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

Running title: Within-host evolution of *Salmonella*

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

Correspondence: rob.kingsley@ifr.ac.uk

Summary Paragraph

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 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 $1x10^{-5}$ 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$^{\Delta 253\text{-}311}$. 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*$^{\Delta 253\text{-}311}$). Serial passage of SW824 *mutS*$^{\Delta 253\text{-}311}$ 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 $1x10^3$ 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 x 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 M☕-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].

Figure 1

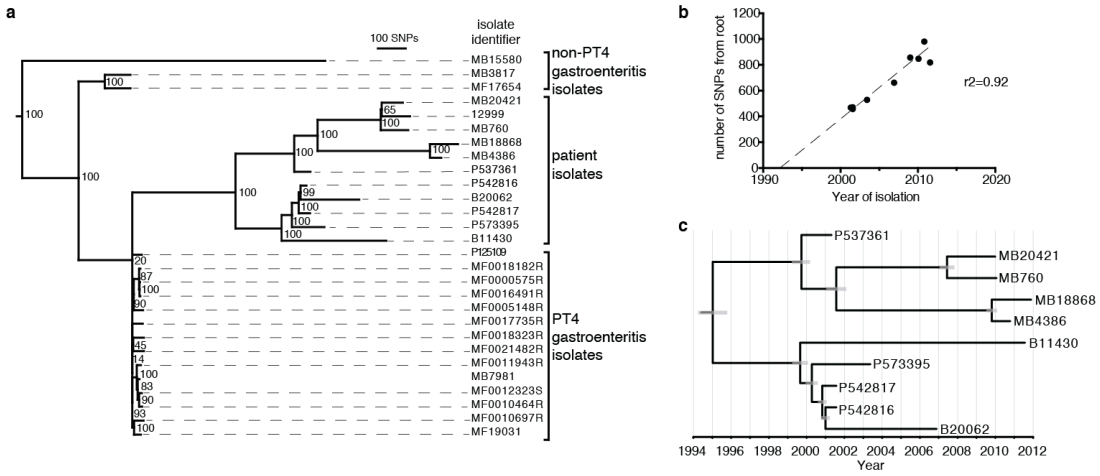


Figure 1. Phylogeny of *S.* Enteritidis gastroenteritis isolates and blood-isolates from the IL12 beta 1 receptor deficient patient. (A) Maximum likelihood tree of 17 *S.* Enteritidis isolates from uncomplicated gastroenteritis cases and eleven isolates from the blood of the patient, constructed using 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.

Figure 2

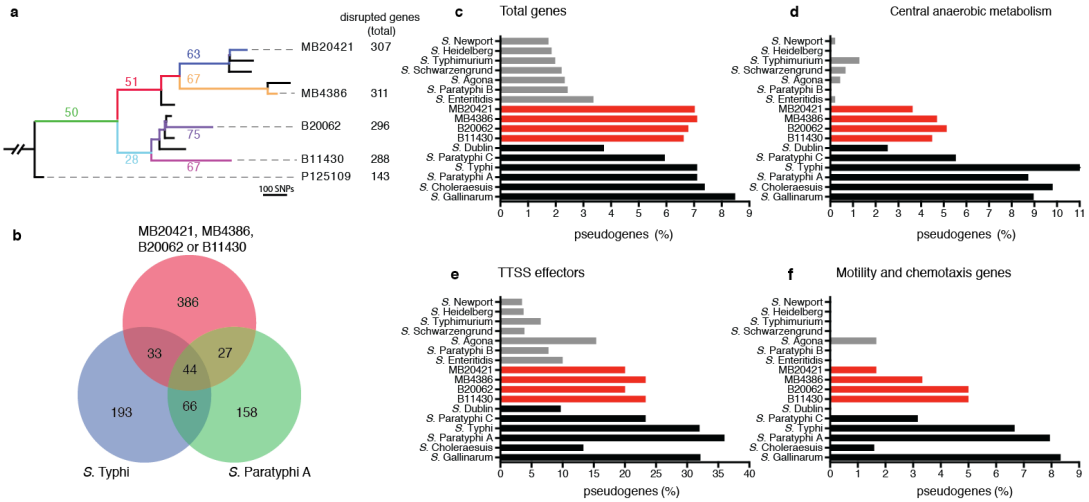


Figure 2. Patient isolates have acquired a large number of pseudogenes. The total number of pseudogenes were determined for four of the patient isolates. (A) The number of pseudogenes acquired along specific branches of the phylogeny are given. (B) The overlap between the pseudogene repertoire in patient 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]

Figure 3

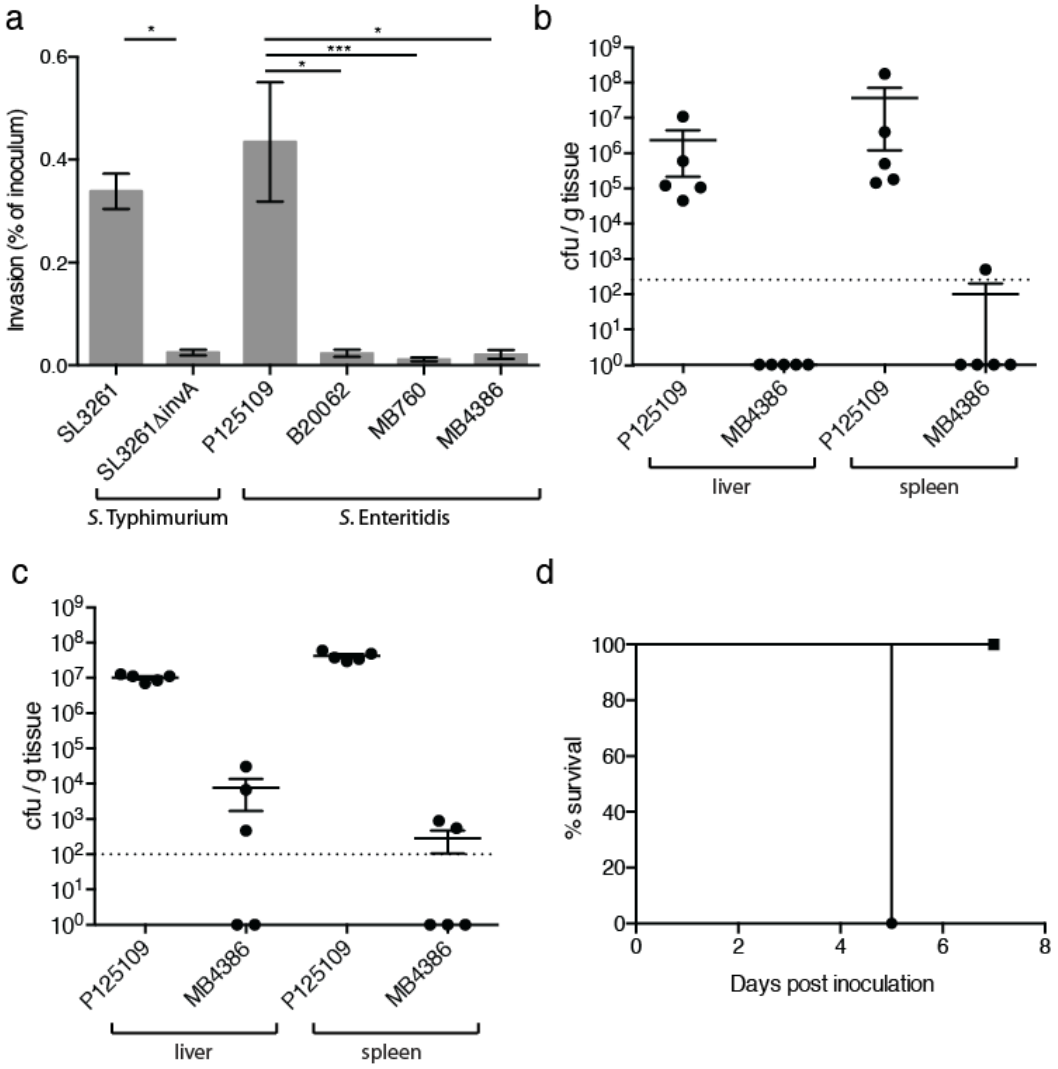


Figure 3. Colonisation of the murine host and interaction with epithelial-like cells, by patient isolates. (A) Invasion of HEp-2 epithelial-like cells inferred by protection from gentamicin killing. Invasion is expressed as the number of viable bacterial cells following lysis of tissue culture cells as a percentage of those added to each well. p-values calculated by Kruskal-Wallis test: * is $p \leq 0.05$, *** is $p \leq 0.001$. 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.* Enterititidis P125109 (circles) n=4 per group.

Correspondence and requests for materials should be addressed to Robert A Kingsley

Acknowledgements. 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.

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.

Author Information.

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.

Gordon Dougan and Robert A Kingsley contributed equally to this work

1 Baumler, A. & Fang, F. C. Host specificity of bacterial pathogens. *Cold Spring Harbor perspectives in medicine* **3**, a010041, doi:10.1101/cshperspect.a010041 (2013).
2 Parkhill, J. *et al.* Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhi CT18. *Nature* **413**, 848-852 (2001).
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* **36**, 1268-1274 (2004).
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* **111**, 12199-12204, doi:10.1073/pnas.1411012111 (2014).
5 Holt, K. E. *et al.* Pseudogene accumulation in the evolutionary histories of Salmonella enterica serovars Paratyphi A and Typhi. *BMC genomics* **10**, 36 (2009).
6 Celum, C. L., Chaisson, R. E., Rutherford, G. W., Barnhart, J. L. & Echenberg, D. F. Incidence of salmonellosis in patients with AIDS. *The Journal of infectious diseases* **156**, 998-1002 (1987).
7 de Beaucoudrey, L. *et al.* Revisiting human IL-12Rbeta1 deficiency: a survey of 141 patients from 30 countries. *Medicine* **89**, 381-402, doi:10.1097/MD.0b013e3181fdd832 (2010).
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* **190**, 1755-1757, doi:10.1086/425021 (2004).
9 Altare, F. *et al.* Impairment of mycobacterial immunity in human interleukin-12 receptor deficiency. *Science* **280**, 1432-1435 (1998).
10 Okoro, C. K. *et al.* Intracontinental spread of human invasive Salmonella Typhimurium pathovariants in sub-Saharan Africa. *Nature genetics* **44**, 1215-1221, doi:10.1038/ng.2423 (2012).
11 Mather, A. E. *et al.* Distinguishable Epidemics of Multidrug-Resistant Salmonella Typhimurium DT104 in Different Hosts. *Science* **341**, 1514-1517, doi:Doi 10.1126/Science.1240578 (2013).
12 Schaaper, R. M. Base selection, proofreading, and mismatch repair during DNA replication in Escherichia coli. *J Biol Chem* **268**, 23762-23765 (1993).
13 Denamur, E. & Matic, I. Evolution of mutation rates in bacteria. *Mol Microbiol* **60**, 820-827, doi:10.1111/j.1365-2958.2006.05150.x (2006).
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* **154**, 959-970 (2000).
15 Rocha, E. P. C. *et al.* Comparisons of dN/dS are time dependent for closely related bacterial genomes. *Journal of Theoretical Biology* **239**, 226-235 (2006).
16 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**, e00565-00513, doi:10.1128/mBio.00565-13 (2013).
17 Nuccio, S. P. & Baumler, A. J. Comparative analysis of Salmonella genomes identifies a metabolic network for escalating growth in the inflamed gut. *mBio* **5**, e00929-00914, doi:10.1128/mBio.00929-14 (2014).
18 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**, 2279-2287, doi:gr.091017.109 [pii] 10.1101/gr.091017.109 (2009).
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* **90**, 542-549, doi:10.1016/j.ygeno.2007.06.009 (2007).

20 Thomson, N. R. *et al.* Comparative genome analysis of Salmonella Enteritidis PT4 and Salmonella Gallinarum 287/91 provides insights into evolutionary and host adaptation pathways. *Genome Res* **18**, 1624-1637 (2008).

21 Kurtz, S. *et al.* Versatile and open software for comparing large genomes. *Genome Biol* **5**, R12 (2004).

22 Kurtz, S. *et al.* REPuter: the manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res* **29**, 4633-4642 (2001).

23 Holt, K. E. *et al.* High-throughput sequencing provides insights into genome variation and evolution in Salmonella Typhi. *Nature genetics* **40**, 987-993 (2008).

24 He, M. *et al.* Evolutionary dynamics of Clostridium difficile over short and long time scales. *Proc Natl Acad Sci U S A* **107**, 7527-7532, doi:10.1073/pnas.0914322107 (2010).

25 Stamatakis, A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688-2690 (2006).

26 Zerbino, D. & Birney, E. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research* **18**, 821-829 (2008).

27 Otto, T. D., Dillon, G. P., Degrave, W. S. & Berriman, M. RATT: Rapid Annotation Transfer Tool. *Nucleic Acids Res* **39**, e57, doi:10.1093/nar/gkq1268 (2011).

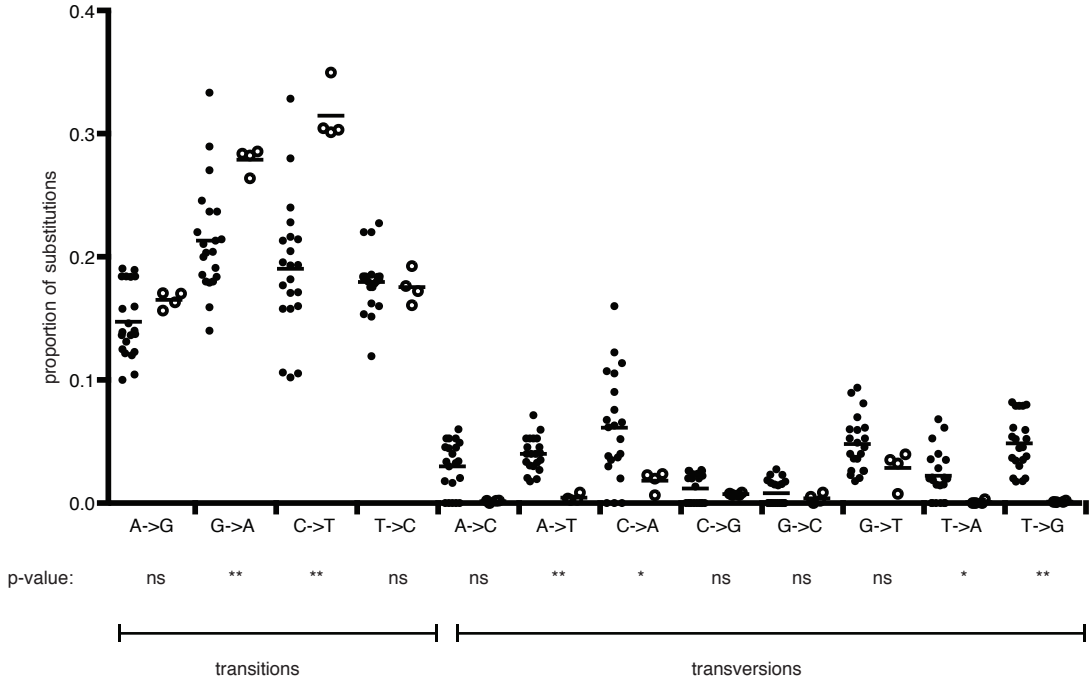
28 Carver, T. J. *et al.* ACT: the Artemis Comparison Tool. *Bioinformatics* **21**, 3422-3423 (2005).

29 Huntley, R. P. *et al.* The GOA database: gene Ontology annotation updates for 2015. *Nucleic Acids Res* **43**, D1057-1063, doi:10.1093/nar/gku1113 (2015).

30 Tsai, C. M. & Frasch, C. E. A sensitive silver stain for detecting lipopolysaccharides in polyacrylamide gels. *Analytical biochemistry* **119**, 115-119 (1982).

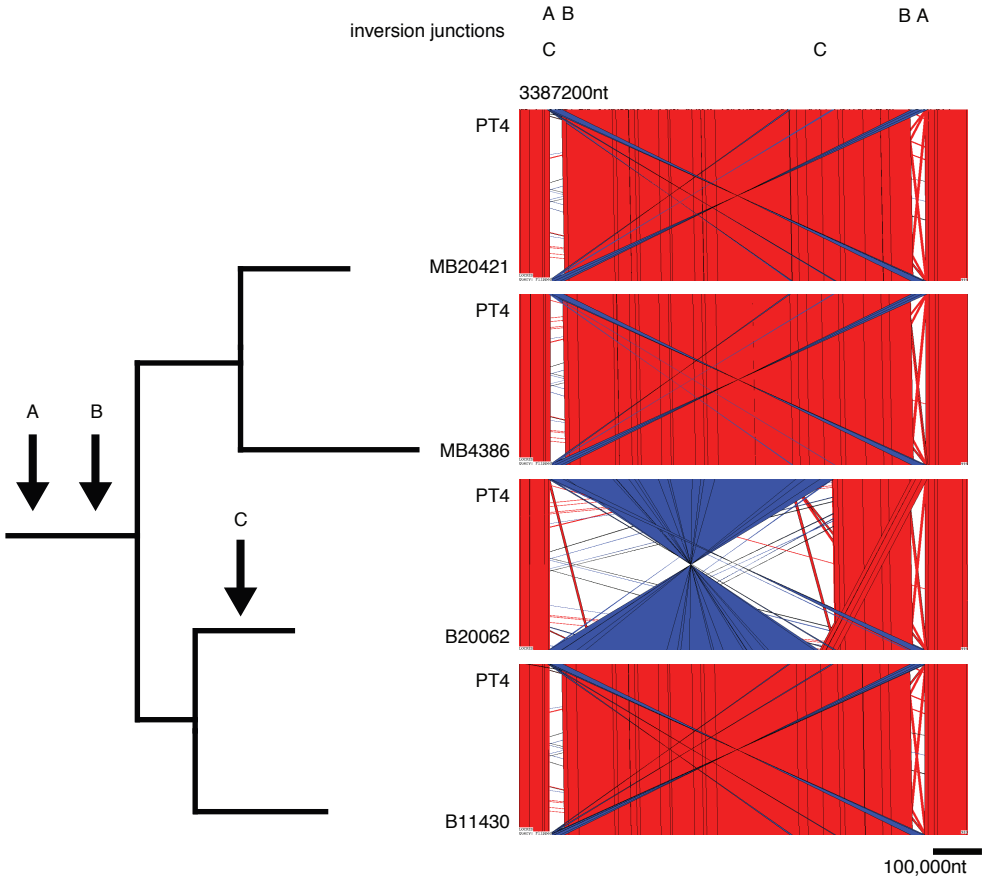
Supplementary Information Table of Contents

Supplementary Information Figure 1



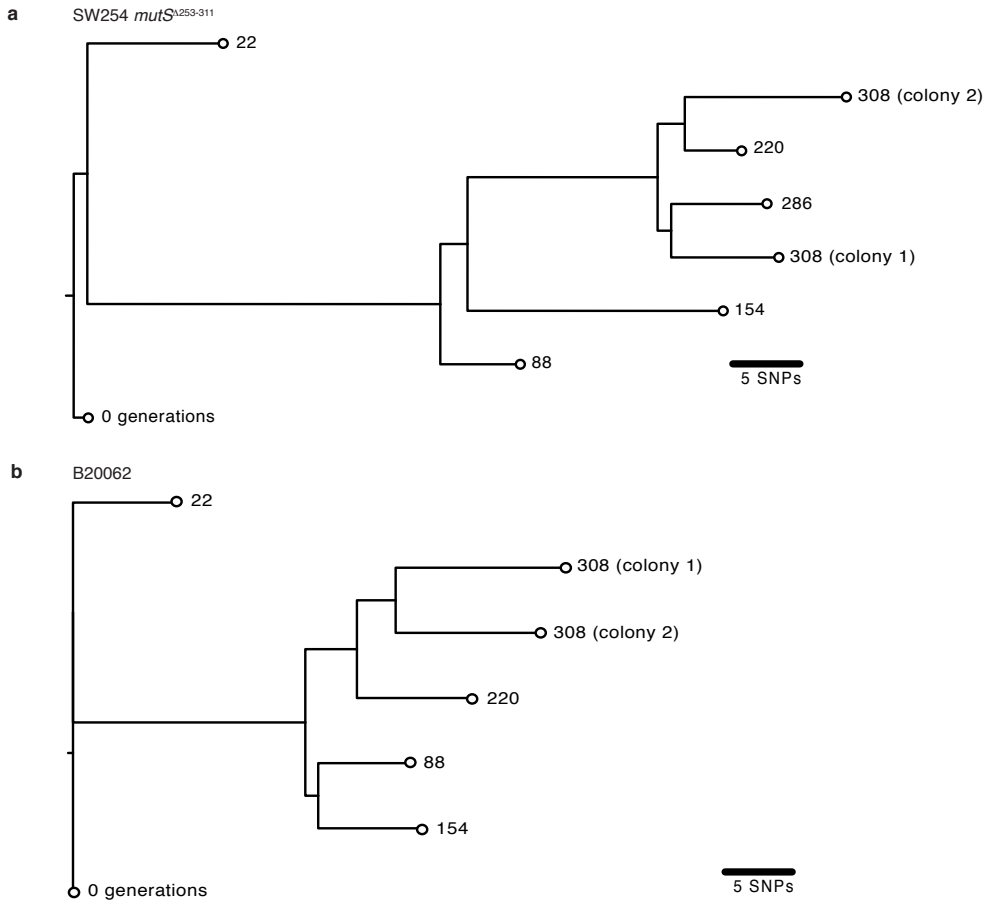
Supplementary Information Figure 1. Substitution frequency in the *S.* Enteritidis phylogenetic tree. Substitutions in lineages leading to classic gastroenteritis isolates (closed circles) or patient isolates (open circles) are plotted (mean values). p-values calculated by Mann-Whitney t-test: * is $p \leq 0.05$ and ** is $p \leq 0.01$.

Supplementary Information Figure 2



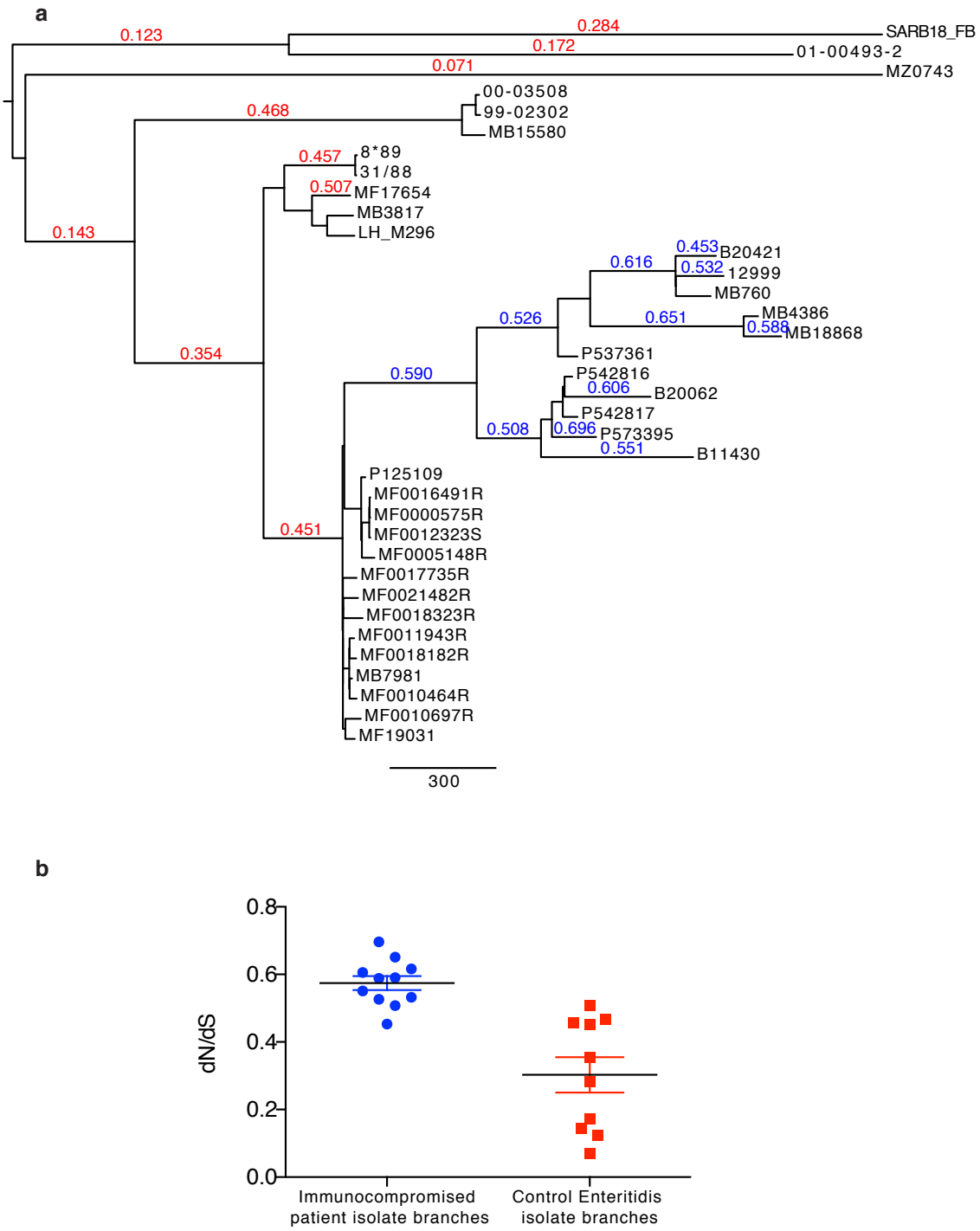
Supplementary Information Figure 2. Genomic rearrangements in patient isolates. We found evidence for multiple inversion events in the patient isolates. Assemblies of four patient isolates were compared to the reference P125109 sequence. Red regions denote synteny and blue regions are reversed. Mapped to the phylogeny, there were two nested inversions (A and B) common to all four isolates. B20062 experienced a third inversion (C) in the same region. Inversion locations are indicated at top and are flanked by rRNA genes. Scale bar indicates number of nucleotides.

Supplementary Information Figure 3



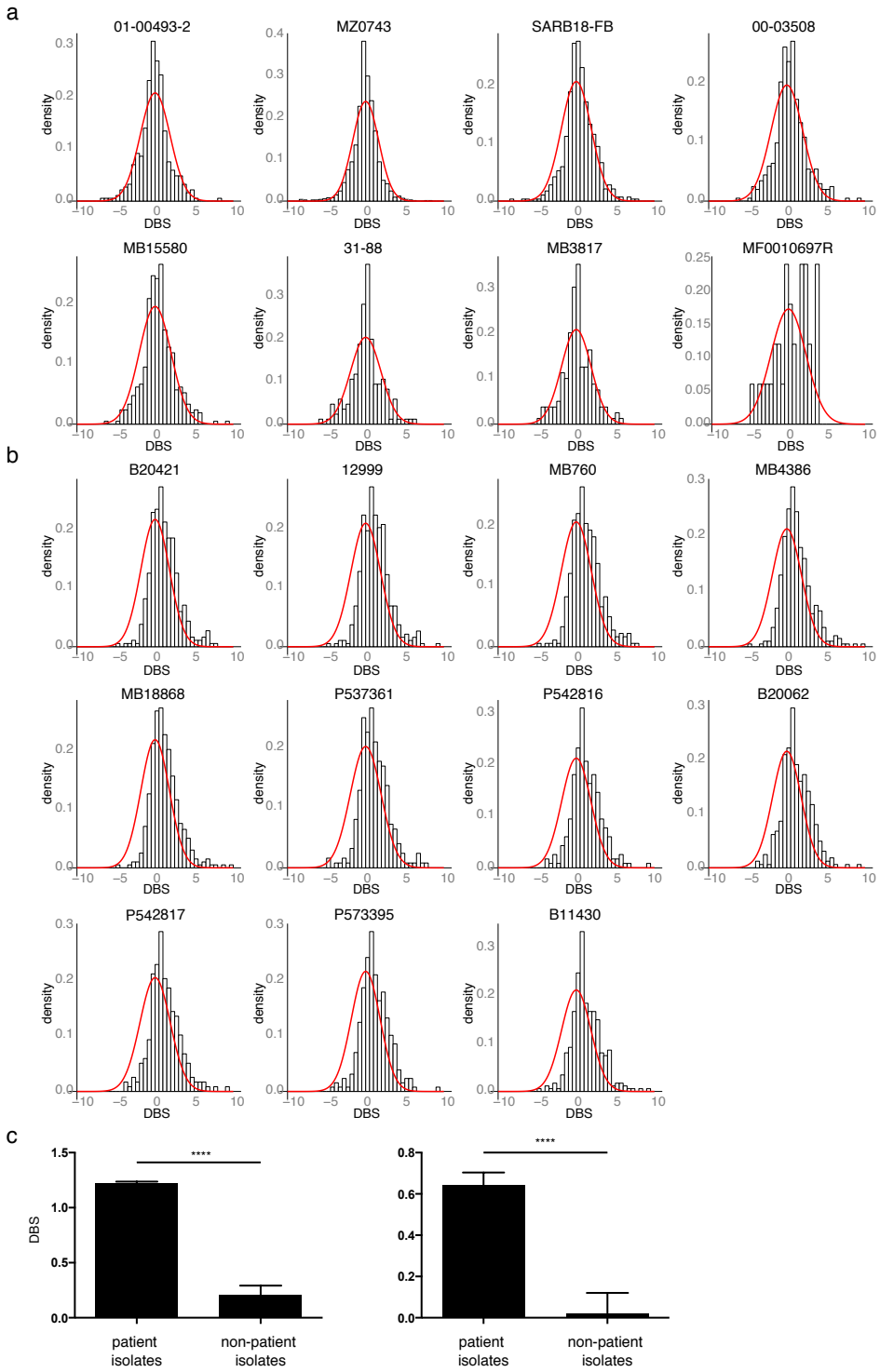
Supplementary Information Figure 3. Phylogeny of *S.* Enteritidis SW824 *mutS*$^{\Delta 253\text{-}311}$ and patient isolate B20062 cultured with serial passage in Luria-Bertani broth. Maximum-likelihood tree constructed using SNPs in whole genome sequence from serial passaged isolates of (A) *S.* Enteritidis SW824 *mutS*$^{\Delta 253\text{-}311}$ or (B) patient isolate B20062, with reference to P125109 whole genome sequence. The bar indicates the estimated distance for 5 SNPs and values at tips indicate the estimated number of generations during passage for each isolate. Two colonies were sequenced from the final culture (308 generations). P125109 only acquired a single SNP during passage so no tree was constructed.

Supplementary Information Figure 4



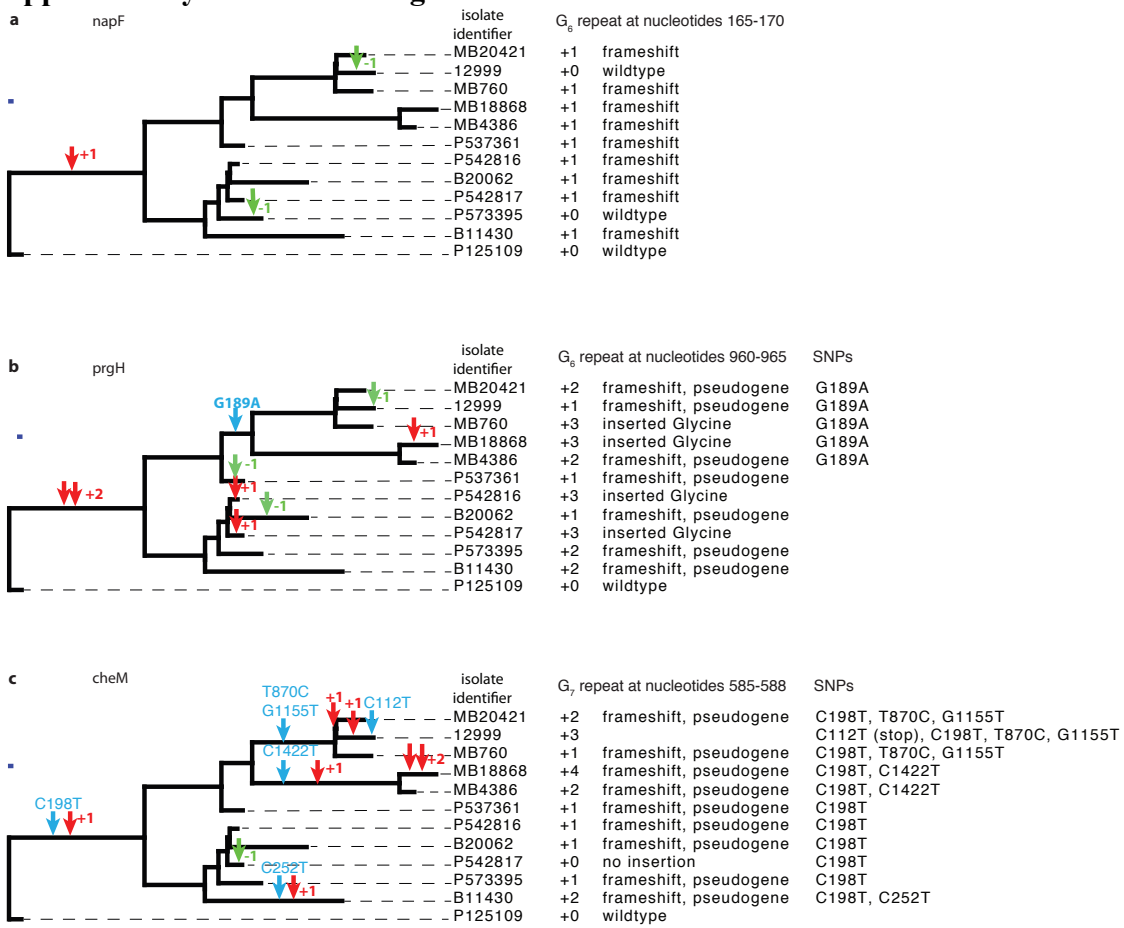
Supplementary Information Figure 4. dN/dS calculations for *S.* Enteritidis phylogeny. (A) dN/dS values were calculated for the major patient isolate (blue) or non-patient isolate (red) branches. (B) Plots of the values from (A) with mean ± SEM. Mean dN/dS values of branches among the patient isolates differ significantly from the control isolates (Mann Whitney test p=0.0001).

Supplementary Information Figure 5



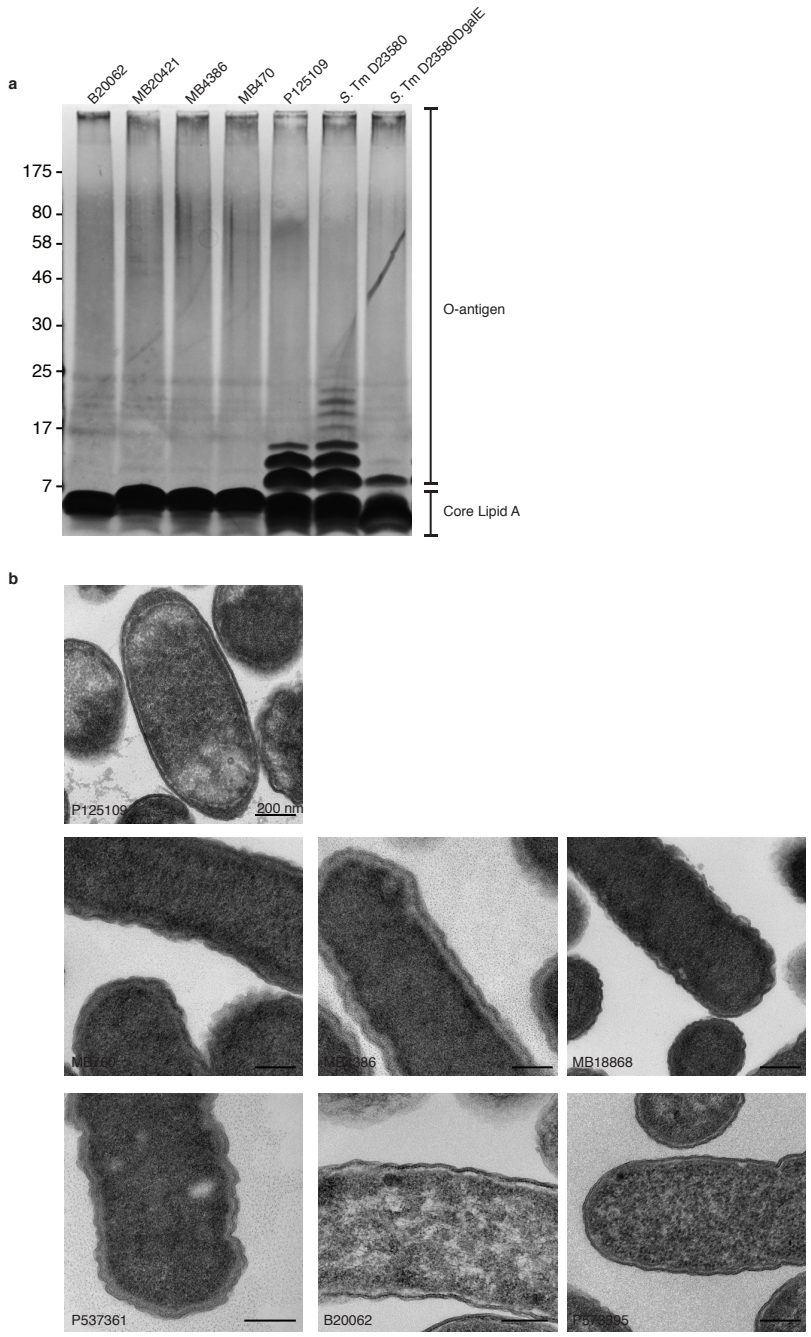
Supplementary Information Figure 5. Delta bitscore analysis. (Aand B) Bitscore values for conserved domains were determined for each strain using hmmpfam and subtracted from P125109 values (delta bitscore, DBS). Histograms of delta bitscore values are plotted (x-axis) for (A) non-patient *S.* Enteritidis isolates and (B) immunocompromised patient isolates. Normal distribution with mean = 0, and with corresponding variance and number of data points for each proteome are plotted in red for comparison. (C) Mean DBS and (D) skewness values were calculated for patient and non-patient isolates (plotted mean ± SEM, **** is p<0.0001 by t-test).

Supplementary Information Figure 6



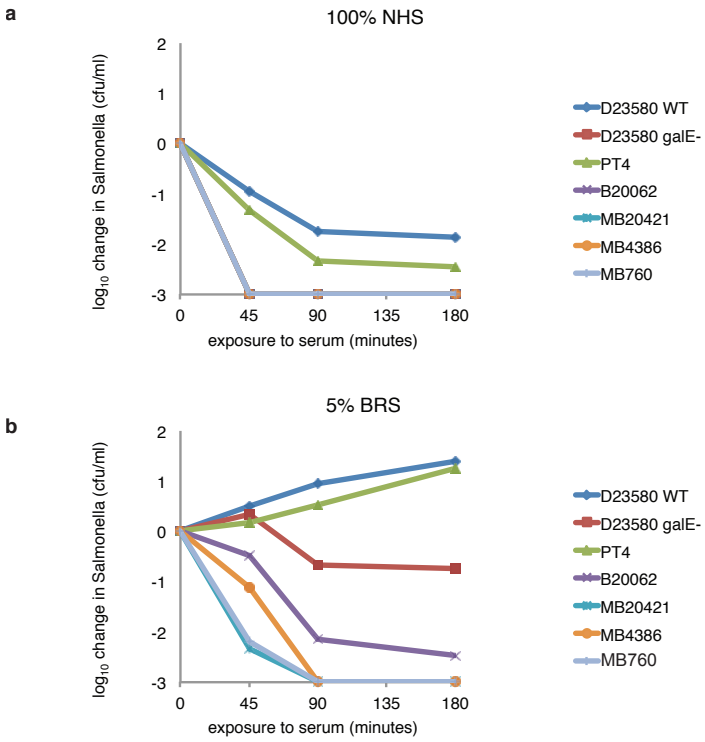
Supplementary Information Figure 6. Examples of patterns of pseudogene persistence and reversion. Indels and SNP mutation events were mapped to the phylogenetic tree of patient isolates for three exemplary genes. Parsimonious models of event timings are indicated with arrows. (A) *napF* has a homopolymeric tract of six Gs which is expanded to seven (+1) in the most recent common ancestor, resulting in a pseudogene. Two patient isolates (12999 and P573395) experienced a reversing indel (-1) at the terminal branches that restored the gene to a functional version. (B) *prgH* experienced several insertion and deletion events at different branches on the tree, resulting in a mixture of lengths of the homopolymeric tract. Those with three inserted Gs (+3) returned to the correct reading frame and encode a protein with an inserted Glycine that is likely tolerated and thus produces a functional protein. (C) *cheM* also experienced several insertion and deletion events, two of which returned the homopolymeric tract to the correct reading frame (+0 in P542817 and +3 in 12999). 12999 has an additional SNP that creates a premature stop upstream of the polymeric tract (C to T at nucleotide 112). *cheM* shows five transition mutations and one transversion.

Supplementary Information Figure 7



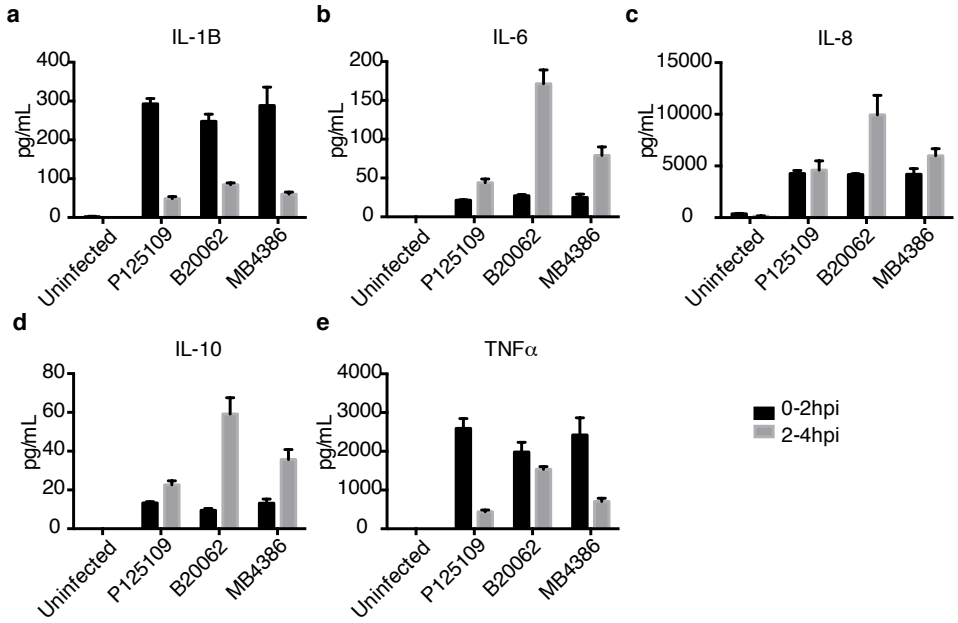
Supplementary Information Figure 7. Lipopolysaccharide profiles and EM. (A) Crude LPS preparations from stationary phase cultures of four patient isolates, *S.* Enteritidis P125109, *S.* Typhimurium D23580 and *galE* mutant separated by SDS PAGE. The *galE* mutant was included for comparison because it is required for O-antigen attachment. LPS preparations were also repeated with log phase cultures that showed similar lack of O-antigen. (B) Positive stain electron microscopy of *S.* Enteritidis P125109 and six patient isolates. Scale bars = 200nm. Images are representative of at least ten fields of view.

Supplementary Information Figure 8



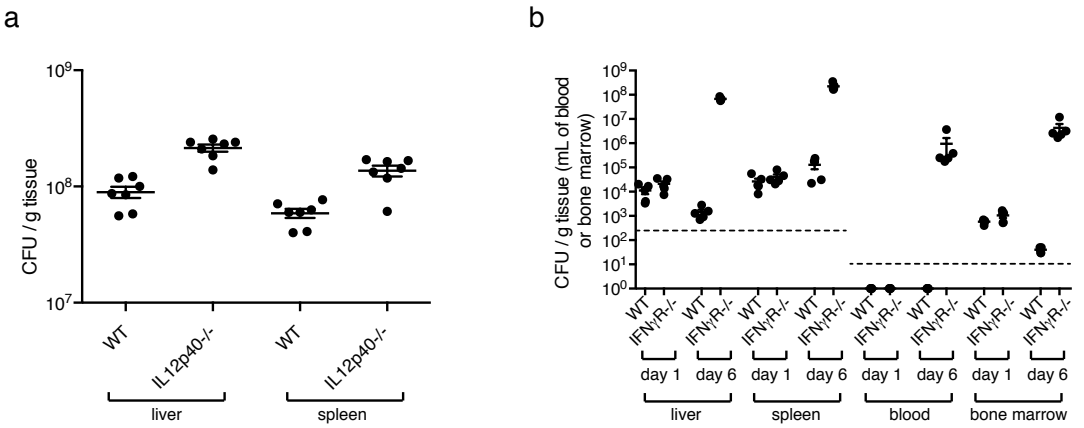
Supplementary Information Figure 8. Patient isolates are highly sensitive to complement-mediated killing. *S.* Enteritidis isolates in log phase were exposed to (A) 100% normal human serum (NHS) and (B) 5% baby rabbit serum (BRS) in PBS and the number of viable bacteria was determined at 45, 90, 120 and 180 minutes and plotted as $\log_{10}$ fold change compared to starting amount of 10^6 cfu/mL. All four patient isolates showed a three $\log_{10}$ fold change at the 45 minute time point after exposure to 100% NHS. The increase in sensitivity is greater than that observed following mutation of *galE*, another gene required for O-antigen attachment, in *S.* Typhimurium. Data in A is representative of five independent experiments. Data in panel B is representative of ten independent experiments performed with BRS concentrations from 1% to 100% (1, 2, 5, 10, 20, 30, 50, 75 and 100%). As with 100% NHS, for all experiment with BRS ≥20%, no viable bacteria from any of the patient isolates could be detected at any of the three time points.

Supplementary Information Figure 9



Supplementary Information Figure 9. Production of cytokines by THP-1 macrophage-like cells following exposure to *S.* Enterititidis P125109 or patient isolates B20062 or MB4386. Cytokines were determined in the cell supernatants collected two hours after the infection (0-2hpi, black bars), media was changed and cell supernatants were collected again two hours later (2-4hpi, grey bars). Mean values (from six replicates per experiment) are plotted with standard deviations. This experiment was repeated twice with similar results.

Supplementary Information Figure 10



Supplementary Information Figure 10. Colonization of immunocompromised mice with *S*. Enteritidis patient isolate. (A) Colonisation of the liver and spleen of C57BL/6 (WT) and $Il12^{btm1Jm}$ ($IL12p40^{-/-}$) mice was determined five days post intravenous inoculation with 10^6 CFUs patient isolate MB4386, n=5 per group. (B) Colonisation of the liver, spleen, blood, and bone marrow of C57BL/6 (WT) and $Ifngr1^{tm1Agt}$ ($IFN\gamma R^{-/-}$) mice was determined one and six days post intravenous inoculation with 10^6 CFUs patient isolate MB4386, n=5 per group. A and B plotted mean ± SEM. Dashed lines indicate limit of detection.

Supplementary Information Table I

Date of isolation	Strain	MIC Ciprofloxacin	MIC Azithromycin	gyrA mutations	gyrB mutations
unknown	12999	ND	ND	D87G	S464F
4/18/2001	P537361	ND	ND	D87G	S464F
7/19/2001	P542817	ND	ND	S83F and D87G	
7/19/2001	P542816	ND	ND	S83F and D87G	
05/06/2003	P573395	ND	ND	D87G	V727G and K728E
11/10/2006	B20062	0.5	1.5	S83F and D87G	
12/17/2009	B20421	0.25	1	D87G	S464F
12/22/2009		0.25	1	ND	ND
1/13/2010	MB760	1	16	D87G	S464F
10/21/2010	B4386	0.25	2	D87G	S464F
3/17/2011		0.25	ND	ND	ND
7/14/2011	B11430	ND	ND	D87G	
08/11/2011		1	4	ND	ND
11/03/2011		1	2	ND	ND
11/11/2011	MB18868	1	2	D87G	S464F
03/08/2012		2	ND	ND	ND

Supplementary Information Table I. Antimicrobial resistance in patient isolates. The minimum inhibitory concentration (MIC) of patient blood samples was determined according to British Society of Antimicrobial Chemotherapy guidelines (British Society for Antimicrobial Chemotherapy, B., B1 3NJ, UK. June 2014. BSAC Methods for Antimicrobial Susceptibility Testing, Version 12). Salmonella with Ciprofloxacin MICs above 0.06μg/mL or Azithromycin MICs above 16μg/mL are considered resistant (The European Committee on Antimicrobial Susceptibility Testing. Breakpoint tables for interpretation of MICs and zone diameters. http://www.eucast.org Version 5.0, 2015.). For isolates with sequence data available, the SNPs in *gyrA*, *gyrB*, *parC* and *parE* are given. *gyrA* S83F, *gyrA* D87G, and *gyrB* S464F mutations are linked to fluoroquinolone resistance. *parC* and *parE* SNPs reported here have not been linked to quinolone resistance. The patient isolates encode three genes associated with macrolide resistance, *carA*, *msrA*, and *srmB*, that are also present in the PT4 isolate P125109. We do not observe the acquisition of any additional genes or SNPs in *rplD*, *rplV* or 23S rRNA known to confer macrolide resistance.

Supplementary Information Table II

			MB20421	12999	MB760	MB4386	MB18868	P537361	P542816	B20062	P542817	P573395	B11430	S. Typhi	S. Paratyphi A
Cell envelope															
rfaL	SEN3535	O-antigen ligase (waaL)	X	X	X	X	X		X	X	X	X	X		
mpl	SEN4185	murein peptide ligase	X	X	X	X	X	X	X	X	X	X	X	X	X
Central anaerobic metabolism															
narK	SEN1274	Nitrate utilization: nitrite extrusion protein	X	X	X	X	X	X	X	X	X	X			
narX	SEN1273	Nitrate utilization: nitrate/nitrite sensor protein		X	X	X	X						X		
ttrB	SEN1660	Tetrathionate utilization: tetrathionate reductase subunit B		X	X			X	X	X	X				X
ttrR	SEN1658	Tetrathionate utilization: two-component response regulator		X		X	X						X		
btuR	SEN1315	COB(I) alamin adenosyltransferase	X	X	X	X	X	X	X	X	X	X	X		
cbiN	SEN2020	Vitamin B12 synthesis: putative cobalt transport protein				X	X								
cbiL	SEN2022	Vitamin B12 synthesis: precorrin-2 C20-methyltransferase				X	X					X	X		
cbiK	SEN2023	Vitamin B12 synthesis: sirohydrochlorin cobaltochelatase							X	X	X	X		X	
cbiJ	SEN2024	Vitamin B12 synthesis: cobalt-precorrin-6a reductase											X	X	
prpR	SEN0350	propionate catabolism operon regulatory protein	X	X	X	X	X	X	X	X	X	X	X		
pduW	SEN2055	acetokinase	X	X	X	X	X	X	X	X	X	X	X		
napF	SEN2244	ferredoxin-type protein	X		X	X	X	X	X	X			X		
dmsC2	SEN4079	putative dimethyl sulfoxide reductase subunit C	X	X	X	X	X	X	X	X	X	X	X		
Motility and chemotaxis															
cheM	SEN1084	methyl-accepting chemotaxis protein II (tar)	X	X	X	X	X	X	X	X		X	X		X
cheR	SEN1085	chemotaxis protein methyltransferase				X	X								
motB	SEN1081	motility protein B		X			X		X		X	X			
Adhesion															
shdA	SEN2493	host colonisation factor	X	X	X	X	X	X						X	X
misL	SEN3580	putative autotransported protein								X				X	
bigA	SEN3305	putative surface-exposed virulence protein								X	X			X	X
lpfC	SEN3461	outer membrane usher protein							X	X	X	X	X	X	X
sefR	SEN4251	fimbrial operon positive regulatory protein	X	X	X	X	X	X	X	X	X	X	X	X	X
fimC	SEN0526	fimbrial chaperone protein		X											
fimD	SEN0527	outer membrane usher protein								X					
Type III secretion system															
prgH	SEN2716	SPI-1 TIIISS apparatus protein	X	X		X		X		X		X	X		
sipA	SEN2723	SPI-1 secreted effector protein	X	X	X	X	X	X	X	X	X	X	X		
sopA	SEN2065	SPI-1 secreted effector protein				X								X	X
sopD	SEN2784	SPI-1 secreted effector protein		X											
sseI	SEN0916	SPI-2 secreted effector protein	X	X		X	X	X	X	X	X	X	X	X	X
sseJ	SEN1422	SPI-2 secreted effector protein	X	X	X	X	X	X		X		X	X	X	X
sspH2	SEN2224	SPI-2 secreted effector protein	X	X	X	X	X	X	X	X	X	X	X		X
steC	SEN1335	SPI-2 secreted effector protein	X	X	X	X	X	X	X	X	X	X	X		X
sseK1	SEN3941	SPI-2 secreted effector protein											X	X	X

Supplementary Information Table II. Genome degradation in patient isolates. Genes with observed disruption in patient isolates, *S.* Typhi and *S.* Paratyphi A are listed, organized by functional class. Common gene name, *S.* Enteritidis locus tag and short description of function are given. Absence of a functional gene in an isolate is indicated by an X.

Supplementary Information Table III

B20421	B4386	B20062	B11430	Typhi	Paratyphi A	gene name	locus	function
X	X	X	X	X	X	sseJ	SEN1422	salmonella translocated effector protein (SseJ)
X	X	X	X	X	X	sefR	SEN4251	fimbrial operon positive regulatory protein
X	X	X	X	X		SEN0031	SEN0031	putative membrane protein
X	X	X	X	X		yecG	SEN1077	conserved hypothetical protein
X	X	X	X	X		dbpA	SEN1376	ATP-independent RNA helicase (DbpA)
X	X	X	X		X	fhuA	SEN0196	ferrichrome-iron receptor
X	X	X	X		X	cheM	SEN1084	methyl-accepting chemotaxis protein II
X	X	X	X		X	SEN1335	SEN1335	hypothetical protein
X	X	X	X		X	ygcY	SEN2806	glucarate dehydratase-related protein
X	X	X	X			SEN4271	SEN4271	probable aspartate racemase
X	X	X	X			SEN4291	SEN4291	putative Type I restriction-modification system methyltransferase
X	X	X	X			phnV	SEN0408	probable membrane component of 2-aminoethylphosphonate transporter
X	X	X	X			SEN2318	SEN2318	conserved hypothetical protein
X	X	X	X			dsbG	SEN0576	thiol:disulfide interchange protein DsbG precursor
X	X	X	X			ycfN	SEN1841	conserved hypothetical protein
X	X	X	X			yceE	SEN1894	putative membrane transport protein
X	X	X	X			yiaE	SEN3469	putative 2-hydroxyacid dehydrogenase
X	X	X	X			SEN2485	SEN2485	hypothetical protein
X	X	X	X			mutS	SEN2748	DNA mismatch repair protein
X	X	X	X			sseI	SEN0916	putative type III secreted protein
X	X	X	X			btuR	SEN1315	COB(I) alamin adenosyltransferase
X	X	X	X			yafS	SEN0264	conserved hypothetical protein
X	X	X	X			mod	SEN0340	type III restriction-modification system enzyme (StyLTI) modification methylase
X	X	X	X			prpR	SEN0350	propionate catabolism operon regulatory protein
X	X	X	X			sbcD	SEN0379	exonuclease SbcD
X	X	X	X			ybcI	SEN0521	putative membrane protein
X	X	X	X			nadA	SEN0701	quinolinate synthetase A protein
X	X	X	X			SEN0802	SEN0802	putative electron transfer flavoprotein (alpha subunit)
X	X	X	X			otsA	SEN1076	trehalose-6-phosphate synthase
X	X	X	X			SEN1789	SEN1789	conserved hypothetical protein
X	X	X	X			SEN1924	SEN1924	phage protein
X	X	X	X			pduW	SEN2055	Acetokinase.
X	X	X	X			sspH2	SEN2224	secreted effector protein
X	X	X	X			napF	SEN2244	ferredoxin-type protein NapF
X	X	X	X			yojN	SEN2251	putative two-component system sensor kinase
X	X	X	X			SEN2324	SEN2324	putative membrane protein
X	X	X	X			prgH	SEN2716	pathogenicity 1 island effector protein
X	X	X	X			sipA	SEN2723	pathogenicity island 1 effector protein (function unknown)
X	X	X	X			aas	SEN2853	2-acylglycerophosphoethanolamine acyl transferase/acyl carrier protein synthetase
X	X	X	X			SEN2961	SEN2961	putative hydrolase
X	X	X	X			yhaN	SEN3079	conserved hypothetical protein
X	X	X	X			SEN3353	SEN3353	putative membrane protein
X	X	X	X			SEN3354A	SEN3354A	putative MFS-family membrane transport protein
X	X	X	X			SEN3370	SEN3370	putative kinase/trancriptional regulatory protein
X	X	X	X			rfaL	SEN3535	O-antigen ligase
X	X	X	X			dmsC2	SEN4079	putative dimethyl sulfoxide reductase subunit C
X	X	X	X			mpl	SEN4185	murein peptide ligase
X	X	X	X			yjiJ	SEN4276	putative membrane protein
X	X	X	X			SEN4290	SEN4290	putative Type I restriction-modification system methyltransferase
X	X	X	X			yjjU	SEN4324	conserved hypothetical protein
X	X	X			X	ydeE	SEN1535	putative membrane protein
X	X	X				rhlE	SEN0766	putative ATP-dependent RNA helicase rhlE
X	X	X				SEN2876	SEN2876	putative membrane protein
X	X	X				narK	SEN1274	nitrite extrusion protein (nitrite facilitator)
X	X	X				ydcW	SEN1458	putative aldehyde dehydrogenase
X	X	X				wcaH	SEN2103	putative O-antigen biosynthesis protein
X	X	X				SEN2609	SEN2609	large repetitive protein
X	X		X		X	SEN3896	SEN3896	serine hydroxymethyltransferase
X	X		X			ygbE	SEN2771	conserved hypothetical protein
X	X		X			SEN0907	SEN0907	putative ion:amino acid symporter
X	X		X			yoaA	SEN1216	putative ATP-dependent helicase
X	X		X			yciR	SEN1329	conserved hypothetical protein
X	X		X			yncD	SEN1468	probable TonB-dependent receptor YncD precursor
X	X		X			ttrR	SEN1658	putative two-component response regulator
X	X		X			SEN2255	SEN2255	putative MR-MLE-family protein
X	X		X			SEN3273	SEN3273	elongation factor Tu
X	X			X	X	SEN0336	SEN0336	putative cation transport atpase (ec 3.6.1.-)
X	X			X	X	ybeV	SEN0627	
X	X			X	X	SEN1500	SEN1500	putative LacI-family transcriptional regulator
X	X			X		SEN0218	SEN0218	putative inner membrane protein
X	X			X		SEN0240	SEN0240	putative secreted chitinase
X	X			X		sthB	SEN4349	Outer membrane fimbrial usher protein
X	X				X	add	SEN1584	adenosine deaminase
X	X					mutY	SEN2953	A/G-specific adenine glycosylase
X	X					SEN1384	SEN1384	putative phage encoded DNA-binding protein
X	X					SEN0216	SEN0216	putative viral enhancing factor
X	X					cdd	SEN2176	cytidine deaminase
X	X					SEN0316	SEN0316	putative lysR family transcriptional regulator
X	X					SEN0580	SEN0580	molybdopterin-containing oxidoreductase catalytic subunit
X	X					SEN0621	SEN0621	putative sigma-54 dependent transcriptional regulator
X	X					yedO	SEN1054	D-cysteine desulfhydrase
X	X					SEN1162	SEN1162	exported phage protein
X	X					SEN1719	SEN1719	conserved hypothetical protein
X	X					yeaC	SEN1760	conserved hypothetical protein
X	X					SEN1765	SEN1765	putative regulatory protein
X	X					SEN1995	SEN1995	conserved hypothetical protein
X	X					baeS	SEN2126	putative two-component system sensor kinase
X	X					yffH	SEN2456	conserved hypothetical protein
X	X					SEN2465	SEN2465	putative membrane protein
X	X					shdA	SEN2493	host colonisation factor (ShdA)
X	X					yfiC	SEN2569	conserved hypothetical protein
X	X					yhdA	SEN3209	putative lipoprotein
X	X					yhgF	SEN3330	putative transcription accessory protein

X	X					SEN3356	SEN3356	putative dihydrodipicolinate synthetase
X	X					SEN3513	SEN3513	putative exported protein
X	X					rfaY	SEN3538	lipopolysaccharide core biosynthesis protein
X	X					pstS	SEN3671	periplasmic phosphate-binding protein
X	X					nfi	SEN3954	putative endonuclease V
X	X					yjaH	SEN3957	conserved hypothetical protein
X	X					SEN4195	SEN4195	conserved hypothetical protein
X	X					SEN1431	SEN1431	putative exported protein
X		X	X	X		SEN3455	SEN3455	putative xanthine permease
X		X	X			pegD	SEN2144A	putative exported protein
X		X	X			rna	SEN0586	ribonuclease I precursor
X		X	X			SEN1006	SEN1006	hypothetical protein
X		X	X			SEN1495	SEN1495	putative hydrolase
X		X	X			ygcB	SEN2783	conserved hypothetical protein
X		X			X	SEN3649	SEN3649	putative membrane transport protein
X		X				bigA	SEN3305	putative surface-exposed virulence protein BigA
X		X				cbiL	SEN2022	precorrin-2 C20-methyltransferase
X		X				pegC	SEN2145	putative outer membrane usher protein
X			X	X		SEN1482	SEN1482	conserved hypothetical protein
X			X		X	SEN1450	SEN1450	putative benzoate membrane transport protein
X			X			yehW	SEN2156	putative permease transmembrane component
X			X			SEN0206	SEN0206	putative exported protein
X			X			yecA	SEN1067	conserved hypothetical protein
X			X			gltS	SEN3567	glutamate permease
X				X		ylbE	SEN0511	conserved hypothetical protein
X					X	safD	SEN0284	fimbrial structural subunit
X					X	asnB	SEN0644	asparagine synthetase B
X						tufA	SEN3930	elongation factor tu (ef-tu) (p-43)
X						eutC	SEN2437	Ethanolamine ammonia-lyase light chain
X						SEN2428	SEN2428	putative lipoprotein
X						SEN2358	SEN2358	putative lipoprotein
X						lasT	SEN4356	putative RNA methyltransferase
X						safB	SEN0282	salmonella atypical fimbria chaperone
X						SEN2555	SEN2555	putative transcriptional regulator
X						SEN2463	SEN2463	putative membrane protein
X						yaiB	SEN0366	conserved hypothetical protein
X						aceE	SEN0156	pyruvate dehydrogenase E1 component
X						SEN0167	SEN0167	conserved hypothetical protein
X						yaiC	SEN0368	adrA protein
X						mdlA	SEN0442	putative ABC transporter ATP-binding membrane protein
X						ybbA	SEN0488	hypothetical ABC transporter ATP-binding protein
X						allB	SEN0504	putative allantoinase
X						potF	SEN0823	putrescine-binding periplasmic protein precursor
X						cbpA	SEN0976	curved DNA-binding protein
X						yebA	SEN1113	conserved hypothetical protein
X						oppF	SEN1293	oligopeptide transport ATP-binding protein (OppF)
X						yncB	SEN1466	putative NADP-dependent oxidoreductase
X						SEN1543A	SEN1543A	putative membrane transport protein
X						SEN1794	SEN1794	putative inner membrane transport protein
X						SEN1938	SEN1938	phage portal protein
X						SEN1951	SEN1951	phage protein
X						rfbP	SEN2081	undecaprenyl-phosphate galactosephosphotransferase
X						SEN2130	SEN2130	hypothetical protein
X						yejB	SEN2210	putative binding-protein-dependent transporter
X						usg	SEN2351	putative semialdehyde dehydrogenase
X						SEN2484	SEN2484	putative membrane protein
X						aroF	SEN2591	phospho-2-dehydro-3-deoxyheptonate aldolase, tyr-sensitive
X						SEN2611	SEN2611	putative type I secretion protein, ATP-binding protein
X						ygcA	SEN2802	putative RNA methyltransferase
X						SEN2929	SEN2929	putative outer membrane lipoprotein
X						gsp	SEN2982	glutathionylspermidine synthetase/amidase
X						sufI	SEN3015	SufI protein
X						yrbD	SEN3144	possible exported protein
X						SEN3354	SEN3354	putative glycerol dehydrogenase
X						yicJ	SEN3571	sodium:galactoside family symporter
X						SEN3610	SEN3610	putative DeoR-family transcriptional regulator
X						ydeY	SEN3865	putative ABC transporter permease protein
X						SEN4029	SEN4029	putative type-I secretion protein
X						SEN4081	SEN4081	putative exported protein
X						SEN4299	SEN4299	putative transcriptional regulatory protein
X						sthA	SEN4350	putative fimbrial chaperone protein
	X	X	X	X	X	mglA	SEN2182	galactoside transport atp-binding protein mgla
	X	X	X		X	SEN3492	SEN3492	hypothetical protein
	X	X	X			ygeA	SEN2858	conserved hypothetical protein
	X	X	X			ampD	SEN0150	AmpD protein (anhydro-N-acetylmuramyl-tripeptide amidase)
	X	X	X			ydhB	SEN1618	putative transcriptional regulator
	X	X	X			plsX	SEN1857	fatty acid/phospholipid synthesis protein
	X	X	X			torA	SEN3638	trimethylamine-N-oxide reductase precursor
	X	X	X			SEN3982	SEN3982	phage-related protein
	X	X		X		ygbJ	SEN2757	conserved hypothetical protein
	X	X		X		SEN4249	SEN4249	outer membrane fimbrial usher protein
	X	X				ygiD	SEN3032	conserved hypothetical protein
	X	X				sgbE	SEN3499	putative sugar isomerase
	X	X				SEN0713	SEN0713	oxaloacetate decarboxylase beta chain (ec 4.1.1.3)
	X	X				yebU	SEN1187	conserved hypothetical protein
	X	X				pegB	SEN2145A	putative fimbrial chaperone protein
	X		X		X	phsA	SEN2063	thiosulfate reductase precursor
	X		X		X	lyxK	SEN3496	putative L-xylulose kinase
	X		X			yaoF	SEN1776	conserved hypothetical protein
	X		X			narX	SEN1273	nitrate/nitrite sensor protein NarX
	X		X			ynfM	SEN1564	putative membrane transport protein
	X			X	X	SEN0335	SEN0335	putative cation efflux pump
	X			X	X	sopA	SEN2065	secreted protein SopA
	X			X		ychH	SEN1257	putative exported protein

	X			X		SEN0325	SEN0325	exported protein
	X			X		ybeM	SEN0600	possible hydrolase
	X			X		fdnG	SEN1485	formate dehydrogenase, nitrate-inducible, major subunit (ec 1.2.1.2)
	X				X	yjeP	SEN4117	putative membrane protein
	X					yebN	SEN1203	putative membrane protein
	X					ydjM	SEN1722	conserved hypothetical protein
	X					ycfL	SEN1843	putative lipoprotein
	X					SEN1996	SEN1996	conserved hypothetical protein
	X					SEN2134	SEN2134	putative membrane protein
	X					SEN2144	SEN2144	conserved hypothetical protein
	X					dinI	SEN1886	damage-inducible protein
	X					bcfB	SEN0021	fimbrial chaperone
	X					crl	SEN0302	curlin genes transcriptional activator
	X					ybaE	SEN0438	putative solute-binding protein
	X					apeE	SEN0546	outer membrane esterase
	X					aroG	SEN0705	phospho-2-dehydro-3-deoxyheptonate aldolase (DAHP synthetase)
	X					SEN0706	SEN0706	putative hydro-lyase
	X					SEN0708	SEN0708	LysR-family transcriptional regulator
	X					SEN0712	SEN0712	oxaloacetate decarboxylase alpha chain
	X					ybhK	SEN0747	conserved hypothetical protein
	X					SEN0996	SEN0996	putative exported protein
	X					cheR	SEN1085	chemotaxis protein methyltransferase
	X					SEN1145	SEN1145	putative phage lysozyme
	X					acnA	SEN1321	aconitate hydratase 1 (citrate hydro-lyase 1)
	X					ydeJ	SEN1537	conserved hypothetical protein
	X					selD	SEN1746	selenophosphate synthase
	X					ycfD	SEN1820	conserved hypothetical protein
	X					SEN1920	SEN1920	phage protein
	X					cbiN	SEN2020	putative cobalt transport protein CbiN
	X					yeeF	SEN2067	putative amino acid transporter protein
	X					wcaC	SEN2109	glycosyltransferase
	X					SEN2232	SEN2232	cytochrome c-type biogenesis protein F1
	X					rcsC	SEN2253	sensor protein RcsC
	X					yfcX	SEN2370	putative fatty acid oxidation complex alpha subunit
	X					xapB	SEN2402	xanthosine permease
	X					ypfI	SEN2464	conserved hypothetical protein
	X					hypF	SEN2683	hydrogenase maturation protein
	X					barA	SEN2803	sensor protein
	X					yrbI	SEN3149	conserved hypothetical protein
	X					kefB	SEN3285	glutathione-regulated potassium-efflux system protein (K(+)/H(+)antiporter)
	X					menA	SEN3880	menaquinone biosynthetic protein
	X					aceA	SEN3966	isocitrate lyase
	X					SEN4028	SEN4028	putative type-I secretion protein
	X					smp	SEN4333	putative membrane protein
		X	X			SEN1134	SEN1134	putative phage membrane protein
		X	X			SEN1142	SEN1142	putative phage lipoprotein
		X	X			yohL	SEN2866	conserved hypothetical protein
		X	X			SEN0030	SEN0030	putative transcriptional regulator (lysR family)
		X	X			fepA	SEN0554	ferrienterobactin receptor precursor
		X	X			SEN0829	SEN0829	possible transport protein
		X	X			motB	SEN1081	motility protein B
		X	X			ynfC	SEN1549	exported protein
		X	X			aadA	SEN1788	aminoglycoside-resistance protein
		X	X			smf	SEN3233	conserved hypothetical protein
		X	X			glpD	SEN3351	aerobic glycerol-3-phosphate dehydrogenase
		X	X			yhjD	SEN3431	putative membrane protein
		X	X			lpfC	SEN3461	outer membrane usher protein (LpfC)
		X	X			malE	SEN3997	periplasmic maltose-binding protein
		X		X	X	SEN1751	SEN1751	Putative metabolite transport protein
		X		X	X	SEN3421	SEN3421	L-asparginase
		X		X		yceB	SEN1884	conserved hypothetical protein
		X		X		ynfD	SEN1550	putative exported protein
		X		X		cbiK	SEN2023	sirohydrochlorin cobaltochelatase (ec 4.99.1.3)
		X		X		speC	SEN2958	ornithine decarboxylase, constitutive (ec 4.1.1.17)
		X		X		misL	SEN3580	putative autotransported protein (MisL)
		X			X	SEN1517	SEN1517	hydrogenase isoenzymes formation protein
		X			X	ttrB	SEN1660	tetrathionate reductase subunit B
		X			X	SEN2985	SEN2985	possible ABC-transport protein, periplasmic-binding component
		X			X	mgtA	SEN4207	Mg(2+) transport ATPase, P-type
		X				SEN2237	SEN2237	cytochrome C biogenesis ATP-binding export protein (ec 3.6.3.41)
		X				SEN1677	SEN1677	putative transporter
		X				msgA	SEN1807	putative virulence protein (MsgA)
		X				SEN1959	SEN1959	phage encoded recombination associated protein
		X				dctA	SEN3437	C4-dicarboxylate transport protein
		X				SEN0273	SEN0273	rhs-associated protein
		X				SEN3183	SEN3183	conserved hypothetical protein
		X				zntR	SEN3240	putative Zn(II)-responsive regulator
		X				yjdB	SEN4064	putative membrane protein
		X				SEN0032	SEN0032	possible sulfatase
		X				mutT	SEN0138	7,8-dihydro-8-oxoguanine-triphosphatase
		X				SEN0141	SEN0141	putative lysR-family transcriptional regulator
		X				phoE	SEN0303	outer membrane pore protein E precursor
		X				fimD	SEN0527	outer membrane usher protein FimD precursor
		X				ybgF	SEN0700	putative exported protein
		X				ybiO	SEN0772	putative membrane protein
		X				yliH	SEN0799	conserved hypothetical protein
		X				pqiB	SEN0929	putative secreted protein
		X				fljB	SEN1049	flagellin
		X				SEN1137	SEN1137	hypothetical phage protein
		X				narG	SEN1275	respiratory nitrate reductase 1 alpha chain
		X				oppA	SEN1289	periplasmic oligopeptide-binding protein precursor (OppA)
		X				pgpB	SEN1323	phosphatidylglycerophosphatase B
		X				astB	SEN1737	succinylarginine dihydrolase
		X				SEN1922	SEN1922	putative phage tail fibre protein

		X				SEN2001	SEN2001	conserved hypothetical protein
		X				pduG	SEN2041	propanediol utilization protein
		X				wcaI	SEN2102	putative glycosyltransferase
		X				yegQ	SEN2131	putative protease
		X				napH	SEN2240	ferredoxin-type protein NapH
		X				srlE	SEN2674	PTS system, glucitol/sorbitol-specific IIBC component
		X				spaO	SEN2732	surface presentation of antigens protein (associated with type III secretion and virulence)
		X				SEN2752	SEN2752	possible permease
		X				ygdH	SEN2813	conserved hypothetical protein
		X				SEN2969	SEN2969	possible amino acid transport protein
		X				nanK	SEN3169	possible kinase
		X				glpR	SEN3348	glycerol-3-phosphate regulon repressor
		X				rfaK	SEN3536	lipopolysaccharide 1,2-n-acetylglucosaminetransferase (ec 2.4.1.56)
		X				glpK	SEN3876	glycerol kinase
		X				aceB	SEN3965	malate synthase A
		X				phnB	SEN4059	conserved hypothetical protein
		X				melB	SEN4071	melibiose carrier protein
		X				aspA	SEN4096	aspartate ammonia-lyase
		X				sgaT	SEN4149	putative transport protein SgaT
		X				ytfM	SEN4178	putative exported protein
		X				yjjI	SEN4327	conserved hypothetical protein
			X	X	X	SEN0756	SEN0756	putative inner membrane protein
			X	X	X	SEN2129	SEN2129	hypothetical protein
			X	X		cbiJ	SEN2024	cobalt-precorrin-6a reductase (ec 1.3.1.-)
			X	X		SEN0018	SEN0018	putative chitinase
			X	X		yedI	SEN1021	conserved hypothetical integral membrane protein
			X	X		yhhJ	SEN3408	putative ABC-2 type superfamily transport protein
			X		X	foxA	SEN0347	ferrioxamine B receptor precursor
			X		X	SEN1504	SEN1504	putative isomerase
			X			araH	SEN1074	conserved hypothetical protein
			X			baeR	SEN2127	putative two-component system response regulator
			X			hilC	SEN2709	possible AraC-family transcriptional regulator
			X			SEN1795	SEN1795	putative inner membrane transport protein
			X			yhcQ	SEN3198	possible exported protein
			X			ydhC	SEN1619	putative integral membrane transport protein
			X			SEN1797	SEN1797	putative lipoprotein
			X			SEN0317	SEN0317	conserved hypothetical protein
			X			yabN	SEN0110	putative ABC transporter periplasmic solute binding protein
			X			citE	SEN0591	citrate lyase beta chain
			X			ybiK	SEN0793	putative L-asparaginase
			X			dacC	SEN0809	D-alanyl-D-alanine carboxypeptidase (penicillin-binding protein 6 precursor)
			X			pqiA	SEN0928	putative inner membrane protein
			X			lonH	SEN0933	conserved hypothetical protein
			X			yedP	SEN1024	conserved hypothetical protein
			X			fliI	SEN1037	flagellum-specific ATP synthase
			X			edd	SEN1118	6-phosphogluconate dehydratase
			X			SEN1279	SEN1279	putative secreted protein
			X			ydcK	SEN1446	putative transferase
			X			narV	SEN1475	respiratory nitrate reductase 2 gamma chain
			X			rnfC	SEN1590	Electron transport complex protein
			X			potA	SEN1823	spermidine/putrescine transport ATP-binding protein PotA (ABC superfamily)
			X			ptsG	SEN1846	PTS system, glucose-specific IIBC component
			X			yegH	SEN2115	putative membrane protein
			X			yehX	SEN2157	ABC transporter ATP-binding protein
			X			yfbB	SEN2290	conserved hypothetical protein
			X			yfcA	SEN2364	putative membrane protein
			X			yfeA	SEN2396	putative membrane protein
			X			yfeD	SEN2398	conserved hypothetical protein
			X			SEN2795	SEN2795	outer membrane usher protein
			X			garR	SEN3088	tartronate semialdehyde reductase (tsar)
			X			yhcM	SEN3179	putative ATP/GTP-binding protein
			X			ilvE	SEN3709	branched-chain amino-acid aminotransferase
			X			yihU	SEN3811	putative oxidoreductase
			X			SEN3857	SEN3857	hypothetical protein
			X			SEN3941	SEN3941	conserved hypothetical protein
			X			yjeM	SEN4115	putative amino acid permease
			X			SEN4214	SEN4214	putative membrane protein
			X			SEN4289	SEN4289	hypothetical protein

Supplementary Information Table III. Genome degradation in patient isolates. Genes with observed disruption in patient isolates, *S.* Typhi and *S.* Paratyphi A are listed. Common gene name, *S.* Enteritidis locus tag and short description of function are given. Absence of a functional gene in an isolate is indicated by an X.

Supplementary Information Table IV

GO term	Description (total annotated genes)	Patient isolates pseudogenes (expected)	Patient isolates p-value	S. Typhi p-value	S. Paratyphi A p-value
GO class Biological Process					
GO:0015891	siderophore transport (7)	3 (0.57)	0.0143	ns	0.0038
GO:0055085	transmembrane transport (332)	36 (26.82)	0.0154	0.01018	0.0071
GO:0042402	cellular biogenic amine catabolic process (3)	2 (0.24)	0.0185	ns	ns
GO:0005992	trehalose biosynthetic process (3)	2 (0.24)	0.0185	ns	ns
GO:0006810	transport (625)	74 (50.49)	0.0248	ns	ns
GO:0043711	pilus organization (15)	4 (1.21)	0.0278	ns	ns
GO:0006528	asparagine metabolic process (4)	2 (0.32)	0.035	ns	0.0143
GO:0009236	cobalamin biosynthetic process (24)	5 (1.94)	0.0393	0.01021	ns
GO:0006835	dicarboxylic acid transport (10)	3 (0.81)	0.0408	ns	ns
GO:0061077	chaperone-mediated protein folding (17)	4 (1.37)	0.0427	ns	ns
GO:0006814	sodium ion transport (25)	5 (2.02)	0.046	ns	0.0347
GO class Cellular Component					
GO:0016021	integral component of membrane (617)	67 (53.14)	0.011	0.0373	0.0047
GO:0009279	cell outer membrane (53)	9 (4.56)	0.033	0.025	ns
GO class Molecular Function					
GO:0004872	receptor activity (45)	9 (3.51)	0.0019	ns	0.0058
GO:0004616	phosphogluconate dehydrogenase (decarboxylating) activity (6)	3 (0.47)	0.0079	ns	ns
GO:0003849	3-deoxy-7-phosphoheptulonate synthase activity (3)	2 (0.23)	0.0173	ns	ns
GO:0030151	molybdenum ion binding (21)	5 (1.64)	0.02	ns	0.00375
GO:0051287	NAD binding (39)	7 (3.04)	0.0286	ns	ns
GO:0051539	4 iron, 4 sulfur cluster binding (67)	10 (5.23)	0.0326	ns	0.02175
GO:0004067	asparaginase activity (4)	2 (0.31)	0.0327	ns	ns
GO:0005283	sodium:amino acid symporter activity (4)	2 (0.31)	0.0327	ns	ns
GO:0005524	ATP binding (366)	38 (28.56)	0.0338	ns	ns
GO:0005215	transporter activity (408)	43 (31.84)	0.0419	0.00776	6.90E-06

Supplementary Information Table IV. GO term enrichment of pseudogenes in immunocompromised patient *S.* Enteritidis isolates. The indicated Gene Ontology terms were enriched in the repertoire of pseudogenes in the immunocompromised patient isolates. The list of 351 pseudogenes acquired in MB20421, MB4386, B20062, or MB11430 since divergence from P125109 were assigned GO terms. Significantly over-represented GO terms with $p < 0.05$ are reported. The p-value for pseudogenes in S. Typhi and S. Paratyphi A with these GO terms is also reported. ns= not significant.

Supplementary Information Table V

Strain / sample	ERS no.	ERR no.	Sequencing Instrument
05: MB7981	ERS037362	ERR048502	Illumina HiSeq 2000
09:MF0000575R	ERS037368	ERR048511	Illumina HiSeq 2000
09:MF0005148R	ERS037369	ERR048512	Illumina HiSeq 2000
09:MF0010306R	ERS037370	ERR048513	Illumina HiSeq 2000
09:MF0011240R	ERS037371	ERR048514	Illumina HiSeq 2000
09:MF0012323S	ERS037373	ERR048515	Illumina HiSeq 2000
09:MF0018182R	ERS037375	ERR048516	Illumina HiSeq 2000
09:MF0019755R	ERS037386	ERR048517	Illumina HiSeq 2000
09:MF0021482R	ERS037387	ERR048518	Illumina HiSeq 2000
09:MF0017735R	ERS037388	ERR048519	Illumina HiSeq 2000
MF17654	ERS037389	ERR048520	Illumina HiSeq 2000
08: MB3817	ERS037363	ERR048503	Illumina HiSeq 2000
MF19031	ERS037390	ERR048521	Illumina HiSeq 2000
B20062	ERS037391	ERR048522	Illumina HiSeq 2000
B20421	ERS037392	ERR048523	Illumina HiSeq 2000
MB760	ERS037393	ERR048524	Illumina HiSeq 2000
B4386	ERS037394	ERR048525	Illumina HiSeq 2000
08: MB15580	ERS037364	ERR048504	Illumina HiSeq 2000
08: MB16115	ERS037365	ERR048505	Illumina HiSeq 2000
08:MF0010464R	ERS037366	ERR048506	Illumina HiSeq 2000
08:MF0010697R	ERS037372	ERR048507	Illumina HiSeq 2000
08:MF0011943R	ERS037374	ERR048508	Illumina HiSeq 2000
08:MF0016491R	ERS037376	ERR048509	Illumina HiSeq 2000
08:MF0018323R	ERS037367	ERR048510	Illumina HiSeq 2000
B11430	ERS055598	ERR064843	Illumina HiSeq 2000
LB_PT4_0	ERS159698	ERR197399	Illumina HiSeq 2000
LB_PT4_1	ERS159699	ERR197400	Illumina HiSeq 2000
LB_PT4_4	ERS159700	ERR197401	Illumina HiSeq 2000
LB_PT4_7	ERS159701	ERR197402	Illumina HiSeq 2000
LB_PT4_10	ERS159702	ERR197403	Illumina HiSeq 2000
LB_PT4_13	ERS159703	ERR197404	Illumina HiSeq 2000
LB_PT4_14	ERS159704	ERR197405	Illumina HiSeq 2000
LB_PT4_14_1	ERS159705	ERR197406	Illumina HiSeq 2000
LB_PT4_14_2	ERS159706	ERR197407	Illumina HiSeq 2000
LB_SW824_0	ERS159707	ERR197408	Illumina HiSeq 2000
LB_SW824_1	ERS159708	ERR197409	Illumina HiSeq 2000
LB_SW824_4	ERS159709	ERR197410	Illumina HiSeq 2000
LB_SW824_7	ERS159710	ERR197411	Illumina HiSeq 2000
LB_SW824_10	ERS159711	ERR197412	Illumina HiSeq 2000
LB_SW824_13	ERS159712	ERR197413	Illumina HiSeq 2000
LB_SW824_14	ERS159713	ERR197414	Illumina HiSeq 2000
LB_SW824_14_1	ERS159714	ERR197415	Illumina HiSeq 2000
LB_SW824_14_2	ERS159715	ERR197416	Illumina HiSeq 2000
LB_B20062_0	ERS159716	ERR197417	Illumina HiSeq 2000
LB_B20062_1	ERS159717	ERR197418	Illumina HiSeq 2000
LB_B20062_4	ERS159718	ERR197419	Illumina HiSeq 2000
LB_B20062_7	ERS159719	ERR197420	Illumina HiSeq 2000
LB_B20062_10	ERS159720	ERR197421	Illumina HiSeq 2000
LB_B20062_13	ERS159721	ERR197422	Illumina HiSeq 2000
LB_B20062_14	ERS159722	ERR197423	Illumina HiSeq 2000
LB_B20062_14_1	ERS159723	ERR197424	Illumina HiSeq 2000
LB_B20062_14_2	ERS159724	ERR197425	Illumina HiSeq 2000
P573395	ERS159741	ERR197442	Illumina HiSeq 2000
P542817	ERS159742	ERR197443	Illumina HiSeq 2000
P542816	ERS159743	ERR197444	Illumina HiSeq 2000
P537361	ERS159744	ERR197445	Illumina HiSeq 2000
MB18868	ERS159745	ERR197446	Illumina HiSeq 2000
12999	ERS055600	ERR064845	Illumina HiSeq 2000
B20062	ERS475420	ERR654502, ERR654489	PacBio RS
MB20421	ERS475421	ERR654503, ERR654490, ERR654492	PacBio RS
MB4386	ERS475422	ERR654504, ERR654491, ERR654493	PacBio RS
B11430	ERS475423	ERR654505, ERR654494	PacBio RS
01-00493-2	ERS400244		Illumina Genome Analyzer II
99-02302	ERS400254		Illumina Genome Analyzer II
MZ0743	ERS400248		Illumina Genome Analyzer II
00-03508	ERS400256		Illumina Genome Analyzer II
LH_M296	ERS400259		Illumina Genome Analyzer II
SARB18_FB	ERS400262		Illumina Genome Analyzer II
31/88	ERS022680		Illumina Genome Analyzer II
8*89	ERS022681		Illumina Genome Analyzer II

Supplementary Information Table V. Accession numbers of *Salmonella* strains used in this study

Supplementary Discussion

Patient history and treatment

In 1995, a 12 year old patient presented with recurrent fever, diffuse lymphadenopathy, splenomegaly and was found to be bacteremic with *S.* Enteritidis. The patient was diagnosed with IL-12 beta-1 receptor deficiency (homozygous mutation deficiency of the IL-12 beta 1 chain) and was started on Interferon γ (IFNγ) treatment and remained on this until the age of 13. At the age of 16 the patient was diagnosed with purulent pericarditis, secondary to *S.* Enteritidis, and IFNγ treatment was restarted. Between the ages of 16 and 21 (1999-2004), the patient was bacteremic with a *S.* Enteritidis-like organism on multiple occasions and was treated with various antibiotics, including Ciprofloxacin, Azithromycin and Ceftriaxone. Every bacteremia episode was characterized by recurrence of vasculitic rash, arthritis and malaise that the patient described was becoming less severe with time. Initially, the vasculitic rash was widespread, involving his face, and later it started becoming limited to smaller and smaller areas until it finally completely disappeared during the last episodes. At the age of 21 an elective laparoscopic cholecystectomy was performed but no evidence of cholecystitis was identified on tissue pathology. Between the ages of 21 and 29 (2004-2012), the patient continued to have multiple bacteremia episodes, sometimes accompanied by a macular rash on trunk as well as progressive hypergammaglobulinemia. Numerous investigations including an FDG-PET, total body CT scans and MRIs, bone scans, bone marrow biopsies and echocardiograms failed to localize a tissue where *Salmonella* was resident. The patient also developed progressive normocytic normochromic anemia as well as chronic weight loss and C-reactive protein elevation. Importantly, *Salmonella* could not be cultured from the stool and there was no evidence of transmission as fecal screens of close relatives were negative. Typing of the isolates yielded inconclusive results leading to their classification as atypical *S.* Enteritidis.

Genetic determination of antimicrobial resistance in patient isolates

We determined if sequence polymorphisms that could be responsible for the increased resistance to the antimicrobials Ciprofloxacin and Azithromycin were present in the patient isolates. At least one of three substitutions previously associated with quinolone resistance, Asp87Gly and Ser83Phe in *gyrA* and Ser464Phe in *gyrB*[1], were present in each of the isolates (Extended Data Table I). No known genetic changes associated with macrolide resistance were identified in the patient isolates.

Classes of pseudogenes in patient isolates

Several anaerobic metabolism pathways have multiple disrupted genes in the patient isolates that could affect their ability to survive in the gut[2,3]. Two genes of the nitrate utilization pathway, *narK* and *narX*, are pseudogenes in the patient isolates. *Salmonella* have evolved to use tetrathionate, a byproduct of gut inflammation, as an additional terminal electron acceptor that allows *Salmonella* to outcompete with the

resident microflora[4]. Two of the genes in the tetrathionate utilization *ttrRS*/*BCA* locus were disrupted in the patient isolates: the response element of the two-component regulatory system, *ttrR* (disrupted in 4 out of 11 patient isolates), and a subunit of the reductase, *ttrB* (disrupted in 6 out of 11 patient isolates). Importantly, the tetrathionate utilization pathway is not required for extra-intestinal infection in the mouse model of systemic infection[4] and degradation of the *ttrRS*/*BCA* gene cluster is also found in *S.* Typhi (*ttrS*), *S.* Paratyphi A (*ttrB*) and *S.* Gallinarum (*ttrB* and *ttrC)*[2,5,6]. Likewise, the ability to respire ethanolamine is also only required for growth in the inflamed gut[7] and the gene encoding the critical enzyme of this pathway, *eutC*, is disrupted in all the patient isolates as well as *S.* Paratyphi C. The anaerobic vitamin B12 synthetic pathway encoded by the *cbi/cob* loci, required for the use of ethanolamine as a carbon source[8], also has multiple disrupted genes in the patient isolates. Four of the twenty genes in this pathway were disrupted in one or more of the patient isolates (*cbiJ, cbiK, cbiL,* and *cbiN*) and most patient isolates have disruptions in more than one gene of the vitamin B12 synthetic pathway (Supplementary Information Table II). Similarly, degradation of the *cbi/cob* locus is found in *S.* Typhi (*cbiM, cbiK, cbiJ, cbiC*), *S.* Paratyphi A (*cbiA*) and *S.* Gallinarum (*cobD, cbiD, cbiO,* and *cbiC*)[3,5,6]. Thus, the genes degraded in host-adapted serovars and the patient isolates are in these same pathways, but sometimes in different genes. Over thirty-five additional anaerobic metabolism genes appear to also be nonessential for growth in this patient (Supplementary Information Table III).

Colonisation of the gut requires adhesion to the mucosal surface mediated by factors including flagella, fimbriae and non-fimbrial adhesins. At least one gene in three of the fimbrial operons is degraded in the patient isolates. The *lpfC* gene, important for binding to Peyer's patches in mice[9], is disrupted in the earlier isolates. Pseudogenes in non-fimbrial adhesins include *shdA*, *misL* and *bigA*. ShdA binds fibronectin and is important for colonisation of the cecum and Peyer's patches[10,11] and is also disrupted in both *S.* Typhi and *S.* Paratyphi A.

S. enterica encode two type III secretion systems (T3SS) that play central roles in host-pathogen interactions. Over half of the isolates have disruptions in *prgH* that encodes a key component of the SPI-1 T3SS inner membrane ring structure [12] and is required for secretion of SPI-1 effector proteins and invasion of epithelial cells [13,14]. Nine of twenty-five SPI-1 and SPI-2 effector proteins contained disruptive mutations. Effector proteins, including *sipA,* which is disrupted in all of the patient isolates, can cause inflammation in the gut to promote a gastrointestinal lifestyle[15]. *sseJ* is disrupted in *S.* Typhi, *S.* Paratyphi A and the majority of patient isolates[3] and has been implicated in replication in macrophages and virulence in mice[16].

Degradation of cell envelope genes in patient isolates and resulting phenotypic changes

Several genes involved in lipopolysaccharide (LPS) biosynthesis, e.g. *mpl*, murein peptide ligase, were pseudogenes in one or more of the patient isolates. As such,

patient isolates had a crenulated morphology revealed by transmission electron microscopy (Supplementary Information Figure 7B). Furthermore, ten of the eleven patient isolates lacked a functional *rfaL* (*waaL)* encoding the O-antigen ligase that links the O-polysaccharide to the completed core of LPS (Supplementary Information Table II). Consistently, patient isolates produced LPS lacking O-antigen (Supplementary Information Figure 7A) that is consistent with the extreme sensitivity to normal human serum and low amounts of complement (Supplementary Information Figure 8).

Supplementary Discussion References

1. Giraud, E., Baucheron, S. & Cloeckaert, A. Resistance to fluoroquinolones in Salmonella: emerging mechanisms and resistance prevention strategies. *Microbes and infection / Institut Pasteur* **8**, 1937-1944 (2006).
2. 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).
3. McClelland, M. *et al.* Comparison of genome degradation in Paratyphi A and Typhi, human-restricted serovars of Salmonella enterica that cause typhoid. *Nat Genet* **36**, 1268-74 (2004).
4. Winter, S.E. *et al.* Gut inflammation provides a respiratory electron acceptor for Salmonella. *Nature* **467**, 426-9 (2010).
5. Parkhill, J. *et al.* Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhi CT18. *Nature* **413**, 848-52 (2001).
6. Thomson, N.R. *et al.* Comparative genome analysis of Salmonella Enteritidis PT4 and Salmonella Gallinarum 287/91 provides insights into evolutionary and host adaptation pathways. *Genome Res* **18**, 1624-37 (2008).
7. 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).
8. Roof, D.M. & Roth, J.R. Functions required for vitamin B12-dependent ethanolamine utilization in Salmonella typhimurium. *J Bacteriol* **171**, 3316-23 (1989).
9. Bäumler, A.J., Tsolis, R.M. & Heffron, F. The *lpf* fimbrial operon mediates adhesion to murine Peyer's patches. *Proc. Natl. Acad. Sci. USA.* **93**, 279-283 (1996).
10. Kingsley, R.A., Santos, R.L., Keestra, A.M., Adams, L.G. & Baumler, A.J. Salmonella enterica serotype Typhimurium ShdA is an outer membrane fibronectin-binding protein that is expressed in the intestine. *Mol Microbiol* **43**, 895-905. (2002).
11. Kingsley, R.A., van Amsterdam, K., Kramer, N. & Baumler, A.J. The shdA gene is restricted to serotypes of Salmonella enterica subspecies I and contributes to efficient and prolonged fecal shedding. *Infect Immun* **68**, 2720-7 (2000).
12. Schraidt, O. *et al.* Topology and organization of the Salmonella typhimurium type III secretion needle complex components. *PLoS Pathog* **6**, e1000824 (2010).
13. Pegues, D.A., Hantman, M.J., Behlau, I. & Miller, S.I. PhoP/PhoQ transcriptional repression of Salmonella typhimurium invasion genes: evidence for a role in protein secretion. *Mol Microbiol* **17**, 169-81 (1995).
14. Behlau, I. & Miller, S.J. A PhoP repressed gene promotes *Salmonella typhimurium* invasion of epithelial cells. *J. Bacteriol.* **175**, 4475-4484 (1993).
15. Hapfelmeier, S. *et al.* Role of the Salmonella pathogenicity island 1 effector proteins SipA, SopB, SopE, and SopE2 in Salmonella enterica subspecies 1 serovar Typhimurium colitis in streptomycin-pretreated mice. *Infect Immun* **72**, 795-809 (2004).
16. Ohlson, M.B., Fluhr, K., Birmingham, C.L., Brumell, J.H. & Miller, S.I. SseJ deacylase activity by Salmonella enterica serovar Typhimurium promotes virulence in mice. *Infect Immun* **73**, 6249-59 (2005).