Europe PMC Funders Author Manuscripts

# Emergence of host-adapted *Salmonella* Enteritidis through rapid evolution in an immunocompromised host

**Elizabeth J Klemm**[1], **Effrossyni Gkrania-Klotsas**[2,3], **James Hadfield**[1], **Jessica L Forbester**[1], **Simon R Harris**[1], **Christine Hale**[1], **Jennifer N Heath**[4], **Thomas Wileman**[1], **Simon Clare**[1], **Leanne Kane**[1], **David Goulding**[1], **Thomas D Otto**[1], **Sally Kay**[1], **Rainer Doffinger**[5], **Fiona J Cooke**[6], **Andrew Carmichael**[2], **Andrew ML Lever**[7], **Julian Parkhill**[1], **Calman A MacLennan**[1,4], **Dinakantha Kumararatne**[5], **Gordon Dougan**[#1,7], and **Robert A Kingsley**[#1,8]

[1]The Wellcome Trust Sanger Institute, The Wellcome Trust Genome Campus, Hinxton, Cambridge, United Kingdom

[2]Department of Infectious Diseases, Cambridge University Hospitals, Cambridge, United Kingdom

[3]Medical Research Council Epidemiology Unit, University of Cambridge School of Clinical Medicine, Cambridge, United Kingdom

[4]School of Immunity and Infection, College of Medicine and Dental Sciences, University of Birmingham, Birmingham, United Kingdom

[5]Department of Clinical Biochemistry and Immunology, Addenbrooke's Hospital, Cambridge, United Kingdom

[6]Clinical Microbiology and Public Health Laboratory, Addenbrooke's Hospital, Cambridge, United Kingdom

[7]Department of Medicine, University of Cambridge, Cambridge, United Kingdom

[8]The Institute of Food Research, Colney, Norwich, United Kingdom

[#] These authors contributed equally to this work.

## Summary

Host adaptation is a key factor contributing to the emergence of new bacterial, viral and parasitic pathogens. Many pathogens are considered promiscuous because they cause disease across a range of host species, while others are host-adapted, infecting particular hosts[1]. Host adaptation can potentially progress to host restriction where the pathogen is strictly limited to a single host species and is frequently associated with more severe symptoms. Host-adapted and host-restricted

Correspondence and requests for materials should be addressed to Robert A Kingsley rob.kingsley@ifr.ac.uk.

**Author contributions.** RAK and GD designed the study. EJK, DK, RD, AC, AMLL, JH, RAK, EGK, JH, JF, SC, LK, SK, DG, TW, and CH collected the data. EJK, RAK, TDO, JF, JP, SH, and CM analysed the data. EJK, RAK, GD, EGK, and JP wrote the manuscript.

Data deposition statement:
Sequencing data is available at www.ebi.ac.uk/ena/data/view/ERP001671

Competing interests declaration:
CAM is a former employee of the Novartis Vaccines Institute for Global Health and recipient of a clinical research fellowship from GlaxoSmithKline. All other authors declare no competing financial interests.

bacterial clades evolve from within a broader host-promiscuous species and sometimes target different niches within their specialist hosts, such as adapting from a mucosal to a systemic lifestyle. Genome degradation, marked by gene inactivation and deletion, is a key feature of host adaptation, although the triggers initiating genome degradation are not well understood. Here, we show that a chronic systemic non-typhoidal *Salmonella* infection in an immunocompromised human patient resulted in genome degradation targeting genes that are expendable for a systemic lifestyle. We present a genome-based investigation of a recurrent blood-borne *Salmonella enterica* serotype Enteritidis (*S.* Enteritidis) infection covering 15 years in an interleukin (IL)-12 β-1 receptor-deficient individual that developed into an asymptomatic chronic infection. The infecting *S.* Enteritidis harbored a mutation in the mismatch repair gene *mutS* that accelerated the genomic mutation rate. Phylogenetic analysis and phenotyping of multiple patient isolates provides evidence for a remarkable level of within-host evolution that parallels genome changes present in successful host-restricted bacterial pathogens but never before observed on this timescale. Our analysis identifies common pathways of host adaptation and demonstrates the role that immunocompromised individuals can play in this process.

Host adaptation is central to pathogen evolution and is a founding element of the emergence of many classical and emerging diseases such as typhoid, whooping cough and bubonic plague. For example, typhoid fever is caused by the human-restricted genetic clades *Salmonella* Typhi and *S.* Paratyphi A, which evolved independently within the broadly promiscuous *Salmonella enterica* species. Genome analysis of host-restricted pathogens linked host adaptation with both gene acquisition and genome degradation[2,3]. Loss of coding capacity by pseudogene formation or deletion is thought to proceed through a combination of neutral or mildly deleterious mutations becoming fixed in the population during evolutionary bottlenecks and positive selection for the loss of function of genes no longer required or disadvantageous in the new host or environmental niche[3-5]. *S. enterica* is a diverse species that harbors both host-promiscuous and host-adapted clades. *S. enterica* serovars such as *S.* Typhimurium and *S.* Enteritidis are predominantly associated with gastroenteritis, a disease normally self-limiting to the intestine. However, bacteremia can occur even with non-typhoidal serovars and this is frequently associated with malnutrition, human immunodeficiency virus infection, or primary immune deficiencies such as mutations in interleukin (IL)-12/-23 and their receptors [6,7,8].

In 1995, a 12 year-old patient presented with a clinically severe systemic *S.* Enteritidis infection and was diagnosed with IL-12β1 receptor deficiency (homozygous mutation of the IL12RB1 gene)[9]. Between the ages of 16 and 29 (1999-2014), the patient had multiple recurrent systemic *S.* Enteritidis infections, but the patient reported less severe symptoms with each subsequent episode (see Extended Discussion). Treatment included Interferon-γ (IFNγ) and various antibiotics, including Ciprofloxacin, Azithromycin and Ceftriaxone. Minimum inhibitory concentrations for the antibiotics used in therapy was monitored (Supplementary Information Table I).

Analysis of whole genome sequence of eleven *S.* Enteritidis isolates from the patient's blood from successive fever episodes revealed that each was due to relapse from the same infection rather than reinfection. A maximum-likelihood tree of the isolates from the

immunocompromised patient and seventeen isolates from other patients with standard gastrointestinal *S.* Enteritidis infections was constructed based on 5,087 SNPs with reference to *S.* Enteritidis phage type 4 (PT4) P125109, excluding SNPs within prophage elements, repetitive sequence and insertion elements (Figure 1A). The immunocompromised patient isolates were in a single clade that shared a most-recent common ancestor (MRCA) within the *S.* Enteritidis PT4 clade. However, the diversity of the immunocompromised patient isolates was much greater than for the main PT4 clade isolates. Isolates within the *S.* Enteritidis PT4 clade differed from one another by ~80 SNPs while the patient isolates differed by as many as 600 SNPs, and the most recent patient isolates contained nearly 1,000 SNPs relative to their MRCA. Accumulation of SNPs along lineages within the patient isolate clade was linear and time-dependent, with a mean of 47 SNPs per genome per year (Figure 1B). A time-dependent phylogenetic reconstruction using Bayesian inference predicted a time of MRCA (tMRCA) around 1995 (+/−0.5 years) (Figure 1C). The molecular clock rate was calculated as $1\times10^{-5}$ SNPs per site per year for the patient lineages, a 25 to 50-fold higher substitution rate than previously reported for *Salmonella*[10,11]. Together these data are consistent with the hypothesis that the patient was infected with a common ancestor that existed around the time of the first detected bacteremia and that since this time the infecting clone had rapidly acquired mutations.

Within the immunocompromised patient clade, there were elevated G➡A and C➡T nucleotide substitutions (Supplementary Information Figure 1) relative to *S.* Enteritidis lineages from other patients and there was a region of frequent inversion (Supplementary Information Figure 2), both of which were consistent with a defective mis-match repair mechanism[12]. Indeed, the *mutS* gene in the patient isolates contained a 174 bp in-frame deletion encoding a truncated protein MutS$^{\Delta253\text{-}311}$. Mutations in *mutS* have previously been implicated in a 100-fold increase of transition rate, a 1,000-fold increased rate of frameshift mutations, frequent chromosomal rearrangements, and single nucleotide indels in homopolymeric tracts[13,14]. To determine the contribution of the *mutS* mutation to the elevated substitution rate of the patient isolates, an identical in-frame deletion was introduced into *S.* Enteritidis P125109 (SW824 *mutS*$^{\Delta253\text{-}311}$). Serial passage of SW824 *mutS*$^{\Delta253\text{-}311}$ in Luria broth resulted in consecutive clones with similar phylogenetic diversity and comparable rate of SNP accumulation (0.19 SNPs per generation) as patient isolate B20062 (0.16 SNPs per generation), nearly 200-fold greater than for *S.* Enteritidis P125109 (0.001 SNPs per generation) (Supplementary Information Figure 3).

Accumulation of SNPs along the branches within the patient clade exhibited a greater dN/dS ratio compared to branches outside this clade (Supplementary Information Figure 4). The significance of this observation is equivocal and may indicate positive selection of a subset of genes in the genome or may indicate there has been insufficient time for purifying selection to reduce the dN/dS ratio. We could not differentiate between these possibilities because the number of mutations per gene is low and purifying selection requires long time periods to act[15]. We also investigated potential signatures of selection using a profile-based method to predict whether polymorphisms in proteins of the patient isolates were likely to have functionally diverged from P125109. Comparison of orthologous proteins from *S.* Enteritidis P125109 with patient isolates or with diverse non-patient *S.* Enteritidis isolates indicated that the patient isolates accumulated more deleterious polymorphisms in conserved

protein domains in the pfam database, giving rise to decreased bitscores relative to P125109 (Supplementary Information Figure 5 A and B). The resulting increase in mean delta bitscore and skew from a normal distribution (Supplementary Information Figure 5 C and D) has been reported in genomes of host-adapted pathogens undergoing genome degradation[16].

To enable us to identify the entire repertoire of hypothetically disrupted genes with a predicted null phenotype (pseudogenes), we constructed single contiguous whole genome sequence assemblies of four patient isolates using the PacBio sequencing platform. Each isolate had acquired, on average, 158 pseudogenes (Figure 2A) since the MRCA with PT4. Combined with the 75 hypothetically disrupted genes and 68 deleted genes already present in *S*. Enteritidis P125109[17], this is degradation of ~300 genes or ~6.9% of the total protein coding capacity. This level of genome degradation is comparable to that reported for host-adapted *Salmonella* serovars (Figure 2C).

While most of the pseudogenes could be readily explained by a single event and were maintained as pseudogenes through the successive generations according to the phylogenetic tree, we noticed that some of the pseudogenes showed evidence of reversion, with the functional gene restored in a terminal branch. This occurred in pseudogenes that arose due to frameshifts caused by changes in the length of homopolymeric tracts, a common occurrence in *mutS*-deficient bacteria[14]. Restoration of the functional gene occurred by one of two mechanisms: deletion of inserted nucleotides or insertion of additional nucleotides that re-established the original reading frame (Supplementary Information Figure 6). Transient genome degradation akin to this has also been proposed for the human-restricted pathogen *S*. Paratyphi A[4] suggesting that this could be a common characteristic of bacteria undergoing adaptation.

Pseudogenes in the patient isolates overlapped with those found in *S*. Typhi and *S*. Paratyphi A, with 44 genes disrupted in *S*. Typhi, *S*. Paratyphi A and at least one of the four patient isolates analyzed (Figure 2B). Importantly, the patient isolates have acquired pseudogenes in a large percentage of central anaerobic metabolism, colonization factor and secreted effector genes, three functional classes that promote growth and cellular invasion in the gut and are linked to pathogenesis (Figure 2D-F). Specifically, the nitrate, tetrathionate, ethanolamine and cobalamin vitamin B12 metabolic pathways were degraded in the patient isolates, along with fimbriae and non-fimbrial adhesins, and type three secretion system (T3SS) apparatus and effector proteins (See Supplementary Discussion and Supplementary Information Tables II and III). Gene Ontology enrichment analysis of the pseudogenes also showed over-representation of terms concerned with certain metabolic processes, including cobalamin vitamin B12, membrane proteins and pilus organization genes (Supplementary Information Table IV). A number of these GO terms are also over-represented in disrupted genes of *S*. Typhi and *S*. Paratyphi A. Thus, the pattern of genome degradation of these host-adapted *Salmonella* exhibited considerable overlap with that observed in the patient isolates and represents convergent evolution.

In many cases the known function of putatively inactivated genes were consistent with the phenotype and virulence of the patient isolates. The patient isolates harbored SNPs in the quinolone resistance-determining region that were consistent with the elevated minimum

inhibitory concentration data for ciprofloxacin in the clinical microbiology data (Supplementary Information Table I). Several genes involved in lipopolysaccharide (LPS) biosynthesis, e.g. *rfaL*, O-antigen ligase, and *mpl*, murein peptide ligase, were pseudogenes in one or more of the patient isolates and may have contributed to the profound changes to cell morphology and sensitivity to serum killing (Supplementary Discussion, Supplementary Information Table II and Supplementary Information Figures 7 and 8). The patient isolates showed aberrant host-pathogen interactions *in vitro* and *in vivo* that may be related to observed mutations in adhesins and T3SS proteins. Patient isolates B20062, MB760 and MB4386 were not able to invade the cultured epithelial-like cell line HEp-2 as efficiently as P125109 (Figure 3A). The reduction in invasion of patient isolates was even greater than for a defined *invA* mutation in *S*. Typhimurium, a key structural component of the SPI-1 T3SS. The cytokine response generated by *Salmonella* infected macrophages elicits a variable and atypical host cell response to infection with individual patient isolates compared to P125109 (Supplementary Information Figure 9). Following oral inoculation of mice, isolate MB4386 was profoundly attenuated compared to P125109. Colonization of the liver and spleen with MB4386 were below the limit of detection in most mice (Figure 3B). Following intravenous inoculation, isolate MB4386 was likewise attenuated with 1,000 to 10,000 fold lower liver and spleen colonization compared to P125109 (Figure 3C) and a higher survival rate. Mice intravenously infected with P125109 were moribund within five days, while all mice infected with MB4386 survived beyond day seven (Figure 3D). Intriguingly, isolate MB4386 was capable of colonizing and replicating in immunocompromised mice following intravenous inoculation. Indeed, both IL12p40-deficient and IFNγR-deficient mice showed higher levels of colonization with isolate MB4386 following intravenous inoculation (Supplementary Information Figure 10).

This example of extensive in-host evolution in a single patient recapitulated that observed during the emergence of pathogens of antiquity and the present day. Although most healthy individuals rapidly clear infections and non-typhoidal *Salmonella* infections are limited to the gut, the extended systemic infection in this patient provided a unique niche for host adaptation. This remarkable clinical case provides insight into the mechanisms driving microbial host adaptation and highlights the potential role of immunocompromised individuals in host adaptation. Likewise, exposing a microbe with a gastro-intestinal lifestyle to an extra-intestinal niche resulted in genome degradation of specific classes of genes, generating the signature of host adaptation we have defined here. Genes required for persistence in the gut appear to be no longer essential to bacteria residing in an extra-intestinal niche and are likely subject to degradation by the process of genome streamlining. Such a process may be occurring in *S*. Typhimurium, ST313, a genotype associated with invasive non-typhoidal *Salmonella* (iNTS) disease in immunocompromised populations in sub-Saharan Africa that has acquired signatures of host adaption in their genome sequence[10,18]. ST313 and other host-adapted pathogens have retained the ability to transmit to other individuals, a feature that has not necessarily been selected for in this patient. Additionally, the rate of evolution is elevated in this case as the rapid genome degradation in the *S*. Enteritidis acquired by the patient was driven by a *mutS* mutation. Reversible mutations to mismatch repair proteins have been previously implicated in *Salmonella* genome diversification and bacterial adaptation[19].

This study captured a rare opportunity to investigate host adaptation in real time during a single infection. Through an ill-fated 'perfect storm' combining a permissive human host with a hyper-mutating *Salmonella*, we were able to observe genome degradation in action as genes and pathways promoting the gastrointestinal lifestyle were lost in an extra-intestinal infection. Crucially, we were able to link the genotypic changes to altered bacterial virulence that drew parallels with the evolution of other host-adapted pathogens including *S.* Typhi and *S.* Paratyphi A. Thus, our analysis contributes to our understanding of conserved evolutionary mechanisms underlying the emergence of host-adapted pathogens, and highlights the remarkable adaptability of pathogens when they enter a new niche.

## Materials and Methods

### Bacterial strains and culture conditions

*S.* Enteritidis P125109 (PT4), *S.* Enteritidis P537361 (Patient isolate, collected 18.04.2001), *S.* Enteritidis P542816 (Patient isolate, collected 19.7.2001), *S.* Enteritidis P542817 (Patient isolate, collected 19.7.2001), *S.* Enteritidis P573395 (Patient isolate, collected 6.5.2003), *S.* Enteritidis B20062 (Patient isolate, collected 10.11.2006), *S.* Enteritidis MB20421 (Patient isolate, collected 17.12.2009), MB760 (Patient isolate, collected 13.1.2010), *S.* Enteritidis MB4386 (Patient isolate, collected 01.10.2010), *S.* Enteritidis MB18868 (Patient isolate, collected 11.11.2011), *S.* Enteritidis B11430 (Patient isolate, collected 14.07.2011). A *S.* Enteritidis strain P125109 in which an in-frame deletion of amino acids 253-311 in the *mutS* gene ($mutS^{\Delta 253\text{-}311}$) was engineered (SW824) using a suicide vector-based approach. A region of the MB4386 genome containing the in-frame deletion of *mutS* was amplified by PCR using primers 5′ ggaAGATCTggataccgggtttatcgg 3′ and 5′ ccgCTCGAGcagtatctcaagctgaaagccc 3′. Restriction enzyme sites *Bgl*I and *Xho*I incorporated into the oligonucleotide primers were used to clone the DNA fragment into the suicide vector pJCB12. The suicide vector construct was introduced into *S.* Enteritidis strain P125109 by electroporation and merodiploid derivatives in which the plasmid had integrated into the bacterial chromosome were selected for by culture on LB+Cm agar medium. The merodiploid was resolved by counter selection of the levansucrase activity encoded by the *sacRB* gene on pJCB12, by culture on LB agar containing 5% sucrose. A resolved merodiploid clone containing the deletion was identified by PCR amplification using primers that annealed flanking the deletion.

For in vitro passage of *S.* Enteritidis Δ*mutS* and patient isolate B20062, bacteria were serially passaged in 50ml LB broth incubated for 12 hours at 37°C with shaking. Each passage was inoculated with $1\times10^3$ CFU from the previous culture. Cultures from passages 1, 4, 7, 10 and 13 were plated on LB agar plates and single colonies selected for inoculation of a 50ml culture of LB broth and incubated for 12 hours at 37°C with shaking. Genomic DNA was prepared from this culture and sequenced using the HiSeq Illumina platform.

### Genome sequencing and sequence analysis

Multiplex libraries with a 300bp insert size were prepared using 96 unique index-tags, and sequenced to generate 75 base-pair (bp) paired-end reads. Cluster formation, primer hybridisation and sequencing reactions were based on reversible terminator chemistry using

the Illumina HiSeq System. Sequence data were submitted to the European Nucleotide Archive (http://www.ebi.ac.uk/ena); accession numbers are indicated in Supplementary Information Table V. For sequence-read alignment and SNP detection, paired-end Illumina sequence data from each isolate were mapped to the reference genome *S*. Enteritidis strain P125109[20] using SMALT (ftp://ftp.sanger.ac.uk/pub4/resources/software/smalt/smalt-manual-0.7.4.pdf). Single nucleotide polymorphisms (SNPs) were identified using samtools mpileup and filtered with a minimum mapping quality of 30 and quality ratio cut-off of 0.75. SNPs identified in prophage elements and repetitive sequence regions of the *S*. Enteritidis reference were excluded from subsequent phylogenetic analysis. Repetitive regions were defined as exact repetitive sequences of ≥ 20 bp, identified using repeat finding programs nucmer[21], REPuter[22] and repeat-match[23,24] A concatenated alignment composed of SNP sites from each sequenced isolate was generated. For phylogenetic analyses, a maximum-likelihood phylogenetic tree was constructed from the SNP alignment with RAxML v7.0.4[25] 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. The raw Illumina HiSeq data were used to generate a *de novo* draft assembly of the genome of each sample using the VELVET v0.7.03 algorithm[26], resulting in multi-contig draft genomes.

Bayesian ancestral state reconstruction of ten patient isolates from a ten-year period was performed using the BEAST v1.7.5 package. The data for these models were the alignment of variable sites (SNPs) in the genome sequence of the ten patient isolates. A discrete gamma distribution was employed to model rate variation among sites and a strict clock with constant size coalescent tree prior was selected as it required the least parameters. The isolation date in years was used to calibrate the time scale of the tree. Three independent Markov Chain Monte Carlo (MCMC) analyses were run, each with a 10 million burn-in and 90 million chain length, sampled every 10,000 states using BEAGLE in conjunction with BEAST. The maximum clade credibility tree from the resulting 3600 trees was summarized using TreeAnnotator and visualized with Figtree.

Bitscores were determined for proteins harboring non-synonymous SNPs using hmmpfam. Delta bitscores were calculated by subtracting the bitscore for a given HMM domain from the bitscore of the orthologous domain in *S*. Enteritidis P125109. Delta bitscores of zero and those greater than five standard deviations from the mean were removed. Remaining delta bitscores were plotted in a frequency distribution with a window size of 0.5 and skewness and mean values were calculated using excel.

For determination of whole genome sequence for *de novo* assembly using SMRT PacBio®, Template Prep Kit (PacBio, Menlo Park, CA, USA) and BluePippin™ Size Selection System protocol were employed to prepare size-selected libraries (20kb) from 5 μg of sheared and concentrated DNA. Sequencing was performed using the magnetic bead collection protocol, a 20,000 bp insert size, stage start, and 180-minute movies. A *de novo* assembly was generated from these reads using the Hierarchical Genome Assembly Process (HGAP) software version 3.0, with the genome size parameter set to 5 Mb.

Pseudogenes were identified by *in silico* determination of annotated features of the draft genome with a decreased length or frameshifts in predicted protein sequence. The annotation was transferred with RATT[27] from the reference onto the new assembly (Parameter PacBio). An ad hoc PERL script compared genes with the transferred annotation to the reference genome and reported frameshifts and changes in length. Pseudogenes were manually confirmed in the aligned genome sequence using the Artemis Comparison Tool (ACT)[28]. Genes with deletions of over 10% of the coding sequence were classified as pseudogenes. Indels in homopolymeric tracts were verified with illumina sequencing data. Total number of pseudogenes in the patient isolates were calculated by adding the number of pseudogenes reported in Nuccio et al.[17] with the newly-identified pseudogenes.

We performed GO enrichment with the R package TopGO (Alexa A and Rahnenfuhrer J (2010). *topGO: topGO: Enrichment analysis for Gene Ontology*. R package version 2.20.0.), using default settings. As databases for the enrichment we downloaded the GO terms for *Salmonella* Enteritidis P125109 from the GOA database[29].

## Code Availability

SMALT v0.7.4 ftp://ftp.sanger.ac.uk/pub4/resources/software/smalt/smalt-manual-0.7.4.pdf

nucmer[21]

REPuter[22]

repeat-match[23,24]

RAxML v7.0.4[25]

VELVET v0.7.03[26]

BEAST v1.7.5

BEAGLE

HGAP v3.0

RATT[27]

TopGO (Alexa A and Rahnenfuhrer J (2010). *topGO: topGO: Enrichment analysis for Gene Ontology*. R package version 2.20.0.)

## Tissue culture, serum sensitivity, cytokine profiling and in vivo infection models

Hep-2 (ECACC #86030501) cells were maintained in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% foetal bovine serum, 2mM L-Glutamine and Non-Essential Amino Acids and tested for mycoplasma. Cells were not further authenticated. The Hep-2 cell line has been reported to contain HeLa cells, another epithelial cell line commonly used in *Salmonella* invasion assays. Hep-2 cells were seeded $1 \times 10^5$ in 24 well plates and incubated at 37°C for 24 hours. Bacterial cells were added at an MOI of 10 to the Hep-2 cells and centrifuged (1500 rpm, 3 mins, 20°C), before incubation at 37°C for 1 hour.

Cells were washed with Dulbecco's phosphate-buffered saline (Sigma-Aldrich) and then incubated for 90 minutes with DMEM/DMEM-F12 containing 100μg/ml gentamicin. Cells were then washed with PBS and lysed with 1% Triton-X100. Cells were serially diluted and plated onto LB agar to determine CFU counts. In each experiment, 3 biological replicates were carried out per isolate and the whole experiment was repeated twice. Percentage invasion was calculated by comparing inoculum CFU/ml to recovered CFU after cell lysing.

Background-matched wild-type C57BL/6, $Il12^{btm1Jm}$ , and $Ifngr1^{tm1Agt}$ (purchased from The Jackson Laboratory) mice 6–14 weeks of age were bred and maintained in accordance with UK Home Office regulations under the project license 80/2596. This license was reviewed by The Wellcome Trust Sanger Institute Animal Welfare and Ethical Review Body (AWERB). We employ 2-tailed analysis of variance to test for statistically significant differences between means in treated mice, and to determine 95% confidence intervals. We use groups of four to five mice as this is a sufficient sample size given the differences between means and within-group variances we typically observe with these infection models. Mice were inoculated orally by gavage or intravenously with stated amounts of *S.* Enteritidis suspended in PBS pH7.4. Although we could not randomize and blind all aspect of the animal experiments, technicians inoculating and checking mice were blinded to the identity of bacterial strain and those determining the bacterial numbers by plating were blind to the animal strain. When mice were moribund (less than 80% body weight compared with day of inoculation) or on day post inoculation indicated in the text, mice were culled and CFUs of each strain in homogenised spleen, liver, and heparinised blood was determined by serial dilution in PBS pH7.4 and plating on LB agar.

The human monocyte-like cell line THP-1 was obtained from ECACC (#88081201) and routinely cultured in RPM1 1640 supplemented with 2mM L-Glutamine and 10% heat-inactivated Foetal calf serum (FCS) and tested for mycoplasma. Cells were not further authenticated. Cells were differentiated into mature Mφ-like cells by stimulating with 100ng/ml Phorbol 12-Myristate 13-acetate (PMA) for 3 days and replaced with medium without PMA for 1 day prior to assay. Bacteria (strains as indicated) were added to the media at indicated MOI=10 and incubated at 37°C for 45 minutes. After incubation, media was replaced with media containing 100ng/ml Gentamicin. At two hours, media was changed to media containing 10ng/ml Gentamicin. After 2 and 4 hours post infection, supernatants were harvested and filtered. Supernatants were analysed for cytokine/chemokine concentrations using the Millipore customised anti-human Milliplex magnetic bead kits as per manufacturer's instruction. Data was acquired on a Luminex FlexMap3D and analysed with either Luminex or Masterplex QT software. In each experiment, 6 biological replicates were carried out per isolate and the whole experiment was repeated twice.

### Electron microscopy, complement-mediated killing assays, and LPS profiling

For Electron Microscopy, individual colonies were picked directly from an agar plate with a plastic loop and fixed in 2.5% glutaraldehyde and 2% paraformaldehyde in 0.1M sodium cacodylate buffer for 1 hour at room temperature. The bacteria were then pelleted and briefly rinsed in buffer before fixing in 1% osmium tetroxide, rinsing again and dehydrating

in an ethanol series. Pellets were then embedded using an Epoxy Embedding Media Kit (Sigma) and 60μm ultrathin sections cut on an EM6 ultramicrotome (Leica). Finally sections were stained with uranyl acetate and lead citrate and imaged on a 120kV FEI Spirit Biotwin transmission electron microscope with a Tietz F4.15 CCD camera. For serum bactericidal assays, isolates in log growth were washed twice with PBS and added at a final concentration of $10^6$ cfu/ml to 100% normal human serum (containing antibodies to *S.* Enteritidis) or 1 to 100% (1, 2, 5, 10, 20, 30, 50, 75 and 100%) baby rabbit serum in PBS. Reaction tubes were incubated at 37°C with rocking at 20 rpm. Viable bacterial counts were determined by plating on LB agar after 45, 90 and 180 minutes. Lipopolysaccharide isolated from stationary phase cultures were separated by SDS-PAGE and visualized by silver staining [30].

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

We would like to thank David Harris, Gemma Langridge and the pathogens informatics team for help with sequencing and bioinformatics and the Sanger Institute Research Support Facility for help with the animal studies. This work was funded by the Wellcome Trust through core funding for the Sanger Institute Pathogen Variation Group.

## References

1. Baumler A, Fang FC. Host specificity of bacterial pathogens. Cold Spring Harbor perspectives in medicine. 2013; 3:a010041.10.1101/cshperspect.a010041 [PubMed: 24296346]
2. Parkhill J, et al. Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhi CT18. Nature. 2001; 413:848–852. [PubMed: 11677608]
3. McClelland M, et al. Comparison of genome degradation in Paratyphi A and Typhi, human-restricted serovars of Salmonella enterica that cause typhoid. Nature genetics. 2004; 36:1268–1274. [PubMed: 15531882]
4. Zhou Z, et al. Transient Darwinian selection in Salmonella enterica serovar Paratyphi A during 450 years of global spread of enteric fever. Proc Natl Acad Sci U S A. 2014; 111:12199–12204.10.1073/pnas.1411012111 [PubMed: 25092320]
5. 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]
6. Celum CL, Chaisson RE, Rutherford GW, Barnhart JL, Echenberg DF. Incidence of salmonellosis in patients with AIDS. The Journal of infectious diseases. 1987; 156:998–1002. [PubMed: 3680999]
7. de Beaucoudrey L, et al. Revisiting human IL-12Rbeta1 deficiency: a survey of 141 patients from 30 countries. Medicine. 2010; 89:381–402.10.1097/MD.0b013e3181fdd832 [PubMed: 21057261]
8. MacLennan C, et al. Interleukin (IL)-12 and IL-23 are key cytokines for immunity against Salmonella in humans. The Journal of infectious diseases. 2004; 190:1755–1757.10.1086/425021 [PubMed: 15499529]
9. Altare F, et al. Impairment of mycobacterial immunity in human interleukin-12 receptor deficiency. Science. 1998; 280:1432–1435. [PubMed: 9603732]
10. Okoro CK, et al. Intracontinental spread of human invasive Salmonella Typhimurium pathovariants in sub-Saharan Africa. Nature genetics. 2012; 44:1215–1221.10.1038/ng.2423 [PubMed: 23023330]
11. Mather AE, et al. Distinguishable Epidemics of Multidrug-Resistant Salmonella Typhimurium DT104 in Different Hosts. Science. 2013; 341:1514–1517.10.1126/Science.1240578 [PubMed: 24030491]

12. Schaaper RM. Base selection, proofreading, and mismatch repair during DNA replication in Escherichia coli. J Biol Chem. 1993; 268:23762–23765. [PubMed: 8226906]
13. Denamur E, Matic I. Evolution of mutation rates in bacteria. Mol Microbiol. 2006; 60:820–827.10.1111/j.1365-2958.2006.05150.x [PubMed: 16677295]
14. Funchain P, et al. The consequences of growth of a mutator strain of Escherichia coli as measured by loss of function among multiple gene targets and loss of fitness. Genetics. 2000; 154:959–970. [PubMed: 10757746]
15. Rocha EPC, et al. Comparisons of dN/dS are time dependent for closely related bacterial genomes. Journal of Theoretical Biology. 2006; 239:226–235. [PubMed: 16239014]
16. Kingsley RA, et al. Genome and transcriptome adaptation accompanying emergence of the definitive type 2 host-restricted Salmonella enterica serovar Typhimurium pathovar. mBio. 2013; 4:e00565–00513.10.1128/mBio.00565-13 [PubMed: 23982073]
17. Nuccio SP, Baumler AJ. Comparative analysis of Salmonella genomes identifies a metabolic network for escalating growth in the inflamed gut. mBio. 2014; 5:e00929–00914.10.1128/mBio.00929-14 [PubMed: 24643865]
18. 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. doi:gr.091017.109 [pii]. 10.1101/gr.091017.109 [PubMed: 19901036]
19. Gong J, et al. Spontaneous conversion between mutL and 6 bpDeltamutL in Salmonella typhimurium LT7: association with genome diversification and possible roles in bacterial adaptation. Genomics. 2007; 90:542–549.10.1016/j.ygeno.2007.06.009 [PubMed: 17689047]
20. Thomson NR, et al. Comparative genome analysis of Salmonella Enteritidis PT4 and Salmonella Gallinarum 287/91 provides insights into evolutionary and host adaptation pathways. Genome Res. 2008; 18:1624–1637. [PubMed: 18583645]
21. Kurtz S, et al. Versatile and open software for comparing large genomes. Genome Biol. 2004; 5:R12. [PubMed: 14759262]
22. Kurtz S, et al. REPuter: the manifold applications of repeat analysis on a genomic scale. Nucleic Acids Res. 2001; 29:4633–4642. [PubMed: 11713313]
23. Holt KE, et al. High-throughput sequencing provides insights into genome variation and evolution in Salmonella Typhi. Nature genetics. 2008; 40:987–993. [PubMed: 18660809]
24. He M, et al. Evolutionary dynamics of Clostridium difficile over short and long time scales. Proc Natl Acad Sci U S A. 2010; 107:7527–7532.10.1073/pnas.0914322107 [PubMed: 20368420]
25. Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics. 2006; 22:2688–2690. [PubMed: 16928733]
26. Zerbino D, Birney E. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. Genome Research. 2008; 18:821–829. [PubMed: 18349386]
27. Otto TD, Dillon GP, Degrave WS, Berriman M. RATT: Rapid Annotation Transfer Tool. Nucleic Acids Res. 2011; 39:e57.10.1093/nar/gkq1268 [PubMed: 21306991]
28. Carver TJ, et al. ACT: the Artemis Comparison Tool. Bioinformatics. 2005; 21:3422–3423. [PubMed: 15976072]
29. Huntley RP, et al. The GOA database: gene Ontology annotation updates for 2015. Nucleic Acids Res. 2015; 43:D1057–1063.10.1093/nar/gku1113 [PubMed: 25378336]
30. Tsai CM, Frasch CE. A sensitive silver stain for detecting lipopolysaccharides in polyacrylamide gels. Analytical biochemistry. 1982; 119:115–119. [PubMed: 6176137]

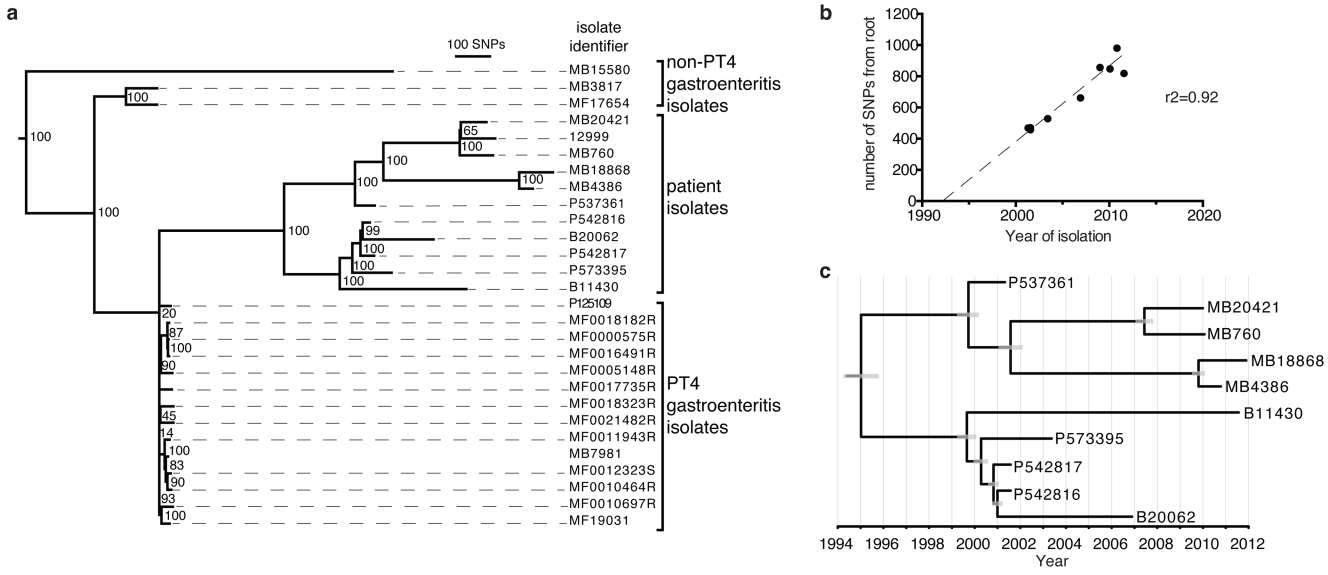


**Figure 1. Phylogeny of *S*. Enteritidis gastroenteritis isolates and blood-isolates from the IL12 beta 1 receptor deficient patient**

(A) Maximum likelihood tree of 17 *S*. Enteritidis isolates from uncomplicated gastroenteritis cases and eleven isolates from the blood of the patient, constructed using 5,087 SNPs with reference to *S*. Enteritidis P125109 whole genome sequence. Bootstrap values are indicated at each node and the scale bar indicates the estimated number of SNPs. (B) The number of SNPs from the root of the clade containing the patient isolates to each tip representing an isolate is plotted against the date of isolation. The broken line indicates the line of best fit for the data points extrapolated to intersect with the x-axis. (C) Bayesian time-dependent maximum clade credibility tree constructed using 2545 SNPs with reference to *S*. Enteritidis P125109 whole genome sequence. Grey bars indicate the 95% confidence intervals for the date of divergence for each ancestral node.




Europe PMC Funders Author Manuscripts

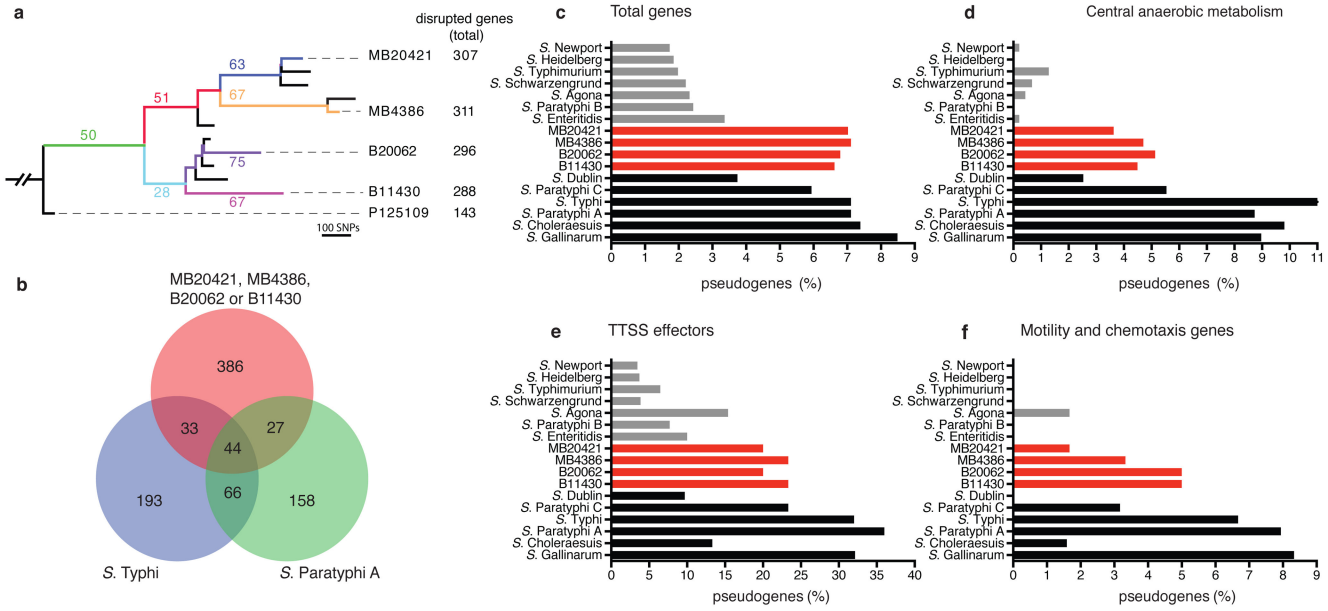


**Figure 2. Patient isolates have acquired a large number of pseudogenes. The total number of pseudogenes were determined for four of the patient isolates**

(A) The number of pseudogenes acquired along specific branches of the phylogeny are given. (B) The overlap between the pseudogene repertoire in patient isolates and human-restricted serovars *S*. Typhi CT18 and *S*. Paratyphi A (ATCC 9150) is represented by a venn diagram where numbers of pseudogenes are indicated. The percentages of pseudogenes in (C) the whole genome, (D) in central anaerobic metabolism genes, and (E) genes encoding TTSS effectors are given for the patient isolates and other *Salmonella* serovars. Host-promiscuous serovars most commonly associated with gastro-intestinal disease are shown with grey bars, patient isolates with red bars and host-adapted serovars causing systemic infection with black bars. Calculations are based on the classification of pseudogenes according to Nuccio et. al.[17]




Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

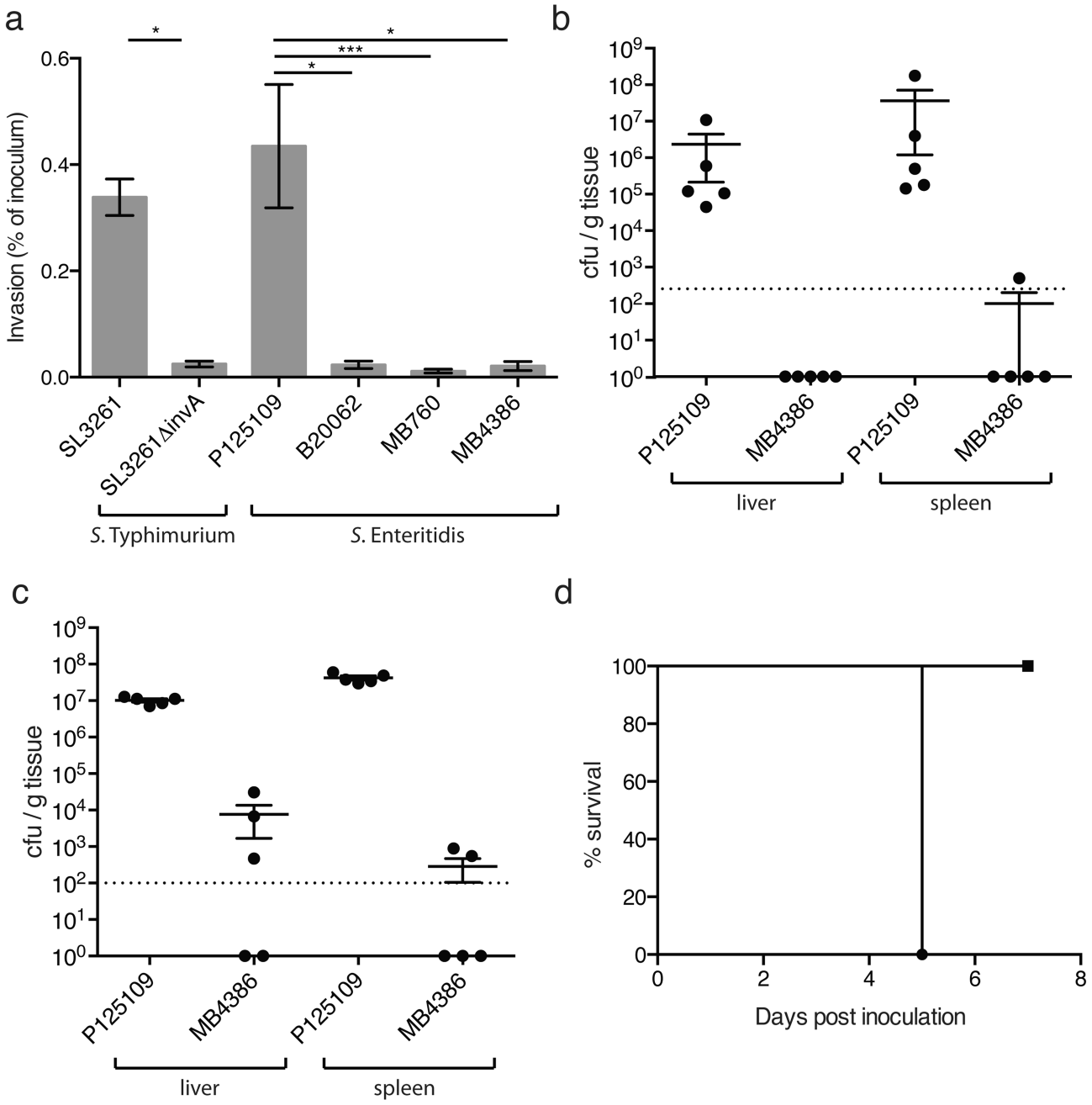


**Figure 3. Colonisation of the murine host and interaction with epithelial-like cells, by patient isolates**

(A) Invasion of HEp-2 epithelial-like cells inferred by protection from gentamicin killing. Invasion is expressed as the number of viable bacterial cells following lysis of tissue culture cells as a percentage of those added to each well. p-values calculated by Kruskal-Wallis test: * is p≤0.05, *** is p≤0.001. Colonisation of the liver, spleen and blood of C57BL/6 mice was determined five days post (B) oral or (C) intravenous inoculation with patient isolate MB4386 or *S*. Enterititidis P125109, n=5 per group. A, B, and C plotted mean ± SEM. Dashed lines indicate limit of detection. (D) Survival of mice following intravenous

inoculation of C57BL/6 mice with either patient isolate MB4386 (squares) or *S*. Enteritititidis P125109 (circles) n=4 per group.



Europe PMC Funders Author Manuscripts