

Published in final edited form as:

*Nat Genet.* 2012 November ; 44(11): 1215–1221. doi:10.1038/ng.2423.

## Intra-continental spread of human invasive *Salmonella* Typhimurium pathovariants in sub-Saharan Africa

Chinyere K. Okoro<sup>1,20</sup>, Robert A. Kingsley<sup>1,20</sup>, Thomas R. Connor<sup>1</sup>, Simon R. Harris<sup>1</sup>, Christopher M. Parry<sup>2,3</sup>, Manar N Al-Mashhadani<sup>3</sup>, Samuel Kariuki<sup>4</sup>, Chisomo L. Msefula<sup>5,6</sup>, Melita A. Gordon<sup>7</sup>, Elizabeth de Pinna<sup>8</sup>, John Wain<sup>8,9</sup>, Robert S. Heyderman<sup>5,10</sup>, Stephen Obaro<sup>11,12</sup>, Pedro L. Alonso<sup>13,14</sup>, Inacio Mandomando<sup>14,15</sup>, Calman A. MacLennan<sup>16,17</sup>, Milagritos D. Tapia<sup>18</sup>, Myron M. Levine<sup>18,19</sup>, Sharon M Tennant<sup>19</sup>, Julian Parkhill<sup>1</sup>, and Gordon Dougan<sup>§,1</sup>

<sup>1</sup>Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton Cambridge, UK. CB10 1SA <sup>2</sup>Mahidol-Oxford Tropical Medicine Research Unit, Faculty of Tropical Medicine, Mahidol University, Bangkok, Thailand <sup>3</sup>Department of Clinical Infection, Microbiology and Immunology, Institute for Infection and Global Health, University of Liverpool, Liverpool, UK <sup>4</sup>Centre for Microbiology Research, Kenya Medical Research Institute, Nairobi, Kenya <sup>5</sup>Malawi-Liverpool-Wellcome Trust Clinical Research Program, University of Malawi College of Medicine, Blantyre, Malawi <sup>6</sup>Department of Microbiology, College of Medicine, University of Malawi, Blantyre, Malawi <sup>7</sup>Department of Gastroenterology, Institute of Translational Medicine, Liverpool University, Liverpool, UK. L69 3GE <sup>8</sup>Health Protection Agency, Laboratory for Gastrointestinal Infections, Centre for Infections, London, United Kingdom <sup>9</sup>Norwich Medical School, University of East Anglia, Norwich, United Kingdom <sup>10</sup>Cellular and Molecular Medicine, School of Medical Sciences, University of Bristol, Bristol, United Kingdom <sup>11</sup>Division of Paediatric Infectious Diseases, Department of Paediatrics and Human Development, Michigan State University, East Lansing, MI 48824, USA <sup>12</sup>National Hospital Abuja, Plot 132 Central District (Phase II), Garki Abuja, Nigeria <sup>13</sup>Barcelona Centre for International Health Research (CRESIB, Hospital Clínic-Universitat de Barcelona), Barcelona, Spain <sup>14</sup>Centro de Investigação em Saúde de Manhiça (CISM), Manhiça, Mozambique <sup>15</sup>Instituto Nacional de Saúde, Ministerio de Saúde, Maputo, Mozambique <sup>16</sup>Novartis Vaccines Institute for Global Health S.r.l. (NVGH), Via Fiorentina 1, 53100 Siena, Italy <sup>17</sup>MRC Centre for Immune Regulation, School of Immunity and Infection,

§Corresponding author .

<sup>20</sup>These authors contributed equally to this work

### URLs.

European Nucleotide Archive (ENA), <http://www.ebi.ac.uk/ena>

MLST database, <http://mlst.ucc.ie/mlst/mlst/dbs/Senterica/>

AIDSInfoOnline.mdb, <http://www.aidsinfoonline.org/>

UNAIDS, <http://www.unaids.org/en/>

UNAIDS Report on the Global AIDS Epidemic 2010, [http://www.unaids.org/globalreport/global\\_report.htm](http://www.unaids.org/globalreport/global_report.htm)

Google Earth, [http://www.google.co.uk/intl/en\\_uk/earth/index.html](http://www.google.co.uk/intl/en_uk/earth/index.html)

**ACCESSION NUMBERS** Referenced accession codes for data deposited in the NCBI Nucleotide database include: FQ312003, FN424405, HE654726, FN432031 and AE006471. The full set of primary accession codes for the Illumina sequence reads of 177 invasive and gastrointestinal *Salmonella* Typhimurium is given in Supplementary Table 1.

**AUTHOR CONTRIBUTIONS** Chinyere K. Okoro & Robert A. Kingsley

CKO, RAK contributed to collecting data and manuscript writing. CKO analyzed sequence data, performed phylogenetic, BEAST and comparative genomics analyses. TRC and SRH wrote the coding scripts for phylogenetic and Bayesian statistical analyses and contributed to manuscript writing. CMP, MNA-M, SK, CLM, MAG, ED-P, RSH, SO, PLA, IM, CAM, JW, MDT, MML and SMT contributed to studies from which isolates were drawn and to manuscript writing. GD, JP and RAK designed the study and GD supervised the study.

**COMPETING FINANCIAL INTERESTS** The authors declare no competing financial interests.

College of Medicine and Dental Sciences, University of Birmingham, Birmingham, B15 2TT, UK  
<sup>18</sup>Center for Vaccine Development, University of Maryland, Baltimore, HSFI 480, 685 West  
 Baltimore St., Baltimore, MD 21201, USA <sup>19</sup>Department of Medicine, University of Maryland,  
 Baltimore, HSFI 480, 685 West Baltimore St., Baltimore, MD 21201, USA

## Abstract

A highly invasive form of non-typhoidal *Salmonella* (iNTS) disease has been recently documented in many countries in sub-Saharan Africa. The most common *Salmonella enterica* serovar causing this disease is Typhimurium. We applied whole-genome sequence-based phylogenetic methods to define the population structure of sub-Saharan African invasive *Salmonella* Typhimurium and compared these to global *Salmonella* Typhimurium isolates. Notably, the vast majority of sub-Saharan invasive *Salmonella* Typhimurium fell within two closely-related, highly-clustered phylogenetic lineages that we estimate emerged independently ~52 and ~35 years ago, in close temporal association with the current HIV pandemic. Clonal replacement of isolates of lineage I by lineage II was potentially influenced by the use of chloramphenicol for the treatment of iNTS disease. Our analysis suggests that iNTS disease is in part an epidemic in sub-Saharan Africa caused by highly related *Salmonella* Typhimurium lineages that may have occupied new niches associated with a compromised human population and antibiotic treatment.

## INTRODUCTION

*S. enterica* is a diverse bacterial species that remains a common cause of infectious disease in humans and animals throughout the world<sup>1</sup>. Human *Salmonella* infections are classically divided into diseases caused by typhoidal or non-typhoidal salmonella (NTS). The former category includes the human restricted *S. enterica* serovars Typhi and Paratyphi that cause the systemic disease typhoid, while NTS is comprised of the majority of other serovars that predominantly cause self-limiting gastroenteritis in humans<sup>2</sup>. *Salmonella* Typhi is a human restricted pathogen that is transmitted from human to human, whereas NTS disease is normally associated with zoonotic reservoirs, typically domesticated animals, with little or no sustained human-to-human transmission.

In contrast to this classical view, NTS are a frequent cause of invasive bacterial disease in many countries in sub-Saharan Africa<sup>3,4</sup>. This invasive form of NTS disease (iNTS) is common both in children with malnutrition, severe anemia, malaria or HIV<sup>4,5</sup> and HIV-infected adults<sup>6</sup>, frequently surpassing *Salmonella* Typhi in many parts of the region as the dominant cause of invasive salmonellosis. The clinical presentation of this disease is distinct from both gastroenteritis and typhoid fever, characterized by a non-specific fever that can be indistinguishable from malaria and in rare cases is accompanied by diarrhea<sup>7</sup>. The frequency of NTS-associated case fatalities can be extremely high in both adults and children (22% – 45%)<sup>6,8-10</sup>.

*S. enterica* serovar Typhimurium (*Salmonella* Typhimurium) is one of the serovars most frequently associated with iNTS in the sub-Saharan region, although other serovars including *S. enterica* serovar Enteritidis have also been implicated<sup>3,4,8</sup>. We previously reported that *Salmonella* Typhimurium isolates from Kenya and Malawi were predominantly of a novel multi-locus sequence type (ST) designated ST313<sup>7</sup>, rarely isolated from outside sub-Saharan Africa. The DNA sequence of representative multi-drug-resistant (MDR) ST313 isolates, D23580 and A130, identified genomic features distinct from previously characterized gastroenteritis-associated strains<sup>7</sup>. These features included

evidence of partial genome degradation, with some parallels to that observed in the *S. enterica* serovars Typhi and Paratyphi A that has been linked to niche adaptation<sup>11,12</sup>.

Here we use single nucleotide polymorphism (SNP)-based phylogenetic methods based on whole genome sequences to determine the population structure of a geographically-diverse collection of invasive *Salmonella* Typhimurium isolates from different sub-Saharan African countries. These data are placed in the phylogenetic context of *Salmonella* Typhimurium isolates from other parts of the world. We provide evidence that two tightly clustered genetic lineages have emerged within the last sixty years to be a dominant cause of an epidemic invasive *Salmonella* Typhimurium disease in the region. We highlight the potential role of antibiotic-resistance acquisition in driving the epidemic and the temporal association with an increased prevalence of HIV.

## RESULTS

### Phylogenetic analysis of *Salmonella* Typhimurium

*Salmonella* Typhimurium represents an unstratified serologically-defined group within the broader species *S. enterica*<sup>13</sup>. Therefore, in order to place the invasive *Salmonella* Typhimurium isolates from sub-Saharan Africa into an evolutionary and phylogenetic context we exploited whole genome sequencing to discover potentially informative SNPs within a collection of 179 *Salmonella* Typhimurium spanning the period 1938 and 2010, collected from different parts of the world. Our collection included 129 invasive *Salmonella* Typhimurium from Malawi, Kenya, Mozambique, Uganda, The Democratic Republic of Congo (DRC), Nigeria and Mali (Supplementary Table 1). Data were available for 10,623 high-quality SNPs, corresponding to approximately 1 SNP for every 407 bp, that were distributed relatively uniformly across the genome of the reference *Salmonella* Typhimurium SL1344. To refine phylogenetic analysis, SNPs associated with repetitive sequences, mobile elements and phage sequences, representing ~4% of the genome, were excluded. We detected no evidence of extensive recombination within the remaining genomic sequences and consequently SNPs mapping to these regions were used to reconstruct a maximum-likelihood phylogenetic tree<sup>14</sup> (Fig. 1).

Notably, invasive *Salmonella* Typhimurium isolates from sub-Saharan Africa fall predominantly into two distinct ST313 phylogenetic lineages designated as lineages I and II. Furthermore, these lineages form distinct and extremely tight clusters on separate branches from other *Salmonella* Typhimurium that were isolated elsewhere in the world. The tight clustering is illustrated by the fact that isolates within either lineage I or lineage II are separated by mean differences as few as 33 and 21 SNPs respectively. The two invasive *Salmonella* Typhimurium lineages are joined to the main tree by relatively long branches, but there is divergence at the branch tips suggesting recent clonal or population expansion. Both lineages are thus more closely related to each other than to any other *Salmonella* Typhimurium within the tree. The two invasive *Salmonella* Typhimurium lineages are joined to the main tree by relatively long branches, but there has been divergence at the branch tips suggesting recent clonal or population expansion. Multi-Locus-Sequence-Typing (MLST) analysis confirmed lineages I and II as ST313, although a single isolate, 5580 from lineage I is ST394, a single locus variant of ST313 (Supplementary Fig. 1). All eight invasive *Salmonella* Typhimurium isolates from sub-Saharan Africa which fall outside of lineages I and II are ST19, a common ST to which 82% (41/50) of the 50 non-African *Salmonella* Typhimurium that we sequenced belong. Other STs represented in the non sub-Saharan *Salmonella* Typhimurium lineages include ST34 (5/50), ST98 (1/50), ST128 (2/50) and ST568 (2/50) (Supplementary Fig. 1).

## Temporal and geographic distribution relative to phylogeny

We performed BEAST<sup>15</sup> analysis on 129 sub-Saharan invasive *Salmonella* Typhimurium isolates from seven sub-Saharan African countries covering a 22 year-period from 1988 to 2010. BEAST is designed to reconstruct evolutionary history within the context of geographical distribution over time from sampled DNA sequences<sup>16</sup>, and has been used extensively in bacterial<sup>17-20</sup>, viral<sup>21,22</sup> and eukaryotic population studies<sup>23</sup>. From this analysis, a single maximum clade credibility tree was produced for each lineage (Fig. 2a & 2b). The mean evolutionary rates, assuming a Bayesian skyline model of population size change and a relaxed molecular clock, were estimated to be  $1.9 \times 10^{-7}$  and  $3.9 \times 10^{-7}$  substitutions<sup>-1</sup>site<sup>-1</sup>year<sup>-1</sup>, for lineage II and I respectively. This corresponds to an accumulation of approximately 1-2 SNPs genome<sup>-1</sup>year<sup>-1</sup>, which is similar to that calculated for the enteric pathogen *Vibrio cholerae* ( $8 \times 10^{-7}$  site<sup>-1</sup> year<sup>-1</sup>)<sup>24</sup> and lies between the rates estimated for *Yersinia pestis* ( $2 \times 10^{-8}$ )<sup>25</sup> and *Staphylococcus aureus* ( $3 \times 10^{-6}$ )<sup>26</sup>. The topologies of the BEAST and maximum-likelihood trees were congruent, and the recovered nodes were supported with high posterior probabilities and bootstrap values, respectively.

A time-dependent phylogeographic reconstruction of lineage I, which is estimated to have emerged ~52 years ago (95% Highest posterior density HPD, 1920.4 -1979.5; Fig. 2a), indicated that in our collection, isolates from Malawi diverged earliest from the last common ancestor for this lineage. Although we cannot completely eliminate the potential bias due the number of Malawian isolates analyzed within this lineage, 25 permutation datasets using 10 randomly selected Malawi isolates (a different set of 10 isolates, an equivalent sample size to those from other countries, was used for each permutation) returned similar results to the complete data set. Thus, we are confident of our estimates of the age and geographical origin of the ancestral node of this lineage (Supplementary Fig. 2a & 2b). Analyses of the distribution of isolates from each country and the tree topology of lineage I are consistent with at least four independent transmission events or movements across Southeastern Africa, with Malawi having served as a potentially important early hub (Fig. 3a, Supplementary Fig. 3a). The earliest identifiable waves or transmissions were from Malawi to Kenya dated ~1982 (95% HPD, 1967.6 – 1990.2) and between Malawi and DRC dated ~1983 (95% HPD, 1974.8 – 1988.3). This same phylogenetically-linked wave was identified in Uganda ~1989 (95% HPD, 1980.0 – 1994.6) and a further onward wave was identifiable in Mozambique in ~1990 (95% HPD, 1981.0 – 1994.4) and this manifested as a second introduction into Uganda in ~2001 (95% HPD, 1981.0 – 1994.4). We cannot identify the specific geographical route that these bacterial lineages followed but the phylogenetic evidence clearly temporally links these outbreaks as a single epidemic. Our results also show evidence of geographical clustering after a transmission event introduced the lineage into a country. This suggests that the epidemic clone was introduced a limited number of times into each country, giving rise to localized epidemics or outbreaks.

Invasive *Salmonella* Typhimurium isolates of lineage I disappeared from our collection between 2003–2005 and were replaced by lineage II and isolates after 2006 were found exclusively in this cluster. Lineage II is estimated to have emerged ~35 years ago (95% HPD, 1957.1 – 1986.8) making it genetically younger than lineage I (Fig. 2b). The spread of lineage II also appears to have occurred in several waves (Fig. 3b, Supplementary Fig. 3b). Our deepest-rooted isolates are from DRC with evidence for transmission onward to Uganda dated ~1985 (95% HPD 1972.6 – 1990.6). This wave was detected in Kenya and Malawi between 1994 and 1996. Malawi likely represents a more recent hub for further dispersal of invasive *Salmonella* Typhimurium lineage II isolates between 1995 and 1998 to several countries, including neighboring Mozambique and further westwards, across the sub-Saharan region, to Mali and Nigeria. A more recent wave of this lineage appears to have spread from Kenya arriving back in Malawi around 2002. We also see evidence for localized

epidemics associated with the lineage II clones, as highlighted by clustering based on geography. Indeed, local epidemiology and molecular typing in Malawi and Kenya<sup>7,8</sup> of invasive *Salmonella* Typhimurium isolates from 1997 to 2006 describes a local clonal replacement event of lineage I by lineage II associated with the emergence of chloramphenicol resistance in an 18-month period from 2001 to 2003.

### Evolution of MDR and potential role of *cat* gene in clonal replacement

Previously, we characterized two distinct composite Tn21-like transposition elements encoding multiple antibiotic resistance determinants located on pSLT, the so-called virulence-associated plasmid, in two representative invasive *Salmonella* Typhimurium isolates, A130 (lineage I) and D23580 (lineage II)<sup>7</sup>. These Tn21 elements are inserted at different sites in the pSLT virulence plasmid in each isolate. Notably, in our phylogenetic analysis, we found these insertion sites to be identical within each lineage but different between lineages, suggesting that Tn21 element acquisition was an independent and early event in each lineage (Fig. 4, Supplementary Fig. 4). Interestingly, only one isolate from lineage I (A24924) and one from lineage II (254DRC) do not have a Tn21-like element (Fig. 4). Comparative analyses of these two isolates, which significantly are the most deeply rooted isolates in each lineage, revealed that although the relevant variant of Tn21 element is absent in both isolates (Fig. 2a & 2b), they share the pSLT plasmid backbone with other isolates of the same lineage. This suggests that in each case they share a common ancestor with other isolates within the same lineage, but importantly the ancestor existed before the acquisition of the composite Tn21-like elements (See supplementary data). With the exception of a deletion in *istA*, a transposase of insertion sequence IS1326, in A16083, the lineage I-specific Tn21 locus is relatively highly conserved in most isolates of lineage I (Fig 4b). In contrast, the Tn21-like locus encoded by lineage II appears somewhat unstable since isolates in different parts of the tree have lost subsets of genes (14DRC, 5582, J17 and A32751) (Fig. 4a; Supplementary fig. 5).

One striking feature of the dataset is the absence of a chloramphenicol resistance (*cat*) gene from all isolates in lineage I. In contrast, the gene was present in > 97% of lineage II, being absent in two isolates (Fig. 4a). These isolates are 254DRC, which does not possess a Tn21 element and 5582, a 2005 Kenyan isolate where the loss of the *cat* gene is due to a simple deletion event (Fig. 4a). This strongly suggests an independent acquisition of the *cat* gene borne on a lineage II-specific Tn21 element early on in the genealogy, most likely about the time of expansion from DRC, as shown in Figure 2b (median node date 1984; 95% HPD, 1972.6 – 1990.6; state posterior probability = 0.78). The analysis of MDR acquisition is consistent with the antibiotic resistance profiles obtained for the isolates. In some of our sampling sites, such as Malawi, the acquisition of resistance to chloramphenicol was observed in invasive *Salmonella* Typhimurium isolates from around 2001-2004, consistent with the arrival of lineage II<sup>7</sup>. At this time, chloramphenicol was the drug of choice for treatment for suspected severe bacterial infections and cases of iNTS infection confirmed by blood culture. The acquisition of chloramphenicol resistance may have afforded lineage II clones a greater opportunity to survive treatment and transmit, which could have in turn contributed to the clonal replacement of lineage I isolates observed between 2003 and 2005 and the subsequent expansion of lineage II clones thereafter.

### Transmission is temporally associated with HIV and the HIV pandemic

Time-dependent phylogeographic analysis revealed the clonal expansion of two distinct invasive *Salmonella* Typhimurium lineages within the last 40-50 years that was accompanied by spread across multiple countries of sub-Saharan Africa. Intriguingly, this emergence temporally coincides with the HIV epidemic in sub-Saharan Africa. Molecular clock analysis of HIV-1 genome sequences suggested that the pandemic began at the start of



the twentieth century<sup>27-29</sup>, with prevalence peaking in the 1990's in many countries, including those represented within our strain collection (2% in Mali to over 15% in Malawi) (Fig. 2c; Supplementary Fig. 6). An association with the HIV status of the patients is also reflected in terms of the samples analyzed in this study. For example, where a test was conducted for HIV, all adult samples were positive. One of the first reported case of HIV infection in Africa was from an adult in DRC<sup>30</sup> and interestingly, the earliest geographic localization of epidemic clones from lineage II is within this country. Thus, the Congo basin provides a potential origin of invasive *Salmonella* Typhimurium lineage II<sup>31</sup>. It therefore appears possible that the epidemic of invasive *Salmonella* Typhimurium and transmission across the sub-Saharan region was potentiated by an increase in the critical population of susceptible and immunocompromised individuals, particularly more mobile adults.

## DISCUSSION

The recent reporting of a remarkably high incidence of invasive *Salmonella* Typhimurium in various parts of the sub-Saharan African region makes it increasingly important to understand the evolutionary origins and spatio-temporal spread of these isolates. Recently, whole-genome sequencing methods have been used to trace intercontinental transmission of different recently emerged and closely related bacterial pathogens<sup>18,24,26,32</sup>, and we have therefore applied this high resolution analysis to determine the phylogenetic structure of invasive *Salmonella* Typhimurium. Here we revealed that the vast majority of *Salmonella* Typhimurium isolates associated with invasive disease from sub-Saharan Africa comprised just two highly conserved lineages of MLST type ST313 that are more closely related to each other than any other known *Salmonella* Typhimurium lineage. This is in contrast to the considerable phylogenetic variation of *Salmonella* Typhimurium isolates associated with gastroenteritis or invasive disease from outside sub-Saharan Africa. Thus, invasive *Salmonella* Typhimurium disease in this region is in part a previously unrecognized epidemic caused by the spread of the clones from these two lineages.

We show how invasive *Salmonella* Typhimurium transmission into a particular country or geographic area occurs as a discrete, temporally defined introduction that is followed by subsequent spread within that particular location (Fig. 2), although some local regions have experienced multiple introduction events. For example, it is evident that two independent introductions occurred in Mali between 1995 and 2000 (Fig. 2b). Considerable clonal expansion has occurred independently in each of these two lineages beginning around 1960. Independent acquisition of a *Tn21* element encoding multiple drug resistance genes by both lineages may have facilitated their successful transmission across the sub-continent, within the susceptible host population. A later acquisition of a *cat* gene on the composite element within lineage II have contributed to a clonal replacement event, which occurred from 2003 – 2005, and resulted in greater spatial dispersion of clones from this lineage over sub-Saharan Africa. An association between acquisition of chloramphenicol resistance and increased transmission has been observed in early epidemiological studies on chloramphenicol-resistant *Salmonella* Typhi in Mexico<sup>33</sup> and is also confirmed by observations reported in Kenya<sup>7</sup> and Malawi<sup>8</sup>.

HIV increases susceptibility to iNTS infections<sup>34</sup> and this form of bacteraemia is an AIDS-defining opportunistic infection in adults<sup>35,36</sup>. Further, animal models of co-infection between iNTS and simian immunodeficiency virus (SIV)<sup>37</sup> or malaria<sup>38</sup> indicate that host immune status plays a critical role in determining the outcome of *Salmonella* infections. Indeed, sporadic human invasive disease is a feature of the non-ST313 lineages of *Salmonella* Typhimurium. Thus, although ST313 is the dominant form of invasive *Salmonella* disease in sub-Saharan Africa<sup>3,39</sup>, it is not unexpected that other *S. enterica* or indeed *Salmonella* Typhimurium lineages can also cause sporadic disease. Notably,

supporting epidemiological evidence indicates that the ST313 *Salmonella* Typhimurium lineages may not have reached some parts of Africa, including the Gambia<sup>40,41</sup> and Ethiopia<sup>42,43</sup> where iNTS has been reported.

It is particularly noteworthy that we see a temporal association of clonal expansion of invasive *Salmonella* Typhimurium with the peaks in HIV prevalence, particularly in adults in the countries included in our study. The rapid expansion and spread of these clones may have been facilitated by the dramatic expansion of a mobile susceptible host population. Previous analysis has shown that HIV-I arrived in East and Central Africa around the 1950's, followed by an eastwards expansion in the 1970's and early 1980's<sup>44</sup>. We find temporal parallels in this estimated HIV-I expansion time frame and our estimate of the earliest detectable transmissions in lineage I around the early 1980's (95% HPD, 1967.6 – 1990.2). The continued expansion of the HIV-susceptible population until the peaks of prevalence in the 1990's (Fig. 2c), together with the acquisition of additional chloramphenicol resistance, is likely contributory to the greater dispersal of lineage II clones. The association of iNTS disease with malaria, anemia and malnourishment in children is well-documented<sup>4,5,45-47</sup> and we have isolates within our collection from children with these underlying conditions (Supplementary Table 1). Malnourished and malarial children thus present an additional ecological niche that co-exists with as well as precedes the HIV-positive population. Notably, we found no evidence of phylogenetic segregation between such isolates and those from HIV-positive children or adults within the two epidemic lineages. This is consistent with immunosuppression as a key predisposing factor in iNTS disease. However, the emergence of a large cohort of HIV infected adults may also have facilitated the spread of the invasive *Salmonella* Typhimurium lineages, as adults are inevitably more mobile. This is especially pertinent since failure of immunological control of iNTS infections in HIV-positive African adults has been well-documented<sup>34,48</sup>.

The resulting large pool of immunosuppressed individuals may also facilitate an unusual human-to-human transmission (anthroponotic) component in invasive *Salmonella* Typhimurium disease, in contrast to most disease caused by NTS outside of Africa, where transmission is predominantly zoonotic<sup>49</sup>. There is a dearth of information on the specifics of NTS transmission in sub-Saharan Africa although independent, country-based studies have shown evidence of non-zoonotic transmission patterns<sup>39,49,50</sup>. It is perhaps noteworthy that we detected a similar pattern of genomic degradation in the form of gene loss and pseudogene formation to the human-adapted serovars *S. Typhi*<sup>12</sup> and *S. Paratyphi*<sup>51</sup> in the two fully sequenced African invasive *Salmonella* Typhimurium isolates, D23580 and A130 representative of lineages I and II respectively<sup>7</sup>. Taken together, these results suggest that the invasive clones may have adapted to facilitate direct person-to-person transmission within the human population. Further comparative studies on the virulence, and transmission potential of different *Salmonella* Typhimurium lineages will be instrumental in closing this critical knowledge gap and are the focus of on-going investigations

These results provide the first whole-genome based transmission study of this kind on iNTS from sub-Saharan Africa, and they highlight the power of these approaches to monitor emergence and spread of clonal bacterial populations associated with epidemics locally or globally over time. The transmission pathways hypothesized here suggest potential routes to the implementation of appropriate clinical intervention strategies.

## ONLINE METHODS

### Isolate Selection and genomic DNA preparation

129 isolates associated with invasive disease from Malawi, Mali, Kenya, and Nigeria were cultured from venous blood, cerebrospinal fluid or stool of febrile adults and children between 1988 and 2010. Gastrointestinal isolates were obtained from collections at the *Salmonella* Genetic Stock centre (SGSC)<sup>52</sup>, Calgary Canada, the Health Protection Agency, London or as indicated in the Supplementary Table 1<sup>13,53-57</sup>. Invasive *Salmonella* Typhimurium isolates were identified by standard serotyping methods, using the O- and H-antigen agglutination, based on the Kauffmann-White Scheme<sup>1</sup>. DNA samples were provided for invasive *Salmonella* Typhimurium isolates from the DRC, Mozambique and Uganda. Isolates were grown on Luria-Bertani (LB) medium; single colonies were incubated in LB Broth overnight at 37°C. Bacteria cells were pelleted by centrifugation (3,700 × g or 4,300 rpm, 5 minutes) and DNA extracted using either the Wizard® Genomic DNA kit (Promega) according to manufacturers' instructions or by phenol/chloroform extraction protocol<sup>18</sup>. DNA quality and quantity were checked by gel electrophoresis and Qubit® quantitation platform (Invitrogen). 20-50 ng/μL of DNA from each isolate was submitted for Illumina sequencing.

### Genomic library preparation and sequencing

Multiplex libraries with a 200bp insert size were prepared using 12 unique index-tags, and sequenced to generate 54 or 76 base-pair (bp) paired-end reads. Cluster formation, primer hybridisation and sequencing reactions were based on reversible terminator chemistry using the Illumina Genome Analyser II System according to standard protocol<sup>26,58</sup>. Sequence data were submitted to the European Nucleotide Archive. Accession numbers are given in Supplementary Table 1.

### Read alignment and SNP detection

Paired-end Illumina sequence data from each isolate was mapped to the reference genome *Salmonella* Typhimurium strain SL1344<sup>57</sup> using SSAHA2<sup>59</sup>. Sequence reads mapped to an average of 97.7% of the reference with a mean depth of 56.5-fold in mapped regions across all isolates (Supplementary Table 1). SNPs were identified using samtools mpileup and filtered with a minimum mapping quality of 30 and quality ratio cut-off of 0.75<sup>18,24,26,59,60</sup>. SNPs called in phage sequences and repetitive regions of the *Salmonella* Typhimurium reference were excluded. Repetitive regions were defined as exact repetitive sequences of 20 bp, identified using repeat finding programs nucmer<sup>61</sup>, REPeuter<sup>62</sup> and repeat-match<sup>12,17</sup>. Recombinant segments of the genome were removed from the whole genome alignment as described previously<sup>18</sup>. Following the removal of recombinant segments, mobile elements and repetitive sequences of the genome, a concatenated alignment composed of 10,623 SNP sites from each sequenced isolates was produced. Small insertions and deletions (indels) were also identified from the SSAHA result output, but were not used for subsequent phylogenetic analyses.

### Phylogenetic analyses

A Maximum-likelihood phylogenetic tree (Fig. 1) was constructed from SNP alignment with RAxML v7.0.4<sup>14</sup> using a general time-reversible (GTR) substitution model with  $\gamma$  correction for among-site rate variation. Support for nodes on the trees were assessed using 100 bootstrap replicates. For the identified lineages 1 and II, 487 and 422 chromosomal SNP loci were identified, respectively. These within-cluster SNP alignments were then used to recalculate individual maximum likelihood trees for each cluster, using the same parameters described above. These were used as input trees for subsequent analyses. The methods



described above were also applied to obtain a maximum-likelihood phylogenetic reconstruction of plasmids from our isolate collection using 1,251 concatenated SNP sites, with the virulence plasmid pSLT-SL1344 from SL1344 as the reference.

### MLST Analyses

Allele coordinates were obtained for the seven housekeeping genes used for the *S. enterica* MLST typing scheme (*aroC*, *dnaN*, *hemD*, *hisD*, *purE*, *sucA* and *thrA*) by manually marking the coordinates in the whole-genome alignment of the isolates. The marked up regions were extracted and a multi-sequence alignment of each gene produced for all the isolates. These were used to determine the sequence type of each isolates using the *Salmonella enterica* MLST database

### Bayesian phylogeny, estimating dates of divergence and phylogeographical analyses of lineages

—Estimation of rates of evolution, divergence times and phylogeography for our isolate collection as well as for each of the identified lineages was performed using the Bayesian MCMC framework, BEAST<sup>15</sup>, on SNP alignments. Various combinations of population size change model and molecular clock model were compared to find the model that best fit the data. In all cases Bayes Factors showed strong support (BF << 200) for the use of skyline<sup>63</sup> model of population size change and a relaxed uncorrelated lognormal clock<sup>64</sup> which allows the evolutionary rates to change among the branches of the tree<sup>24</sup> and a GTR substitution model with gamma correction for amongst site rate variation.

Using the same parameters, the geographical locations of ancestral nodes were estimated using the discrete geospatial model implemented in BEAST (Supplementary Table 1)<sup>16</sup>. In all cases, three independent chains were run for 250 million steps each, and sampled every 10,000 steps. The three chains were combined with LogCombiner<sup>15</sup> with the initial 25 million steps removed from each as a burn-in. Maximum clade credibility (MCC) trees were created and annotated using TreeAnnotator and viewed in FigTree<sup>15</sup>. We report estimates as median values within 95% highest posterior density (HPD) and report posterior probability values as support for identified ancestral node age and geographical location. For the latter, we report values greater than 0.7. Spatial reconstruction of MCC trees was carried out using the SPREAD<sup>65</sup> software and visualized with Google Earth (Supplementary figure 3).

**HIV prevalence data extrapolation**—HIV prevalence data for the sampled countries was modelled with a generalized logistic (or Richards')<sup>66</sup> curve using the grofit R package<sup>67</sup>. Curves were fit to all data points from the beginning of monitoring until stabilization or decline of the HIV-positive population. We then used these fitted models to extrapolate possible past population sizes.

### Validation tests for the origin of lineage I

25 permutation datasets made up of 10 randomly selected Malawian isolates, together with the 7 DRC, 8 Kenyan, 8 Mozambican and 7 Ugandan isolates, were used to reconstruct Bayesian maximum clade credibility (MCC) phylogenetic trees. Each of the 25 datasets included a different set of 10 randomly selected Malawian isolates. The same parameters described above were applied in making the trees. Malawi was the ancestral state of all resulting 25 MCC trees with posterior probability values ranging from 0.58 - 0.92. The resulting phylogenetic trees and their root location state probability distributions are shown in supplementary figure 2b.

### Plasmid sequence analyses

Paired-end sequence reads of each isolates were mapped to the multi-fasta sequence features including the *Tn21* locus of pSLT-BT, the reference plasmid from invasive strain, D23580

using the Burrows-Wheeler Aligner software BWA<sup>68</sup>, with minimum base call quality of 50, minimum mapping quality of 30, and minimum read depth of 4. Isolates from each of the three clusters were analyzed separately by cluster. Isolates with <30% of reads mapping to the length of the feature were interpreted as not having the feature, and those with >70% of reads mapping to the feature were interpreted as having the region of interest. A heatmap of the analysis based on the selected cut-off values was generated and aligned to the BEAST MCC tree of each cluster.

### **De novo sequence assembly and plasmid genome comparisons**

Paired-end Illumina sequence data was assembled *de novo* using Velvet<sup>69</sup> and the parameters optimized to give the highest N50 value. The multi-contig draft genomes generated for each isolate were ordered using either pSLT or pSLT-BT to confirm plasmid structure using Abacas<sup>70</sup>. The draft plasmid genomes were used to query pSLT and/or pSLT-BT sequences using BLASTN<sup>71</sup> and comparison files generated and viewed using Artemis Comparison Tool (ACT)<sup>72</sup>

### **Supplementary Material**

Refer to Web version on PubMed Central for supplementary material.

### **Acknowledgments**

We thank J. Cheesborough for providing the DRC isolates, M. Okong, N. French and the Medical Research Council, Uganda, for providing the Uganda isolates, S. Nair for providing the Health Protection Agency (HPA) isolates, L. Barquist for modeling the pre-1990 HIV prevalence data and the Sequencing team at the Wellcome Trust Sanger Institute. This work was funded by a Wellcome Trust grant (098051).

C.A.M. was supported by a Tropical Research Fellowship from the Wellcome Trust and a Clinical Research Fellowship from GlaxoSmithKline.

### **REFERENCES**

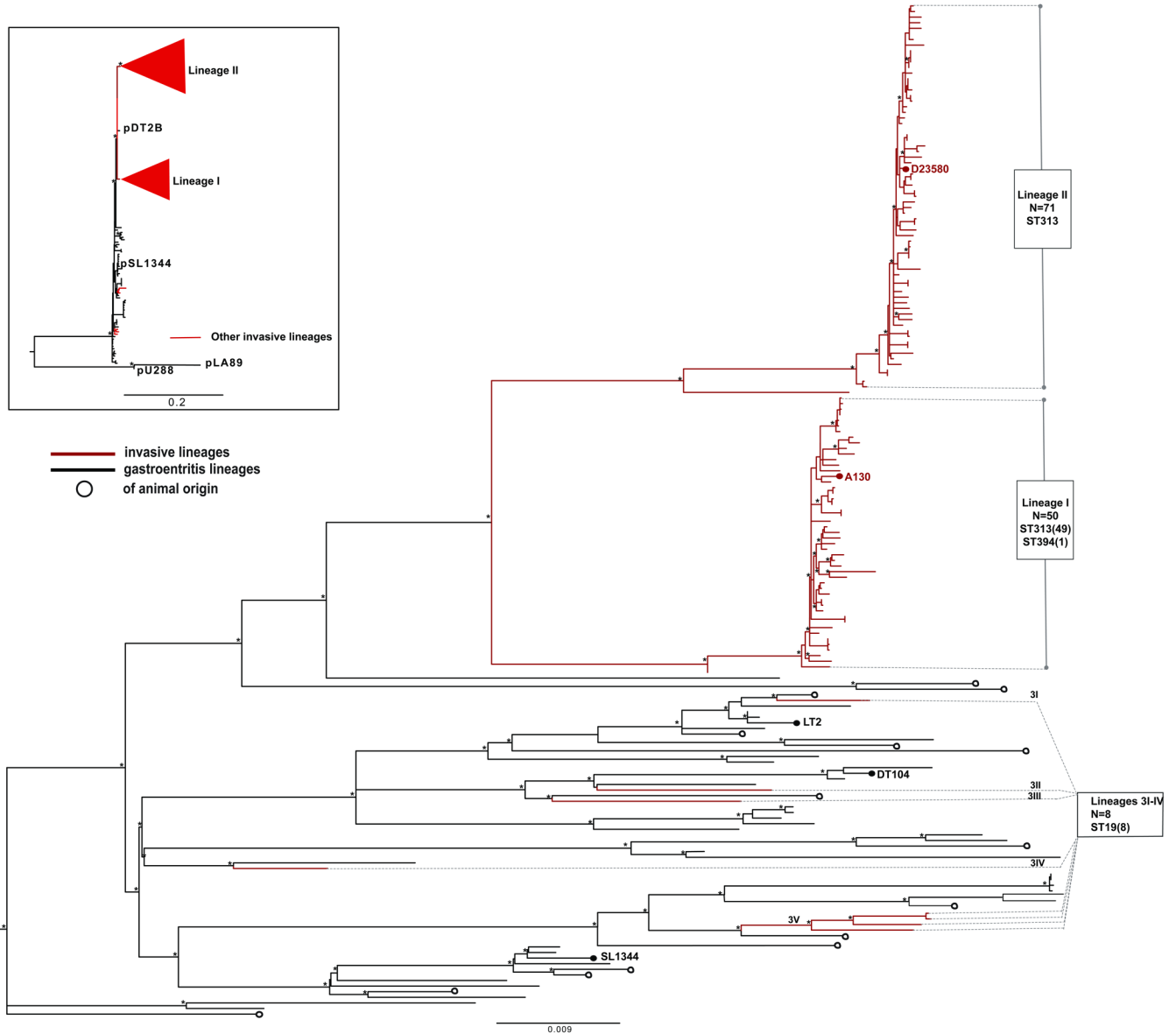
1. Popoff MY, Bockemuhl J, Gheesling LL. Supplement 2002 (no. 46) to the Kauffmann-White scheme. *Res Microbiol.* 2004; 155:568–570. [PubMed: 15313257]
2. Langridge GC, Nair S, Wain J. Nontyphoidal Salmonella serovars cause different degrees of invasive disease globally. *The Journal of infectious diseases.* 2009; 199:602–603. [PubMed: 19281309]
3. 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]
4. Graham SM. Nontyphoidal salmonellosis in Africa. *Curr Opin Infect Dis.* 2010; 23:409–414. [PubMed: 20736739]
5. Berkley JA, et al. HIV infection, malnutrition, and invasive bacterial infection among children with severe malaria. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America.* 2009; 49:336–343. [PubMed: 19548833]
6. Gordon MA, et al. Non-typhoidal salmonella bacteraemia among HIV-infected Malawian adults: high mortality and frequent recrudescence. *AIDS (London, England).* 2002; 16:1633–1641.
7. 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]
8. Gordon MA, et al. Epidemics of invasive Salmonella enterica serovar enteritidis and S. enterica Serovar typhimurium infection associated with multidrug resistance among adults and children in Malawi. *Clin Infect Dis.* 2008; 46:963–969. [PubMed: 18444810]
9. Gordon MA. Salmonella infections in immunocompromised adults. *The Journal of infection.* 2008; 56:413. [PubMed: 18474400]

10. Cheesbrough JS, Taxman BC, Green SD, Mewa FI, Numbi A. Clinical definition for invasive *Salmonella* infection in African children. *Pediatr Infect Dis J*. 1997; 16:277–283. [PubMed: 9076815]
11. Parkhill J, et al. Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar Typhi CT18. *Nature*. 2001; 413:848. [PubMed: 11677608]
12. Holt KE, et al. High-throughput sequencing provides insights into genome variation and evolution in *Salmonella* Typhi. *Nat Genet*. 2008; 40:987–993. [PubMed: 18660809]
13. Beltran P, et al. Reference collection of strains of the *Salmonella typhimurium* complex from natural populations. *J Gen Microbiol*. 1991; 137:601–606. [PubMed: 2033380]
14. Stamatakis A. RAXML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*. 2006; 22:2688–2690. [PubMed: 16928733]
15. Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol*. 2007; 7:214. [PubMed: 17996036]
16. Lemey P, Rambaut A, Drummond AJ, Suchard MA. Bayesian phylogeography finds its roots. *PLoS Comput Biol*. 2009; 5:e1000520. [PubMed: 19779555]
17. He M, et al. Evolutionary dynamics of *Clostridium difficile* over short and long time scales. *Proceedings of the National Academy of Sciences of the United States of America*. 2010; 107:7527–7532. [PubMed: 20368420]
18. Croucher NJ, et al. Rapid pneumococcal evolution in response to clinical interventions. *Science*. 2011; 331:430–434. [PubMed: 21273480]
19. Holt KE, et al. Temporal fluctuation of multidrug resistant salmonella typhi haplotypes in the mekong river delta region of Vietnam. *PLoS Negl Trop Dis*. 2011; 5:e929. [PubMed: 21245916]
20. den Bakker HC, Bundrant BN, Fortes ED, Orsi RH, Wiedmann M. A population genetics-based and phylogenetic approach to understanding the evolution of virulence in the genus *Listeria*. *Appl Environ Microbiol*. 2010; 76:6085–6100. [PubMed: 20656873]
21. Smith GJ, et al. Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic. *Nature*. 2009; 459:1122–1125. [PubMed: 19516283]
22. Smith GJ, et al. Dating the emergence of pandemic influenza viruses. *Proceedings of the National Academy of Sciences of the United States of America*. 2009; 106:11709–11712. [PubMed: 19597152]
23. Endicott P, Ho SY, Stringer C. Using genetic evidence to evaluate four palaeoanthropological hypotheses for the timing of Neanderthal and modern human origins. *J Hum Evol*. 2010; 59:87–95. [PubMed: 20510437]
24. Mutreja A, et al. Evidence for several waves of global transmission in the seventh cholera pandemic. *Nature*. 2011; 477:462–465. [PubMed: 21866102]
25. Morelli G, et al. *Yersinia pestis* genome sequencing identifies patterns of global phylogenetic diversity. *Nat Genet*. 2010; 42:1140–1143. [PubMed: 21037571]
26. Harris, SR., et al. *Science*. Vol. 327. New York, N.Y.: 2010. Evolution of MRSA during hospital transmission and intercontinental spread; p. 469
27. Korber B, et al. Timing the ancestor of the HIV-1 pandemic strains. *Science*. 2000; 288:1789–1796. [PubMed: 10846155]
28. Lemey P, et al. The molecular population genetics of HIV-1 group O. *Genetics*. 2004; 167:1059–1068. [PubMed: 15280223]
29. Worobey M, et al. Direct evidence of extensive diversity of HIV-1 in Kinshasa by 1960. *Nature*. 2008; 455:661–664. [PubMed: 18833279]
30. Nahmias AJ, et al. Evidence for human infection with an HTLV III/LAV-like virus in Central Africa, 1959. *Lancet*. 1986; 1:1279–1280. [PubMed: 2872424]
31. Sharp ER, et al. Immunodominance of HIV-1 specific CD8+ T-cell responses is related to disease progression rate in vertically infected adolescents. *PloS one*. 2011; 6:e21135. [PubMed: 21818255]
32. Harris SR, et al. Whole-genome analysis of diverse *Chlamydia trachomatis* strains identifies phylogenetic relationships masked by current clinical typing. *Nat Genet*. 2012

33. Gangarosa EJ, et al. An epidemic-associated episome? *The Journal of infectious diseases*. 1972; 126:215–218. [PubMed: 4559853]
34. MacLennan CA, et al. Dysregulated humoral immunity to nontyphoidal *Salmonella* in HIV-infected African adults. *Science*. 2010; 328:508–512. [PubMed: 20413503]
35. Smith PD, et al. *Salmonella typhimurium* enteritis and bacteremia in the acquired immunodeficiency syndrome. *Ann Intern Med*. 1985; 102:207–209. [PubMed: 3881075]
36. Levine WC, Buehler JW, Bean NH, Tauxe RV. Epidemiology of nontyphoidal *Salmonella* bacteremia during the human immunodeficiency virus epidemic. *The Journal of infectious diseases*. 1991; 164:81–87. [PubMed: 2056220]
37. Raffatellu M, et al. Simian immunodeficiency virus-induced mucosal interleukin-17 deficiency promotes *Salmonella* dissemination from the gut. *Nature medicine*. 2008; 14:421–428.
38. Roux CM, et al. Both hemolytic anemia and malaria parasite-specific factors increase susceptibility to Nontyphoidal *Salmonella enterica* serovar typhimurium infection in mice. *Infect Immun*. 2010; 78:1520–1527. [PubMed: 20100860]
39. Keddy KH, et al. Genotypic and demographic characterization of invasive isolates of *Salmonella Typhimurium* in HIV co-infected patients in South Africa. *Journal of infection in developing countries*. 2009; 3:585. [PubMed: 19801800]
40. Ikumapayi UN, et al. Molecular epidemiology of community-acquired invasive non-typhoidal *Salmonella* among children aged 2–29 months in rural Gambia and discovery of a new serovar, *Salmonella enterica* Dingiri. *J Med Microbiol*. 2007; 56:1479. [PubMed: 17965348]
41. Dione MM, et al. Clonal differences between Non-Typhoidal *Salmonella* (NTS) recovered from children and animals living in close contact in the Gambia. *PLoS Negl Trop Dis*. 2011; 5:e1148. [PubMed: 21655353]
42. Beyene G, et al. Multidrug resistant *Salmonella* Concord is a major cause of salmonellosis in children in Ethiopia. *Journal of infection in developing countries*. 2011; 5:23–33. [PubMed: 21330737]
43. Sibhat B, et al. *Salmonella* serovars and antimicrobial resistance profiles in beef cattle, slaughterhouse personnel and slaughterhouse environment in Ethiopia. *Zoonoses Public Health*. 2011; 58:102–109. [PubMed: 20042064]
44. Gray RR, et al. Spatial phylodynamics of HIV-1 epidemic emergence in east Africa. *AIDS*. 2009; 23:F9–F17. [PubMed: 19644346]
45. Brent AJ, et al. *Salmonella* bacteremia in Kenyan children. *Pediatr Infect Dis J*. 2006; 25:230–236. [PubMed: 16511385]
46. Mandomando I, et al. Invasive non-typhoidal *Salmonella* in Mozambican children. *Trop Med Int Health*. 2009; 14:1467. [PubMed: 19793081]
47. Rosanova MT, Paganini H, Bologna R, Lopardo H, Ensinck G. Risk factors for mortality caused by nontyphoidal *Salmonella* sp. in children. *Int J Infect Dis*. 2002; 6:187–190. [PubMed: 12718833]
48. Gordon MA, et al. Invasive non-typhoid salmonellae establish systemic intracellular infection in HIV-infected adults: an emerging disease pathogenesis. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*. 2010; 50:953. [PubMed: 20180702]
49. Kariuki S, et al. Invasive multidrug-resistant non-typhoidal *Salmonella* infections in Africa: zoonotic or anthroponotic transmission? *J Med Microbiol*. 2006; 55:585–591. [PubMed: 16585646]
50. Fashae K, Ogunsola F, Aarestrup FM, Hendriksen RS. Antimicrobial susceptibility and serovars of *Salmonella* from chickens and humans in Ibadan, Nigeria. *Journal of infection in developing countries*. 2010; 4:484–494. [PubMed: 20818100]
51. Holt KE, et al. Pseudogene accumulation in the evolutionary histories of *Salmonella enterica* serovars Paratyphi A and Typhi. *BMC genomics*. 2009; 10:36. [PubMed: 19159446]
52. Zinder ND, Lederberg J. Genetic exchange in *Salmonella*. *Journal of bacteriology*. 1952; 64:679–699. [PubMed: 12999698]
53. Helm RA, et al. Pigeon-associated strains of *Salmonella enterica* serovar Typhimurium phage type DT2 have genomic rearrangements at rRNA operons. *Infect Immun*. 2004; 72:7338. [PubMed: 15557664]

54. Beltran P, et al. Toward a population genetic analysis of Salmonella: genetic diversity and relationships among strains of serotypes *S. choleraesuis*, *S. derby*, *S. dublin*, *S. enteritidis*, *S. heidelberg*, *S. infantis*, *S. newport*, and *S. typhimurium*. *Proceedings of the National Academy of Sciences of the United States of America*. 1988; 85:7753–7757. [PubMed: 3051004]
55. Cooke FJ, et al. Characterization of the genomes of a diverse collection of *Salmonella enterica* serovar Typhimurium definitive phage type 104. *Journal of Bacteriology*. 2008; 190:815. [PubMed: 18849424]
56. Andrews-Polymenis HL, et al. Host restriction of *Salmonella enterica* serotype Typhimurium pigeon isolates does not correlate with loss of discrete genes. *Journal of Bacteriology*. 2004; 186:2619. [PubMed: 15090502]
57. Hoiseth SK, Stocker BA. Aromatic-dependent *Salmonella typhimurium* are non-virulent and effective as live vaccines. *Nature*. 1981; 291:238–239. [PubMed: 7015147]
58. Bentley DR, et al. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*. 2008; 456:53–59. [PubMed: 18987734]
59. Ning Z, Cox AJ, Mullikin JC. SSAHA: a fast search method for large DNA databases. *Genome Res*. 2001; 11:1725–1729. [PubMed: 11591649]
60. Li H, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009; 25:2078–2079. [PubMed: 19505943]
61. Kurtz S, et al. Versatile and open software for comparing large genomes. *Genome Biol*. 2004; 5:R12. [PubMed: 14759262]
62. Kurtz S, et al. REPuter: the manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res*. 2001; 29:4633–4642. [PubMed: 11713313]
63. Drummond AJ, Rambaut A, Shapiro B, Pybus OG. Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol Biol Evol*. 2005; 22:1185–1192. [PubMed: 15703244]
64. Drummond AJ, Ho SY, Phillips MJ, Rambaut A. Relaxed phylogenetics and dating with confidence. *PLoS Biol*. 2006; 4:e88. [PubMed: 16683862]
65. Bielejec F, Rambaut A, Suchard MA, Lemey P. SPREAD: spatial phylogenetic reconstruction of evolutionary dynamics. *Bioinformatics*. 2011; 27:2910–2912. [PubMed: 21911333]
66. Richards FJ. A flexible growth function for empirical use. *Journal of Experimental Botany*. 1959; 10:290–301.
67. Kahm M, Hasenbrink G, Lichtenberg-Fraté H, Ludwig J, Kschischo M. grofit: Fitting Biological Growth Curves with R. *Journal of Statistical Software*. 2010; 33
68. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009; 25:1754–1760. [PubMed: 19451168]
69. Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res*. 2008; 18:821–829. [PubMed: 18349386]
70. Assefa S, Keane TM, Otto TD, Newbold C, Berriman M. ABACAS: algorithm-based automatic contiguation of assembled sequences. *Bioinformatics*. 2009; 25:1968–1969. [PubMed: 19497936]
71. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *Journal of molecular biology*. 1990; 215:403–410. [PubMed: 2231712]
72. Carver T, et al. Artemis and ACT: viewing, annotating and comparing sequences stored in a relational database. *Bioinformatics*. 2008; 24:2672–2676. [PubMed: 18845581]



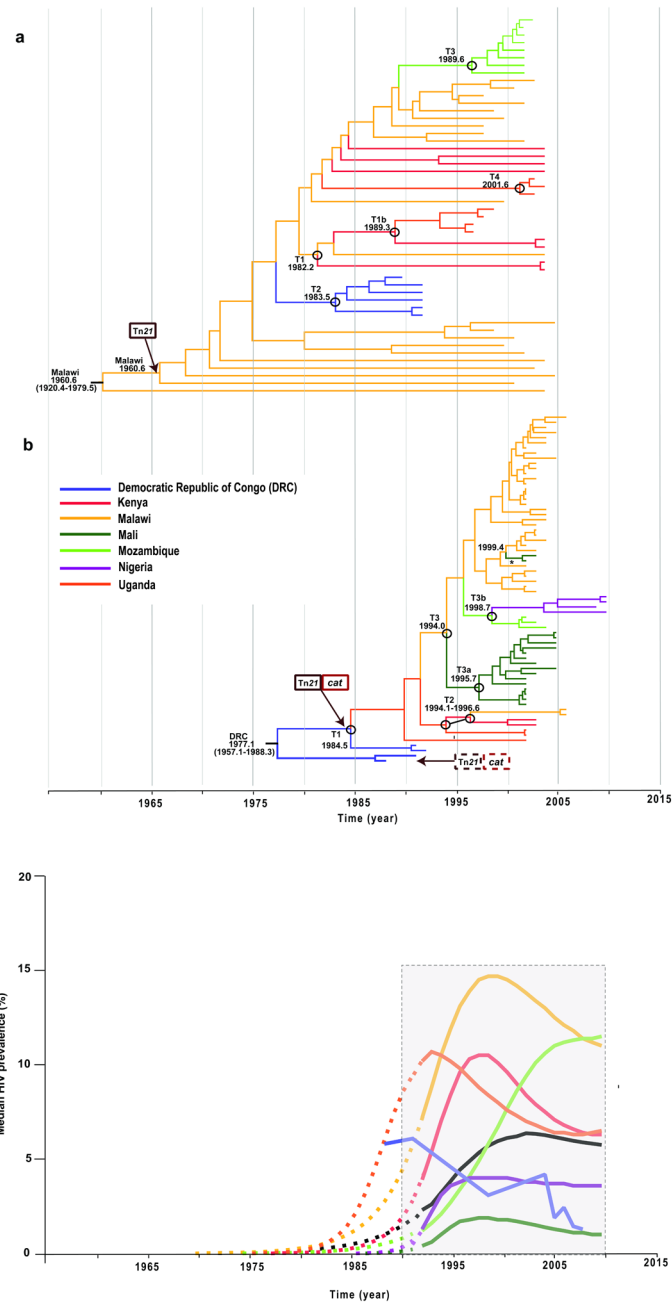


**Figure 1. Population structure of *Salmonella* Typhimurium isolates**

Unrooted maximum likelihood tree showing the relationship between isolates associated with invasive disease and gastroenteritis. Lineages of human invasive *Salmonella* Typhimurium are shown in red and labeled I–3IV.

The Phylogenetic positions of invasive strains A130 (ref. 7) and D23580 (ref. 7) and gastroenteritis-associated strains DT104, LT2 and SL1344 are indicated. Branch lengths are indicative of the estimated substitution rate per variable site. Scale bar, 0.009 substitutions per variable site.

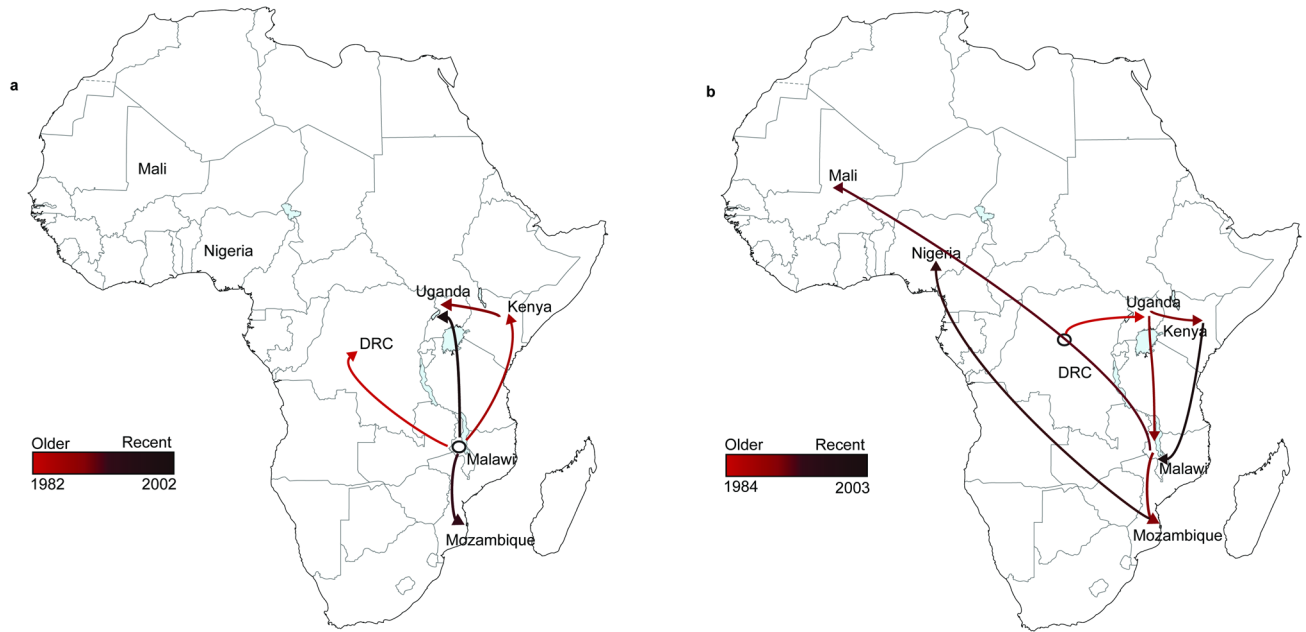
The numbers of isolates (in parentheses) and MLST groups are indicated in boxes. Top left, unrooted maximum-likelihood tree of plasmid sequences showing congruence with the chromosomal tree. Asterisks indicate nodes with 100% bootstrap support.



**Figure 2. Bayesian-based analyses of spatial and temporal distribution of Sub-Saharan African lineages of invasive *Salmonella* Typhimurium**

(a,b) MCC trees from BEAST showing phylogeographic reconstruction of lineage I (a) and lineage II (b) with estimated sampling intervals of 43.0 years (1960.6-2003.6) and 32.3 years (1977.1-2009.4), respectively. Estimated ages of nodes where transmissions occurred (black circles) are reported as the median values, with 95% HPD given in parentheses. Asterisks indicates the second introduction of invasive *Salmonella* Typhimurium in Mali. Posterior probability values for all geographical locations at the ancestral nodes other than the second spread into Uganda (0.56)  $>0.9$  for lineage I in a and  $>0.7$  for lineage II in b were recovered for all the geographical locations at the ancestral nodes. Branches and nodes are

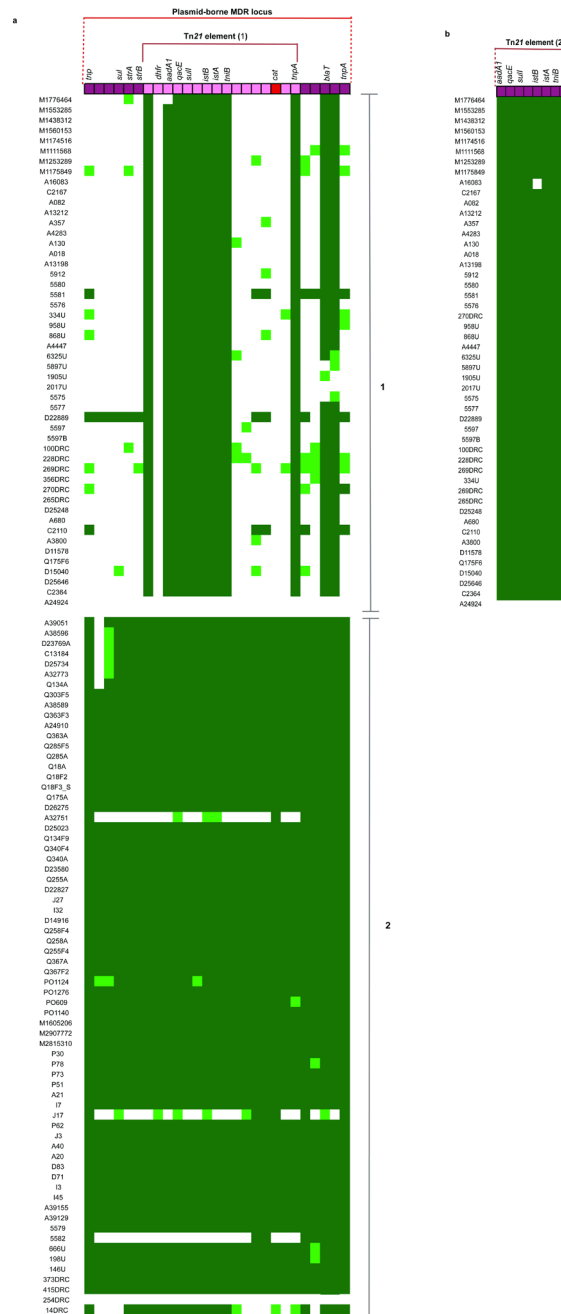
colored according to the location that had the highest posterior probability value. Arrows indicate the estimated points of insertion of independently acquired *Tn21* and *cat* loci within the plasmids in both lineages. (c) Percentage of HIV prevalence in sampled countries from 1960 to the present. HIV prevalence is defined as the percentage of men and women between the ages of 15 and 49 who are HIV positive (UNAIDS Report on the Global AIDS Epidemic 2010; see URLs). Dashed lines show predicted HIV prevalence values before monitoring and reportage for the different countries extrapolated backwards in time to 1960. The block outlined by dashed lines indicates the time when HIV prevalence monitoring was temporally associated with the expansion of invasive *Salmonella* Typhimurium clones across sub-Saharan Africa.



**Figure 3. Geospatial transmission of invasive *Salmonella* Typhimurium isolates in sub-Saharan Africa**

**(a,b)** Phylogeographical diffusion of lineages I **(a)** and II **(b)** across sub-Saharan Africa over time based on a discrete geospatial model with associated geographical coordinates.

Countries shown here represent discrete locations annotated at the tree nodes taken from the BEAST analyses, and branches that indicate location changes are represented on the map as the transmission lines. The color gradient shows the ages of transmission lines.



**Figure 4. Distribution of MDR loci in invasive *Salmonella* Typhimurium in relation to phylogeny**  
**A** Tn21 elements in lineage I (panel 1) and lineage II (panel 2) mapped to Tn21 element from strain D23580 (on top). **B** lineage I elements mapped to assembled Tn21 sequence in strain A130. Sequence reads mapping to the complete sequence length is represented as a heatmap with dark green color indicating > 90% (high) coverage, light green indicating >30<90% coverage and white indicating < 30% (low) coverage.