), D23580 (ST313 L2), BKQZM9 (ST313 L3) and U2 (UK strains). Red blocks represent the extent of conservation between lineage reference genomes, with white gaps indicating missing regions. Coloured squares on the pSLT plasmid represents different AMR cassettes.

Europe PMC Funders Author Manuscripts
Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

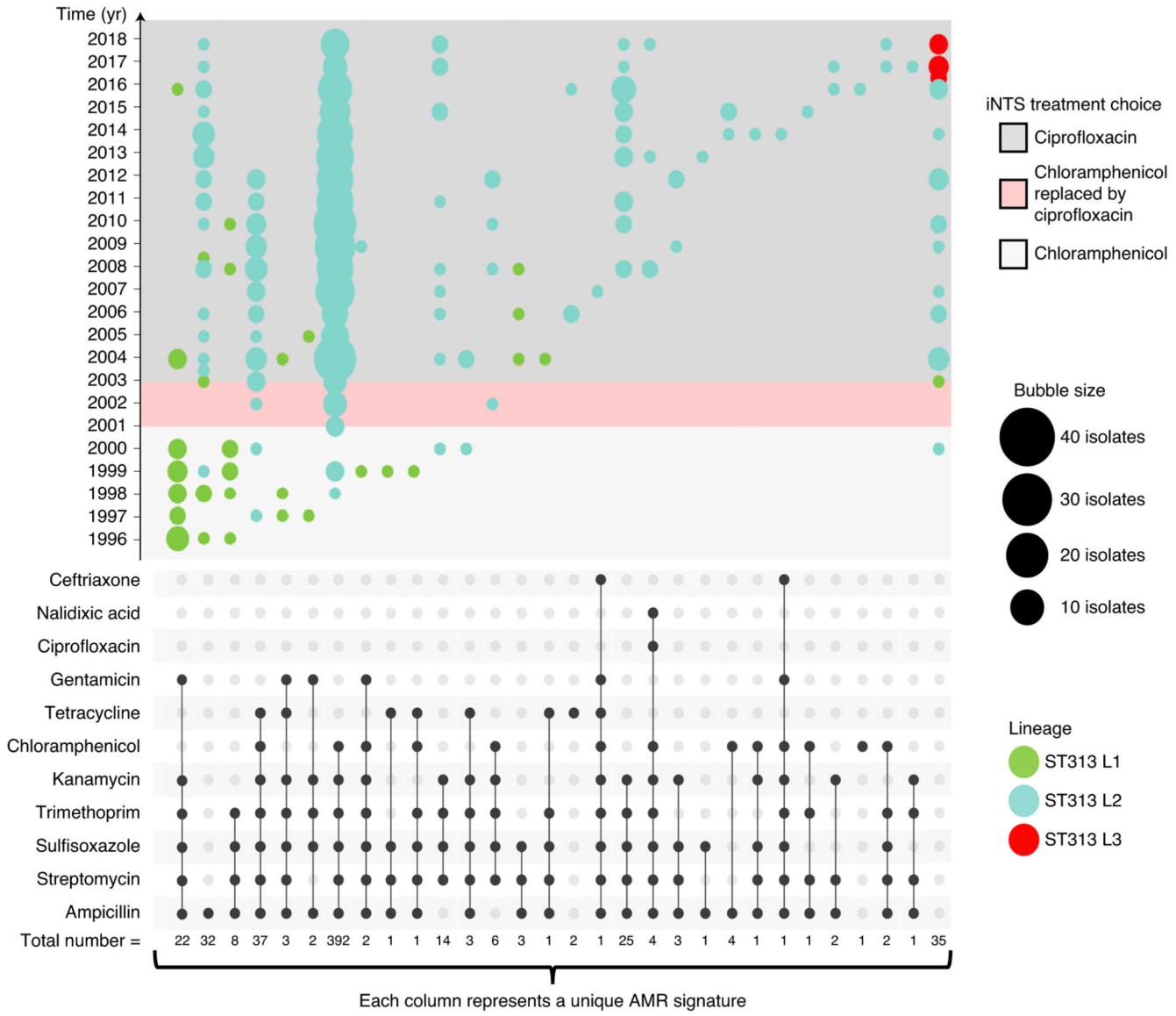


Figure 3. Temporal AMR Trends in *S.* Typhimurium lineages (1996 – 2018)

Combination matrix (lower) depicts predicted AMR patterns of *S.* Typhimurium. Isolates collected before 1996 were only sampled sporadically, and thus were excluded from analysis. Within the combination matrix, dark grey circles indicate genome-predicted resistance and vertical combination of grey circles represents resistance profile. Total number of isolates with each resistance profile is indicated below matrix. The combination matrix was created using the UpSetR package[87]. Bubble plot (above) depicts the relative number of isolates (bubble size) with each resistance profile (combination matrix) per year (y axis). Lineage assignments are shown using bubble colour to identify lineage-specific AMR trends. Local Malawi antimicrobial usage policy[9] is highlighted in background colours overlying the timeline in the bubble plot. Bubble plot was created using R ggplot2[88].

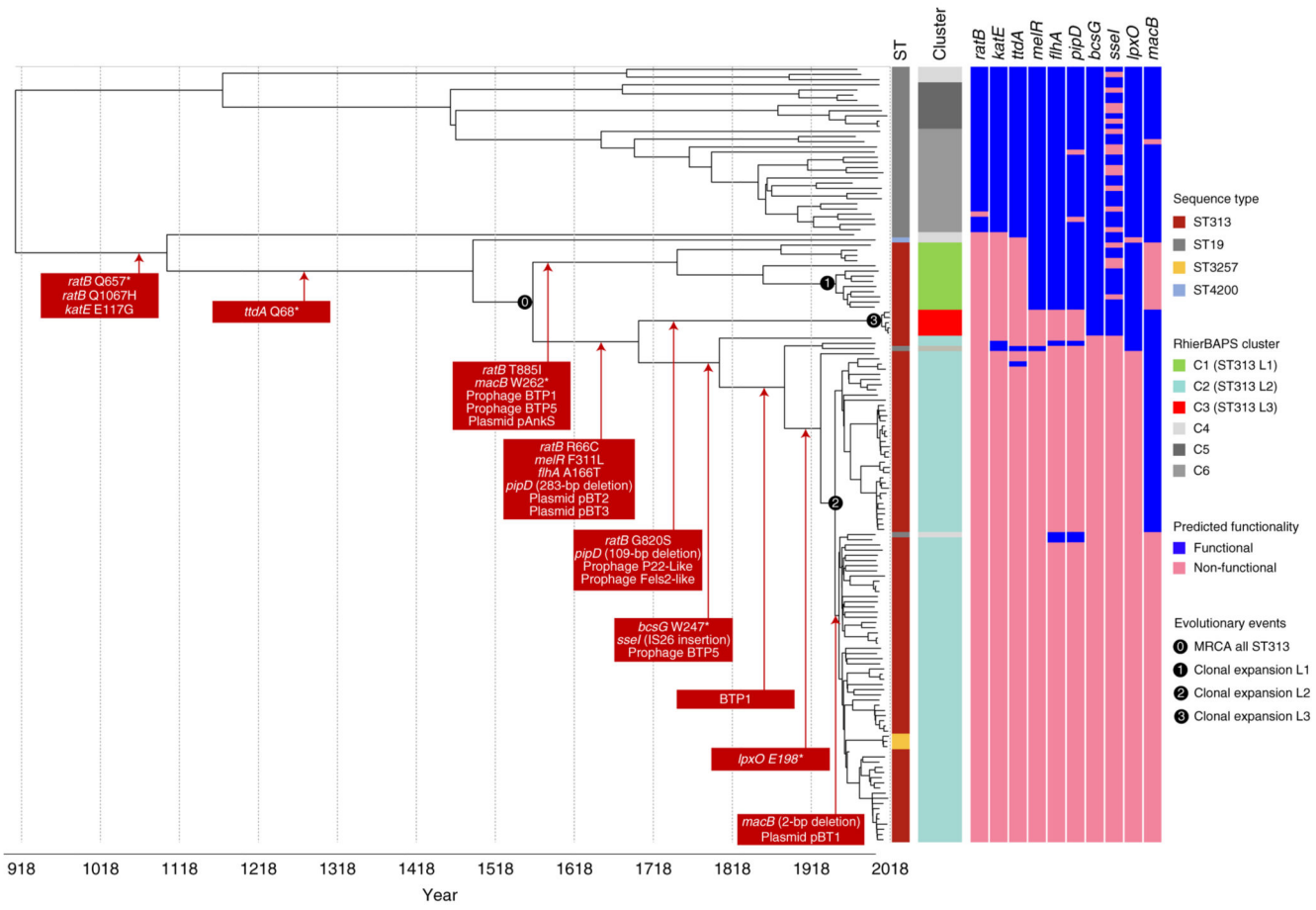


Figure 4. Stepwise evolution of *S*. Typhimurium responsible for BSI in Africa

Chronograph of 150 *S.* Typhimurium strains isolated from bloodstream of human iNTS disease patients. The choice of the 150 isolates from both the Malawian and contextual datasets is described in Methods. The figure shows a maximum clade credibility tree. Adjacent colour strips are as follows (from left to right); ST, lineage assignment (rHierBAPs) with the three major ST313 lineages highlighted in colour. Predicted functionality (Methods) is depicted as a colour strip for each gene and is based on whole genome-based predictions of SNPs likely to play a functional role. Genome degradation events that generate functionally relevant pseudogenes are displayed in red boxes overlying the chronograph. An asterix (*) represents a premature stop codon. The numbers indicate four key evolutionary events. Figure visualised using iTOL[69].

Europe PMC Funders Author Manuscripts
Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

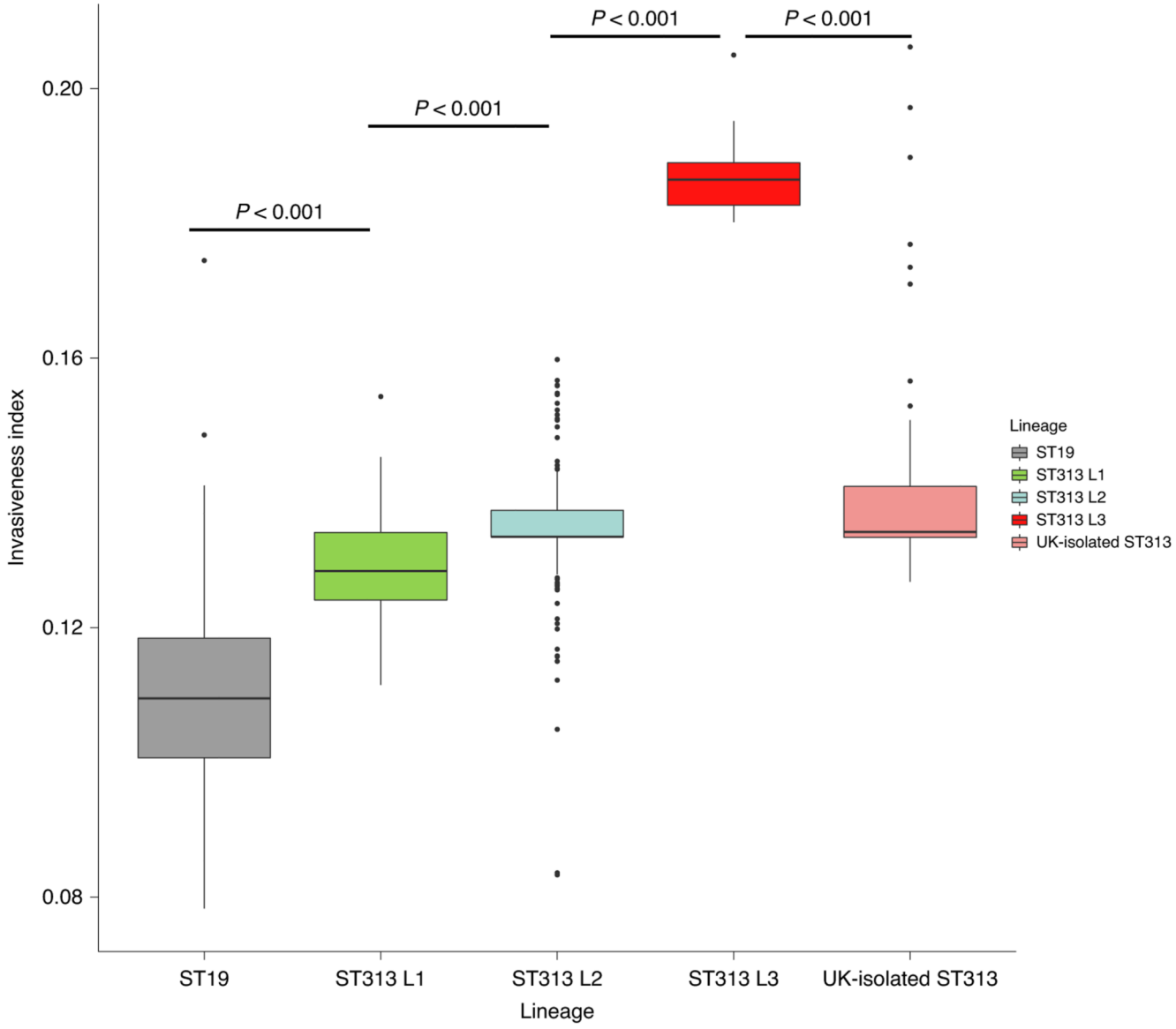


Figure 5. Invasiveness index of ST19 and ST313 lineages

Box plot representing the distribution of invasiveness index values for all genome sequences included in this study summarised by lineage assignment. UK-isolated ST313 refer to those published by Ashton *et al.* (2017)[7]. Groups were compared using a two-sided Wilcox Mann Whitney test[46] and the resultant *p*-values were all <0.001. Number of isolates in each group; ST19 (*n* = 66), ST313 L1 (*n* = 52), ST313 L2 (*n* = 550), ST313 L3 (*n* = 9), UK-isolated ST313 (*n* = 59). Boxplot centre lines represent median values, boxplot limits represent upper and lower quartiles, boxplot whiskers represent the 1.5 interquartile range and individual points represent outliers. Boxplot was created using R ggplot2[88].

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Table 1

Summary of the phenotypic impact of pseudogenes in ST313 lineages*

Gene Name	Description	Phenotype
ratB **(STM2514)**	**Secreted outer membrane protein**	**Inactivation of *ratB* is associated with gut persistence in mouse model and is likely to have reduced the enteric potential of ST313 [14].**
katE **(STM1318)**	**Stationary phase catalase**	**Catalase activity is reduced across all ST313 L1 & L2 isolates [7,15]. We confirmed that ST313 L3 has a lower catalase activity than ST19.**
ttdA **(STM3355)**	**L(+)-tartrate dehydratase**	**L-tartaric acid and dihydroxyacetone cannot be used as sole carbon sources by ST313 L1 & L2 [14]. We confirmed experimentally that ST313 L3 are unable to grow on L-tartaric acid as a sole carbon source.**
melR **(STM4297)**	**Melibiose operon response regulator**	**ST313 L2 cannot grow on melibiose as a sole carbon source [17,82]. We confirmed experimentally that ST313 L3 was unable to grow on melibiose as a sole carbon source.**
flhA **(STM1913)**	**Flagella biosynthesis protein**	**We previously showed that the A166T SNP carried by ST313 L2 and ST313 L3 causes reduced motility, by comparing the ST313 L2 D23580 *flhA*^{474} mutant to wild type [17].**
pipD **(STM1094)**	**Pathogenicity island encoded protein D**	**Although a causal relationship is unproven, the PipD effector protein of ST19 contributes to macrophage persistence in murine models [14,17,43].**
bcsG **(STM3624)**	**Cellulose biosynthetic enzyme**	**Biofilm formation is impaired in ST313 L2 due to a mutation in *bcsG* [15]. Here, we showed that biofilm formation in ST313 L3 is also impaired. The genetic basis is unknown.**
sseI **(STM1051)**	**Type III secretion effector protein (SPI-2)**	**Studies in mice demonstrate that an accumulation of SNPs within the *sseI* gene of ST313 L2 increases the ability of the bacteria to disseminate rapidly from the gut to the draining lymph nodes [16].**
lpxO **(STM4286)**	**Putative membrane-bound β-hydroxylase**	**We compared the lipid A of ST313 L2 (D23580) and ST19 (4/74) during growth in SPI2-inducing media. By mass spectrometry, we confirmed that the lipid A of ST313 L2 lacks the LpxO-mediated modification associated with ST19.**
macB **(STM0942)**	**Putative ABC transport protein**	**MacAB is involved in oxidative stress resistance [13] and antibiotic resistance to macrolides [81]. In ST313 L1 and L2, variants of the MacAB-TolC tripartite efflux pump affect replication in macrophages and influence fitness during colonisation of the murine gastrointestinal tract [18].**

*Associated phenotypic data are presented in Supplementary Table 4.