# Distinct *Salmonella* Enteritidis lineages associated with enterocolitis in high-income settings and invasive disease in low-income settings


## Authors

Nicholas A. Feasey [1,2,3,4], James Hadfield [2], Karen H. Keddy [5,6], Timothy J Dallman [7],Jan Jacobs [8,9], Xiangyu Deng [10,11], Paul Wigley [4], Lars Barquist [12], Gemma C. Langridge [2], Theresa Feltwell [2], Simon R. Harris [2], Alison E. Mather [2]*, Maria Fookes [2], Martin Aslett [2], Chisomo Msefula [3,13], Samuel Kariuki [14], Calman A. Maclennan [2,15], Robert S. Onsare [14], François-Xavier Weill [16], Simon Le Hello [16], Anthony M. Smith [5], Michael McClelland [17], Prerak Desai [17], Christopher M. Parry [1,18], John Cheesbrough [20], Neil French [4], Josefina Campos [21], Jose A. Chabalgoity [22], Laura Betancor [22], Katie L. Hopkins [23], Satheesh Nair [7], Tom J. Humphrey [24], Octavie Lunguya [25,26], Tristan A. Cogan [27], Milagritos D. Tapia [28], Samba O. Sow [29], Sharon M. Tennant [28], Kristin Bornstein [28], Myron M. Levine [28], Lizeth Lacharme-Lora [4], Dean B. Everett [4], Robert A. Kingsley [2,30], Julian Parkhill [2], Robert S. Heyderman [3,31], Gordon Dougan [2], Melita A Gordon† [3,4] & Nicholas R. Thomson† [2,18]

## Affiliations

1. Liverpool School of Tropical Medicine, Liverpool, UK
2. Wellcome Trust Sanger Institute, Cambridge, UK
3. Malawi Liverpool Wellcome Trust Clinical Research Programme, University of Malawi College of Medicine, Blantyre, Malawi
4. Institute of Infection and Global Health, University of Liverpool, Liverpool, UK
5. Centre for Enteric Diseases, National Institute for Communicable Diseases, Johannesburg, South Africa
6. Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa
7. Gastrointestinal Bacteria Reference Unit, Public Health England, Colindale, UK
8. Institute of Tropical Medicine, Antwerp, Belgium
9. Department of Microbiology and Immunology, University of Leuven, Belgium
10. Center for Food Safety, Department of Food Science and Technology, University of Georgia
11. Centers for Disease Control and Prevention, Atlanta, USA
12. Institute for Molecular Infection Biology, University of Würzburg, Würzburg, Germany
13. University of Malawi, The College of Medicine, Blantyre, Malawi
14. Centre for Microbiology Research, Kenya Medical Research Institute, Nairobi, Kenya

15. Jenner Institute, Nuffield Department of Medicine, University of Oxford, Oxford, UK
16. Institut Pasteur, Paris, France
17. Department of Microbiology and Molecular Genetics, University of California, Irvine, CA, USA
18. London School of Hygiene & Tropical Medicine, London, UK
19. School of Tropical Medicine and Global Health, Nagasaki University, Japan
20. Department of Epidemiology and Population Health, University of Liverpool, Liverpool, UK
21. Enteropathogen Division, Administración Nacional de Laboratorios e Institutos de Salud (ANLIS) Carlos G. Malbran Institute, Buenos Aires, Argentina
22. Instituto de Higiene, Facultad de Medicina, Universidad de la Republica, Uruguay
23. Antimicrobial Resistance and Healthcare-Associated Infections Reference Unit, Public Health England, Colindale, UK
24. Swansea Medical School, Swansea University, Swansea, UK
25. National Institute of Biomedical Research, Kinshasa, the Democratic Republic of the Congo
26. University Hospital of Kinshasa, Kinshasa, the Democratic Republic of the Congo
27. School of Veterinary Sciences, University of Bristol, Bristol, UK
28. Center for Vaccine Development, University of Maryland School of Medicine, Baltimore, Maryland, USA
29. Centre pour le Développement des Vaccins, Bamako, Mali
30. Institute of Food Research, Colney, Norwich, UK
31. Division of Infection and Immunity, University College London, London, UK

* Current address: Department of Veterinary Medicine, University of Cambridge, Cambridge, UK

†These authors contributed equally

Corresponding Author
Nicholas Feasey
Liverpool School of Tropical Medicine
5 Pembroke Place
Liverpool
L3 5QA
Tel: 0151 705 3214
E-mail: Nicholas.feasey@lstmed.ac.uk

Running Title: Emergence of distinct lineages of *S*. Enteritidis

Key words: *Salmonella* Enteritidis, global, multidrug resistant, bacteraemia, gastroenteritis

**Abstract**

An epidemiological paradox surrounds *Salmonella enterica* serovar Enteritidis. In high-income settings, it has been responsible for an epidemic of poultry-associated, self-limiting enterocolitis, whilst in sub-Saharan Africa it is a major cause of invasive nontyphoidal *Salmonella* disease, associated with high case-fatality. Whole-genome sequence analysis of 675 isolates of *S*. Enteritidis from 45 countries reveals the existence of a global epidemic clade and two novel clades of *S*. Enteritidis that are each geographically restricted to distinct regions of Africa. The African isolates display genomic degradation, a novel prophage repertoire and have an expanded, multidrug resistance plasmid. *S*. Enteritidis is a further example of a *Salmonella* serotype that displays niche plasticity, with distinct clades that enable it to become a prominent cause of gastroenteritis in association with the industrial production of eggs, and of multidrug resistant, bloodstream invasive infection in Africa.

## Introduction

*Salmonella enterica* serovar Enteritidis (hereafter referred to as *S*. Enteritidis) has been a global cause of major epidemics of enterocolitis, which have been strongly associated with intensive poultry farming and egg production[1]. The serovar is usually considered to be a generalist in terms of host range and has a low human invasiveness index, typically causing self-limiting enterocolitis[2]. Following a number of interventions in the farming industry involving both improved hygiene and poultry vaccination, epidemic *S*. Enteritidis has been in decline in many countries including the United Kingdom and USA[3,4]. *S.* Enteritidis has also been used extensively since the early 1900s as a rodenticide (named the "Danysz virus"), following development at Institut Pasteur, France. Although by the 1960s, *Salmonella*-based rodenticides had been banned in the US, Germany and the UK, *S*. Enteritidis is still produced as a rodenticide in Cuba, under the name Biorat®[5].

Serovars of *Salmonella* that cause enterocolitis in industrialised settings are strongly associated with life-threatening invasive nontyphoidal *Salmonella* (iNTS) disease in sub-Saharan Africa (SSA). *S*. Enteritidis and *Salmonella enterica* serovar Typhimurium (*S*. Typhimurium) are the two leading causes of iNTS disease in SSA[6] and both are associated with multidrug resistance (MDR)[7]. The clinical syndrome iNTS disease is associated with immunosuppression in the human host, particularly malnutrition, severe malaria and advanced HIV in young children and advanced HIV in adults[8]. It has been estimated to cause 681,000 deaths per year[9].

*Salmonella* is a key example of a bacterial genus in which there is a recognizable genomic signature that distinguishes between a gastrointestinal and an extra-intestinal/invasive lifestyle[10], whereby functions required for escalating growth in an inflamed gut are lost when the lineage becomes invasive[11]. In order to investigate whether there were distinct bacterial characteristics explaining the very different epidemiological and clinical profile of epidemic isolates of serotype *S*. Typhimurium

from SSA and industrialised settings, whole-genome sequence (WGS) investigations of this serovar were previously undertaken. These revealed a novel pathotype of multilocus sequence type (MLST) ST313 from SSA, which differed from clades that cause enterocolitis in industrialised settings, by showing patterns of genomic degradation potentially associated with more invasive disease and differential host adaptation[12-17].

In relation to *S.* Enteritidis, there is a growing body of literature on the evolutionary history, phylogeny and utility of WGS for surveillance of *S*. Enteritidis outbreaks[18-20]. The broadest study of the phylogeny to date revealed five major lineages, but contained only two African isolates[21]. There have also been limited reports of isolates of *S.* Enteritidis from African patients living in Europe that are MDR and which display a distinct phage type (PT 42)[22,23]. We therefore hypothesized that there are distinct lineages of *S*. Enteritidis circulating in both the industrialised and developing world with different origins, likely distinct routes of spread and that are associated with different patterns of disease, which will display the distinct genomic signatures characteristic of differential adaptation. To investigate this we have collected a highly diverse global collection of *S*. Enteritidis isolates and compared them using whole-genome sequencing, the highest possible resolution typing methodology.

## Results

### Isolate collection

In total, 675 isolates of *S.* Enteritidis isolated between 1948 and 2013 were sequenced. The collection originated from 45 countries and six continents (Table 1). 496/675 isolates were from Africa, with 131 from the Republic of South Africa (RSA), a further 353 from the rest of SSA, and 12 from North Africa (Table 1). There were 343 isolates from normally sterile human sites (invasive), 124 non-invasive human isolates (predominantly stool samples) and 40 from animal, food or environmental sources. The full metadata are described in Supplementary Table 1 and have been uploaded to the publically available database Enterobase.

### Phylogeny

675 *S.* Enteritidis genomes and one *Salmonella enterica* serovar Gallinarum were mapped to the *S.* Enteritidis strain P125109 reference sequence, variable regions excluded and the remaining sites were screened for single nucleotide polymorphisms (SNPs). This left an alignment containing a total of 42,373 variable sites, from which a maximum likelihood (ML)-phylogeny was constructed using *S.* Gallinarum, which is a closely related serovar, as an out-group (Figure 1). HierBAPS was run over two rounds, which provided clear distinction between clades/clusters[24]. The phylogeny of *S.* Enteritidis revealed evidence of three clades associated with epidemics, one which we have termed the 'global epidemic clade' and includes the reference PT4 isolate P125109 and two African clades: one predominantly composed of West African isolates (labeled the 'West African clade') and a second composed of isolates predominantly originating in Central and Eastern Africa, called the 'Central/Eastern African clade'). Figure 1 also shows the other clades and clusters predicted by HierBAPS, the largest of which is a paraphyletic

cluster from which the global epidemic clade emerged (Outlier Cluster in Figure 1), and a further five smaller clades or clusters predicted by HierBAPS.

The global epidemic clade contains isolates of multiple phage types, including 4 and 1, which have been linked to the global epidemic of poultry associated human enterocolitis[25]. It comprised 250 isolates from 28 countries, including 43 from Malawi and 82 from RSA. They were isolated from across a 63-year period (1948-2013). Antimicrobial susceptibility testing had been performed on 144 isolates and 104 were susceptible to all antimicrobials tested, five were multidrug resistant (MDR: resistant to 3 or more antimicrobial classes), one was nalidixic acid resistant and none were extended-spectrum beta-lactamase (ESBL)-producing isolates. Database comparison of the genomes from this clade revealed that 221 (88%) of them contained no predicted antimicrobial resistance (AMR) genes apart from the cryptic resistance gene *aac(6')-Iy*[26].

The global epidemic clade has emerged from a diverse cluster previously described by Zheng[27], which encompassed 131 isolates (Figure 1: 'Outlier Cluster'). In addition to being paraphyletic, this group was geographically and temporally diverse, and predominantly drug susceptible (59/71 isolates). Whilst the majority of the diversity of phage typed isolates was contained within the global epidemic clade, this cluster alone contained isolates of phage type 14b, which was recently associated with a multi-country outbreak of *S*. Enteritidis enterocolitis in Europe associated with chicken eggs from Germany[28]. There were also 41 isolates from RSA in this clade, where it has been a common cause of bloodstream infection, and 39 bloodstream isolates from Malawi. Database comparison of the genomes from this clade revealed that 122 (82%) of these genomes contained no predicted AMR genes apart from the cryptic resistance gene *aac(6')-Iy*.

There were two related, but phylogenetically and geographically distinct, epidemic clades that largely originated from SSA. The Central/Eastern African clade included

166 isolates, all but two of which (from RSA) came from this region. Of these, 126/155 (82%) were MDR and 148/153 (97%) displayed phenotypic resistance to between one and four antimicrobial classes. All of these genomes contained at least five predicted resistance genes and 128 (77%) contained nine (Table 2 and Supplementary Table 2). 155/165 (94%) of these isolates were cultured from a normally sterile compartment of a human (i.e. blood or cerebrospinal fluid) and were considered to be causing invasive disease (Table 2). The second African epidemic clade was significantly associated with West Africa with 65/66 isolates coming from this region and one isolate from USA. This clade was also associated with drug resistance (62 [94%] resistant to ≥1 antimicrobial class by phenotype and genotype) and human invasive disease (61 [92%]). It also included two isolates that were subtyped as phage type 4.

The remaining 58 isolates included in this study were extremely diverse, phylogenetically, temporally and geographically. Only two displayed any phenotypic AMR, one of which was MDR. Inspection of the genome revealed that five had predicted AMR genes in addition to *aac(6')-Iy*, four of which were isolated in sub-Saharan Africa. Twenty were associated with invasive human disease, and six were recovered from stool. Three isolates were from stocks of rodenticide and these were phylogenetically remote from both global-epidemic and the two African epidemic clades.

To add further context to these findings we screened the entire publically available Public Health England (PHE) sequenced *Salmonella* routine surveillance collection, which includes 2,986 *S*. Enteritidis genomes, 265 of which were associated with travel to Africa (Supplementary Figure 1). Within this huge collection, including 61 (2.0%) bloodstream isolates and 2670 (89.4%) stool isolates, only 6 isolates (4 from blood culture, 1 from stool) fell within to the West African clade and 1 (from stool) belonged to the Central/Eastern African clade. Notably, these isolates were all either associated with travel to Africa and/or taken from patients of African origin.

231 It is apparent from the location of the archetypal reference isolate and archetypal
232 phage types in the phylogeny (Supplementary Figure 2) that the majority of *S*.
233 Enteritidis studied previously belonged to the global epidemic clade associated with
234 enterocolitis in industrialised countries. Furthermore, its also clear that two
235 additional, previously unrecognized *S*. Enteritidis lineages have emerged, largely
236 restricted to Africa, that are strongly associated with MDR and invasive disease.
237
238 To understand how recently these Africa-associated lineages emerged we used
239 Bayesian Evolutionary Analysis by Sampling Trees (BEAST) to reconstruct the
240 temporal history of the epidemic clades[29]. These data (Supplementary Figure 3)
241 estimate the most recent common ancestor (MRCA) of the Central/Eastern African
242 clade dates to 1945 (95% Credible Interval [CrI]: 1924-1951) and for the West
243 African clade it was 1933 (95% CrI: 1901-1956). We estimate the MRCA of the
244 global epidemic clade originated around 1918 (95% CrI: 1879-1942 –
245 Supplementary Figure 4), with a modern expansion occurring in 1976 (95% CrI:
246 1968-1983), whereas the paraphyletic cluster from which it emerged dates to
247 approximately 1711 (95% CrI: 1420-1868).
248
249 **Contribution of the accessory genome**
250
251 Prophages have the potential to carry non-essential "cargo" genes, which suggests
252 they confer a level of specialization to their host bacterial species, whilst plasmids
253 may confer a diverse array of virulence factors and AMR [30,31]. Therefore it is critical
254 to evaluate the accessory genome in parallel with the core. 622 sequenced genomes
255 were used to determine a pangenome, which yielded a core genome comprising
256 4,076 predicted genes present in ≥90% isolates, including all 12 recognised
257 *Salmonella* Pathogenicity Islands as well as all 13 fimbrial operons found in the
258 P125109 reference[32]. The core gene definition was set to minimize stochastic loss of
259 genes from the core due to errors in individual assemblies across such a large
260 dataset. The accessory genome consisted of 14,015 predicted genes. Of the

accessory genes, 324 were highly conserved across the global and two African epidemic clades, as well as the outlier cluster. Almost all were associated with the acquisition or loss of mobile genetic elements (MGEs) such as prophage or plasmids. Prophage regions have been shown to be stable in *Salmonella* genomes and are potential molecular markers, the presence of which has previously been used to distinguish specific clades[13,33].

The lineage-specific whole gene differences of the major clades are summarized in Figure 2 and plotted against the representatives of the four major clades in Supplementary Figure 5. The lineage specific sequence regions include 57 predicted genes found to be unique to the global epidemic clade (Figure 2), all of which were associated with prophage ϕSE20, a region shown to be essential for invasion of chicken ova and mice in one previous study[34]. There were a further 39 genes conserved in the global epidemic and the paraphyletic outlying cluster, which were absent from both African clades, 26 of which correspond to region of difference (ROD) 21[32]. The Central/Eastern Africa clade contained 77 predicted genes that were absent in the other clades. 33 were associated with the virulence plasmid and a further 40 chromosomal genes were associated with a novel, Fels-2 like prophage region (ϕ*fels-BT*). The West African clade had only 15 distinct predicted genes, 11 of which were plasmid-associated. The two African clades shared a further 102 genes: 48, including a leucine-rich repeat region, were associated with a novel prophage region closely related to *Enterobacter* phage P88, 44 were associated with a Gifsy-1 prophage found in *S*. Bovismorbificans and eight were associated with a Gifsy-2 prophage which has degenerated in the reference P125109.

The *S.* Enteritidis plasmid is the smallest of the generic *Salmonella* virulence plasmids at 58 kb and is unusual in that it contains an incomplete set of *tra* genes that are responsible for conjugative gene transfer. The phylogeny of the *S*. Enteritidis virulence plasmid backbone was reconstructed using reads that mapped to the *S.* Enteritidis reference virulence plasmid, pSENV. 120/675 (18%) genomes

lacked pSENV. The virulence plasmid phylogeny is similar to that of the chromosome, suggesting that they have been stably maintained by each lineage and diversified alongside them (Supplementary Figure 6).

The virulence plasmids from the African clades were much larger than those held in the other clades at ~90 kb. A representative example was extracted from Malawian isolate D7795, sequenced using long read technology to accurately reconstruct it (PacBio; see methods) and denoted pSEN-BT (Accession number LN879484). pSEN-BT is composed of a backbone of pSENV with additional regions that are highly similar to recently sequenced fragments of an novel *S.* Enteritidis virulence plasmid (pUO-SeVR) isolated from an African patient presenting with MDR invasive *S.* Enteritidis in Spain[22]. Plasmid pSEN-BT harbours nine AMR genes (full list in Supplementary Table 2), plus additional genes associated with virulence and a toxin/antitoxin plasmid addiction system. Of note, plasmids from the West African isolates carry resistance gene chloramphenicol acetyl transferase A1 (*cat*A1), whereas the Central/Eastern African strains carry *cat*A2 and tetracycline resistance gene *tet*(A). Like pSENV, the African virulence plasmid contained an incomplete set of *tra* genes and so is not self-transmissible. This was confirmed by conjugation experiments and is consistent with previous reports[22,23]. These observations suggest that the evolution of the *S.* Enteritidis plasmid mirrors that of the chromosome; it is thus not a 'novel' plasmid, but in different SSA locations has acquired different AMR genes.

**Multiple signatures of differential host adaptation**

It has been observed in multiple serovars of *Salmonella* including *S*. Typhi, *S*. Gallinarum and *S*. Typhimurium ST313 that the degradation of genes necessary for the utilization of inflammation-derived nutrients is a marker of that lineage having moved from an intestinal to a more invasive lifestyle[13,14,32,35]. Accordingly, we have looked for similar evidence within a representative example of a MDR, invasive,

Central/Eastern African clade isolate, D7795, that was isolated from the blood of a Malawian child in 2000. The draft genome sequence of D7795 closely resembles that of P125109, however, in addition to the novel prophage repertoire and plasmid genes described above, it harbours a number of predicted pseudogenes or hypothetically disrupted genes (HDGs)[11].

In total, there were 42 putative HDGs in D7795, many of which are found in genes involved in gut colonisation and fecal shedding as well as various metabolic processes such as cobalamine biosynthesis which is a cofactor for anaerobic catabolism of inflammation-derived nutrients, such as ethanolamine, following infection[36]. Curation of the SNPs and insertions or deletions (indels) predicted to be responsible for pseudogenisation across the Central/Eastern African clade and West African clade revealed 37/42 predicted HDGs were fixed in other representatives of the Central/East African clade, with 27 of them being present in over 90% of isolates from that clade. Relatively fewer HDGs in D7795 (19/42) were present in representatives of the West African clade, although 13 were present in ≥90% of isolates (Supplementary Table 3).

In addition to this evidence of reductive evolution in D7795, there were 363 genes containing non-synonymous (NS)-SNPs, which change the amino acid sequence and so may have functional consequences[37]. The two African clades were screened for the presence of these NS-SNPs and 131 were found to be present and completely conserved across both clades, including NS-SNPs in 43 genes encoding predicted membrane proteins, 36 metabolic genes and 23 conserved hypothetical genes (Supplementary Table 4). Furthermore many of these NS-SNPs fall in genes within the same metabolic pathways as the HDGs (see Supplementary Note for detailed description). Supplementary Table 5 provides a list of some of the common traits identified amongst the functions of genes lost independently by D7795, *S.* Typhi and *S*. Gallinarum. The disproportionate clustering of mutations in membrane structures

observed in the African clades is yet another sign of differential host adaptation analogous to that reported in both *S*. Typhi[35] and *S*. Gallinarum[32].

## Biolog™ growth substrate platform profiling

The Biolog™ platform was utilized to generate a substrate growth utilisation profile for selected *S*. Enteritidis isolates. Corresponding signal values of replicate pairs of a Central/Eastern African isolate (D7795) and a global epidemic isolate (A1636) were compared using principal component analysis and found to be highly consistent. In total, 80 metabolites showed evidence of differential metabolic activity (Figure 3). Evaluation of data from the Central/Eastern African isolate using Pathway Tools software revealed that 14/27 (52%) of pathways with evidence of decreased metabolic activity at 28°C had a corresponding component of genomic degradation. This was also true for 12/30 (40%) of pathways with evidence of decreased metabolic activity at 37°C.

Instances of reduced metabolic activity in a Central/Eastern African strain (D7795) compared to a global epidemic strain (A1636) included dulcitol and glycolic acid in the glycerol degradation pathway, propionic acid in the propanediol pathway and ethylamine and ethanolamine. These are all vitamin B12 (cobalamin) dependent reactions, for which there was a corresponding signature of genomic degradation. Also there was reduced activity in response to three forms of butyric acid, alloxan and allantoic acid metabolism. Allantoin can be found in the serum of birds, but not humans and is utilised as a carbon source during *S.* Enteritidis infection of chickens[38], and HDGs relating to allantoin have been noted in *S.* Typhimurium ST313[13]. The full list of differences is detailed in Supplementary Table 6 and 7. This is a further sign of decreased metabolism of the Central/Eastern African isolate in the anaerobic environment of the gut.

## Chick infection model suggests divide in host range

Given the phenotypic differences observed in the genotypically distinct global and African clades, we hypothesized that these lineages could have differing infection phenotypes in an *in vivo* challenge model. We compared the infection profile of a member of the Central/Eastern African clade (D7795) to the reference global epidemic strain P125109 in an avian host. The chicken group infected with P125109 showed mild hepatosplenomegaly consistent with infection by this *Salmonella* serovar and cecal colonization (Figure 4A-C). In contrast, the Central/Eastern African strain displayed significantly reduced invasion at 7 dpi of both liver ($p=0.027$) and spleen ($p=0.007$), however cecal colonization was not significantly reduced ($p=0.160$). This is in marked contrast to the behavior of *S*. Typhimurium ST313, which is more invasive in a chick infection model[12].

## Discussion

*S.* Enteritidis is an example of a successful *Salmonella* lineage with the apparent ability to adapt to different hosts and transmission niches as and when opportunities for specialization have presented. Langridge *et al* recently evaluated the Enteritidis/Gallinarum/Dublin lineage of *Salmonella*, revealing components of the nature and order of events associated with host-range and restriction[39]. In the present study, we have highlighted the plasticity of *S*. Enteritidis, providing evidence of three distinct epidemics of human disease. In addition we show multiple additional clades and clusters that demonstrate the huge reservoir of diversity amongst *S.* Enteritidis from which future epidemics might emerge.

An important question posed by this study is why have distinct clades of *Salmonella* emerged to become prominent causes of iNTS disease in Africa, from a serotype normally considered to be weakly invasive? The presence of a highly immunosuppressed population due to the HIV pandemic is clearly a key host factor

that facilitates the clinical syndrome iNTS disease[40,41]. In addition to human host factors, there are two distinct African epidemic lineages that have emerged in the last 90 years. Both lineages are significantly associated with a novel prophage repertoire, an expanded, MDR-augmented virulence plasmid, and patterns of genomic degradation with similarity to other host-restricted invasive *Salmonella* serotypes including *S*. Typhi and *S*. Gallinarum and to clades of *S.* Typhimurium associated with invasive disease in Africa[13,32,35]. This pattern of genomic degradation is concentrated in pathways specifically associated with an enteric lifestyle, however it is noteworthy that in the chick infection model, the African *S*. Enteritidis invaded the chick liver and spleen less well than the global pandemic clade. This raises the possibility that the two clades occupy different ecological niches outside the human host or that they behave differently within the human host and screening of the huge *S.* Enteritidis collection from routine *Salmonella* surveillance by PHE supports the assertion that these lineages are geographically restricted to Africa. This study therefore indicates a need to understand what these ecological niches might be, and then to define the transmission pathways of African clades of *S*. Enteritidis, in order to facilitate public health interventions to prevent iNTS disease.

The evolution of the *S*. Enteritidis virulence plasmid is intriguing; pSENV is the smallest of the known *Salmonella* virulence-associated plasmids, but in SSA, the plasmid has nearly doubled in size partly through the acquisition of AMR genes. The absence of *tra* genes necessary for conjugal transfer either indicates that MDR status has evolved through acquisition of MGEs multiple times or through clonal expansion and vertical transmission of the plasmid to progeny. The available data suggest that the former scenario has happened twice, once in West Africa, and once in Central/Eastern Africa.

Despite *S*. Enteritidis being reported as a common cause of bloodstream infection (BSI) in Africa[6,7] the Global Enteric Multicenter Study (GEMS) found that *Salmonella*

serotypes were an uncommon cause of moderate to severe diarrhoea in African children less than 5-years of age[42]. Our data associating the African lineages with invasive disease is also consistent with data presented in a recent Kenyan study comparing a limited number and diversity of *S*. Enteritidis isolates from blood and stool. Applying the lineages defined in this study to the genome data reported from Kenya showed that 20.4% of isolates from that study belonging to the global clade were associated with invasive disease, whereas 63.2% of the isolates in that study belonging to our Central/Eastern African clade were associated with invasive disease[43]. The remaining isolates were associated with cases of enterocolitis or asymptomatic carriage, confirming that the Central/Eastern African clade can also cause enterocolitis. The association of *S*. Enteritidis clades circulating in sub-Saharan Africa with iNTS disease may reflect the fact that their geographical distribution permits them to behave as opportunistic invasive pathogens in a setting where advanced immunosuppressive disease is highly prevalent in human populations.

In summary, two clades of *S.* Enteritidis have emerged in Africa, which have different phenotypes and genotypes to the strains of *S*. Enteritidis circulating in the industrial world. These strains display evidence of changing host adaptation, different virulence determinants and multi-drug resistance, a parallel situation to the evolutionary history of *S*. Typhimurium ST313. They may have different ecologies and/or host ranges to global strains and have caused epidemics of BSI in at least three countries in SSA, yet are rarely responsible for disease in South Africa. An investigation into the environmental reservoirs and transmission of these pathogens is warranted and urgently required.

**URLs:**

Enterobase: https://enterobase.warwick.ac.uk/

**Data access:** Accession numbers for sequencing data including both raw sequencing reads and assembled sequences are available in Supplementary Table 1.

**ACKNOWLEDGEMENTS**

This work was supported by The Wellcome Trust. We would like to thank the members of the Pathogen Informatics Team and the core sequencing teams at the Wellcome Trust Sanger Institute (Cambridge, UK). We are grateful to David Harris for work in managing the sequence data.

**FUNDING**

This work was supported by a number of organizations. The Wellcome Trust Sanger Institute authors were funded by Wellcome Trust Award 098051; NF was supported by the Wellcome Trust Research Fellowship WT092152MA. NF, RSH and this work were supported by a strategic award from the Wellcome Trust for the MLW Clinical Research Programme (101113/Z/13/Z). The Institut Pasteur (IP) authors were funded by the IP, the Institut de Veille Sanitaire, and by the French Government "Investissement d'Avenir" program (Integrative Biology of Emerging Infectious Diseases" Laboratory of Excellence, grant no. ANR-10-LABX-62-IBEID). JJ was supported by the antibiotic resistance surveillance project in DR Congo, funded by Project 2.01 of the Third Framework Agreement between the Belgian Directorate General of Development Cooperation and the Institute of Tropical Medicine, Antwerp, Belgium. SK was supported by the NIH Grant Number R01 AI099525-02. AEM was supported by Wellcome Trust grant 098051 whilst at the Wellcome Trust Sanger Institute, and Biotechnology and Biological Sciences Research Council grant BB/M014088/1 at the University of Cambridge. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**AUTHOR CONTRIBUTIONS**

Study design: NAF, NRT, MAG, GD, RAK, JP. Data analysis: NAF, NRT, JH, TF, LB, PQ LLL, GL, SRH, AEM, MF, MA. Isolate acquisition and processing and clinical data collection: NAF, KHK, JJ, XD, CMe, SK, CMl, RSO, FXW, SLH AMS, MM, PD, CMP, JC, NF, JC, JAC, LBe, KLH, TJH, OL, TAC, M T, SS, SMT, KB, MML, DBE, RSH. Manuscript writing: NAF, JH, NRT, MAG. All authors contributed to manuscript editing.

## References


1. Ward, L.R., Threlfall, J., Smith, H.R. & O'Brien, S.J. Salmonella enteritidis epidemic. *Science* **287**, 1753-4; author reply 1755-6 (2000).

2. Jones, T.F. *et al.* Salmonellosis outcomes differ substantially by serotype. *J Infect Dis* **198**, 109-14 (2008).

3. O'Brien, S.J. The "decline and fall" of nontyphoidal salmonella in the United kingdom. *Clin Infect Dis* **56**, 705-10 (2013).

4. Braden, C.R. Salmonella enterica serotype Enteritidis and eggs: a national epidemic in the United States. *Clin Infect Dis* **43**, 512-7 (2006).

5. Friedman, C.R., Malcolm, G., Rigau-Perez, J.G., Arambulo, P., 3rd & Tauxe, R.V. Public health risk from Salmonella-based rodenticides. *Lancet* **347**, 1705-6 (1996).

6. Reddy, E.A., Shaw, A.V. & Crump, J.A. Community-acquired bloodstream infections in Africa: a systematic review and meta-analysis. *Lancet Infect Dis* **10**, 417-32 (2010).

7. Feasey, N.A. *et al.* Three Epidemics of Invasive Multidrug-Resistant Salmonella Bloodstream Infection in Blantyre, Malawi, 1998-2014. *Clin Infect Dis* **61 Suppl 4**, S363-71 (2015).

8. Feasey, N.A., Dougan, G., Kingsley, R.A., Heyderman, R.S. & Gordon, M.A. Invasive non-typhoidal salmonella disease: an emerging and neglected tropical disease in Africa. *Lancet* **379**, 2489-99 (2012).

9. Ao, T.T. *et al.* Global burden of invasive nontyphoidal salmonella disease, 2010(1). *Emerg Infect Dis* **21**(2015).

10. Nuccio, S.P. & Baumler, A.J. Reconstructing pathogen evolution from the ruins. *Proc Natl Acad Sci U S A* **112**, 647-8 (2015).

11. Nuccio, S.P. & Baumler, A.J. Comparative analysis of Salmonella genomes identifies a metabolic network for escalating growth in the inflamed gut. *MBio* **5**, e00929-14 (2014).

12. Parsons, B.N. *et al.* Invasive non-typhoidal Salmonella typhimurium ST313 are not host-restricted and have an invasive phenotype in experimentally infected chickens. *PLoS Negl Trop Dis* **7**, e2487 (2013).

13. Kingsley, R.A. *et al.* Epidemic multiple drug resistant Salmonella Typhimurium causing invasive disease in sub-Saharan Africa have a distinct genotype. *Genome Res* **19**, 2279-87 (2009).

14. Okoro, C.K. *et al.* Signatures of adaptation in human invasive Salmonella Typhimurium ST313 populations from sub-Saharan Africa. *PLoS Negl Trop Dis* **9**, e0003611 (2015).

15. Herrero-Fresno, A. *et al.* The role of the st313-td gene in virulence of Salmonella Typhimurium ST313. *PLoS One* **9**, e84566 (2014).

16. Ramachandran, G., Perkins, D.J., Schmidlein, P.J., Tulapurkar, M.E. & Tennant, S.M. Invasive Salmonella Typhimurium ST313 with naturally attenuated flagellin elicits reduced inflammation and replicates within macrophages. *PLoS Negl Trop Dis* **9**, e3394 (2015).

17. Okoro, C.K. *et al.* Intracontinental spread of human invasive Salmonella Typhimurium pathovariants in sub-Saharan Africa. *Nat Genet* **44**, 1215-21 (2012).

18. Allard, M.W. *et al.* On the evolutionary history, population genetics and diversity among isolates of Salmonella Enteritidis PFGE pattern JEGX01.0004. *PLoS One* **8**, e55254 (2013).

19. den Bakker, H.C. *et al.* Rapid whole-genome sequencing for surveillance of Salmonella enterica serovar enteritidis. *Emerg Infect Dis* **20**, 1306-14 (2014).

20. Taylor, A.J. *et al.* Characterization of Foodborne Outbreaks of Salmonella enterica Serovar Enteritidis with Whole-Genome Sequencing Single Nucleotide Polymorphism-Based Analysis for Surveillance and Outbreak Detection. *J Clin Microbiol* **53**, 3334-40 (2015).

21. Deng, X. *et al.* Genomic epidemiology of Salmonella enterica serotype Enteritidis based on population structure of prevalent lineages. *Emerg Infect Dis* **20**, 1481-9 (2014).

22. Rodriguez, I., Guerra, B., Mendoza, M.C. & Rodicio, M.R. pUO-SeVR1 is an emergent virulence-resistance complex plasmid of Salmonella enterica serovar Enteritidis. *J Antimicrob Chemother* **66**, 218-20 (2011).

566 23. Rodriguez, I., Rodicio, M.R., Guerra, B. & Hopkins, K.L. Potential international
567 spread of multidrug-resistant invasive Salmonella enterica serovar enteritidis.
568 *Emerg Infect Dis* **18**, 1173-6 (2012).

569 24. Cheng, L., Connor, T.R., Siren, J., Aanensen, D.M. & Corander, J. Hierarchical
570 and spatially explicit clustering of DNA sequences with BAPS software. *Mol Biol*
571 *Evol* **30**, 1224-8 (2013).

572 25. Rodrigue, D.C., Tauxe, R.V. & Rowe, B. International increase in Salmonella
573 enteritidis: a new pandemic? *Epidemiol Infect* **105**, 21-7 (1990).

574 26. Magnet, S., Courvalin, P. & Lambert, T. Activation of the cryptic aac(6')-Iy
575 aminoglycoside resistance gene of Salmonella by a chromosomal deletion
576 generating a transcriptional fusion. *J Bacteriol* **181**, 6650-5 (1999).

577 27. Zheng, J. *et al.* Genetic diversity and evolution of Salmonella enterica serovar
578 Enteritidis strains with different phage types. *J Clin Microbiol* **52**, 1490-500
579 (2014).

580 28. Inns, T. *et al.* A multi-country Salmonella Enteritidis phage type 14b outbreak
581 associated with eggs from a German producer: 'near real-time' application of
582 whole genome sequencing and food chain investigations, United Kingdom, May
583 to September 2014. *Euro Surveill* **20**(2015).

584 29. Drummond, A.J. & Rambaut, A. BEAST: Bayesian evolutionary analysis by
585 sampling trees. *BMC Evol Biol* **7**, 214 (2007).

586 30. Thomson, N. *et al.* The role of prophage-like elements in the diversity of
587 Salmonella enterica serovars. *J Mol Biol* **339**, 279-300 (2004).

588 31. Rotger, R. & Casadesus, J. The virulence plasmids of Salmonella. *Int Microbiol*
589 **2**, 177-84 (1999).

590 32. Thomson, N.R. *et al.* Comparative genome analysis of Salmonella Enteritidis
591 PT4 and Salmonella Gallinarum 287/91 provides insights into evolutionary and
592 host adaptation pathways. *Genome Res* **18**, 1624-37 (2008).

593 33. Cooke, F.J. *et al.* Prophage sequences defining hot spots of genome variation in
594 Salmonella enterica serovar Typhimurium can be used to discriminate between
595 field isolates. *J Clin Microbiol* **45**, 2590-8 (2007).

596 34. Shah, D.H., Zhou, X., Kim, H.Y., Call, D.R. & Guard, J. Transposon mutagenesis
597 of Salmonella enterica serovar Enteritidis identifies genes that contribute to
598 invasiveness in human and chicken cells and survival in egg albumen. *Infect*
599 *Immun* **80**, 4203-15 (2012).

600 35. Parkhill, J. *et al.* Complete genome sequence of a multiple drug resistant
601 Salmonella enterica serovar Typhi CT18. *Nature* **413**, 848-52 (2001).

36. Thiennimitr, P. *et al.* Intestinal inflammation allows Salmonella to use ethanolamine to compete with the microbiota. *Proc Natl Acad Sci U S A* **108**, 17480-5 (2011).

37. Kingsley, R.A. *et al.* Genome and transcriptome adaptation accompanying emergence of the definitive type 2 host-restricted Salmonella enterica serovar Typhimurium pathovar. *MBio* **4**, e00565-13 (2013).

38. Dhawi, A.A. *et al.* Adaptation to the chicken intestine in Salmonella Enteritidis PT4 studied by transcriptional analysis. *Vet Microbiol* **153**, 198-204 (2011).

39. Langridge, G.C. *et al.* Patterns of genome evolution that have accompanied host adaptation in Salmonella. *Proc Natl Acad Sci U S A* **112**, 863-8 (2015).

40. Feasey, N.A. *et al.* Modelling the Contributions of Malaria, HIV, Malnutrition and Rainfall to the Decline in Paediatric Invasive Non-typhoidal Salmonella Disease in Malawi. *PLoS Negl Trop Dis* **9**, e0003979 (2015).

41. Gordon, M.A. *et al.* Non-typhoidal salmonella bacteraemia among HIV-infected Malawian adults: high mortality and frequent recrudescence. *AIDS* **16**, 1633-41 (2002).

42. Kotloff, K.L. *et al.* Burden and aetiology of diarrhoeal disease in infants and young children in developing countries (the Global Enteric Multicenter Study, GEMS): a prospective, case-control study. *Lancet* **382**, 209-22 (2013).

43. Kariuki, S. & Onsare, R.S. Epidemiology and Genomics of Invasive Nontyphoidal Salmonella Infections in Kenya. *Clin Infect Dis* **61 Suppl 4**, S317-24 (2015).

633 **Figures**

634

635 Figure 1: Maximum likelihood phylogeny of *S*. Enteritidis based on 675 isolates
636 rooted to *S*. Gallinarum. There are 3 epidemic clades; 2 African epidemic clades and
637 a global epidemic clade. Scale bar indicates nucleotide substitutions per site.

638

639 Figure 2: Differences in accessory genomes of 4 major clades. Approximate position
640 of prophages in chromosome is depicted, although prophages are not drawn to scale

641

642 Figure 3: Heat map revealing changes in metabolic activity of Central/Eastern
643 African clade isolate D7795 when compared to global epidemic isolate A1636 at 28
644 and 37˚C. The figure also displays whether there are corresponding mutations in
645 genes related to the affected metabolic pathway. (NSSNP=Non-synonymous single
646 nucleotide polymorphism, HDG = Hypothetically disrupted gene)

647

648 Figure 4: *Salmonella* isolation from a chick infection model demonstrates failure of
649 Central/Eastern African clade isolate to invade chicken spleen (4A) and liver (4B) or
650 to colonize chicken caeca (4C) at 7 days post infection (dpi) (n=24 at this time point)
651 compared to the global epidemic clade. Numbers are expressed as colony forming
652 units (CFU) per gram of tissue

653

654

655

656

657

658

659

## Tables

Table 1: Summary of metadata (n) by region in numbers

| Region | Total | Site of isolation | | | Antimicrobial resistance phenotype | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Human Invasive | Human non-invasive | Food/Animal/ Environment | Drug susceptible | Resistant to 1-2 1st line | MDR* | Fluoroquinolone | ESBL† |
| **Asia** | 11 | 5 | 5 | 1 | 0 | 0 | 0 | 0 | 0 |
| **Europe** | 61 | 0 | 16 | 24 | 2 | 0 | 0 | 0 | 0 |
| **South America** | 27‡ | 3 | 6 | 7 | 8 | 0 | 0 | 0 | 0 |
| **North Africa** | 12 | 9 | 1 | 1 | 9 | 0 | 0 | 2 | 0 |
| **Sub-Saharan Africa** | 353 | 269 | 22 | 7 | 99 | 64 | 14 9 | 0 | 3 |
| **Republic of South Africa** | 131 | 57 | 74 | 0 | 83 | 44 | 4 | 0 | 0 |

*Multidrug resistant: resistant to ≥3 antimicrobials

†Extended spectrum beta lactamase producing

‡Uruguay strains previously characterised by Betancor [44]

Table 2: Metadata summarised by clade

| Major Clade/cluster | Site of isolation N (%) | | | | Number (%) of antimicrobial resistance genes* | | |
|---|---|---|---|---|---|---|---|
| | Human Invasive | Human non-invasive | Food/Animal/ Environment | Unknown | 1-3 | 4-6 | 7-9 |
| **West African** | 61 (92) | 1 (2) | 0 (0) | 4 (6) | 22 (33) | 9 (14) | 35 (66) |
| **Central/Eastern African** | 155 (93) | 7 (4) | 0 (0) | 5 (3) | 0 (0) | 11 (7) | 156 (93) |
| **Global epidemic** | 94 (38) | 95 (38) | 31 (12) | 30 (12) | 243 (97) | 7 (3) | 0 (0) |
| **Outlier cluster** | 51 (38) | 36 (27) | 27 (20) | 20 (15) | 128 (96) | 3 (2) | 3 (2) |

*All isolates contained cryptic aminoglycoside acetyltransferase gene *aac(6')-ly*[26]

## Online Methods

### Bacterial Isolates

*S*. Enteritidis isolates were selected on the basis of six factors; date of original isolation, antimicrobial susceptibility pattern, geographic site of original isolation, source (human [invasive vs stool], animal or environmental), phage type (where available), and multilocus variable number tandem repeat (MLVA) type (where available). *S.* Enteritidis P125109 (EMBL accession no. AM933172) isolated from a poultry farm from the UK was used as a reference[32]. The full metadata are in Supplementary Table 1. Isolates have been attributed to region according to United Nations statistical divisions.

### Sequencing, SNP-calling, construction of phylogeny and comparative genomics

PCR libraries were prepared from 500 ng of DNA as previously described[45]. Isolates were sequenced using Illumina GA II, HiSeq 2000 and MiSeq machines (Illumina, San Diego, CA, USA) and 150 bp paired-end reads were generated. The strains were aligned to *Salmonella* Enteritidis reference genome P125109 using a pipeline developed in-house at the Wellcome Trust Sanger Institute (WTSI). For each isolate sequenced, the raw sequence read pairs were split to reduce the overall memory usage and allow reads to be aligned using more than one CPU. The reads were then aligned using SMALT, a hashing based sequence aligner. The aligned and unmapped reads were combined into a single BAM file. Picard was used to identify and flag optical duplicates generated during the making of a standard Illumina library, which reduces possible effects of PCR bias. All of the alignments were created in a standardized manner, with the commands and parameters stored in the header of each BAM file, allowing for the results to be easily reproduced.

The combined BAM file for each isolate was used as input data in the SAMtools

mpileup program to call SNPs and small indels, producing a BCF file describing all of the variant base positions[46]. A pseudo-genome was constructed by substituting the base call at each variant or non-variant site, defined in the BCF file, in the reference genome. Only base calls with a depth of coverage >4 or quality >50 were considered in this analysis. Base calls in the BCF file failing this quality control filter were replaced with the "N" character in the pseudo-genome sequence.

All of the software developed is freely available for download from GitHub under an open source license, GNU GPL 3.

Phylogenetic modelling was based on the assumption of a single common ancestor, therefore variable regions where horizontal genetic transfer occurs were excluded[47,48]. A maximum likelihood (ML) phylogenetic tree was then built from the alignments of the isolates using RAxML (version 7.0.4) using a GTR+I+G model[49]. The maximum-likelihood phylogeny was supported by 100 bootstrap pseudo-replicate analyses of the alignment data. Clades were predicted using Hierarchical Bayesian Analysis of Population Structure (HierBAPS)[24]. This process was repeated to construct the plasmid phylogeny, using reads that aligned to pSENV.
To ascertain the presence of the clusters defined by HierBAPs in the Public Health England (PHE) routine *Salmonella* surveillance collection, seventeen isolates representing the diversity of the collection were compared against 2986 *S*. Enteritidis PHE genomes. Single linkage SNP clustering was performed as previously described[50]. A maximum-likelihood phylogeny showing the integration of the seventeen isolates with 50-SNP cluster representatives of the PHE *S*. Enteritidis collection was constructed as above. FASTQ reads from all PHE sequences in this study can be found at the PHE Pathogens BioProject at the National Center for Biotechnology Information (Accession PRJNA248792).

Temporal reconstruction was performed using Bayesian Evolutionary Analysis Sampling Trees (BEAST version 1.8.2)[51]. A relaxed lognormal clock model was

initially employed. The results of this model indicated that a constant clock model was not appropriate, as the posterior of the standard deviation of the clock rate did not include zero. A range of biologically plausible population models (constant, exponential and skyline) was investigated. Skyline models can be biased by non-uniform sampling and we observed a strong similarity between reconstructed skyline population and the histogram of sampling dates and so this model was excluded. The exponential models consistently failed to converge and were excluded. Thus, for all datasets, lognormal clock and constant population size models were used. The computational expense required for this analysis precluded running estimators for model selection. However, we note that Deng et al used the same models in their analysis of 125 *S*. Enteritidis isolates. Default priors were used except for ucld.mean, Gamma(0.001,1000), initial: 0.0001; exponential.popSize, LogNormal(10,1.5), initial: 1[21].

Three chains of 100 million states were run in parallel for each clade of the four major HierBAPS clades, as well as a fourth chain without genomic data to examine the influence of the prior, which in all cases was uninformative. The final results, as used here, all had effective sample sizes (ESS) of over 200 and had convergence between all three runs. For the Global and Global Outlier lineages, the datasets were not computationally feasible to analyse. We thus created 3 further random subsets of the data by drawing n isolates from each sampled year where n was sampled from a Poisson distribution where $\lambda=2$. The posteriors of all subsets were extremely similar and runs were combined to produce the final most recent common ancestor (MRCA) estimates.

In order to gain a detailed insight into genomic differences, a single high quality sequence from Malawian *S*. Enteritidis isolate D7795 was aligned against the P125109 using ABACAS and annotated[52]. Differences were manually curated against the reference using the Artemis Comparison Tool (ACT)[53]. Sections of contigs which were incorporated into the alignment, but which did not align with P125109 were

767 manually inspected and compared to the public databases using BLASTn. When
768 these regions appeared to be novel prophages, they were annotated using the phage
769 search tool PHAST and manually curated[54]. In order to investigate whether the SNPs
770 and/or indels that were predicted to be responsible for pseudogene formation in
771 D7795 were distinct to that isolate or conserved across both African epidemic
772 clades, all isolates were aligned to P12509 and the relevant SNPs/indels
773 investigated using *in-silico* PCR of the aligned sequences. Manual curation was
774 performed to confirm the nature of all pseudogene-associated SNPs/indels. NS-SNPs
775 identified in D7795 were sorted throughout the African clades by extracting and
776 aligning the appropriate gene sequences from P125109 and D7795. The coordinates
777 of the NS-SNPs were then used to identify the relevant sequence and determine the
778 nature of the base.
779
780 Accessory genome
781 The pangenome for the dataset was predicted using ROARY [55]. Genes were
782 considered to be core to *S*. Enteritidis if present in ≥90% of isolates. A relaxed
783 definition of core genome was used as assemblies were used to generate it and the
784 more assemblies one uses, the more likely it is that a core gene will be missed in one
785 sample due to an assembly error. The remaining genes were considered to be core
786 to the clades/clusters predicted by HierBAPS if present in ≥75% if isolates from
787 within each clade/cluster. These genes were then curated manually using ACT to
788 search for their presence and position in P125109 or the improved draft assembly
789 of representative isolates of each of the other clades if not present in P125109. Any
790 large accessory regions identified were blasted against the assembled genomes of
791 the entire collection to confirm they were grossly intact.
792
793 Plasmid identification
794 Plasmid DNA was extracted from isolate D7795 using the Kado & Liu method and
795 separated by gel-electrophoresis alongside plasmids of known size, to estimate the
796 number and size of plasmids present[56]. Plasmid conjugation was attempted by

mixing 100 µL of overnight culture of donor and recipient strains (rifampicin resistant *Escherichia coli* C600) on Luria-Bertani agar plates and incubating overnight at 26°C and 37°C. The plasmid was sequenced using the PacBio platform to gain long reads and a single improved draft assembly, which was aligned against P125109 plasmid pSENV (Accession Number HG970000). For novel regions of the plasmid from isolate D7795, genes were predicted using GLIMMER and manual annotations applied based on homology searches against the public databases, using both BLASTn and FASTA. The plasmid phylogeny was reconstructed using the same methodology as the chromosome; a maximum likelihood (ML) phylogenetic tree was built from the alignments of the isolates using RAxML (version 7.0.4) using a GTR+I+G model

Identification of AMR genes

A manually curated version of the Resfinder database was used to investigate the isolates for the presence of AMR genes[57]. To reduce redundancy, the database was clustered using CD-HIT-EST[58], with the alignment length of the shorter sequence required to be 90% the length of the longer sequence. All other options were left as the defaults. The representative gene of each cluster was then mapped with SMALT to the assemblies of each isolate to identify and matches with an identity of 90% or greater were considered significant, in line with the default clustering parameters of CD-HIT-EST. Where partial matches were identified at the ends of contigs, having an identity of 90% or greater to the matched region of the gene, potential AMR gene presence was recorded. To confirm presence of these partial matches, raw sequencing reads of the pertinent isolates were mapped using SMALT to these genes to check for 90% identity across the entire gene.

Biolog™ growth substrate platform profiling

826 The Biolog™ platform enables the simultaneous quantitative measurement of a
827 number of cellular phenotypes, and therefore the creation of a phenotypic profile of
828 a variety of assay conditions[59]. Incubation and recording of phenotypic data were
829 performed using an OmniLog® plate reader.
830
831 In these experiments, two replicates of D7795 were compared to two replicates of a
832 PT4 like strain at 28˚C and 37˚C to represent environmental and human
833 temperatures. Biolog™ plates PM1-4 and 9 (Carbon source [PM1,PM2], nitrogen
834 source [PM3] and phosphor and sulphur source [PM4] metabolism and osmotic
835 pressure [PM9]) were used. Each well was inoculated as described below, thereby
836 testing 475 conditions at once (each plate has one negative control well).
837

838 The isolates were cultured overnight on LB-agar at 37°C in air to exclude
839 contamination. Colonies were scraped off plates and dispensed into IF-0a solution
840 (Biolog) to a cell density corresponding to 81% transmittance. For each plate used,
841 880 µL of this cell suspension was added to 10 mL IF-10b GP/GP solution (Biolog)
842 and 120 µL dye mix G (Biolog). This was then supplemented with a 1 mL solution of
843 7.5 mM D-ribose (Sigma), 2 mM magnesium chloride, 1 mM calcium chloride, 2 mM
844 sodium pyrophosphate (Sigma), 25 µM L-arginine (Sigma), 25 µM L-methionine
845 (Sigma), 25 µM hypoxanthine (Sigma), 10 µM lipoamide (Sigma), 5 µM nicotine
846 adenine dinucleotide (Sigma), 0.25 µM riboflavin (Sigma), 0.005% by mass yeast
847 extract (Fluka) and 0.005% by mass Tween 80 (Sigma). 100µl of this mixture was
848 dispensed into each well on the assay plate. Plates were then allowed to equilibrate
849 in air for 5 min prior to being sealed in airtight bags and loaded into the Omnilog
850 machine (Biolog). Plates were scanned every 15 min for 48 hours while incubated at
851 28°C and 37°C in air. Culture under anaerobic conditions was unavailable. Two
852 paired replicates were performed for each of the two isolates.

853 After completion of the run, the signal data were compiled and analysed using the
854 limma package in 'R' described previously[60]. A log-fold change of 0.5 controlling for

855 a 5% false discovery rate was used as a cut-off for investigating a specific metabolite
856 further using Pathway Tools[61] and whether the metabolic change was related to
857 pseudogenes and non synonymous(NS)-SNPs in genes in the respective genomes.
858
859 *In vivo* Infection Model
860
861 Two isolates were used in the animal models: *S*. Enteritidis P125109 and D7795.
862 Unvaccinated commercial female egg-layer Lohmann Brown chicks (Domestic Fowl
863 [*Gallus gallus*]) were obtained from a commercial hatchery and housed in secure
864 floor pens at a temperature of 25°C. Eight chicks per strain per time point were
865 inoculated by gavage at 10 days (d) of age and received a dose of ~$10^8$ *Salmonella*
866 colony forming units (CFU) in a volume of 0.2 mL. Subsequently, four to five birds
867 from each group were humanely killed at 3, 7 or 21 d post-infection (p.i.). At post
868 mortem, the liver, spleen, and caecal contents were removed aseptically,
869 homogenised, serially diluted and dispensed onto Brilliant Green agar (Oxoid) to
870 quantify colony forming units (CFU) as described previously[62]. Statistical analysis
871 was performed using SPSS, version 20 (IBM). Kruskal-Wallis was used to compare
872 bacterial loads between infected groups.
873
874 All work was conducted in accordance with the UK legislation governing
875 experimental animals, Animals (Scientific Procedures) Act 1986, under project
876 licence 40/3652 and was approved by the University of Liverpool ethical review
877 process prior to the award of the project license. The licensing procedure requires
878 power calculations to determine minimal group sizes for each procedure to ensure
879 results are significant. For these experiments a group size of 8 birds per time point
880 was chosen, based on a variation in 1.0 $\log_{10}$ in bacterial count between groups as
881 being significant along with prior experience of *Salmonella* infection studies. Groups
882 were randomly selected on receipt from the hatchery and investigators conducting
883 animal experiments were not blinded, as the current UK code of practice requires all
884 cages or pens to be fully labeled with experimental details. No animals were

885 excluded from the analysis. All animals were checked a minimum of twice daily to
886 ensure their health and welfare.
887
888 Code availability & URLs
889
890 Software is referenced and URLS are provided below. All software is open source.
891
892 BEAST: http://beast.bio.ed.ac.uk/
893 Biolog™: http://www.biolog.com
894 BLASTn: http://blast.ncbi.nlm.nih.gov
895 limma package: www.bioconductor.org
896 PacBio platform: http://www.pacificbiosciences.com/
897 Picard: https://broadinstitute.github.io/picard
898 'R': www.R-project.org
899 SMALT: www.sanger.ac.uk/science/tools/smalt-0
900 United Nations statistical divisions:
901 www.unstats.un.org/unsd/methods/m49/m49regin.htm

902 Supplementary References


903 44. Betancor, L. *et al.* Genomic and phenotypic variation in epidemic-spanning
904 Salmonella enterica serovar Enteritidis isolates. *BMC Microbiol* **9**, 237 (2009).

905 45. Harris, S.R. *et al.* Whole-genome sequencing for analysis of an outbreak of
906 meticillin-resistant Staphylococcus aureus: a descriptive study. *Lancet Infect Dis*
907 **13**, 130-6 (2013).

908 46. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics*
909 **25**, 2078-9 (2009).

910 47. Holt, K.E. *et al.* High-throughput sequencing provides insights into genome
911 variation and evolution in Salmonella Typhi. *Nat Genet* **40**, 987-93 (2008).

912 48. Croucher, N.J. *et al.* Rapid pneumococcal evolution in response to clinical
913 interventions. *Science* **331**, 430-4 (2011).

914 49. Stamatakis, A. RAxML-VI-HPC: maximum likelihood-based phylogenetic
915 analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688-90
916 (2006).

917 50. Philip Ashton, S.N., Tansy Peters, Rediat Tewolde, Martin Day, Michel
918 Doumith, Jonathan Green, Claire Jenkins, Anthony Underwood, Catherine
919 Arnold, Elizabeth de Pinna, Tim Dallman, Kathie Grant. Revolutionising
920 Public Health Reference Microbiology using Whole Genome Sequencing:
921 Salmonella as an exemplar. *bioRxiv* (2015).

922 51. Drummond, A.J., Suchard, M.A., Xie, D. & Rambaut, A. Bayesian phylogenetics
923 with BEAUti and the BEAST 1.7. *Mol Biol Evol* **29**, 1969-73 (2012).

924 52. Assefa, S., Keane, T.M., Otto, T.D., Newbold, C. & Berriman, M. ABACAS:
925 algorithm-based automatic contiguation of assembled sequences. *Bioinformatics*
926 **25**, 1968-9 (2009).

927 53. Carver, T.J. *et al.* ACT: the Artemis Comparison Tool. *Bioinformatics* **21**, 3422-3
928 (2005).

929 54. Zhou, Y., Liang, Y., Lynch, K.H., Dennis, J.J. & Wishart, D.S. PHAST: a fast
930 phage search tool. *Nucleic Acids Res* **39**, W347-52 (2011).

931 55. Page, A.J. *et al.* Roary: Rapid large-scale prokaryote pan genome analysis.
932 *Bioinformatics* (2015).

933 56. Kado, C.I. & Liu, S.T. Rapid procedure for detection and isolation of large and
934 small plasmids. *J Bacteriol* **145**, 1365-73 (1981).

57. Zankari, E. *et al.* Identification of acquired antimicrobial resistance genes. *J Antimicrob Chemother* **67**, 2640-4 (2012).

58. Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658-9 (2006).

59. Tracy, B.S., Edwards, K.K. & Eisenstark, A. Carbon and nitrogen substrate utilization by archival Salmonella typhimurium LT2 cells. *BMC Evol Biol* **2**, 14 (2002).

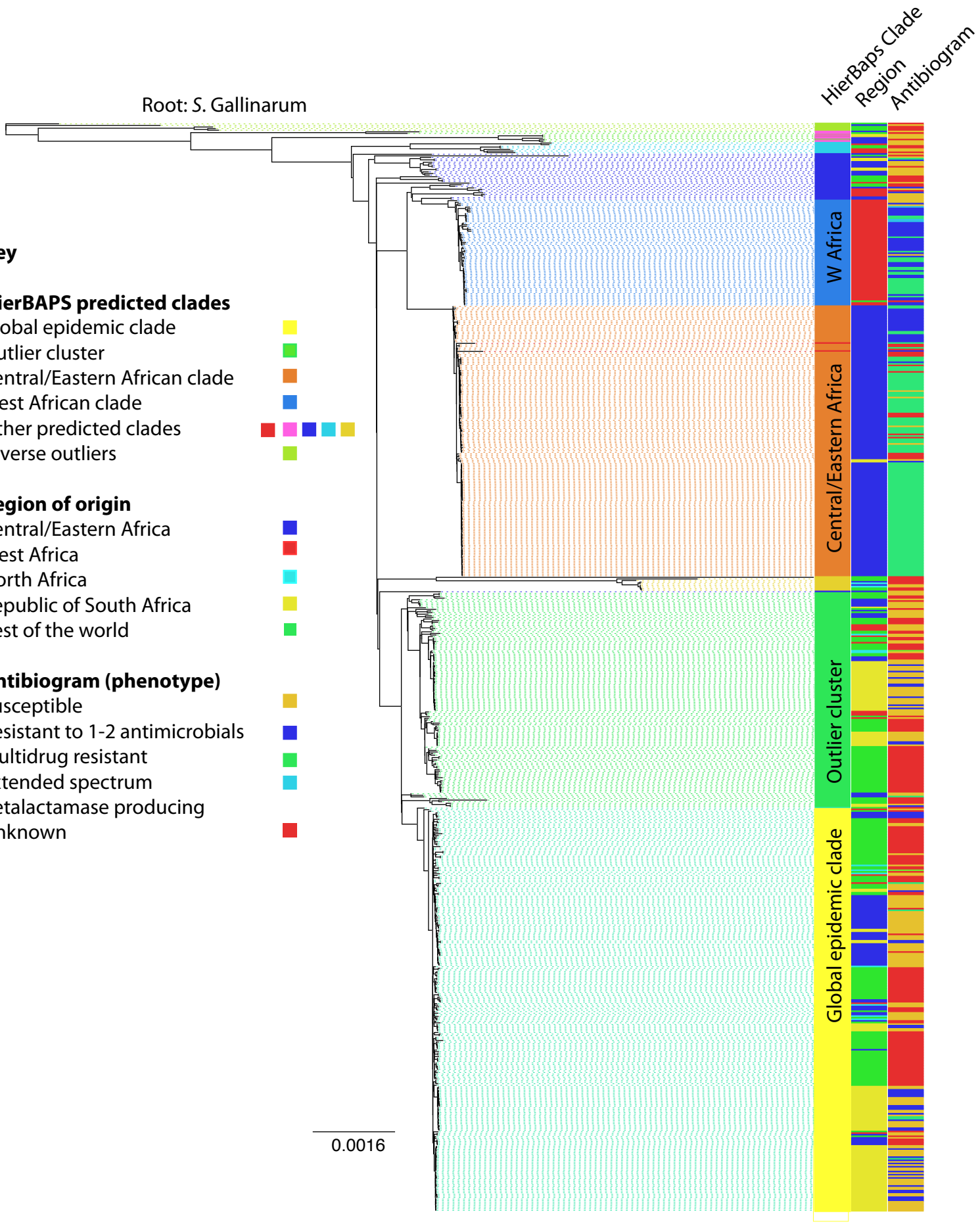
60. Croucher, N.J. *et al.* Dominant role of nucleotide substitution in the diversification of serotype 3 pneumococci over decades and during a single infection. *PLoS Genet* **9**, e1003868 (2013).

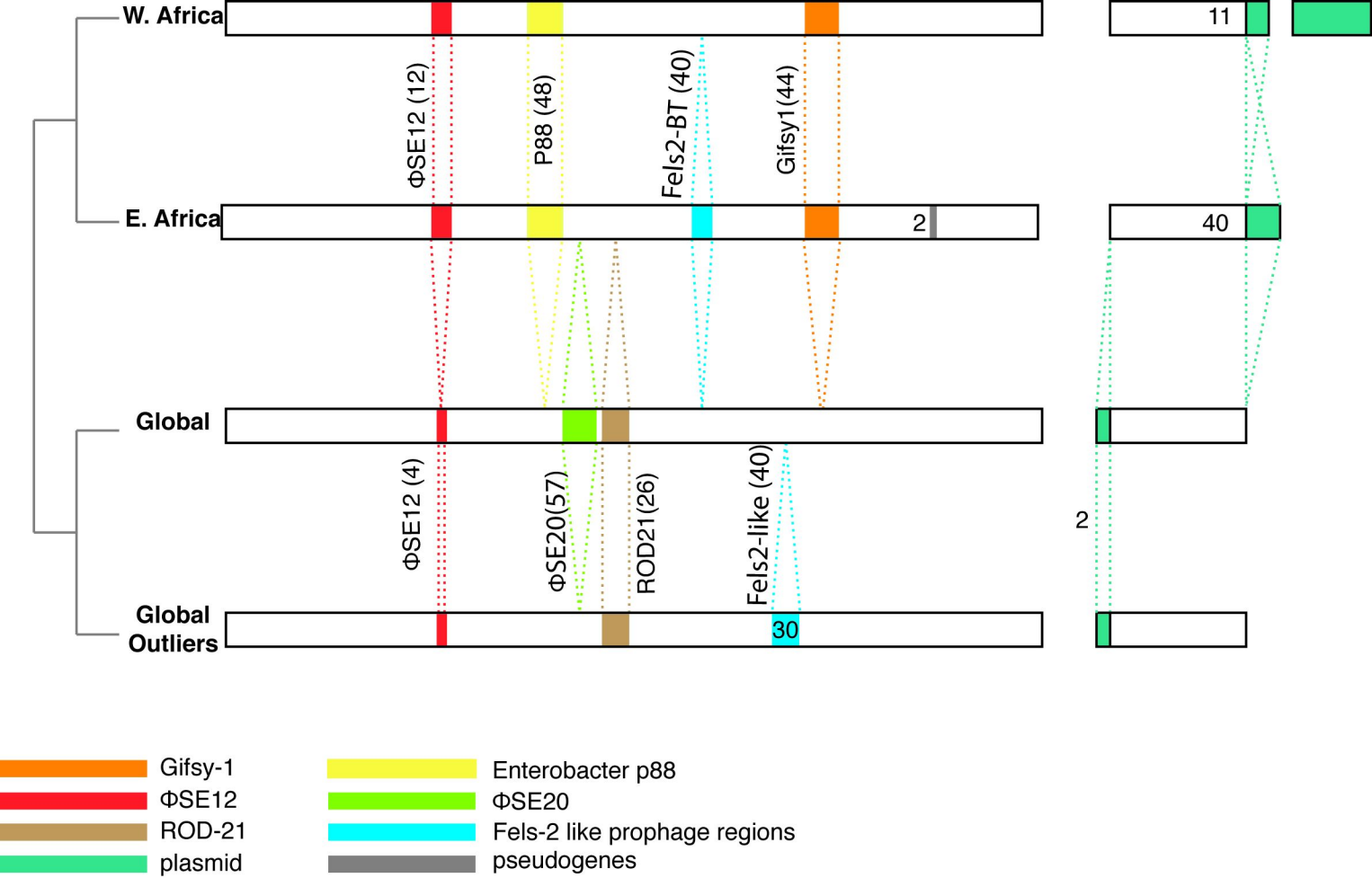
61. Karp, P.D. *et al.* Pathway Tools version 13.0: integrated software for pathway/genome informatics and systems biology. *Brief Bioinform* **11**, 40-79 (2010).

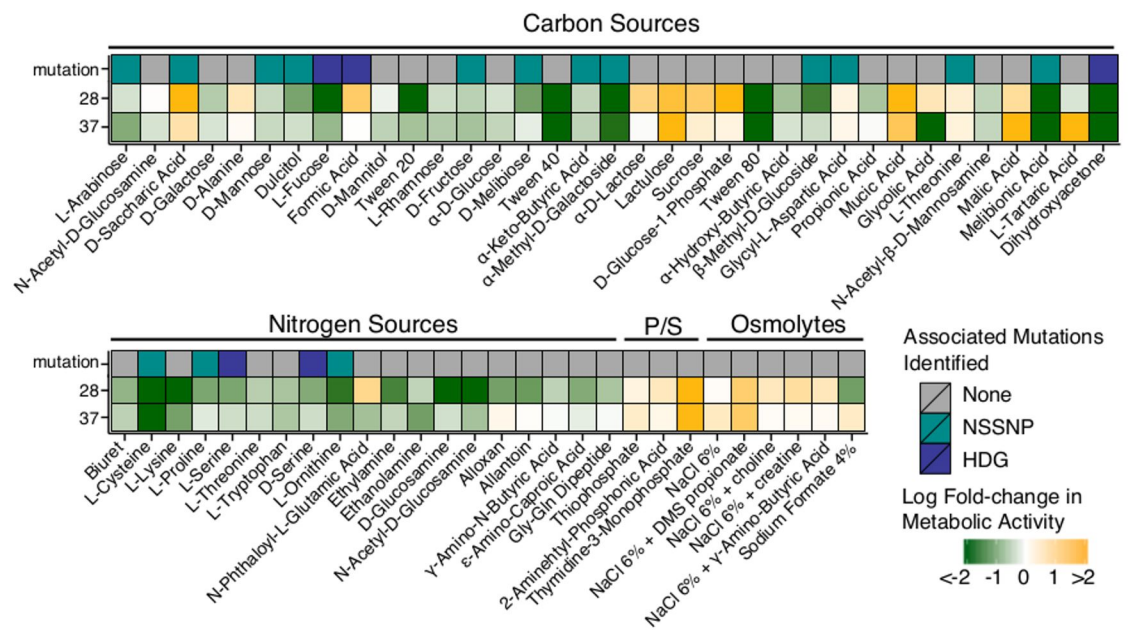
62. Salisbury, A.M., Bronowski, C. & Wigley, P. Salmonella Virchow isolates from human and avian origins in England--molecular characterization and infection of epithelial cells and poultry. *J Appl Microbiol* **111**, 1505-14 (2011).

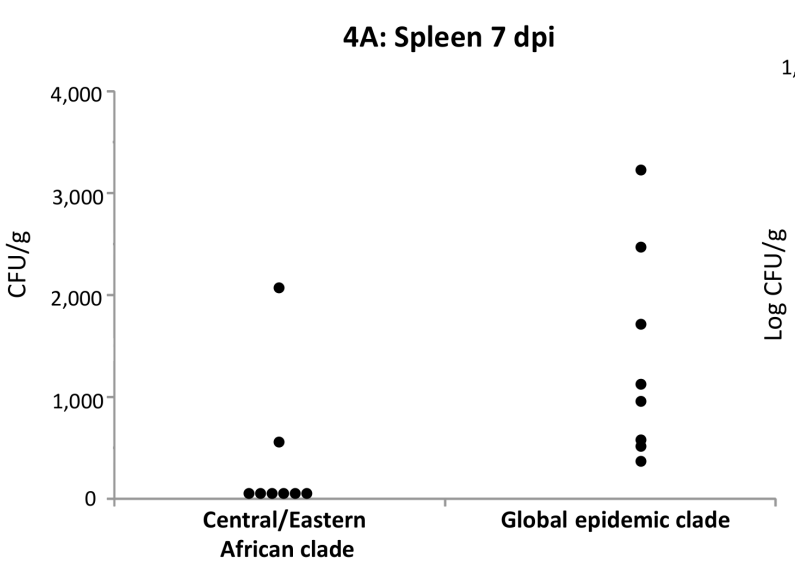
Competing financial interests:

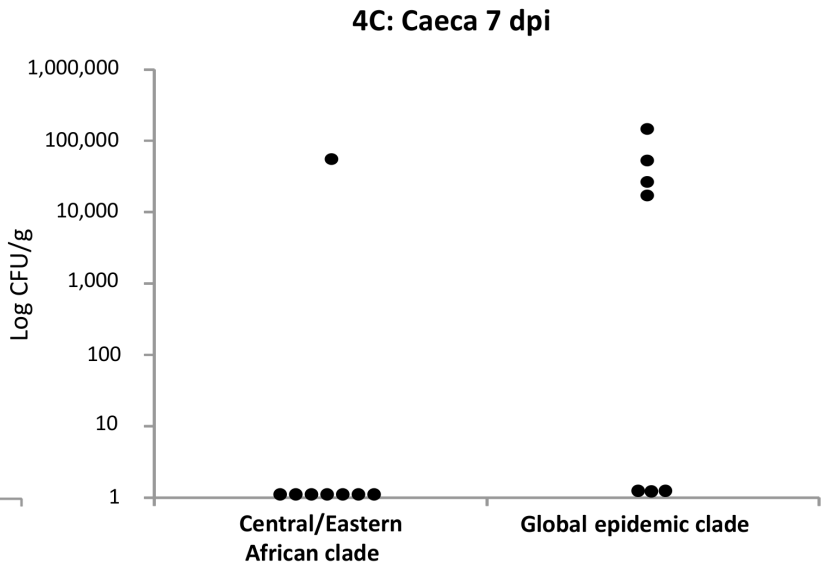
The authors declare no competing financial interests.

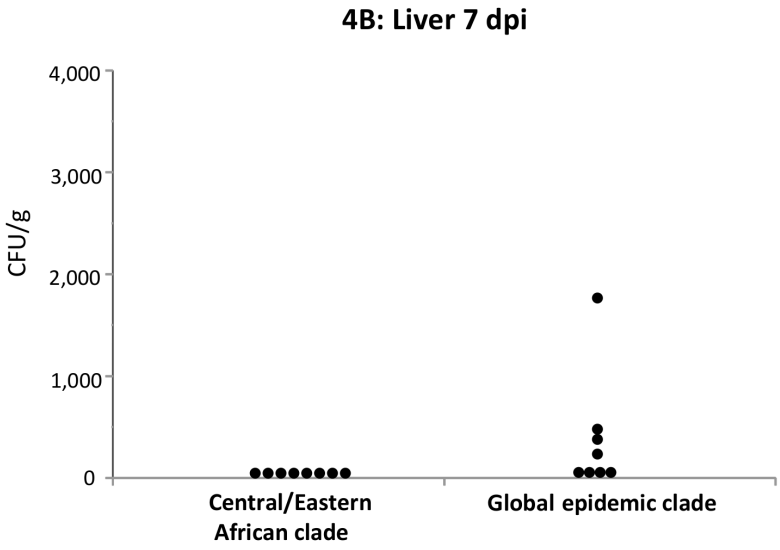

HierBaps Clade
Region
Antibiogram
Root: S. Gallinarum
Key
HierBAPS predicted clades
Global epidemic clade
Outlier cluster
Central/Eastern African clade
West African clade
Other predicted clades
Diverse outliers
Region of origin
Central/Eastern Africa
West Africa
North Africa
Republic of South Africa
Rest of the world
Antibiogram (phenotype)
Susceptible
Resistant to 1-2 antimicrobials
Multidrug resistant
Extended spectrum betalactamase producing
Unknown
W Africa
Central/Eastern Africa
Outlier cluster
Global epidemic clade
0.0016


W. Africa
E. Africa
Global
Global Outliers
ΦSE12 (12)
P88 (48)
Fels2-BT (40)
Gifsy1(44)
ΦSE12 (4)
ΦSE20(57)
ROD21(26)
Fels2-like (40)
11
2
40
2
30
Gifsy-1
ΦSE12
ROD-21
plasmid
Enterobacter p88
ΦSE20
Fels-2 like prophage regions
pseudogenes


Carbon Sources
mutation
28
37
L-Arabinose
N-Acetyl-D-Glucosamine
D-Saccharic Acid
D-Galactose
D-Alanine
D-Mannose
Dulcitol
L-Fucose
Formic Acid
D-Mannitol
Tween 20
L-Rhamnose
D-Fructose
α-D-Glucose
D-Melibiose
Tween 40
α-Keto-Butyric Acid
α-Methyl-D-Galactoside
α-D-Lactose
Lactulose
Sucrose
D-Glucose-1-Phosphate
Tween 80
α-Hydroxy-Butyric Acid
β-Methyl-D-Glucoside
Glycyl-L-Aspartic Acid
Propionic Acid
Mucic Acid
Glycolic Acid
L-Threonine
N-Acetyl-β-D-Mannosamine
Malic Acid
Melibionic Acid
L-Tartaric Acid
Dihydroxyacetone
Nitrogen Sources
P/S
Osmolytes
mutation
28
37
Biuret
L-Cysteine
L-Lysine
L-Proline
L-Serine
L-Threonine
L-Tryptophan
D-Serine
L-Ornithine
N-Phthaloyl-L-Glutamic Acid
Ethylamine
Ethanolamine
D-Glucosamine
N-Acetyl-D-Glucosamine
Alloxan
Allantoin
γ-Amino-N-Butyric Acid
ε-Amino-Caproic Acid
Gly-Gln Dipeptide
Thiophosphate
Phosphonic Acid
2-Aminethyl-Phosphonic Acid
Thymidine-3-Monophosphate
NaCl 6%
NaCl 6% + DMS propionate
NaCl 6% + choline
NaCl 6% + creatine
NaCl 6% + γ-Amino-Butyric Acid
Sodium Formate 4%
Associated Mutations Identified
None
NSSNP
HDG
Log Fold-change in Metabolic Activity
<-2 -1 0 1 >2


4A: Spleen 7 dpi
CFU/g
4,000
3,000
2,000
1,000
0
Central/Eastern African clade
Global epidemic clade


4C: Caeca 7 dpi
Log CFU/g
1,000,000
100,000
10,000
1,000
100
10
1
Central/Eastern African clade
Global epidemic clade


4B: Liver 7 dpi
CFU/g
4,000
3,000
2,000
1,000
0
Central/Eastern African clade
Global epidemic clade